Tạp chí Khoa học Đại học Huế: Kinh tế và Phát triển;
pISSN: 2588–1205; eISSN: 2615–9716
Tập 133, Số 5A, 2024, Tr. 53–72, DOI: 10.26459/hueunijed.v133i5A.7300

ĐÁNH GIÁ HIỆU NĂNG TÍCH HỢP HỆ THỐNG TRÍ TUỆ NHÂN TẠO CHUYỂN VĂN BẢN THÀNH GIỌNG NÓI HỖ TRỢ SINH VIÊN KHIẾM THỊ TRONG MÔ HÌNH ĐẠI HỌC THÔNG MINH

Nguyễn Thanh Tuấn[1, *], Nguyễn Đình Hoa Cương[2]

[1] Trường Đại học Kinh tế, Đại học Huế, 99 Hồ Đắc Di, Huế, Việt Nam
[2] Trường Đại học Phú Xuân, 28 Nguyễn Tri Phương, Huế, Việt Nam

* Tác giả liên hệ: Nguyễn Thanh Tuấn <nguyenthanhtuan@hueuni.edu.vn>
(*Ngày nhận bài: 7-9-2023; Ngày chấp nhận đăng: 3-12-2023)*

Tóm tắt. Người khuyết tật đã và đang gặp những khó khăn và rào cản trong việc hòa nhập giáo dục, đặc biệt là giáo dục đại học. Trong những năm gần đây, việc xây dựng và ứng dụng mô hình đại học thông minh dựa trên sự phát triển của khoa học công nghệ, kỹ thuật đang dần mở ra những cơ hội học tập cho người khuyết tật. Nghiên cứu này đánh giá các hệ thống chuyển văn bản thành giọng nói và thực hiện thí nghiệm về hiệu năng tích hợp với các mô hình đại học thông minh để phát huy khả năng hỗ trợ sinh viên khiếm thị trong các trường đại học Việt Nam. Cùng với đó, nghiên cứu cũng chỉ ra lộ trình phát triển đại học thông minh tích hợp hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói một cách phù hợp cho các trường đại học Việt Nam.

Từ khóa: hệ thống trí tuệ nhân tạo, chuyển văn bản thành giọng nói, hòa nhập giáo dục, đại học thông minh, chuyển đổi số, sinh viên khiếm thị

Evaluate the integration performance of text-to-speech artificial intelligence systems for supporting visually impaired students in the smart university model

Nguyen Thanh Tuan[1, *], Nguyen Đinh Hoa Cuong[2]

[1] University of Economics, Hue University, 99 Ho Dac Di St., Hue, Vietnam
[2] Phu Xuan University, 28 Nguyen Tri Phuong St., Hue, Vietnam

* Correspondence to Nguyen Thanh Tuan <nguyenthanhtuan@hueuni.edu.vn>
(*Received: September 7, 2023; Accepted: December 3, 2023*)

Abstract. People with disabilities have been facing difficulties and barriers in terms of inclusive education, especially at the undergraduate level. In recent years, building and applying the smart university model based on the development of science, technology, and engineering has gradually opened up learning opportunities for disabled people. This study evaluated text-to-speech systems and conducted performance experiments on the integration ability of the smart university model in order to better support visually impaired students in Vietnam's universities. Along with this, it also showed a suitable roadmap for integrating text-to-speech artificial intelligence systems into the smart university model for Vietnam's universities.

Keywords: artificial intelligence system, text-to-speech, inclusive education , smart university, digital transformation, visually impaired students

1 Đặt vấn đề

Giáo dục tương quan chặt chẽ với chênh lệch về tỷ lệ tham gia lực lượng lao động giữa nhóm người khuyết tật và không khuyết tật. Do không được đến trường cũng như việc thụ hưởng giáo dục còn thấp nên người khuyết tật thiếu kiến thức và kỹ năng sống, dẫn đến mất cơ hội việc làm. Theo kết quả Điều tra Quốc gia người khuyết tật tại Việt Nam năm 2016[1] do Tổng cục Thống kê thực hiện vào cuối năm 2016 đến năm 2017, với sự trợ giúp kỹ thuật của UNICEF, hơn 7% dân số từ 2 tuổi trở lên (khoảng hơn 6,2 triệu người) là người khuyết tật. Chênh lệch về tỷ lệ đi học giữa người khuyết tật và không khuyết tật tăng lên ở các cấp học cao hơn. Đại đa số học sinh khuyết tật đang tham gia học tập trong lớp học thông thường ở các trường học thông thường. Trên thực tế, chỉ có 0,5% người khuyết tật học trong lớp học chuyên biệt và 1,0% học sinh ở trường chuyên biệt.

Trong các nghiên cứu đề cập đến môi trường học đường, Pivik và cs. [1] xác định ba khía cạnh cần được giải quyết với tư cách là người hỗ trợ: điều chỉnh môi trường, thay đổi chính sách và nguồn lực thể chế. Về những thay đổi về môi trường, họ cho rằng điều quan trọng là phải bao gồm các nguồn lực công nghệ và điều chỉnh cơ sở hạ tầng theo nhu cầu của học sinh, và về các chính sách, họ khuyến nghị giáo dục người dân và điều chỉnh chương trình giảng dạy.

Đại học thông minh (ĐHTM) là một khái niệm liên quan đến việc hiện đại hóa toàn diện tất cả các quy trình giáo dục với sự xuất hiện của các công nghệ như bảng thông minh, màn hình thông minh và truy cập Internet không dây ở mọi nơi mà trung tâm là các lớp học thông minh. Một lớp học thông minh tích hợp các công nghệ nhận dạng giọng nói, thị giác máy tính, công nghệ IoT kết hợp với các kỹ thuật phân tích xã hội và hành vi cũng như các các công nghệ khác, được gọi chung là các tác nhân thông minh, để cung cấp trải nghiệm giáo dục từ xa tương tự như trải nghiệm lớp học truyền thống cho mọi đối tượng. Với những khó khăn mà sinh viên khuyết tật gặp phải trong cuộc sống và trong trường học, lớp học thông minh sẽ mang lại nhiều lợi ích và giúp sinh viên học tập hiệu quả hơn. Đối với những sinh viên khiếm thị, việc sử dụng các hệ thống chuyển đổi văn bản thành giọng nói ứng dụng trí tuệ nhân tạo (Text - to - Speech Artificial

[1] https://www.gso.gov.vn/wp-content/uploads/2019/04/Baocao-nguoikhuyet-tat.pdf

Intelligence - TTS AI) đang được nghiên cứu áp dụng để hỗ trợ khả năng học tập cũng như giảng dạy trong giáo dục thông minh.

Việc xuất hiện ngày càng nhiều các hệ thống TTS AI đã hỗ trợ nhiều mặt trong cuộc sống. Nhiều trường đại học đã và đang triển khai các công nghệ cũng như các hệ thống phần mềm hỗ trợ người học nói chung và sinh viên khuyết tật nói riêng [2, 3]. Việc đánh giá chất lượng các hệ thống TTS dựa trên tính năng trong các hoạt động học tập nói chung [4, 5] cũng như trong trường đại học [6] đã được thực hiện. Tuy nhiên, việc đánh giá hiệu quả của các hệ thống này khi tích hợp vào đại học thông minh vẫn chưa được thực hiện [3, 6]. Điều này cũng tương tự đối với nhiều công nghệ, kỹ thuật được ứng dụng, triển khai trong quá trình chuyển đổi từ đại học truyền thống sang đại học thông minh bởi tính chất phức tạp và mới mẻ của vấn đề. Yêu cầu cấp thiết đặt ra hiện nay là cần phải có mô hình đại học thông minh với những công nghệ, kỹ thuật tích hợp phù hợp và lộ trình chuyển đổi từ đại học truyền thống sang đại học thông minh thích hợp cho các trường đại học. Vì vậy, nghiên cứu sẽ tập trung giải quyết các vấn đề sau:

– Nghiên cứu tích hợp hệ thống TTS AI vào mô hình đại học thông minh phù hợp với thực tiễn nhằm định hướng cho tiến trình công nghệ hóa giáo dục trong các trường đại học ở Việt Nam hiện nay.

– Đánh giá hiệu năng của các hệ thống TTS AI tích hợp vào mô hình đại học thông minh trong quá trình chuyển đổi nhằm hỗ trợ lựa chọn đúng giải pháp, tối đa hóa lợi ích, tối thiểu hóa chi phí.

– Khuyến nghị lộ trình tích hợp hệ thống TTS AI vào mô hình đại học thông minh phù hợp với các trường đại học.

Sự chuyển đổi và triển khai các mô hình đại học thông minh sẽ giúp các trường đại học Việt Nam trang bị, áp dụng những thành tựu của khoa học, kỹ thuật, công nghệ nhằm hỗ trợ, đem lại nhiều lợi ích không chỉ cho người học bình thường mà cả những người khuyết tật trong quá trình dạy - học. Phần còn lại của bài báo được tổ chức như sau: Các công trình liên quan đến các đại học thông minh, trong đó tập trung vào các hệ thống TTS AI được trình bày trong tiểu mục 2. Tiểu mục 3 trình bày kiến trúc tích hợp các hệ thống TTS AI vào mô hình đại học thông minh. Thí nghiệm kiểm chứng được mô tả trong tiểu mục 4. Lộ trình chuyển đổi và triển khai đại học thông minh được đề xuất trong tiểu mục 5. Tiểu mục 6 nêu kết luận và định hướng nghiên cứu tiếp theo.

2 Các công trình liên quan

Những bước nhảy vọt của Cách mạng công nghiệp 4.0 (CMCN 4.0) đặt ra nhiều thách thức [7, 8], và phát sinh thêm rất nhiều ngành nghề mới trên thị trường lao động [9, 10]. Sự thay đổi này đòi hỏi giáo dục phải đem lại cho người học năng lực thích ứng với thách thức [11, 12] và những yêu cầu mới mà các phương pháp giáo dục truyền thống không thể đáp ứng [11, 13]. Do đó, trường đại học là nơi cung cấp nguồn nhân lực bậc cao cho xã hội [8, 9] cũng phải thay đổi toàn diện cả về mô hình [14, 15], nội dung chương trình [9, 12] và phương thức đào tạo [16, 17].

Các ý tưởng về giáo dục thông minh (Smart Education - SmE), trường đại học thông minh (Smart University - SmU), lớp học thông minh (Smart Classroom - SmC), môi trường học tập thông minh (Smart Learning Environments - SLE) đã trở thành chủ đề chính của nhiều sự kiện, dự án quốc tế và quốc gia. Xu hướng chuyển từ mô hình đại học truyền thống sang đại học thông minh đang trở thành tất yếu, đáp ứng yêu cầu xã hội hiện nay. Đóng vai trò là nhân tố cơ bản, trung tâm trong việc xây dựng, triển khai các mô hình đại học thông minh, lớp học thông minh cần tối ưu hóa việc trình bày nội dung giảng dạy, truy cập thuận tiện các tài nguyên học tập, tính tương tác sâu sắc của việc dạy và học, nhận thức và phát hiện theo ngữ cảnh, bố trí và quản lý lớp học dựa trên các tác nhân thông minh, một trong số đó là các hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói.

Một trong những công nghệ đang được nghiên cứu triển khai trong các lớp học thông minh là các hệ thống TTS AI nhằm hỗ trợ cho hoạt động dạy - học nói chung và sinh viên khiếm thị nói riêng trong môi trường đại học. Chuyển văn bản thành giọng nói (Text - to - Speech - TTS) là một cơ chế giúp chuyển đổi văn bản viết thành lời nói. Quá trình này tuân theo hai bước chính, đó là phân tích văn bản và tạo dạng sóng giọng nói. Trong vài thập kỷ qua, công nghệ chuyển văn bản thành giọng nói đã phát triển nhanh với sự trợ giúp của nhiều công nghệ khác nhau như trí tuệ nhân tạo và học máy [18]. Sử dụng các công nghệ học máy, việc tổng hợp giọng nói trong hệ thống chuyển văn bản thành giọng nói hỗ trợ tạo ra các hệ thống máy tính có giọng nói giống con người. Các thành phần chính của hệ thống chuyển văn bản thành giọng nói bao gồm [19]:

– Mô-đun Xử lý ngôn ngữ tự nhiên (Natural Language Processing - NLP) tạo ra phiên âm của văn bản đầu vào và ngữ điệu trong một số trường hợp (ngữ điệu, thời lượng, biên độ và cường độ).

– Đầu ra dữ liệu từ mô-đun Xử lý ngôn ngữ tự nhiên (NLP) được mô-đun Xử lý tín hiệu số (Digital Signal Processing - DSP) chuyển đổi thành lời nói.

Các hệ thống chuyển văn bản thành giọng nói của các ngôn ngữ khác nhau dựa trên các chỉ số chất lượng như tỷ lệ nhận dạng, tính chính xác, điểm TTS, độ chính xác, khả năng gợi nhớ và điểm F1.

Sự thành công của hệ thống chuyển văn bản thành giọng nói dựa trên các yếu tố khác nhau như: lĩnh vực sở trường của giọng nói, tính rõ ràng, tính tự nhiên và các yếu tố thuộc về con người như tính dễ hiểu. Tính năng độ rõ âm thanh tạo ra chất lượng hoặc số lượng từ được hình thành trong một câu [20]. Thuật ngữ tính tự nhiên xác định tính ưu việt của việc tạo lời nói trong bối cảnh phát âm, thể hiện cảm xúc và cấu trúc thời gian của nó. Lựa chọn loại giọng nói tùy theo sở thích của người nghe là tính năng vượt trội của hệ thống chuyển văn bản thành giọng nói. Tính tự nhiên và sở thích đều bị ảnh hưởng thông qua tính năng chuyển văn bản thành giọng nói cũng như thông qua chất lượng giọng nói và âm thanh ảnh hưởng đến thông điệp nhận được [21].

Học sâu (Deep Learning - DL) là một hướng nghiên cứu mới trong lĩnh vực học máy những năm gần đây. Nó có thể nắm bắt hiệu quả các cấu trúc ẩn bên trong của dữ liệu và sử dụng các khả năng lập mô hình mạnh mẽ hơn để mô tả dữ liệu [22]. Dựa trên phương pháp học sâu để

tổng hợp tiếng nói, nhiều mô hình đã được đề xuất như máy Boltzmann giới hạn (Restrictive Boltzmann Machine - RBM), mạng niềm tin sâu (Deep Belief Network - DBN), mạng mật độ hỗn hợp sâu (Deep Mixture Density Network - DMDN), bộ nhớ ngắn dài hạn hai chiều sâu (Deep Bidirectional Long Short-Term Memory - DBLSTM), WaveNet, Tacotron và mạng thần kinh tích chập (Convolutional Neural Network - CNN).

WaveNet [23] được phát triển từ mô hình PixelCNN [24] hoặc PixelRNN [25] áp dụng trong lĩnh vực tạo hình ảnh, đây là mô hình tạo các dạng sóng âm thanh thô mạnh mẽ. Mặc dù mô hình WaveNet có thể tạo ra âm thanh chất lượng cao nhưng vẫn gặp phải các vấn đề sau: (1) quá chậm vì dự đoán của từng điểm lấy mẫu luôn phụ thuộc vào các điểm lấy mẫu đã dự đoán trước đó; (2) phụ thuộc vào các đặc điểm ngôn ngữ từ giao diện người dùng TTS hiện có và các lỗi từ phân tích văn bản giao diện người dùng sẽ ảnh hưởng trực tiếp đến hiệu quả tổng hợp.

Để giải quyết những vấn đề này, WaveNet song song được đề xuất để cải thiện hiệu quả lấy mẫu. Nó có khả năng tạo ra các mẫu giọng nói có độ trung thực cao với tốc độ nhanh hơn 20 lần [26]. Một mô hình nơ-ron khác là Deep Voice [27] cũng được đề xuất để thay thế từng thành phần bao gồm giao diện người dùng phân tích văn bản, mô hình âm thanh và bộ tổng hợp giọng nói bằng một mạng nơ-ron tương ứng. Tuy nhiên, vì mỗi thành phần được đào tạo độc lập nên nó không phải là sự tổng hợp từ đầu đến cuối thực sự.

Tacotron [28] là một mô hình tổng hợp tiếng nói hoàn chỉnh từ đầu đến cuối. Nó có khả năng đào tạo một mô hình tổng hợp giọng nói với các cặp <văn bản, âm thanh>, do đó, giảm bớt nhu cầu về kỹ thuật tính năng tốn nhiều công sức. Ngoài ra, vì nó dựa trên cấp độ ký tự nên có thể áp dụng cho hầu hết các loại ngôn ngữ bao gồm cả tiếng Quan Thoại của Trung Quốc. Vì Tacotron là một mô hình hoàn chỉnh từ đầu đến cuối ánh xạ trực tiếp văn bản đầu vào sang biểu đồ phổ Mel nên nhận được nhiều sự chú ý của các nhà nghiên cứu và nhiều phiên bản cải tiến khác nhau đã được đề xuất.

Các tác giả trong [29] cũng đề xuất hệ thống Tacotron2 để tạo tín hiệu âm thanh dẫn đến điểm ý kiến trung bình (Mean Opinion Score - MOS) rất cao, có thể so sánh với lời nói của con người [30]. Mặc dù hệ thống đầu cuối dựa trên Tacotron đạt hiệu suất đầy hứa hẹn trong thời gian gần đây nhưng vẫn có một nhược điểm là có nhiều đơn vị lặp lại.

Nhiều công trình đã được đề xuất nhằm giải quyết vấn đề có nhiều đơn vị lặp lại. Các tác giả trong [31] đề xuất một mạng tích chập sâu (Deep Convolutional Network - DCN) với sự chú ý rằng, hướng dẫn có thể được đào tạo nhanh hơn nhiều so với hệ thống thần kinh tiên tiến nhất dựa trên RNN. Khác với mô hình WaveNet sử dụng cấu trúc tích chập hoàn toàn như một loại bộ phát âm hoặc phần phụ trợ, Ref. đúng hơn là một frond-end (và hầu hết quá trình xử lý back-end) có thể tổng hợp một phổ. Các tác giả trong [32, p. 3] đã đề xuất một kiến trúc từ ký tự đến phổ tích chập hoàn toàn mới, có tên là Deep Voice 3, để tổng hợp giọng nói, cho phép tính toán song song hoàn toàn để làm cho quá trình đào tạo nhanh hơn so với sử dụng các đơn vị lặp lại.

Cùng với các thuật toán và mô hình xử lý TTS ngày càng phát triển, các hãng phần mềm cũng như các công ty công nghệ lớn đã và đang phát triển các nền tảng, hệ thống chuyển văn bản thành giọng nói phục vụ cho nhiều mục đích khác nhau [4, 33, 34].

Ngoài những tiến bộ tiềm năng trong công nghệ TTS, có thể có nhiều tiến bộ trong việc tích hợp công nghệ này với các công nghệ khác [34]. Chất lượng của các hệ thống TTS và khả năng tích hợp vào mô hình đại học thông minh cũng đã được nghiên cứu [2, 6], khẳng định vai trò to lớn của các công nghệ AI nói chung và các hệ thống TTS nói riêng trong việc hỗ trợ các hoạt động giảng dạy và học tập trong trường đại học.

3 Kiến trúc tích hợp hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói vào mô hình đại học thông minh

Việc chuyển đổi từ giáo dục truyền thống sang giáo dục thông minh là tất yếu nhằm đáp ứng những yêu cầu mới của cuộc cách mạng công nghiệp 4.0 [11, 14]. Một trong những người tiên phong xây dựng mô hình ĐHTM là Vladimir Uskov [35] với cái nhìn toàn cảnh về thành phần cũng như về mặt kỹ thuật, công nghệ của mô hình ĐHTM. Theo đó, ĐHTM và giáo dục thông minh thể hiện sự tích hợp của 1) Các hệ thống trí tuệ và thông minh [35], đối tượng thông minh [36] và môi trường thông minh [37]; 2) Công nghệ thông minh [38], các ngành khác nhau của khoa học máy tính và kỹ thuật máy tính [10]; 3) Phần mềm giáo dục thông minh hiện đại, hệ thống phần cứng [7, 35], tác nhân, công cụ [38, 39] và 4) Phương pháp sư phạm sáng tạo [16, 36], các chiến lược giảng dạy [9, 17], và phương pháp học tập dựa trên công nghệ tiên tiến [9, 40].

Tuy nhiên, việc đề xuất kiến trúc tích hợp các hệ thống trí tuệ và thông minh vào mô hình đại học thông minh như thế nào cho phù hợp vẫn chưa thực sự được quan tâm [2, 3]. Dựa trên các nghiên cứu ở mục 2 kết hợp với phương pháp *tiếp cận và phân tích sáng tạo có hệ thống* [41], nghiên cứu này đề xuất một kiến trúc tích hợp hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói vào mô hình đại học thông minh được trình bày trong Hình 1.

Trong kiến trúc này có sự tích hợp của hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói (khung màu xanh lá cây) với các thiết bị của mô hình đại học thông minh (khung màu hồng), cụ thể là các thiết bị đầu cuối, hạ tầng kỹ thuật công nghệ thông tin, truyền thông, các kho dữ liệu của mô hình đại học thông minh.

Hình 1. Mô hình đại học thông minh tích hợp hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói (T2S AI)

3.1 Thu thập dữ liệu văn bản

Với việc triển khai mô hình đại học thông minh mà trung tâm là lớp học thông minh, môi trường học tập thông minh, các thiết bị, công nghệ, cảm biến nhận dạng được lắp đặt trong khuôn viên lớp học cũng như toàn trường. Các thiết bị kỹ thuật số như camera, webcam, máy ảnh, máy chiếu, điện thoại thông minh đóng vai trò là các thiết bị thu nhận dữ liệu văn bản đầu vào cho hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói từ nhiều nguồn khác nhau như sách giáo khoa, báo chí, chữ viết tắt, các ghi chú cũng như các loại chữ trên các hình ảnh.

Các thiết bị nhập dữ liệu này thông qua một hệ thống giao diện người dùng để đưa các dữ liệu văn bản cần chuyển thành giọng nói vào các kho dữ liệu dưới dạng hình ảnh. Hệ thống giao diện người dùng nhập liệu bao gồm các ứng dụng hoặc các API cho điện thoại, máy tính và các thiết bị khác hỗ trợ người dùng có thể sử dụng nhiều phương thức khác nhau để đưa văn bản yêu cầu vào kho dữ liệu. Các kho dữ liệu bao gồm các hệ thống máy chủ phục vụ cho các hoạt động khác nhau của mô hình đại học thông minh cũng như các dịch vụ đám mây của các hệ thống trong hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói.

3.2 Kho dữ liệu

Trong các hệ thống trí tuệ nhân tạo chuyển các văn bản thành giọng nói (Text-to-Speech Artificial Intelligence - TTS AI), kho dữ liệu đóng vai trò rất quan trọng trong việc lưu trữ các tập dữ liệu đào tạo cũng như các dữ liệu văn bản đầu vào.

Kho dữ liệu văn bản dùng để lưu trữ các văn bản đầu vào của hệ thống TTS AI được hệ thống nhập liệu chuyển đến. Ngoài ra, kho dữ liệu văn bản còn lưu trữ các tập dữ liệu đào tạo của các loại ngôn ngữ khác nhau, bao gồm các đặc trưng về mặt ngữ nghĩa, ngôn ngữ, cũng như các đặc tính về từ, ngữ và câu. Kho dữ liệu văn bản phục vụ cho hệ thống nhận dạng, xử lý văn bản đầu vào.

Kho dữ liệu giọng nói phục vụ cho hệ thống tổng hợp giọng nói. Từ kết quả xử lý, phân tích văn bản của giai đoạn trước, dữ liệu được chuyển đến kho dữ liệu giọng nói để hệ thống tổng hợp giọng nói có thể xử lý và đưa ra kết quả phát âm phù hợp với văn bản yêu cầu.

Kho dữ liệu sóng âm lưu trữ tập dữ liệu đào tạo liên quan đến việc tạo sóng âm cũng như dữ liệu kết quả từ việc tổng hợp giọng nói chuyển đến. Nó bao gồm cả các dữ liệu giọng nói tự nhiên của các ngôn ngữ từ các đối tượng khác nhau để có thể đưa ra giọng nói phù hợp với yêu cầu của hệ thống tạo giọng nói.

3.3 Xử lý văn bản

Trong giai đoạn tiền xử lý văn bản của chu trình chuyển văn bản thành giọng nói, văn bản đầu vào được nhận dạng và chuyển thành định dạng chuẩn hóa để phần còn lại của hệ thống có thể hiểu được. Nó cũng có thể liên quan đến việc tăng số lượng và chia văn bản đầu vào thành các câu và ký tự, từ riêng biệt. Sau đó, văn bản được xử lý trước được chia thành thông tin có ý nghĩa bằng cách sử dụng các quy tắc diễn đạt và mô hình ngôn ngữ trong giai đoạn diễn đạt theo giai điệu. Các thuật toán nhận dạng, xử lý ảnh kết hợp các mô hình học máy và dữ liệu đào tạo lưu trữ trong kho dữ liệu văn bản được sử dụng để xác định chính xác các ký tự, từ, câu trong văn bản yêu cầu.

Để thực hiện phân đoạn âm vị và chú thích trên ngữ liệu, các hệ thống xử lý, phân tích văn bản thường sử dụng năm cấp độ sau:

Cấp độ âm vị: các ký hiệu ngữ âm của âm trước trước, âm trước, âm sau, âm sau hoặc âm sau sau; khoảng cách tiến hoặc lùi của âm vị hiện tại trong âm tiết.

Cấp độ âm tiết: âm tiết trước đó, hiện tại hay âm tiết tiếp theo được nhấn mạnh; số lượng âm vị chứa trong âm tiết trước, hiện tại hoặc tiếp theo; khoảng cách tiến hoặc lùi của âm tiết hiện tại trong từ hoặc cụm từ; số lượng âm tiết được nhấn trước hoặc sau âm tiết hiện tại trong cụm từ; khoảng cách từ âm tiết hiện tại đến âm tiết được nhấn mạnh gần nhất về phía trước hoặc phía sau; ngữ âm nguyên âm của âm tiết hiện tại.

Cấp độ từ: phần lời nói (Part of Speech - POS) của từ trước, từ hiện tại hoặc từ tiếp theo; số lượng âm tiết của từ trước, từ hiện tại hoặc từ tiếp theo; vị trí tiến hoặc lùi của từ hiện tại trong cụm từ; từ nội dung tiến hoặc lùi của từ hiện tại trong cụm từ; khoảng cách từ từ hiện tại đến từ nội dung gần nhất về phía trước hoặc phía sau; POS của từ trước đó, từ hiện tại hoặc từ tiếp theo.

Cấp độ cụm từ: số lượng âm tiết của cụm từ trước đó, hiện tại hoặc tiếp theo; số lượng từ của cụm từ trước đó, hiện tại hoặc tiếp theo; vị trí tiến hoặc lùi của cụm từ hiện tại trong câu; chú thích prosodic của cụm từ hiện tại.

Cấp độ câu: Số lượng âm tiết, từ hoặc cụm từ trong câu hiện tại.

Văn bản yêu cầu sau khi được xử lý, phân tích sẽ được đưa vào kho dữ liệu văn bản để lưu trữ, phục vụ cho các hoạt động xử lý, phân tích tiếp theo. Nó cũng được chuyển đến hệ thống tổng hợp giọng nói để xử lý trong giai đoạn tiếp theo của chu trình.

3.4 Tổng hợp giọng nói

Hệ thống tổng hợp giọng nói dựa trên kết quả nhận dạng, phân tích văn bản ở bước trước kết hợp các quy luật, dữ liệu đào tạo sẵn có cùng với các phương pháp tổng hợp giọng nói phù hợp (đã trình bày ở mục 2) để đưa ra các âm thanh phù hợp với từng ký tự, từ, cụm từ và câu của văn bản yêu cầu.

Sau đó, hệ thống sẽ tổng hợp âm thanh của từng từ, cụm từ, và câu, kết hợp với các ghép nối, ngữ điệu để cho ra câu nói hoàn chỉnh. Kết quả xử lý của quá trình này sẽ là đầu vào cho hệ thống tạo giọng nói cũng như lưu trữ lại để sử dụng cho những lần tổng hợp tiếp theo.

3.5 Tạo giọng nói

Với kết quả của hệ thống tổng hợp giọng nói, hệ thống tạo giọng nói sẽ lựa chọn ngôn ngữ, giọng điệu tự nhiên phù hợp với văn bản yêu cầu ban đầu để tạo ra các sóng âm tương thích với câu nói cũng như đoạn văn đầu vào.

Ngoài ra, hệ thống này còn tạo ra các tín hiệu đầu ra phù hợp với các thiết bị sẵn có trong mô hình đại học thông minh, thể hiện sự đồng bộ hóa về mặt kỹ thuật và công nghệ, đáp ứng yêu cầu của các đối tượng sử dụng các thiết bị khác nhau trong môi trường giáo dục thông minh.

4 Thí nghiệm kiểm chứng việc tích hợp hệ thống TTS AI vào mô hình đại học thông minh

4.1 Mô tả về dữ liệu và cách thức thu thập, phân tích

Để đánh giá hiệu năng của việc sử dụng và tích hợp các hệ thống TTS AI vào mô hình đại học thông minh, nghiên cứu lựa chọn một số hệ thống TTS AI và cách thức sử dụng chúng trong thực tế với hai hình thức là sử dụng ứng dụng và tích hợp API được trình bày trong Bảng 1.

Có thể thấy một số nền tảng sử dụng ứng dụng như IBM, Google, Amazon đều sử dụng dịch vụ giao diện lập trình ứng dụng đám mây (Cloud Application Programming Interface - Cloud API). Vì vậy, nghiên cứu sử dụng Cloud API để đánh giá hiệu năng của các nền tảng TTS trong quá trình tích hợp với mô hình đại học thông minh. Như vậy, tất cả 15 nền tảng TTS AI sẽ được đưa vào đánh giá trong nghiên cứu này.

Các yếu tố được lựa chọn để đánh giá hiệu năng của các hệ thống TTS AI bao gồm: *sự chính xác* (khả năng đọc đúng các tên/ danh từ riêng thuộc ngôn ngữ khác tiếng Anh) [42]; *sự trôi chảy, dễ nghe* (từ ngữ, câu và văn bản được tổng hợp liền mạch, giọng nói có âm điệu đúng với ngữ cảnh,

Bảng 1. Các hệ thống Text-to-Speech (TTS) và hình thức sử dụng

Nền tảng chuyển văn bản thành giọng nói	Sử dụng ứng dụng	Tích hợp API
IBM Watson TTS	☒ Watson Assistant	☒ Dịch vụ đám mây API
Microsoft Speech	☒ Lync Assistant	☒ Dịch vụ đám mây API
Google Cloud TTS	☒ Home Assistant	☒ Dịch vụ đám mây API
Mac TTS	☒ AppleScipt	☒ Dịch vụ đám mây API
Naturaltts		☒ Dịch vụ đám mây API
VocaliD's	☒ MyVocaliD	☒ Dịch vụ đám mây API
LOVO Studio		☒ Dịch vụ đám mây API
Respeecher		☒ Dịch vụ đám mây API
Woord - TTS	☒ Woord Text to Speech	☒ Dịch vụ đám mây API
Amazon Polly	☒ PollyPlayer	☒ Dịch vụ đám mây API
ReadSpeaker		☒ Dịch vụ đám mây API
Nuance TTS		☒ Dịch vụ đám mây API
Acapela VaaS		☒ Dịch vụ đám mây API
Resemble	☒ Resembleai App Workflow	☒ Dịch vụ đám mây API
Murf TTS	☒ Murf App	☒ Dịch vụ đám mây API

kết hợp các ghép nối chính xác) [5, 43]; *sự tự nhiên, thân thiện với người dùng* (các từ ngữ và câu được phát ra giống với giọng nói tự nhiên) [44]; *tốc độ* (tốc độ phát âm phù hợp với người nghe) [5]; *chất lượng* (các từ ngữ trong văn bản được phát âm rõ ràng, dễ hiểu) [4]. Các yếu tố này sẽ tương ứng phản ánh hiệu năng của các hệ thống con trong hệ thống TTS AI ở Hình 1.

Các yếu tố trên được đánh giá dựa trên thang đo Điểm ý kiến trung bình (Mean Opinion Score - MOS) với 5 mức độ xếp hạng từ Tệ, Kém, Bình thường, Tốt, Xuất sắc được ánh xạ thành các số từ 1 đến 5.

Thí nghiệm được thực hiện dựa trên việc lựa chọn 35 sinh viên học tiếng Anh tham gia nghiên cứu. Sinh viên sẽ nghe ngẫu nhiên 30 văn bản được lấy ra từ một tập dữ liệu đầu vào gồm 1.000 văn bản tiếng Anh qua hệ thống TTS AI bất kỳ, mỗi hệ thống sẽ phát 3 văn bản với các giọng đọc của cả nam và nữ. Mỗi sinh viên sẽ nghe 2 trong 3 đoạn văn bản được chọn ngẫu nhiên của một hệ thống TTS. Như vậy, mỗi hệ thống TTS sẽ phát 70 văn bản. Sinh viên sẽ không được biết mình đang sử dụng hệ thống nào. Do các hệ thống TTS AI được sử dụng trong hệ thống đều sử dụng dịch vụ Cloud API nên sinh viên có thể sử dụng các phương tiện, thiết bị khác nhau trong lớp học cũng như của cá nhân để nghe các giọng nói được tổng hợp từ các hệ thống này.

Sau đó, sinh viên được đọc các văn bản mà các hệ thống TTS AI đã tổng hợp giọng nói. Đối chiếu văn bản với kết quả nghe được, sinh viên thực hiện đánh giá mức độ hài lòng về các yếu tố cần khảo sát thông qua một bảng hỏi với các yêu cầu cụ thể như ở Bảng 2.

Bảng 2. Các yếu tố đánh giá hiệu năng của các hệ thống TTS AI

Yếu tố	Nhận định, đánh giá	Thang điểm quy đổi				
		1	**2**	**3**	**4**	**5**
Sự chính xác	Chính xác trong các danh từ riêng, các từ viết tắt, các từ ghép nối: cho thấy mức độ chính xác của âm thanh lời nói so với từ trong văn bản.	Rất không chính xác	Không chính xác	Tạm chấp nhận	Chính xác	Rất chính xác
	Phát âm chính xác: nhận thấy bất kỳ điểm bất thường nào trong cách phát âm câu tự nhiên hay không?	Rất không chính xác	Không chính xác	Tạm chấp nhận	Chính xác	Rất chính xác
Sự trôi chảy, dễ nghe	Tính biểu cảm: cho biết cảm nhận về âm thanh giọng nói đơn điệu hay rất biểu cảm?	Rất đơn điệu	Đơn điệu	Bình thường	Biểu cảm	Rất biểu cảm
	Phát âm âm thanh lời nói: cho biết giọng nói có thể phân biệt rõ ràng các từ ngữ và câu, ngắt câu hay không?	Không thể nào phân biệt	Không phân biệt rõ	Phân biệt khá rõ	Phân biệt rõ	Phân biệt rất rõ
Sự tự nhiên, thân thiện với người dùng	Độ dễ chịu của giọng nói: giọng nói nghe thấy dễ chịu hay không?	Rất không dễ chịu	Không dễ chịu	Bình thường	Dễ chịu	Rất dễ chịu
	Tính nhân tạo: có nghĩ rằng đây là giọng nói của người máy không?	Hoàn toàn giống	Tương đối giống	Hơi giống	Tương đối không giống	Hoàn toàn không giống
	Sự phù hợp: có nghĩ rằng giọng nói này phù hợp với ngữ cảnh này không?	Hoàn toàn không phù hợp	Tương đối không phù hợp	Khá phù hợp	Tương đối phù hợp	Hoàn toàn phù hợp
Tốc độ	Tốc độ nói: cho biết tốc độ truyền tải thông điệp phù hợp hay không.	Quá nhanh	Quá chậm	Hơi nhanh	Hơi chậm	Phù hợp
	Nỗ lực theo kịp: cho biết mức độ nỗ lực phải thực hiện để theo kịp thông điệp.	Rất nỗ lực nhưng không theo kịp	Nỗ lực nhiều mới theo kịp	Nỗ lực vừa phải thì theo kịp	Cần ít nỗ lực để theo kịp	Không cần nỗ lực vẫn theo kịp
	Thoải mái tiếp nhận: mức độ thoải mái trong tiếp nhận thông điệp.	Rất không thoải mái	Không thoải mái	Bình thường	Thoải mái	Rất thoải mái
Chất lượng	Nỗ lực lắng nghe: mức độ nỗ lực phải thực hiện để hiểu thông điệp.	Rất nỗ lực nhưng không hiểu	Nỗ lực nhiều mới hiểu	Nỗ lực vừa phải thì hiểu	Cần ít nỗ lực để hiểu	Không cần nỗ lực vẫn hiểu
	Vấn đề về hiểu: nhận thấy những từ đơn lẻ khó hiểu hay không.	Tất cả các từ	Nhiều từ	Một số	Ít từ	Không có

Yếu tố	Nhận định, đánh giá	Thang điểm quy đổi				
		1	2	3	4	5
	Tính dễ hiểu: có thể hiểu rõ giọng nói từ hệ thống này không?	Hoàn toàn khó hiểu	Tương đối khó hiểu	Bình thường	Dễ hiểu	Hoàn toàn dễ hiểu
	Ấn tượng chung: đánh giá chất lượng âm thanh của giọng nói đã nghe.	Tệ	Kém	Bình thường	Tốt	Xuất sắc

Số liệu thu thập được phân tích dựa trên phần mềm SPSS 20 với các công cụ như phân tích độ tin cậy thang đo Cronbach's Alpha, thống kê mô tả để tính toán giá trị trung bình về điểm số MOS của mỗi hệ thống TTS AI, từ đó tiến hành phân tích sự khác biệt về giá trị trung bình giữa các hệ thống này.

4.2 Kết quả phân tích và đánh giá

Tác giả tiến hành phân tích độ tin cậy của các thang đo bằng cách sử dụng hệ số tin cậy Cronbach's Alpha. Giá trị Cronbach's Alpha của các nhân tố được thể hiện trong Bảng 3.

Các nhân tố đều có hệ số Cronbach's > 0,6 nên đảm bảo độ tin cậy. Tiếp tục phân tích giá trị trung bình của các hệ thống TTS AI dựa trên thống kê mô tả, kết quả được trình bày ở Bảng 4.

Nhìn vào Bảng 4, có thể nhận thấy giá trị trung bình của các hệ thống TTS AI là tương đối đồng đều, không có sự khác biệt quá nhiều giữa các hệ thống này. Các hệ thống của các công ty công nghệ lớn như Google, Microsoft, Mac, Amazon vẫn được đánh giá cao. Bên cạnh đó, một số hệ thống TTS AI của một số công ty chỉ sử dụng công nghệ điện toán đám mây API vẫn nhận được sự đánh giá cao từ những người trải nghiệm như LOVO Studio, Read Speaker.

Bảng 3. Giá trị Cronbach's alpha của các nhân tố trong nghiên cứu

Biến quan sát	Cronbach's alpha
Sự chính xác	0,708
Sự trôi chảy	0,766
Sự tự nhiên	0,689
Tốc độ	0,747
Chất lượng	0,713

Bảng 4. Giá trị trung bình của các hệ thống TTS AI

Hệ thống TTS AI	Giá trị trung bình	Số quan sát	Độ lệch chuẩn
IBM Watson TTS	3,3426	70	0,65330
Microsoft Speech	3,4196	70	0,65058
Google Cloud TTS	3,5110	70	0,61271
Mac TTS	3,4128	70	0,63990
Natural TTS	3,3000	70	0,67708

Hệ thống TTS AI	Giá trị trung bình	Số quan sát	Độ lệch chuẩn
VocalID's	3,3393	70	0,64164
LOVO Studio	3,4682	70	0,62515
Respeecher	3,3360	70	0,57356
Woord TTS	3,3128	70	0,68009
Amazon Polly	3,4238	70	0,62309
ReadSpeaker	3,4202	70	0,61947
Nuance TTS	3,3408	70	0,62418
Acapela VaaS	3,2878	70	0,63951
Resemble	3,2884	70	0,64373
Murf TTS	3,3241	70	0,59290
Tổng	**3,3685**	**1.050**	**0,63299**

Để kiểm định sự khác biệt giữa các giá trị này, nghiên cứu tiến hành phân tích Anova và thu được kết quả ở Bảng 5.

Các hệ số Sig của kiểm định Levene > 0,05 nghĩa là không có sự khác biệt phương sai giữa các hệ thống TTS AI tham gia khảo sát. Tiếp tục xem xét các kết quả khác trong phân tích Anova ở Bảng 6.

Hệ số Sig. của nhân tố TROI_CHAY là 0,003 < 0,05, nghĩa là có sự khác biệt trung bình về sự trôi chảy giữa các hệ thống TTS AI (Bảng 7). Đối với các yếu tố còn lại, kết quả phân tích cho thấy, không có sự khác biệt giữa các hệ thống TTS AI về các yếu tố này, chứng tỏ có sự đồng đều về mặt chất lượng, tốc độ, sự chính xác và sự tự nhiên trong các hệ thống TTS AI tham gia thử nghiệm.

Qua những phân tích trên, có thể đánh giá các hệ thống TTS AI tham gia khảo sát có sự tương đồng về các yếu tố đánh giá, chỉ có sự khác biệt tương đối về tính trôi chảy giữa các hệ thống. Như vậy, khi tích hợp các hệ thống TTS AI vào mô hình đại học thông minh, các trường đại học có thể có nhiều lựa chọn phù hợp, đặc biệt là việc sử dụng các hệ thống hỗ trợ công nghệ điện toán đám mây API. Vấn đề quyết định còn lại là chi phí bỏ ra để sử dụng các hệ thống TTS AI này.

Bảng 5. Kiểm định Homogeneity of Variances

Nhân tố	Thống kê Levene	df1	df2	Sig.
Sự chính xác	0,216	14	1035	0,999
Sự trôi chảy	0,717	14	1035	0,759
Sự tự nhiên	1,656	14	1035	0,059
Tốc độ	0,734	14	1035	0,741
Chất lượng	0,553	14	1035	0,901

Bảng 6. Kiểm định giá trị trung bình (Robust Tests of Equality of Means)

Nhân tố		Statistic[a]	df1	df2	Sig.
Sự chính xác	Welch	0,864	14	394,272	0,599
Sự trôi chảy	Welch	2,415	14	394,252	0,003
Sự tự nhiên	Welch	0,462	14	394,248	0,952
Tốc độ	Welch	0,180	14	394,249	1,000
Chất lượng	Welch	0,190	14	394,270	1,.000

a. Asymptotically F distributed

Bảng 7. Sự khác biệt về giá trị trung bình của Sự trôi chảy giữa các hệ thống TTS AI

Nhân tố	Hệ thống TTS AI	Giá trị trung bình	Độ lệch chuẩn	Sai số chuẩn	95% khoảng tin cậy cho giá trị trung bình	
					Giới hạn dưới	*Giới hạn trên*
Sự trôi chảy	IBM Watson TTS	3,2357	0,79273	0,09475	3,0467	3,4247
	Microsoft Speech	3,5214	0,87822	0,10497	3,3120	3,7308
	Google Cloud TTS	**3,6429**	0,82595	0,09872	3,4459	3,8398
	Mac TTS	3,4929	0,80978	0,09679	3,2998	3,6859
	Natural TTS	**3,2000**	0,85719	0,10245	2,9956	3,4044
	VocalID's	3,3357	0,95826	0,11453	3,1072	3,5642
	LOVO Studio	3,6143	0,79933	0,09554	3,4237	3,8049
	Respeecher	3,2643	0,77423	0,09254	3,0797	3,4489
	Woord TTS	3,2643	0,85855	0,10262	3,0596	3,4690
	Amazon Polly	3,4357	0,92047	0,11002	3,2162	3,6552
	ReadSpeaker	3,5500	0,77155	0,09222	3,3660	3,7340
	Nuance TTS	3,2429	0,80191	0,09585	3,0516	3,4341
	Acapela VaaS	3,2214	0,85819	0,10257	3,0168	3,4261
	Resemble Speech	3,2643	0,92361	0,11039	3,0441	3,4845
	Murf TTS	3,3000	0,78204	0,09347	3,1135	3,4865

5 Lộ trình tích hợp hệ thống TTS AI vào mô hình đại học thông minh

Nghiên cứu này áp dụng phương pháp *tiếp cận và phân tích sáng tạo có hệ thống* [41] kết hợp với chính sách của Chính phủ Việt Nam để đề xuất một mô hình tổng thể đại học thông minh bao quát các lĩnh vực: Đại học thông minh, Giáo dục thông minh, Môi trường học tập thông minh, Khuôn viên thông minh, Giáo viên thông minh, Lớp học thông minh, trong đó, Lớp học thông minh sẽ là yếu tố cơ bản nhất trong quá trình xây dựng và triển khai đại học thông minh.

Một lớp học thông minh liên quan đến việc tối ưu hóa việc trình bày nội dung giảng dạy, truy cập thuận tiện các tài nguyên học tập, tính tương tác sâu sắc của việc dạy và học, nhận thức

và phát hiện theo ngữ cảnh, bố trí và quản lý lớp học dựa trên các công nghệ thông minh. Các lớp học thông minh tích hợp nhận dạng giọng nói, thị giác máy tính và các công nghệ khác, được gọi chung là các tác nhân thông minh, để cung cấp trải nghiệm giáo dục từ xa tương tự như trải nghiệm lớp học truyền thống.

Để liên kết các lĩnh vực trên, mô hình sẽ sử dụng các công nghệ tiên tiến dựa trên nền tảng của Internet, IoT, Cloud, thực tế ảo cũng như các cảm biến, công nghệ nhận dạng, nhận biết. Kết hợp với đó là hệ thống phần mềm hỗ trợ các hoạt động học tập, giảng dạy, ghi âm, ghi hình, nhận dạng, nhận biết, giám sát, phân tích cùng các công cụ phân tích, dự báo, suy luận nhằm đưa ra những đánh giá, nhận định theo thời gian thực, hỗ trợ các hoạt động dạy và học, quản lý và điều hành.

Với việc nghiên cứu, tìm hiểu về mô hình đại học thông minh trong những năm gần đây đã chứng tỏ các trường đại học Việt Nam đã có sự chuẩn bị cũng như định hướng chiến lược cho sự phát triển của giáo dục thông minh, đại học thông minh trong những năm tới. Để hiện thực hoá mô hình đề xuất, nghiên cứu đã đề ra quy trình chuyển đổi từ đại học truyền thông sang đại học thông minh, tập trung vào ba khía cạnh: quản trị, nội dung chương trình và phương thức đào tạo. Những khía cạnh này sẽ được cụ thể hóa bằng việc ứng dụng các hệ thống quản trị tiên tiến đã thành công trong các doanh nghiệp lớn, cùng với đó là việc xây dựng lớp học thông minh kiểu mẫu kết hợp với việc thay đổi nội dung chương trình cũng như phương thức đào tạo, từ đó tiến hành nhân rộng mô hình lớp học thông minh ra phạm vi toàn trường.

Chuyển đổi mô hình quản trị đại học

Hiện nay, nhiều nghiên cứu lý thuyết cũng như thực tiễn và thực tế về quản trị đại học đều chỉ ra rằng, xu hướng trường đại học hoạt động như một doanh nghiệp để đảm bảo hiệu quả đầu tư kết hợp với hương vị "cận thị trường" nhằm thích ứng với nền kinh tế thị trường nhưng tránh thương mại hóa dưới sự hỗ trợ, giám sát và điều tiết của nhà nước là mô hình hoạt động tối ưu nhất trên thế giới hiện nay. Do có nhiều điểm tương đồng giữa các trường đại học và các doanh nghiệp, các trường đại học cũng sẽ phải đối mặt với nhiều vấn đề rất phổ biến với hầu hết các tổ chức hiện đại.

Để đối phó với những vấn đề trên, một xu hướng nổi bật trong những năm gần đây là giáo dục đại học chuyển sang áp dụng hệ thống Hoạch định tài nguyên doanh nghiệp ERP, với hy vọng thích ứng với những thay đổi của môi trường đầy cạnh tranh. Từ sự thành công của việc triển khai ERP vào trường đại học ở Mỹ, Úc, Anh, hàng loạt các trường đại học ở các nước châu Á đã nghiên cứu ứng dụng mô hình này vào công tác quản lý như Trung Quốc, Ấn Độ, Hàn Quốc, Nhật Bản.

Sau khi triển khai tương đối đầy đủ các quy trình hỗ trợ cũng như đảm bảo việc lập kế hoạch và quản lý các nguồn lực dựa trên các hệ thống thông tin tích hợp như ERP, xóa bỏ rào cản về mặt quản lý tổ chức, các trường đại học sẽ có cơ sở vững chắc để tiến hành các hoạt động chuyển đổi số khác cũng như triển khai việc cải tiến, đổi mới nội dung chương trình đào tạo cũng như phương thức đào tạo. Các đối tượng trong trường đại học như giảng viên và người học sẽ

đồng hành với những tiến bộ khoa học kỹ thuật, sử dụng các sản phẩm phần cứng, phần mềm hỗ trợ cho hoạt động dạy và học, tạo tiền đề thích nghi với những phương thức đào tạo mới.

Chuyển đổi phương thức đào tạo

Từ những thay đổi về mặt nội dung chương trình đào tạo, một vấn đề khác đặt ra cho các cơ sở đào tạo bậc cao là cách thức tổ chức để chuyển tải nội dung chương trình đào tạo đến người học. CMCN 4.0 đòi hỏi phương thức và phương pháp đào tạo thay đổi với sự ứng dụng mạnh mẽ của công nghệ thông tin, công nghệ kỹ thuật số và hệ thống mạng. Các hình thức đào tạo online, đào tạo ảo, mô phỏng, số hóa bài giảng sẽ là xu hướng đào tạo nghề nghiệp trong tương lai. Điều này đòi hỏi các cơ sở đào tạo phải có sự chuẩn bị tốt nguồn lực tổ chức giảng dạy, đặc biệt là đội ngũ giảng viên, xây dựng không gian học tập, trang thiết bị phục vụ cho việc dạy và học.

Để chuyển tải những nội dung phù hợp với xã hội thông minh như hiện nay, ĐHTM cần ứng dụng các hệ thống thực - ảo (Cyber Physical System - CPS) và IoT. Các hệ CPS là đặc trưng tiêu biểu của môi trường công nghiệp 4.0, là cơ sở để thiết kế và xây dựng các mô hình nhà máy thông minh. CPS thường được thiết kế với cấu trúc 5C (Connection - kết nối thông tin, Conversion - chuyển đổi thông tin, Cyber - phân tích, Cognition - nhận diện và Configuration - cấu hình hóa).

Bên cạnh đó, các trang thiết bị, công nghệ hỗ trợ, nhận dạng cũng được ứng dụng vào trong lớp học thông minh, đảm bảo việc chuyển tải nội dung đào tạo đến với nhiều đối tượng khác nhau ở các phạm vi địa lý khác nhau, bao gồm cả sinh viên khiếm thị. Các hệ thống TTS AI, nhận dạng giọng nói, nhận dạng khuôn mặt, cử chỉ, thiết bị trỏ thông minh và bảng để điều hướng, chỉnh sửa và hiển thị thông tin trên bảng thông minh cùng các thiết bị, công nghệ, kỹ thuật khác sẽ cung cấp cho sinh viên những tiện ích và sự hỗ trợ ngay tại phòng học cũng như các sinh viên từ xa được trải nghiệm giống như học tập trực tiếp.

Việc trang bị các thiết bị, công nghệ, kỹ thuật trong lớp học thông minh sẽ được tiến hành thí điểm từng lĩnh vực, phụ thuộc và khả năng đầu tư của từng trường cũng như nhu cầu của người học đối với từng lĩnh vực cụ thể, từng ngành nghề đào tạo trọng tâm của nhà trường.

Với việc có sẵn cơ sở hạ tầng công nghệ thông tin, hệ thống thông tin và truyền thông trong bước chuyển đổi mô hình quản trị, việc gắn kết và tích hợp các hệ thống mới phục vụ phương thức đào tạo mới sẽ dễ dàng thực hiện được. Những người trực tiếp tham gia vào phương thức đào tạo mới là giảng viên và người học cần có sự chuẩn bị, làm quen với các hệ thống hỗ trợ đã được triển khai trước đó, giảm thời gian thích nghi, tăng hiệu quả và chất lượng đào tạo.

Chuyển đổi nội dung chương trình đào tạo

Để đáp ứng nhu cầu xã hội và các bên liên quan, bản thân các trường đại học cần phải thay đổi, màtrước tiên là thay đổi về nội dung chương trình đào tạo. Các trường đại học đã có những điều chỉnh trong nội dung chương trình, đi theo xu hướng tiếp cận liên ngành để giải quyết các vấn đề thuộc nhiều lĩnh vực khác nhau bằng công nghệ. Các trường đại học sẽ phải thực hiện

hoạt động đào tạo theo hai hướng: một mặt phải đáp ứng tính định hướng xã hội, mặt khác đào tạo cung cấp nguồn nhân lực đáp ứng yêu cầu của thị trường lao động.

Tuy nhiên, áp lực đối với các trường đại học càng lớn khi chương trình đào tạo vừa đáp ứng tính chuyên môn cao trong lĩnh vực nhất định, vừa đáp ứng tính liên ngành (công nghệ thông tin, kỹ thuật số, mạng, kiến thức chuyên ngành) và các kỹ năng khác không thể thiếu, như khả năng suy nghĩ có hệ thống, khả năng tổng hợp, khả năng liên kết giữa thế giới thực và ảo, khả năng sáng tạo, kỹ năng làm việc nhóm, khả năng hợp tác liên ngành. Như vậy, CMCN 4.0 đã tạo áp lực lớn trong hoạt động đào tạo đối với các trường đại học, từ xây dựng chương trình đào tạo, cập nhật nội dung chương trình cho đến đào tạo kỹ năng cho người học để đáp ứng yêu cầu công nghiệp hóa.

Trên cơ sở đã chuyển đổi mô hình quản trị đại học ở bước trước, các trường đại học sẽ có điều kiện thuận lợi trong việc nắm bắt được những thay đổi về khoa học, kỹ thuật, công nghệ, từ đó dễ dàng tiếp cận với việc xây dựng những chương trình đào tạo có khả năng đáp ứng được những yêu cầu của xã hội cũng như cung cấp nguồn nhân lực có chất lượng cao cho thị trường lao động.

6 Kết luận

Nghiên cứu về mô hình đại học thông minh và các hệ thống TTS AI cho thấy sự tiến bộ trong các hệ thống trí tuệ nhân tạo chuyển văn bản thành giọng nói với rất nhiều giải pháp cho lĩnh vực này cũng như chất lượng của các giải pháp là khá tương đồng và được người sử dụng đánh giá cao. Cùng với đó là việc chuyển đổi, triển khai mô hình đại học thông minh đã và đang diễn ra ở nhiều nước tiên tiến sẽ tạo điều kiện cho việc chuyển đổi sang mô hình đại học thông minh của các trường đại học Việt Nam. Do đó, các trường đại học cần có sự chuẩn bị cũng như định hướng chiến lược cho sự phát triển của giáo dục thông minh, đại học thông minh trong những năm tới.

Việc chuyển đổi sang mô hình đại học thông minh có thể thí điểm ở mô hình lớp học thông minh thông qua việc trang bị các công nghệ, thiết bị, kỹ thuật tiên tiến, đáp ứng nhu cầu đào tạo của nhà trường cũng như của người học nhằm đáp ứng yêu cầu của xã hội. Hệ thống TTS AI được ứng dụng trong giai đoạn này và hình thức tích hợp phù hợp nhất là thông qua tích hợp API vì hầu như tất cả các hệ thống TTS AI trên thị trường hiện nay đều sử dụng hình thức này. Vấn đề còn lại là lựa chọn hệ thống nào sẽ phụ thuộc vào yếu tố chi phí, khả năng đầu tư của từng trường đại học. Mở rộng ra, để có lộ trình chuyển đổi, triển khai đại học thông minh có sự tối ưu chi phí, các trường đại học cần có chiến lược hợp tác với các công ty công nghệ để lựa chọn mô hình phù hợp với khả năng tài chính và nhu cầu thực tiễn của mình.

Việc chuyển đổi này đem lại cho người khuyết tật Việt Nam nói chung và sinh viên khiếm thị nói riêng những cơ hội học tập dựa trên áp dụng các công nghệ, phần cứng, phần mềm phù hợp. Bước tiếp theo của nghiên cứu này sẽ chú trọng vào chi tiết các giai đoạn cũng như các

phương pháp thực hiện quy trình chuyển đổi với mục tiêu đạt hiệu quả cao, đem lại nhiều cơ hội học tập hơn nữa cho người khuyết tật Việt Nam.

Tài liệu tham khảo

1. Pivik, J., McComas, J. and Laflamme, M. (2002), Barriers and facilitators to inclusive education, *Exceptional children*, 69(1), 97–107.
2. Bakken, J. P. et al. (2018), Smart university: software systems for students with disabilities, *Smart Universities: Concepts, Systems, and Technologies 4*, 87–128.
3. Bakken, J. P., Varidireddy, N. and Uskov, V. L. (2019), Analysis and classification of university centers for students with disabilities, Smart Education and e-Learning 2019, *Springer*, 445–459.
4. Cambre, J. et al. (2020), Choice of voices: A large-scale evaluation of text-to-speech voice quality for long-form content, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–13.
5. Dai, L. et al. (2022), Evaluating the usage of Text-To-Speech in K12 education, *Proceedings of the 2022 6th International Conference on Education and E-Learning*, 182–188.
6. Bakken, J. P. et al. (2019), The quality of text-to-voice and voice-to-text software systems for smart universities: perceptions of college students with disabilities, *Smart Education and e-Learning 2018 5*, *Springer*, 51–66.
7. Alper, A. E. and Alper, F. Ö. (2020), Industry 4.0 revolution and its impacts on labor market, *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 29(3), 441–460.
8. Vu, T. L. A. and Le, T. Q. (2019), Development orientation for higher education training programme of mechanical engineering in industrial revolution 4. 0: a perspective in vietnam, *J. Mech. Eng. Res. Dev*, 42(1), 68–70.
9. Lukita, C. et al. (2020), Curriculum 4.0: adoption of industry era 4.0 as assessment of higher education quality, *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(3), 297–308.
10. Sima, V. et al. (2020), Influences of the industry 4.0 revolution on the human capital development and consumer behavior: A systematic review, *Sustainability*, 12(10), 4035.
11. Sitepu, E. S., Rangkuti, A. E. and Fachrizal, F. (2020), Analysis of the competency of fresh graduated higher education in supporting industrial era 4.0, *IJIET (International Journal of Indonesian Education and Teaching)*, 4(1), 82–101.
12. Law, M. Y. (2022), A Review of Curriculum Change and Innovation for Higher Education, *Journal of Education and Training Studies*, 10(2), 16.
13. Prasolova-Førland, E. et al. (2018), *Practicing interprofessional team communication and collaboration in a smart virtual university hospital*, Smart Universities: Concepts, Systems, and Technologies 4, 191–224.

14. Shahroom, A. A. and Hussin, N. (2018), Industrial revolution 4.0 and education, *International Journal of Academic Research in Business and Social Sciences*, 8(9), 314–319.
15. Brand, B. S. et al. (2022), Sapientia: a Smart Campus model to promote device and application flexibility, *Advances in Computational Intelligence*, 2(1), 18.
16. Lai, C., Chundra, U. and Lee, M. (2020), Teaching and learning based on IR 4.0: Readiness of attitude among polytechnics lecturers, *Journal of Physics: Conference Series*, IOP Publishing, 032105.
17. Romero-Rodríguez, J. -M. et al. (2020), Mobile learning in higher education: Structural equation model for good teaching practices, *Ieee Access*, 8, 91761–91769.
18. Alsharhan, E. and Ramsay, A. (2019), Improved Arabic speech recognition system through the automatic generation of fine-grained phonetic transcriptions, *Information Processing & Management*, 56(2), 343–353.
19. Barkana, B. D. and Patel, A. (2020), Analysis of vowel production in Mandarin/Hindi/American-accented English for accent recognition systems, *Applied Acoustics*, 162, 107203.
20. Bhuyan, M. and Sarma, S. (2019), A Higher-Order N-gram Model to enhance automatic Word Prediction for Assamese sentences containing ambiguous Words, *International Journal of Engineering and Advanced Technology*, 8(6), 2921–2926.
21. Bhuyan, M., Sarma, S. and Rahman, M. (2020), Natural language processing based stochastic model for the correctness of assamese sentences, International Conference on Communication and Electronics Systems (ICCES), *IEEE*, 1179–1182.
22. Yang, J. et al. (2014), Deep learning theory and its application in speech recognition, *Commun. Countermeas*, 33, 1–5.
23. Oord, A. van den et al. (2016), Wavenet: A generative model for raw audio, *arXiv preprint arXiv:1609.03499* [Preprint].
24. Van den Oord, A. et al. (2016), *Conditional image generation with pixelcnn decoders*, Advances in neural information processing systems, 29.
25. Van Den Oord, A., Kalchbrenner, N. and Kavukcuoglu, K. (2016), *Pixel recurrent neural networks*, International conference on machine learning, PMLR, 1747–1756.
26. Oord, A. et al. (2018), *Parallel wavenet: Fast high-fidelity speech synthesis*, International conference on machine learning, PMLR, 3918–3926.
27. Arık, S. Ö. et al. (2017), *Deep voice: Real-time neural text-to-speech*, International conference on machine learning, PMLR, 195–204.
28. Wang, Y. et al. (2017), Tacotron: Towards end-to-end speech synthesis', *arXiv preprint arXiv:1703.10135* [Preprint].
29. Shen, J. et al. (2018), Natural tts synthesis by conditioning wavenet on mel spectrogram predictions, IEEE international conference on acoustics, speech and signal processing (ICASSP), *IEEE*, 4779–4783.

30. Yasuda, Y. et al. (2019), Investigation of enhanced Tacotron text-to-speech synthesis systems with self-attention for pitch accent language, ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), *IEEE*, 6905–6909.
31. Tachibana, H., Uenoyama, K. and Aihara, S. (2018), Efficiently trainable text-to-speech system based on deep convolutional networks with guided attention, IEEE international conference on acoustics, speech and signal processing (ICASSP), *IEEE*, 4784–4788.
32. Ping, W. et al. (2017), Deep voice 3: Scaling text-to-speech with convolutional sequence learning, *arXiv preprint arXiv:1710.07654* [Preprint].
33. Khanam, F. et al. (2022), Text to Speech Synthesis: A Systematic Review, Deep Learning Based Architecture and Future Research Direction, *Journal of Advances in Information Technology*, 13(5).
34. Kumar, Y., Koul, A. and Singh, C. (2023), A deep learning approaches in text-to-speech system: A systematic review and recent research perspective, *Multimedia Tools and Applications*, 82(10), 15171–15197.
35. Uskov, V. L. et al. (2018), *Smart university: conceptual modeling and systems' design*, Smart Universities: Concepts, Systems, and Technologies 4, 49–86.
36. Kato, T., Kambayashi, Y. and Kodama, Y. (2018), *Using a Programming Exercise Support System as a Smart Educational Technology*, Smart Universities: Concepts, Systems, and Technologies 4, 295–324.
37. Villegas-Ch, W., Palacios-Pacheco, X. and Luján-Mora, S. (2019), Application of a smart city model to a traditional university campus with a big data architecture: A sustainable smart campus, *Sustainability*, 11(10), 2857.
38. Jurva, R. et al. (2020), Architecture and operational model for smart campus digital infrastructure, *Wireless Personal Communications*, 113, 1437–1454.
39. Utami, R. et al. (2019), *Teacher Professional Development in Education 4.0: Awareness of Digital Literacy*, in Proceedings of the 1st International Conference on Business, Law And Pedagogy, ICBLP 2019, 13–15 February 2019, Sidoarjo, Indonesia.
40. Hayudiyani, M. and Arifin, I. (2020), Reorientation of Curriculum in the Face of Industrial Revolution 4.0, in. 1st International Conference on Information Technology and Education (ICITE 2020), *Atlantis Press*, 659–664.
41. Heinemann, C. and Uskov, V. L. (2018), Smart university: literature review and creative analysis, *Smart Universities: Concepts, Systems, and Technologies* 4, 11–46.
42. Tan, X. et al. (2021), A survey on neural speech synthesis, *arXiv preprint arXiv:2106.15561* [Preprint].
43. Gundle, P. and Chavan, R. (2019), Survey on Text to Speech Synthesis Models and Methods, *International Journal of Scientific & Engineering Research*, 10(7).
44. Alonso Martin, F. et al. (2020), Four-features evaluation of text to speech systems for three social robots, *Electronics*, 9(2), 267.