

KHUNG LÝ THUYẾT TÍCH HỢP VỀ SỬ DỤNG MINH CHỨNG DO TRÍ TUỆ NHÂN TẠO TẠO SINH HỖ TRỢ QUYẾT ĐỊNH SỰ PHẠM TRONG PHÁT TRIỂN KỸ NĂNG NÓI TIẾNG ANH: MỘT NGHIÊN CỨU TỔNG THUẬT LÝ THUYẾT

Trịnh Thị Thu Hằng¹

Email: Trinhthuhanghanu@gmail.com. ORCID: 0009-0006-9945-8039

Ngày tòa soạn nhận được bài báo: 15/06/2026

Ngày phản biện đánh giá: 22/06/2026

Ngày bài báo được duyệt đăng: 14/09/2026

DOI: 10.59266/houjs.2026.1426

Tóm tắt: Phần lớn nghiên cứu xem trí tuệ nhân tạo (AI) như công cụ đánh giá hoặc thay thế giáo viên; vai trò của AI trong hỗ trợ quyết định giảng dạy cá nhân hóa còn bỏ ngỏ. Bằng tổng thuật lý thuyết, bài viết tích hợp ba mô hình năng lực nói (đa dạng - chính xác - trôi chảy (CAF), năng lực ngôn ngữ giao tiếp (CLA) và năng lực tương tác (IC)), với thiết kế dựa trên minh chứng (Evidence-Centred Design - ECD) để đề xuất khung lý thuyết xem AI là hệ thống tạo sinh minh chứng, không phải hệ thống đánh giá. Khung xác định thành tố năng lực nói, đánh giá mức độ AI quan sát đáng tin cậy từng thành tố, và đề xuất quy trình giúp giáo viên diễn giải minh chứng để chẩn đoán nhu cầu và thiết kế can thiệp. Khung là cơ sở khái niệm cho nghiên cứu thực nghiệm về AI trong phát triển kỹ năng nói cá nhân hóa cho người học tiếng Anh.

Từ khóa: Trí tuệ nhân tạo tạo sinh, minh chứng học tập, quyết định sự phạm, kỹ năng nói cá nhân hóa

I. Đặt vấn đề

Sự bùng nổ của AI, đặc biệt trong công nghệ nhận dạng giọng nói tự động (ASR), xử lý ngôn ngữ tự nhiên (NLP) và mô hình ngôn ngữ lớn (LLMs), đang thay đổi mạnh mẽ việc giảng dạy và đánh giá kỹ năng nói, một lĩnh vực vốn tiêu tốn nhiều thời gian, nguồn lực và chuyên môn để theo dõi sự phát triển của người học.

Trong thực tế giảng dạy, giáo viên thường gặp khó khăn trong việc thu thập đầy đủ minh chứng về quá trình phát triển kỹ năng nói của từng cá nhân, đặc biệt trong các lớp học đông đúc; AI mở ra cơ hội thu thập và phân tích dữ liệu lời nói liên tục để hỗ trợ việc theo dõi này. (Nguyễn và Phạm, 2026).

Mặc dù vậy, phần lớn nghiên cứu hiện tại vẫn chủ yếu nhìn nhận AI từ góc độ

¹ Học viện Hành chính và Quản trị công, Hà Nội, Việt Nam

đánh giá. Các hệ thống như SpeechRater, Versant hay Duolingo English Test được xem xét chủ yếu dựa trên độ chính xác chấm điểm tự động, còn các nghiên cứu về tác nhân hội thoại thông minh tập trung vào phản hồi do AI cung cấp (Nguyễn và cộng sự, 2026), vô tình giới hạn vai trò của AI ở vị trí người đánh giá hoặc người phản hồi. Tuy nhiên, giá trị thực sự của đánh giá không nằm ở con số hay nhận xét cuối cùng, mà ở khả năng tạo ra minh chứng hỗ trợ quá trình học tập.

Từ khía cạnh sư phạm, giáo viên không chỉ cần biết người học đạt kết quả ra sao mà còn cần hiểu nhu cầu học tập và can thiệp phù hợp để thúc đẩy tiến bộ, điều này càng cấp thiết trong môi trường học tập cá nhân hóa. Xuất phát từ khoảng trống này, bài viết đề xuất một khung lý thuyết tái định hình vai trò của AI từ hệ thống đánh giá tự động sang hệ thống tạo sinh minh chứng, AI phân tích dữ liệu và tạo ra minh chứng về quá trình phát triển kỹ năng nói, còn giáo viên đánh giá độ tin cậy của minh chứng, diễn giải nhu cầu học tập và thiết kế can thiệp phù hợp để phát triển kỹ năng nói cá nhân hóa.

Để xây dựng khung lý thuyết này, bài viết tập trung trả lời hai câu hỏi nghiên cứu trọng tâm. Việc xác định thành tố năng lực nói và chuyển hóa chúng thành minh chứng quan sát được là bước xây dựng nền tảng khái niệm, được trình bày trong Mục III và Mục 4.1-4.2:

1. Những yếu tố nào của năng lực nói có thể được quan sát đáng tin cậy bởi AI?

2. Các minh chứng do AI tạo ra có thể hỗ trợ giáo viên chẩn đoán nhu cầu học tập và đưa ra quyết định sư phạm như thế nào?

Qua đó, nghiên cứu hướng tới xây dựng nền tảng lý thuyết giải thích mối quan hệ giữa năng lực nói, minh chứng do AI sản xuất và quá trình ra quyết định sư phạm, đóng góp vào cả nghiên cứu lẫn thực hành giảng dạy kỹ năng nói trong môi trường học tập hỗ trợ bởi AI.

II. Cơ sở lý thuyết

Năng lực nói là một khái niệm đa chiều, được tiếp cận từ nhiều góc độ trong nghiên cứu ngôn ngữ ứng dụng. Để xây dựng khung lý thuyết về sử dụng minh chứng do AI tạo ra trong phát triển kỹ năng nói, nghiên cứu này lựa chọn ba mô hình có ảnh hưởng lớn: Đa dạng - Chính xác - Trôi chảy (CAF), Năng lực ngôn ngữ giao tiếp (CLA) và Năng lực tương tác (IC), đại diện cho ba góc nhìn bổ sung nhau về hiệu suất ngôn ngữ, hiệu quả giao tiếp và bản chất tương tác của hoạt động nói.

2.1. Mô hình Đa dạng - Chính xác - Trôi chảy (CAF)

Mô hình CAF do Skehan (1998) đề xuất là một trong những khung lý thuyết được sử dụng rộng rãi nhất trong nghiên cứu kỹ năng nói, phản ánh hiệu suất nói qua ba chiều kích: sự đa dạng ngôn ngữ, tính chính xác và độ trôi chảy, thông qua các chỉ báo có thể quan sát và định lượng được. Tuy nhiên, CAF chủ yếu hướng đến đặc điểm ngôn ngữ của sản phẩm lời nói, chưa xem xét đầy đủ hiệu quả giao tiếp và tương tác xã hội.

2.2. Mô hình Năng lực ngôn ngữ giao tiếp (CLA)

Để mở rộng cách hiểu về năng lực nói vượt ra ngoài hiệu suất ngôn ngữ, Bachman và Palmer (2010) đề xuất mô hình Năng lực Ngôn ngữ Giao tiếp (CLA),

nhấn mạnh rằng giao tiếp thành công đòi hỏi người học sử dụng ngôn ngữ phù hợp với mục đích và bối cảnh xã hội, đặc biệt qua năng lực ngữ dụng và năng lực chiến lược, hai thành tố giúp người học điều chỉnh cách diễn đạt, duy trì giao tiếp và xử lý khó khăn phát sinh.

2.3. Mô hình Năng lực tương tác (IC)

Mô hình Năng lực Tương tác (IC) của Young (2011) xem giao tiếp là quá trình đồng kiến tạo ý nghĩa giữa các bên tham gia, nhấn mạnh các khía cạnh như quản lý lượt nói, phản hồi, phát triển chủ đề và xử lý sự cố giao tiếp, những yếu tố phản ánh bản chất tương tác của kỹ năng nói.

Ba mô hình trên bổ sung cho nhau nhưng được phát triển trong những bối cảnh nghiên cứu riêng biệt, chưa được kết nối thành một khuôn khổ thống nhất để tạo sinh, diễn giải và sử dụng minh chứng trong môi trường học tập hỗ trợ bởi AI. Để lấp đầy khoảng trống này, nghiên cứu sử dụng Thiết kế dựa trên minh chứng (Evidence-Centred Design - ECD) (Mislevy và cộng sự, 2003) làm khung lý thuyết nền tảng, giúp chuyển hóa ba mô hình năng lực nói nêu trên thành các minh chứng có thể quan sát được, cơ chế này được trình bày chi tiết ở Mục 4.2.

III. Phương pháp nghiên cứu

Nghiên cứu được thực hiện theo hướng nghiên cứu lý thuyết nhằm xây dựng một khung khái niệm về việc sử dụng minh chứng do AI tạo ra trong phát triển kỹ năng nói cá nhân hóa. Dữ liệu nghiên cứu bao gồm các công trình liên quan đến năng lực nói, thiết kế dựa trên minh chứng (ECD), phân tích dữ liệu học tập (Learning Analytics), đánh giá phức

vụ học tập (AfL), ra quyết định dựa trên dữ liệu (DBDM), dạy học dựa trên nhiệm vụ (TBLT) và ứng dụng AI trong giáo dục ngôn ngữ. Cụ thể, đây là một nghiên cứu tổng thuật lý thuyết nhằm phân tích khái niệm và tích hợp các mô hình lý thuyết sẵn có, không phải là một tổng thuật hệ thống có phân tích định lượng. Phạm vi lý thuyết của bài viết giới hạn ở người học tiếng Anh như một ngoại ngữ ở bậc đại học, bao gồm cả hoạt động nói độc thoại và hội thoại được thể hiện bằng tiếng Anh.

Quy trình nghiên cứu gồm ba bước:

(1) xác định các thành tố của năng lực nói và chuyển hóa chúng thành các minh chứng có thể quan sát được; (2) phân tích khả năng của AI trong việc tạo sinh và xử lý các minh chứng đó; và (3) xây dựng khung lý thuyết mô tả quá trình sử dụng minh chứng do AI tạo ra nhằm hỗ trợ chẩn đoán nhu cầu học tập và ra quyết định sư phạm.

Việc lựa chọn tài liệu được thực hiện có định hướng trên các cơ sở dữ liệu học thuật Scopus, ERIC và Google Scholar, trong khoảng thời gian 1998-2026, sử dụng kết hợp các từ khóa “speaking competence”, “evidence-centred design”, “AI-generated evidence”, “automated speaking assessment”, “learning analytics”, “data-based decision making” với tên từng mô hình lý thuyết (CAF, CLA, IC, ECD, AfL, DBDM, TBLT), ưu tiên công trình nền tảng, được trích dẫn rộng rãi hoặc liên quan trực tiếp đến khung lý thuyết. Ba mô hình CAF, CLA, IC được chọn cho Mục 4.1 vì đại diện ba chiều kích bổ sung nhau: hiệu suất ngôn ngữ, hiệu quả giao tiếp, bản chất tương tác. Việc tích hợp thành khung chín thành tố ở Bảng 1 theo ba nguyên tắc: giữ thành

tổ không trùng lặp, gộp thành tổ tương đồng, và sắp xếp theo ba nhóm tương ứng, để mỗi thành tố có thể truy nguyên về mô hình xuất xứ.

IV. Kết quả nghiên cứu và thảo luận

4.1. Mô hình năng lực nói tích hợp

Dựa trên việc tổng hợp ba mô hình CAF, CLA và IC, nghiên cứu đề xuất một

mô hình năng lực nói tích hợp gồm chín thành tố cốt lõi. Mô hình này kết hợp các khía cạnh về hiệu suất ngôn ngữ, hiệu quả giao tiếp và năng lực tương tác nhằm phản ánh bản chất đa chiều của hoạt động nói. Chín thành tố được xem là nền tảng để xác định các minh chứng có thể quan sát được và làm cơ sở cho quá trình phân tích hỗ trợ bởi AI.

Bảng 1. Khung năng lực nói tích hợp (tổng hợp và tích hợp từ Skehan, 1998; Bachman và Palmer, 2010; Young, 2011)

Mô hình	Thành tố	Mô tả
CAF	Sự đa dạng	Sự tinh vi và đa dạng của các nguồn lực ngôn ngữ
	Tính chính xác	Độ chính xác trong việc sử dụng ngôn ngữ
	Độ trôi chảy	Sự mượt mà và hiệu quả trong sản xuất lời nói
CLA	Năng lực ngữ dụng	Sử dụng ngôn ngữ phù hợp trong các bối cảnh khác nhau
	Năng lực chiến lược	Sử dụng các chiến lược giao tiếp để duy trì tương tác
IC	Quản lý lượt nói	Điều phối các lượt nói trong cuộc hội thoại
	Phản hồi	Phản ứng phù hợp với người đối thoại
	Quản lý chủ đề	Duy trì và phát triển các chủ đề trong cuộc hội thoại
	Sửa lỗi giao tiếp	Giải quyết các sự cố, gián đoạn trong giao tiếp

Tuy nhiên, các thành tố năng lực nói vẫn là những cấu trúc tiềm ẩn và không thể quan sát trực tiếp. Do đó cần một cơ chế lý thuyết giúp chuyển hóa các cấu trúc này thành các minh chứng có thể quan sát được. Trong nghiên cứu này, ECD được lựa chọn để thực hiện chức năng đó.

4.2. Thiết kế dựa trên minh chứng trong đánh giá năng lực nói

Theo ECD, việc đánh giá chính xác đòi hỏi lập luận rõ ràng về cách các hành vi quan sát được trở thành minh chứng cho khả năng tiềm ẩn, thông qua ba thành

phần: thành tố năng lực, minh chứng và nhiệm vụ. Trong khuôn khổ này, chín thành tố năng lực nói cấu thành mô hình năng lực người học, các chỉ báo quan sát được tạo ra từ mô hình minh chứng, còn các nhiệm vụ tạo nên mô hình nhiệm vụ.

ECD nhấn mạnh rằng đánh giá nên tập trung vào thu thập minh chứng thay vì đo lường trực tiếp các thành phần trừu tượng, nguyên tắc càng quan trọng trong học tập hỗ trợ bởi AI, vì AI không thể trực tiếp theo dõi năng lực nói mà chỉ phân tích các biến số được thu thập làm minh chứng cho khả năng tiềm ẩn.

Bảng 2: Bảng mô phỏng mối liên hệ giữa thành tố năng lực, minh chứng và biến số quan sát được (xây dựng dựa trên logic Thành tố - Minh chứng - Biến quan sát của Mislevy và cộng sự, 2003)

Thành tố năng lực	Minh chứng	Các biến quan sát được
Sự đa dạng	Sử dụng ngôn ngữ tinh vi	Vốn từ phong phú, cấu trúc câu phức tạp.
Tính chính xác	Sử dụng ngôn ngữ đúng	Tỷ lệ sai sót thấp, các câu không mắc lỗi.
Độ trôi chảy	Sản xuất lời nói liên tục	Tốc độ nói và phát âm, tần suất của các khoảng dừng.

Thành tố năng lực	Minh chứng	Các biến quan sát được
Năng lực ngữ dụng	Sử dụng ngôn ngữ phù hợp	Lựa chọn giọng điệu và hành vi giao tiếp.
Năng lực chiến lược	Quản lý giao tiếp	Yêu cầu làm rõ thông tin, diễn giải lại ý kiến.
Quản lý lượt nói	Quản lý tương tác	Khởi động lượt nói và phân bổ quyền phát biểu.
Phản hồi	Tham gia hội thoại	Phản hồi có liên quan và các bước theo dõi trong cuộc trò chuyện.
Quản lý chủ đề	Phát triển chủ đề	Duy trì và mở rộng nội dung thảo luận.
Sửa lỗi giao tiếp	Khôi phục giao tiếp	Tự sửa lỗi và chuỗi làm rõ để đảm bảo sự hiểu biết.

4.3. AI trong phân tích và đánh giá kỹ năng nói

Câu hỏi tiếp theo là mức độ mà AI có thể ghi nhận các minh chứng này, vì không phải minh chứng nào cũng được AI xử lý với cùng độ tin cậy. Các nghiên cứu về hệ thống đánh giá nói tự động như SpeechRater, Versant và TOEFL cho thấy AI phân tích khá chính xác các đặc điểm ngôn ngữ như độ trôi chảy, tính chính xác và sự đa dạng ngôn ngữ, tốc độ nói, độ dài khoảng dừng, lỗi ngữ pháp, vốn từ vựng (Bernstein và cộng sự, 2010), cho thấy các thành tố hiệu suất ngôn ngữ là lĩnh vực AI hiện đạt mức độ quan sát đáng tin cậy cao. Kết quả này được củng cố bởi nghiên cứu thực nghiệm gần đây hơn của Aryadoust và cộng sự (2026), trong đó một tác nhân hội thoại dựa trên mô hình ngôn ngữ lớn (GPT) được sử dụng để đánh giá kỹ năng nói tiếng Anh học thuật của sinh viên năm nhất, cho thấy điểm số do AI chấm không khác biệt có ý nghĩa thống kê so với điểm

số của giám khảo con người ở phần lớn tiêu chí.

Các mô hình ngôn ngữ lớn (LLMs) mở rộng khả năng quan sát của AI sang khía cạnh giao tiếp cao hơn như phát triển chủ đề, tính mạch lạc và một số chiến lược giao tiếp cơ bản (Settles và cộng sự, 2020), dù vẫn chịu ảnh hưởng bởi ngữ cảnh. Hạn chế rõ rệt hơn ở năng lực tương tác và ngữ dụng: AI có thể nhận diện một số hiện tượng như yêu cầu làm rõ hay tự sửa lỗi, nhưng đánh giá tính phù hợp phụ thuộc vào bối cảnh xã hội, nên chỉ được AI quan sát ở mức hạn chế và vẫn cần sự diễn giải của giáo viên. Nghiên cứu thực nghiệm của Karatay và Xu (2025), sử dụng hệ thống hội thoại do AI vận hành để mô phỏng giám khảo trong bài thi nói IELTS, cũng cho thấy hệ thống này còn hạn chế trong việc khơi gợi đầy đủ các đặc điểm của năng lực tương tác so với giám khảo con người, củng cố nhận định rằng thành tố tương tác vẫn cần sự diễn giải của giáo viên.

Bảng 3: So sánh mức độ khả năng quan sát của từng thành tố (tổng hợp và diễn giải dựa trên Bernstein và cộng sự, 2010; Settles và cộng sự, 2020)

Thành tố	Mức độ AI quan sát được
Sự đa dạng	Tin cậy
Tính chính xác	Tin cậy
Độ trôi chảy	Tin cậy
Quản lý lượt nói	Trung bình
Phản hồi	Trung bình
Quản lý chủ đề	Trung bình
Sửa lỗi giao tiếp	Trung bình
Năng lực ngữ dụng	Hạn chế
Năng lực chiến lược	Hạn chế

Nhìn chung, AI không quan sát tất cả các thành tố năng lực nói với cùng một mức độ tin cậy. Trong khi các đặc điểm ngôn ngữ có thể tạo ra nguồn minh chứng tương đối đáng tin cậy, các thành tố giao tiếp và tương tác thường chỉ cung cấp những minh chứng mang tính gợi ý. Sự khác biệt này cho thấy các kết quả do AI tạo ra cần được xem như những minh chứng với mức độ chắc chắn khác nhau, thay vì những kết luận sư phạm hoàn chỉnh.

Cần lưu ý rằng việc phân loại mức độ tin cậy (cao/trung bình/hạn chế) ở Bảng 3 chủ yếu là suy luận có cơ sở lý thuyết, được rút ra từ việc diễn giải các nghiên cứu tổng quan hiện có (Bernstein và cộng sự, 2010; Settles và cộng sự, 2020), chứ chưa dựa trên một minh chứng thực nghiệm cụ thể nào được thu thập hay phân tích trong phạm vi bài viết này. Do đó, các mức độ tin cậy nêu ra cần được xem là giả thuyết làm việc cần được kiểm chứng qua các nghiên cứu thực nghiệm tiếp theo, chứ không phải là kết luận đã được xác nhận bằng thực nghiệm.

4.4. Từ minh chứng do AI tạo ra đến chẩn đoán sư phạm

Nếu AI có thể tạo ra minh chứng về sự phát triển kỹ năng nói, câu hỏi tiếp theo là làm thế nào để minh chứng đó được chuyển hóa thành hiểu biết sư phạm có ý nghĩa, giúp giáo viên hiểu nhu cầu học tập và đưa ra quyết định phù hợp.

Phân tích dữ liệu học tập (Learning Analytics) là quá trình thu thập, phân tích và ứng dụng thông tin về người học nhằm nâng cao kết quả học tập (Ferguson, 2012), mở rộng vai trò của AI từ công cụ đánh giá thành hệ thống cung cấp minh chứng sinh động về diễn biến theo thời gian, như cải thiện độ trôi chảy hay gia

tăng vốn từ vựng, phản ánh quỹ đạo phát triển của người học, qua đó giúp giáo viên hiểu sâu hơn về quá trình học tập thay vì trực tiếp đưa ra quyết định sư phạm.

Không phải mọi minh chứng đều có giá trị như nhau: giáo viên cần đánh giá độ tin cậy của từng loại dựa trên khả năng quan sát đã đề cập ở Mục 4.3, dùng trực tiếp làm cơ sở chẩn đoán với chỉ báo tin cậy cao, và kiểm tra thêm với chỉ báo kém tin cậy. Theo lý thuyết Đánh giá phục vụ học tập (AfL) của Black và Wiliam (1998), minh chứng do AI tạo ra giúp trả lời người học đang ở đâu và cần phát triển gì tiếp theo. Ví dụ, tần suất ngắt quãng chỉ ra nhu cầu cải thiện độ trôi chảy, còn dấu hiệu tương tác cần kết hợp quan sát của giáo viên trước khi kết luận.

Sau khi xác định nhu cầu học tập, quyết định sư phạm được hình thành theo lý thuyết Ra quyết định dựa trên dữ liệu (DBDM), kết hợp dữ liệu, chuyên môn giáo viên, mục tiêu và bối cảnh giảng dạy (Mandinach và Gummer, 2016). Giáo viên sử dụng minh chứng do AI cung cấp để thiết kế can thiệp như điều chỉnh nhiệm vụ, mức độ hỗ trợ hay phản hồi cá nhân hóa; AI không thay thế vai trò ra quyết định của giáo viên mà chỉ hỗ trợ họ có cơ sở vững chắc hơn.

4.5. Hiện thực hóa quyết định sư phạm thông qua dạy học dựa trên nhiệm vụ

Để hiện thực hóa các quyết định sư phạm thành cơ hội học tập cụ thể, khung lý thuyết gợi ý Dạy học dựa trên nhiệm vụ (TBLT) như một cơ chế giảng dạy phù hợp, vì các nhiệm vụ trong TBLT không cố định theo trình tự mà được thiết kế linh hoạt dựa trên nhu cầu học tập đã phân tích từ minh chứng, giúp mỗi người học nhận

được nhiệm vụ và lộ trình phát triển riêng. Cá nhân hóa ở đây không nằm ở việc AI trực tiếp giảng dạy, mà ở việc AI cung

cấp minh chứng để giáo viên diễn giải và chuyển hóa thành can thiệp phù hợp với từng cá nhân.

Bảng 4. Minh họa một số ví dụ tiêu biểu về quá trình chuyển hóa minh chứng thành can thiệp sư phạm đối với các nhóm năng lực thường gặp (xây dựng dựa trên khung ECD và mô hình TBLT)

Minh chứng từ AI	Độ tin cậy	Chẩn đoán sư phạm	Can thiệp
Tốc độ nói chậm	Cao	Hạn chế độ trôi chảy	Nhiệm vụ nói lặp lại (Repeated Tasks) hoặc kỹ thuật 4/3/2 (người học trình bày cùng một nội dung trong 4 phút, sau đó 3 phút và cuối cùng 2 phút nhằm tăng tốc độ xử lý ngôn ngữ)
Tỷ lệ lỗi cao	Cao	Hạn chế tính chính xác	Nhiệm vụ giao tiếp tập trung vào hình thức ngôn ngữ (Focused Tasks), trong đó người học thực hiện nhiệm vụ giao tiếp nhưng được định hướng sử dụng cấu trúc hoặc từ vựng cụ thể
Vốn từ hạn chế	Cao	Hạn chế đa dạng	Nhiệm vụ giải quyết vấn đề và thương lượng ý nghĩa (Problem-solving/Negotiation Tasks) nhằm khuyến khích người học diễn giải, làm rõ và duy trì giao tiếp
Thiếu phản hồi	Trung bình	Hạn chế tương tác	Thảo luận nhóm (Discussion Tasks) hoặc đóng vai (Roleplay Tasks) nhằm tăng cơ hội tương tác và phối hợp hội thoại
Ngôn ngữ chưa phù hợp bối cảnh	Thấp	Cần quan sát bổ sung	Nhiệm vụ mô phỏng tình huống thực tế (Simulation Tasks) như phỏng vấn xin việc, họp nhóm, tư vấn khách hàng hoặc giao tiếp trong môi trường nghề nghiệp

V. Kết luận và kiến nghị

Nghiên cứu này giới thiệu một khung lý thuyết tích hợp làm sáng tỏ cách AI có thể hỗ trợ phát triển kỹ năng nói cá nhân hóa thông qua quyết định sư phạm của giáo viên: các yếu tố năng lực nói được chuyển thành chỉ số quan sát được qua ECD, AI phân tích các chỉ số này để tạo hồ sơ minh chứng, nhưng do mức độ quan sát của AI khác nhau giữa các yếu tố, độ tin cậy của minh chứng cũng khác nhau nên trước khi dùng cho mục đích sư phạm, giáo viên cần đánh giá, diễn giải minh chứng kết hợp với chuyên môn và bối cảnh giảng dạy để xác định nhu cầu học tập, ra quyết định và thiết kế nhiệm vụ phù hợp.

Đóng góp lý thuyết chính của nghiên cứu là đề xuất tái định hình vai trò của AI trong giáo dục ngôn ngữ, thay vì hệ thống đánh giá tự động, khung lý thuyết này gợi ý nhìn nhận AI như hệ thống tạo sinh minh chứng, hỗ trợ chứ không thay thế quyết định sư phạm của giáo viên. Khung tích hợp kết hợp ba mô hình năng lực nói sẵn có với logic ECD và khả năng quan sát của AI cung cấp nền tảng khái niệm ban đầu cho nghiên cứu thực nghiệm trong tương lai, chứ chưa phải một khung lý thuyết hoàn toàn mới hay đã được kiểm chứng; các nghiên cứu tiếp theo có thể khảo sát khả năng ứng dụng của khung trong lớp học thực tế và tác động của minh chứng do AI tạo ra đối với quá trình chẩn đoán nhu cầu học tập, ra quyết định sư phạm và phát triển kỹ năng nói theo thời gian.

Tài liệu tham khảo

- Aryadoust, V., Wong, J., Zhang, W., Qian, J., Sai, Z., Liu, T., & Han, Y. (2026). Automatic conversational assessment using a GPT-based model for EAP speaking. *English for Specific Purposes*, 84, 181-200. <https://doi.org/10.1016/j.esp.2026.06.005>
- Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their use in the real world*. Oxford University Press.
- Bernstein, J., Van Moore, A., & Cheng, J. (2010). Validating automated speaking tests. *Language Testing*, 27(3), 355-377. <https://doi.org/10.1177/0265532210364404>
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. <https://doi.org/10.1080/0969595980050102>
- Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5-6), 304-317. <https://doi.org/10.1504/IJTEL.2012.051816>
- Karatay, Y., & Xu, J. (2025). Exploring the potential of conversational AI for assessing second language oral proficiency. *TESOL Quarterly*, 59(S1), S220-S250. <https://doi.org/10.1002/tesq.70003>
- Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, 60, 366-376. <https://doi.org/10.1016/j.tate.2016.07.011>
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1(1), 3-62. https://doi.org/10.1207/S15366359MEA0101_02
- Nguyen, T. B., & Pham, K. L. (2026). Cải thiện kỹ năng nói cho sinh viên năm thứ hai ngành Tiếng Anh tại Trường Đại học Mở Hà Nội thông qua phân tích lỗi trong bài thi nói. *Tạp chí Khoa học Trường Đại học Mở Hà Nội*, 135, 54. <https://doi.org/10.59266/houjs.2026.1119>
- Nguyen, T. B., Dang, T. H. T., & Tran, T. V. (2026). Khảo sát thực trạng và đề xuất ứng dụng ELSA Speak nhằm hỗ trợ kỹ năng nói cho sinh viên năm thứ nhất. *Tạp chí Khoa học Trường Đại học Mở Hà Nội, Số đặc biệt 12A*, 59. <https://doi.org/10.59266/houjs.2026.1266>
- Settles, B., LaFlair, G. T., & Hagiwara, M. (2020). Machine learning-driven language assessment. *Transactions of the Association for Computational Linguistics*, 8, 247-263. https://doi.org/10.1162/tacl_a_00310
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford University Press.
- Young, R. F. (2011). Interactional competence in language learning, teaching, and testing. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (Vol. 2, pp. 426-443). Routledge.

AN INTEGRATED FRAMEWORK FOR USING AI-GENERATED EVIDENCE TO SUPPORT TEACHER DECISION-MAKING IN ENGLISH SPEAKING DEVELOPMENT: A THEORETICAL REVIEW

Trinh Thi Thu Hang¹

Abstract: *Most current research treats AI mainly as an evaluator or a substitute for teachers, leaving open how AI can support teachers' instructional decisions for personalised speaking development. Using a theoretical review, this paper integrates three speaking-competence models (CAF, CLA, IC) with Evidence-Centred Design (ECD) to propose a framework that reconceives AI as an evidence-generating system rather than an automated evaluator. The framework identifies core components of speaking competence, evaluates the extent to which AI can reliably observe each component, and outlines a process for teachers to interpret this evidence in order to diagnose learning needs and design interventions. The framework offers a conceptual foundation for future empirical research on AI-supported, personalised speaking development for learners of English.*

Keywords: *Generative AI, learning evidence, pedagogical decision-making, personalised speaking development*

¹ Academy of Public Administration and Governance, Hanoi, Vietnam