HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



CHU ĐỨC CHÍNH

NGHIÊN CỨU PHÁT TRIỂN HỆ THỐNG NGƯỜI ĐẠI DIỆN ẢO TRONG HOẠT ĐỘNG TƯ VẤN LỄ TÂN, CUNG CẤP THÔNG TIN CHO SINH VIÊN HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(*Theo định hướng ứng dụng*)

HÀ NỘI - NĂM 2025

HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG



CHU ĐỨC CHÍNH

NGHIÊN CỨU PHÁT TRIỂN HỆ THỐNG NGƯỜI ĐẠI DIỆN ẢO TRONG HOẠT ĐỘNG TƯ VẤN LỄ TÂN, CUNG CẤP THÔNG TIN CHO SINH VIÊN HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

Chuyên ngành: Hệ thống thông tin

Mã số: 8.48.01.04

ĐỀ ÁN TỐT NGHIỆP THẠC SĨ KỸ THUẬT

(*Theo định hướng ứng dụng*)

NGƯỜI HƯỚNG DẪN KHOA HỌC: TS. TRẦN TIẾN CÔNG

HÀ NỘI - NĂM 2025

LỜI CAM ĐOAN

Tôi cam đoan đây là công trình nghiên cứu của riêng tôi.

Các số liệu, kiếu quả nêu trong đề án tốt nghiệp là trung thực và chưa từng được ai công bố trong bất kỳ công trình nào khác.

Tác giả đề án tốt nghiệp

Chu Đức Chính

LỜI CẢM ƠN

Lời đầu tiên, cho tôi xin phép gửi lời cảm ơn chân thành nhất tới các thầy, cô trong khoa Sau đại học nói riêng cũng như toàn thể các thầy cô trong Học viện Công nghệ Bưu chính Viễn thông đã dành trọn tâm huyết để truyền đạt các kiến thức hữu ích, giúp tôi có một môi trường học tập và kiến thức vững chắc.

Đặc biệt tôi xin gửi lời cảm ơn sâu sắc nhất tới TS. Trần Tiến Công. Một người thầy luôn tận tâm và nhiệt huyết vô tận với việc truyền đạt các kiến thức và các nghiên cứu mới cho nhiều thế hệ sinh viên. Những hướng dẫn, góp ý cùng lời khuyên của thầy cho em đã giúp tôi rất nhiều trong suốt quá trình từ lựa chọn đề tài đến việc thực hiện đề án. Giúp tôi mở mang thêm được nhiều kiến thức chuyên sâu cũng như đam mê trong lĩnh vực nghiên cứu khoa học.

Trong quá trình thực hiện đề án, tuy bản thân tôi đã cố gắng hết sức trong việc tìm các tài liệu và phương pháp giải quyết bài toán. Tuy nhiên, do kiến thức và năng lực còn nhiều hạn chế và thiếu sót. Tôi kính mong nhận được những lời góp ý từ các thầy cô để có thể hoàn thiện đề án một cách tốt nhất.

Tôi xin chân thành cảm ơn!

Hà Nội, tháng 12 năm 2025

Học viên

Chu Đức Chính

MỤC LỤC

LỜI CAM ĐOAN ..... **i**
LỜI CẢM ƠN ..... **ii**
DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT ..... **v**
DANH MỤC BẢNG BIỂU ..... **vi**
DANH MỤC HÌNH VẼ ..... **vii**
MỞ ĐẦU **1**
Chương 1. CƠ SỞ LÝ THUYẾT VÀ TỔNG QUAN **3**
1.1. Tổng quan về hệ thống người đại diện ảo **3**
1.1.1. Khái niệm và định nghĩa người đại diện ảo (Virtual Agent) 3
1.1.2. Phân loại và đặc điểm của các loại người đại diện ảo 4
1.2. Cơ sở lý thuyết về trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên **7**
1.2.1. Tổng quan trí tuệ nhân tạo (AI) và học máy (Machine Learning) 7
1.2.2. Xử lý ngôn ngữ tự nhiên 9
1.2.3. Công nghệ xử lý tiếng việt 11
1.2.4. Lý thuyết tương tác người – máy và trải nghiệm người dùng 12
1.3. Ứng dụng trong hoạt động tư vấn lễ tân **20**
1.3.1. Tổng quan hệ thống 20
1.3.2. Công nghệ và kiến trúc hệ thống 20
1.3.3. Tính năng và khả năng của hệ thống 22
1.4. Kết chương **22**
Chương 2. PHÂN TÍCH THIẾT KẾ VÀ XÂY DỰNG HỆ THỐNG **24**
2.1. Phân tích yêu cầu hệ thống **24**
2.1.1. Phân tích các bên liên quan 24
2.1.2. Phân tích yêu cầu chức năng 25
2.1.3. Phân tích yêu cầu phi chức năng 26
2.2. Thiết kế kiến trúc hệ thống **28**
2.2.1. Kiến trúc tổng quan của hệ thống 28
2.2.2. Luồng chính trong hệ thống 29

2.2.3. Thiết kế chi tiết các mô-đun 32
2.3. Thiết kế thuật toán và mô hình AI 38
2.4. Cài đặt và phát triển hệ thống 39
2.5. Kết chương 39
Chương 3. THỬ NGHIỆM VÀ ĐÁNH GIÁ 41
3.1. Tập dữ liệu 41
3.2. Môi trường thử nghiệm 41
3.3. Xây dựng ứng dụng thử nghiệm 42
3.3.1. Giới thiệu hệ thống 42
3.3.2. Xây dựng hệ thống 42
3.4. Kết quả thử nghiệm và đánh giá 42
3.4.1. Thiết lập thử nghiệm 42
3.4.2. Giao diện hệ thống 44
3.5. Kết chương 46
KẾT LUẬN 47
TÀI LIỆU THAM KHẢO 48

DANH MỤC CÁC THUẬT NGỮ, CHỮ VIẾT TẮT

STT	Viết tắt	Tiếng Anh	Tiếng Việt
1	LLM	Large Language Model	
2	ASR	Automatic Speech Recognition	
3	TTS	Text-to-Speech	
4	AI	Artificial Intelligence	
5	UX	User experience	
6	UI	User interface	
7	TTNT		Trí tuệ nhân tạo
8	NLP		Xử lý ngôn ngữ tự nhiên
9	CTSV		Công tác sinh viên
10	SV		Sinh viên
11	GV		Giảng viên
12	HCI	Human-Computer Interaction	
13	ACM SIGCHI	Special Interest Group on Computer-Human Interaction	

DANH MỤC BẢNG BIỂU

Bảng 1.1: Infographic so sánh Virtual Agent với các hình thức hỗ trợ truyền thống 4
Bảng 1.2: Tech Stack Overview 21
Bảng 2.1: Ánh xạ cảm xúc 38
Bảng 3.1: Thông số môi trường thử nghiệm 41

DANH MỤC HÌNH VẼ

Hình 1.1: Sơ đồ minh họa cơ chế hoạt động của Virtual Agent 3
Hình 1.2: Minh họa 3 loại hình thức Virtual Agent 5
Hình 1.3: Biểu đồ kim tự tháp thể hiện 3 cấp độ Virtual Agent 5
Hình 1.4: Sơ đồ mindmap thể hiện 3 nhóm chức năng chính của Virtual Agent 6
Hình 1.5: Sơ đồ cây thể hiện các nhánh chính của Trí tuệ nhân tạo 7
Hình 1.6: Infographic so sánh 3 loại Machine Learning 8
Hình 1.7: Minh họa kiến trúc Neural Network và Transformer 9
Hình 1.8: Timeline tiến hóa của các mô hình ngôn ngữ 10
Hình 1.9: Mô hình tương tác người – máy trong Virtual Agent 13
Hình 1.10: 4 nguyên tắc Grice trong thiết kế hội thoại 14
Hình 1.11: Mô hình tổ ong UX (UX Honeycomb) của Peter Morville 15
Hình 1.12: Dashboard đo lường hiệu quả Virtual Agent 16
Hình 1.13: Responsive design của Virtual Agent trên các thiết bị 17
Hình 1.14: Các yếu tố xây dựng niềm tin trong Virtual Agent 19
Hình 1.15: Sơ đồ chung của hệ thống 20
Hình 2.1: Luồng xử lý cơ bản của hệ thống 24
Hình 2.2: Luồng khởi tạo và kết nối session 29
Hình 2.3: Luồng ghi âm ở client 29
Hình 2.4: Luồng gọi LLM 30
Hình 2.5: Luồng gửi audio tới client 30
Hình 2.6: Luồng lưu trữ hội thoại 31
Hình 2.7: Luồng xử lý lỗi 31
Hình 2.8: Luồng proactive speaking 32
Hình 3.1: Tập dữ liệu câu hỏi 41
Hình 3.2: Trang chủ 44
Hình 3.3: Khung chat 44
Hình 3.4: Thanh công cụ 45
Hình 3.5: Cài đặt 45
Hình 3.6: Thanh công cụ biểu cảm 46

MỞ ĐẦU

Trong bối cảnh công nghệ số phát triển mạnh mẽ và xu hướng chuyển đổi số diễn ra trên toàn cầu, việc ứng dụng trí tuệ nhân tạo (AI) vào các hoạt động dịch vụ khách hàng đã trở thành một xu thế tất yếu . Đặc biệt, hoạt động tư vấn lễ tân – một khâu quan trọng trong việc tạo ấn tượng đầu tiên với khách hàng – đang đối mặt với nhiều thách thức như: chi phí nhân sự cao, tính nhất quán trong chất lượng dịch vụ, khả năng phục vụ 24/7, và nhu cầu ngày càng tăng về trải nghiệm khách hàng cá nhân hóa.

Với phạm vi trong trường đại học như Học viện Công nghệ Bưu chính Viễn thông thì việc ứng dụng công nghệ số được phát triển rộng rãi, cùng với sự phát triển của công nghệ nhu cầu cung cấp thông tin cho sinh viên cũng là vấn đề cần chú ý tới làm sao có thể đáp ứng cung cấp thông tin được nhanh chóng và chính xác.

Hệ thống người đại diện ảo (Virtual Agent) với khả năng tương tác thông minh, xử lý ngôn ngữ tự nhiên và mô phỏng hành vi con người đang được coi là giải pháp hiệu quả để giải quyết những vấn đề trên. Việc nghiên cứu và phát triển một hệ thống như vậy không chỉ có ý nghĩa thực tiễn cao trong việc nâng cao chất lượng dịch vụ mà còn góp phần vào sự phát triển của lĩnh vực công nghệ AI tại Việt Nam.

Chính vì lý do đó tôi đã lựa chọn và nghiên cứu đề tài: "***Nghiên cứu phát triển hệ thống người đại diện ảo trong hoạt động tư vấn lễ tân, cung cấp thông tin cho sinh viên Học viện Công nghệ Bưu chính Viễn thông***". Nhằm đóng góp vào việc ứng dụng công nghệ tiên tiến để cải thiện trải nghiệm nhỏ hơn là với sinh viên Học viện Công nghệ Bưu chính Viễn thông và tương lại rộng hơn là với khách hàng và tối ưu hóa hoạt động dịch vụ.

Nội dung đề án được chia thành 3 chương chính như sau:

Chương 1: Cơ sở lý thuyết và tổng quan

Giới thiệu tổng quan về hệ thống người đại diện ảo, cơ sở lý thuyết về trí tuệ nhân tạo (AI), và xử lý ngôn ngữ tự nhiên và ứng dụng vào trong hoạt động lễ tân. Nội dung chính của chương này sẽ giới thiệu về các khái niệm, các kiến thức mới liên quan đến người đại diện ảo, cơ sở lý thuyết về TTNT và xử lý ngôn ngữ tự nhiên.

Ngoài ra, chương này cũng đưa ra phương thức ứng dụng trong hoạt động tư vấn lễ tân trong việc cung cấp thông tin tới sinh viên.

Chương 2: Phân tích thiết kế và xây dựng hệ thống

Nội dung chính của chương này sẽ tập trung vào việc phân tích, thiết kế và cài đặt triển khai hệ thống.

Chương 3: Thực nghiệm và đánh giá kết quả

Trong lĩnh vực trí tuệ nhân tạo nói chung và học sâu nói riêng, tập dữ liệu đóng vai trò quan trọng trong việc phát triển và đánh giá các thuật toán, ứng dụng. Việc phân tích và đánh giá chất lượng của tập dữ liệu cũng là một yếu tố quan trọng để đảm bảo tính đúng đắn và đáng tin cậy của các kết quả thu được. Trong chương 3 này, tôi sẽ trình bày chi tiết về thực nghiệm và đánh giá kết quả của hệ thống.

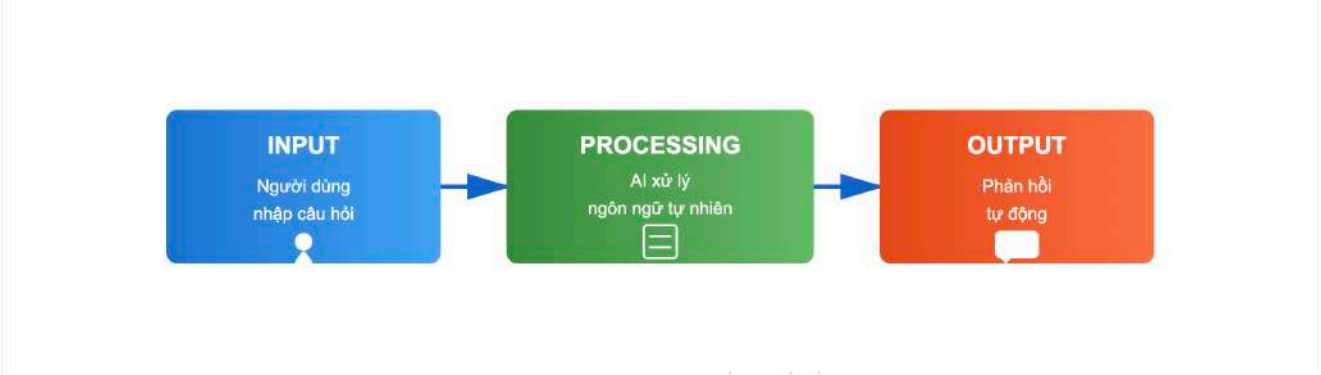
Chương 1. CƠ SỞ LÝ THUYẾT VÀ TỔNG QUAN

1.1. Tổng quan về hệ thống người đại diện ảo

1.1.1. Khái niệm và định nghĩa người đại diện ảo (Virtual Agent)

a) Khái niệm:

Người đại diện ảo (Virtual Agent) là giải pháp phần mềm dựa trên trí tuệ nhân tạo, cho phép tự động hóa quy trình tương tác với khách hàng thông qua giao diện đối thoại. Hệ thống được trang bị khả năng nhận diện và phân tích ngôn ngữ tự nhiên, giúp xử lý các yêu cầu và phản hồi người dùng một cách độc lập, giảm thiểu nhu cầu can thiệp thủ công.



Hình 1.1: Sơ đồ minh họa cơ chế hoạt động của Virtual Agent

b) Định nghĩa chi tiết

Về mặt kỹ thuật, Virtual Agent là hệ thống sử dụng AI có tích hợp công nghệ xử lý ngôn ngữ tự nhiên (NLP), học máy (Machine Learning) và đôi khi cả giọng nói tổng hợp, cho phép tự động hóa cuộc hội thoại qua văn bản hoặc giọng nói. Khả năng này giúp hệ thống hiểu ý định người dùng từ câu hỏi bằng ngôn ngữ tự nhiên, không bị giới hạn bởi cụm từ khóa cố định.

Về mặt chức năng, Virtual Agent đóng vai trò trợ lý số có khả năng trả lời câu hỏi, cung cấp thông tin, hướng dẫn thực hiện tác vụ, giải quyết khiếu nại và hỗ trợ 24/7. Điều này giảm tải công việc cho cán bộ và đảm bảo sự liên tục không gián đoạn.

Về mặt ứng dụng, Virtual Agent thường được triển khai trong dịch vụ cần tương tác thường xuyên như: khách hàng, bán hàng, hỗ trợ kỹ thuật. Hệ thống hoạt động đa nền tảng: website, mobile app, mạng xã hội hoặc điện thoại.

Bảng 1.1: Infographic so sánh Virtual Agent với các hình thức hỗ trợ truyền thống

Tiêu chí	Hỗ trợ trực tiếp	Chatbot đơn giản	Virtual Agent
Thời gian phản hồi	Phụ thuộc nhân viên	Tức thì	Tức thì
Chi phí vận hành	Cao (Lương + đào tạo)	Thấp	Trung bình
Khả năng mở rộng	Hạn chế	Trung bình	Cao
Độ chính xác	Cao (Con người)	Trung bình	Cao (AI)
Khả năng học hỏi	✓ (Kinh nghiệm)	×	✓ (Machine Learning)
Hoạt động 24/7	×	✓	✓

c) Đặc điểm phân biệt

Virtual Agent khác với chatbot đơn giản ở chỗ có khả năng học hỏi, cải thiện theo thời gian và xử lý các tình huống phức tạp hơn, không chỉ dựa vào kịch bản có sẵn mà còn có thể đưa ra quyết định dựa trên ngữ cảnh và lịch sử tương tác.

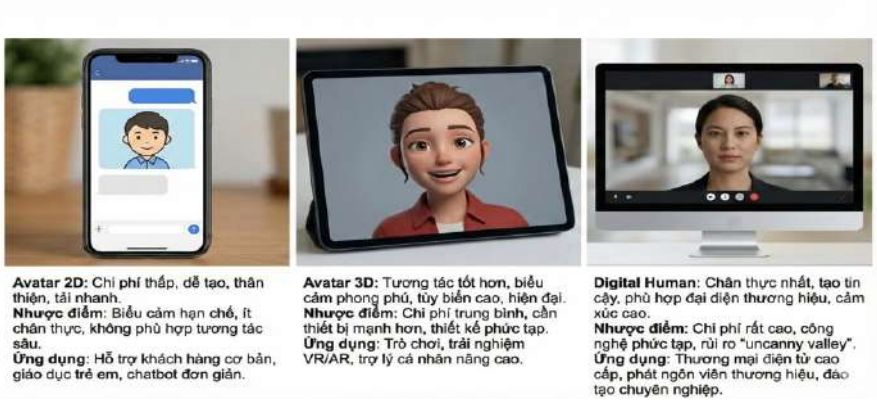
1.1.2. Phân loại và đặc điểm của các loại người đại diện ảo

a) Theo hình thức thể hiện

Avatar 2D là nhân vật hoạt hình phẳng với thiết kế đơn giản, chi phí phát triển thấp và dễ tích hợp. Phù hợp với website hoặc app mobile có cấu hình khiêm tốn.

Avatar 3D sử dụng mô hình ba chiều với chuyển động tự nhiên, yêu cầu nền tảng cao cấp (Unity, Unreal Engine). Tạo cảm giác tương tác chân thực, phù hợp với kiosk và màn hình tương tác lớn.

Digital Human sử dụng hình ảnh/video người thật kết hợp AI, tạo độ chân thực từ đó người dùng tin tưởng cao nhưng chi phí cao do yêu cầu công nghệ deepfake/AI video synthesis. Phù hợp với tư vấn quan trọng và giới thiệu chính thức.



Hình 1.2: Minh họa 3 loại hình thức Virtual Agent

b) Theo khả năng tương tác

Người đại diện tĩnh (Static Agent) chỉ hiển thị thông tin có sẵn, không xử lý ngôn ngữ tự nhiên. Phù hợp với hỏi đáp đơn giản và menu điều hướng.

Người đại diện phản ứng (Reactive Agent) phản hồi theo kịch bản lập trình sẵn, sử dụng keyword matching hoặc decision tree. Phù hợp với câu hỏi thường gặp (FAQ) với độ chính xác cao.

Người đại diện thông minh (Intelligent Agent) tích hợp AI/NLP và học máy, hiểu ngữ cảnh, xử lý ngôn ngữ tự nhiên linh hoạt và có khả năng học hỏi từ tương tác. Đây là lựa chọn tối ưu cho tư vấn phức tạp và cá nhân hóa cao.



Hình 1.3: Biểu đồ kim tự tháp thể hiện 3 cấp độ Virtual Agent

c) Theo chức năng chính

Agent tư vấn thông tin cung cấp thông tin về chương trình đào tạo, học phí, thủ tục và hướng dẫn dịch vụ. Đây là loại phổ biến nhất trong tổ chức giáo dục.

Agent hỗ trợ thủ tục chuyên biệt hóa trong hướng dẫn quy trình hành chính như đăng ký học phần, tra cứu kết quả, lịch thi. Giúp giảm tải đáng kể cho bộ phận hành chính và tăng tính chính xác.

Agent giải trí và định hướng tập trung vào giới thiệu hoạt động ngoại khóa, tư vấn nghề nghiệp và tour ảo cơ sở vật chất. Tạo trải nghiệm tích cực cho sinh viên mới và khách tham quan.



Hình 1.4: Sơ đồ mindmap thể hiện 3 nhóm chức năng chính của Virtual Agent

d) Đặc điểm cần có cho hệ thống tại học viện

Về công nghệ**:** Hệ thống cần hỗ trợ tiếng Việt tốt (xử lý giọng địa phương), tích hợp với cơ sở dữ liệu hiện có của Học viện để đảm bảo thống nhất thông tin, và hoạt động đa nền tảng (web, mobile, kiosk).

Về nội dung: Phải được cập nhật liên tục đảm bảo thông tin chính xác và kịp thời. Cần có cơ chế phân quyền truy cập thông tin phù hợp với từng đối tượng, đồng thời lưu trữ lịch sử tương tác để phân tích và cải thiện dịch vụ.

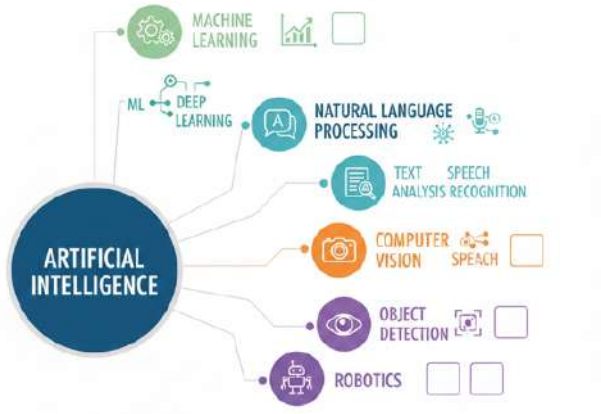
Về trải nghiệm: Giao diện thân thiện, dễ sử dụng phù hợp với sinh viên. Thời gian phản hồi nhanh để không gián đoạn trải nghiệm. Quan trọng nhất, hệ thống cần nhận biết giới hạn và chuyển đổi sang tư vấn viên người thật khi gặp tình huống phức tạp vượt khả năng xử lý tự động.

1.2. Cơ sở lý thuyết về trí tuệ nhân tạo và xử lý ngôn ngữ tự nhiên

1.2.1. Tổng quan trí tuệ nhân tạo (AI) và học máy (Machine Learning)

a) Khái niệm trí tuệ nhân tạo

Trí tuệ nhân tạo (Artificial Intelligence - AI) là khả năng của máy tính mô phỏng các quá trình tư duy và hành vi thông minh của con người. Theo ACM SIGCHI, AI là "môn học liên quan đến thiết kế, đánh giá và triển khai hệ thống máy tính tương tác dành cho con người sử dụng" [1]. Lĩnh vực này kết hợp kiến thức từ khoa học máy tính, tâm lý học nhận thức, thiết kế và nhiều ngành khác để tạo ra các hệ thống thông minh.AI bao gồm các nhánh chính: Machine Learning cho phép hệ thống học từ dữ liệu, Deep Learning sử dụng mạng nơ-ron nhiều lớp, Natural Language Processing tập trung xử lý ngôn ngữ con người, và Computer Vision giúp máy tính "nhìn" và hiểu hình ảnh. Trong Virtual Agent, AI đóng vai trò then chốt ở việc nhận dạng ý định người dùng, sinh phản hồi tự động và cá nhân hóa trải nghiệm dựa trên lịch sử tương tác [2].



Hình 1.5: Sơ đồ cây thể hiện các nhánh chính của Trí tuệ nhân tạo

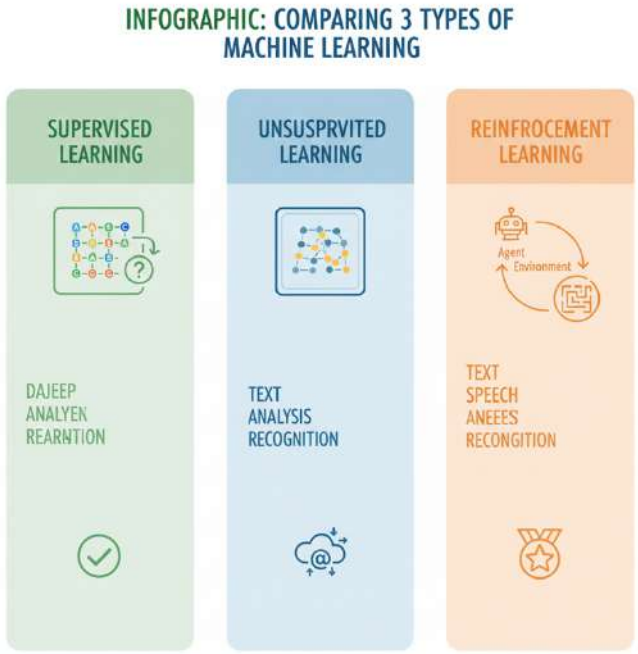
b) Machine Learning cơ bản

Machine Learning được phân thành ba loại chính theo phương pháp học. Supervised Learning (Học có giám sát) huấn luyện với dữ liệu đã gán nhãn, đặc biệt

hiệu quả cho phân loại câu hỏi và dự đoán ý định. Ví dụ: hệ thống học cách phân biệt câu hỏi về học phí, lịch thi hay thủ tục hành chính từ hàng ngàn mẫu đã phân loại.

Unsupervised Learning (Học không giám sát) làm việc với dữ liệu không có nhãn, hữu ích trong phát hiện pattern ẩn. Trong Virtual Agent, phương pháp này phân nhóm người dùng dựa trên hành vi tương tác, giúp hệ thống điều chỉnh phản hồi phù hợp với từng nhóm.

Reinforcement Learning (Học tăng cường) học thông qua thử và sai, dựa trên phản hồi từ môi trường. Virtual Agent tối ưu chiến lược hội thoại theo thời gian bằng cách học hành động nào dẫn đến kết quả tích cực và hành động nào cần tránh.



Hình 1.6: Infographic so sánh 3 loại Machine Learning

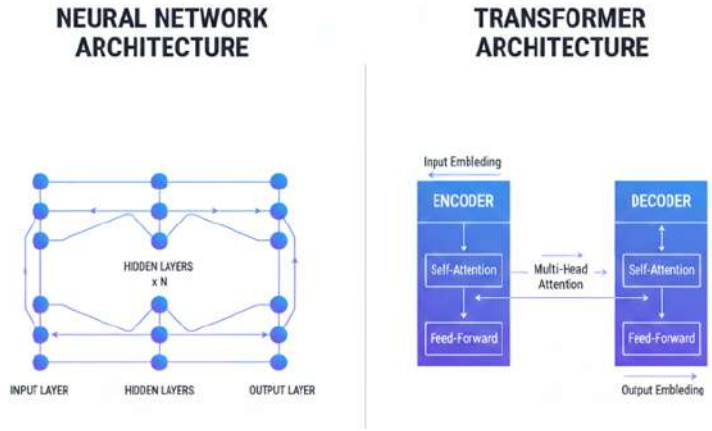
c) Deep Learning

Deep Learning ứng dụng mạng nơ-ron đa tầng với khả năng tự học đặc trưng từ dữ liệu, loại bỏ nhu cầu feature engineering thủ công của các mô hình ML truyền thống.

Transformer Architecture là nền tảng cho các mô hình ngôn ngữ hiện đại như BERT và GPT. Kiến trúc này vượt trội trong xử lý văn bản nhờ cơ chế "attention"

cho phép mô hình tập trung vào phần quan trọng khi đưa ra dự đoán, giúp hiểu ngữ cảnh và mối quan hệ giữa các từ hiệu quả hơn [3].

Mô hình huấn luyện sẵn (Pre-trained Models) đã được training trên corpus lớn, cho phép tùy chỉnh cho domain cụ thể với dữ liệu hạn chế. Cách tiếp cận transfer learning này tiết kiệm đáng kể chi phí và thời gian, phù hợp với ngân sách các tổ chức giáo dục khi xây dựng Virtual Agent.



Hình 1.7: Minh họa kiến trúc Neural Network và Transformer

1.2.2. Xử lý ngôn ngữ tự nhiên

a) Các tác vụ cơ bản

Tokenization là bước tách văn bản thành đơn vị nhỏ hơn (từ, cụm từ, ký tự), tạo nền tảng cho xử lý tiếp theo. Part-of-Speech Tagging gán nhãn từ loại (danh từ, động từ, tính từ) giúp hiểu cấu trúc câu. Named Entity Recognition (NER) nhận diện thực thể có tên như tên người, địa danh, tổ chức. Trong Virtual Agent của Học viện, NER giúp nhận diện tên ngành học, giảng viên, phòng ban. Dependency Parsing phân tích mối quan hệ phụ thuộc giữa các từ, giúp nắm bắt chính xác ý nghĩa câu phức tạp.

b) Các tác vụ nâng cao

Intent Recognition (Nhận diện ý định) xác định mục đích thực sự của người dùng - đang hỏi thông tin, yêu cầu hỗ trợ hay khiếu nại. Đây là bước quan trọng nhất quyết định cách xử lý tiếp theo.

Entity Extraction (Trích xuất thực thể) không chỉ nhận diện mà còn trích xuất thông tin cụ thể. Ví dụ từ "Em muốn đăng ký học phần Lập trình Java học kỳ 1", hệ thống trích xuất: môn học (Lập trình Java), hành động (đăng ký), thời gian (học kỳ 1).

Sentiment Analysis hiểu cảm xúc và thái độ người dùng qua văn bản, đặc biệt quan trọng để nhận biết khi người dùng không hài lòng và cần chuyển sang nhân viên thực.

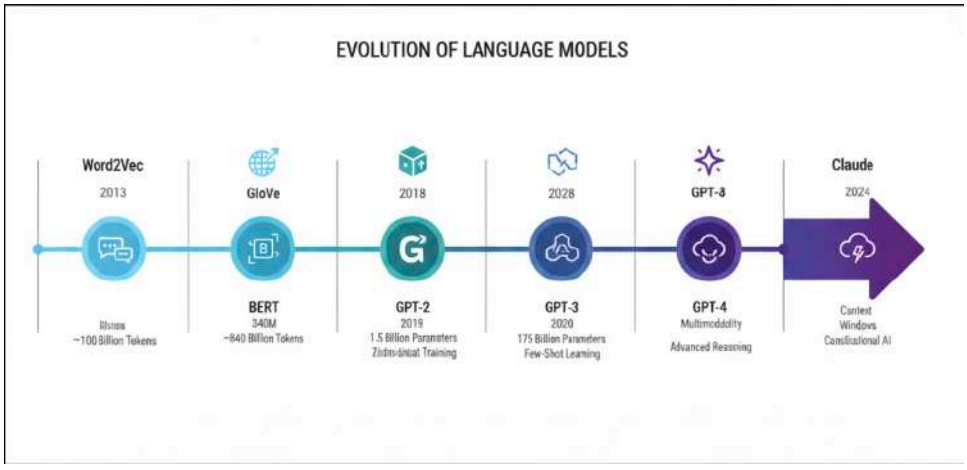
Text Generation tạo văn bản tự nhiên dựa trên ngữ cảnh, cho phép hệ thống sinh câu trả lời đa dạng thay vì chỉ dùng template cố định. Question Answering trả lời câu hỏi dựa trên cơ sở tri thức, yêu cầu hiểu và tổng hợp thông tin [4].

c) Mô hình ngôn ngữ hiện đại

Word Embeddings (Word2Vec, GloVe) chuyển từ thành vector số, các từ có nghĩa gần nhau có vector gần nhau. Ví dụ: vector "sinh viên" gần "học sinh" hơn là "giảng viên". Hạn chế: mỗi từ chỉ có một vector bất kể ngữ cảnh.

Contextual Embeddings (BERT, PhoBERT) tạo biểu diễn từ phụ thuộc ngữ cảnh. Cùng một từ có vector khác nhau trong ngữ cảnh khác nhau. BERT sử dụng Transformer và được huấn luyện theo cả hai hướng (bidirectional), hiểu toàn bộ ngữ cảnh xung quanh mỗi từ. PhoBERT là BERT cho tiếng Việt, mang lại hiệu suất vượt trội cho các tác vụ NLP tiếng Việt [5].

Large Language Models như GPT-3/GPT-4 với hàng trăm tỷ parameters được training trên corpus văn bản quy mô lớn. LLM tạo ra nội dung tự nhiên và mạch lạc, giúp Virtual Agent cải thiện đáng kể chất lượng hiểu ngữ cảnh và sinh phản hồi.



Hình 1.8: Timeline tiến hóa của các mô hình ngôn ngữ

1.2.3. Công nghệ xử lý tiếng việt

a) Đặc thù của tiếng Việt

Tiếng Việt là ngôn ngữ đơn lập, các từ không biến hình theo ngữ pháp. Từ tiếng Việt có thể gồm nhiều tiếng (từ ghép, từ láy), tạo ra vấn đề tách từ phức tạp: "học sinh" (1 từ) vs "học sinh học" (3 từ). Ranh giới từ không rõ ràng như tiếng Anh.

Dấu thanh điệu ảnh hưởng nghĩa: "ma, mà, má, mã, mả, mạ" - cùng âm cơ bản nhưng 6 nghĩa hoàn toàn khác. Điều này tạo thách thức khi xử lý văn bản thiếu dấu - phổ biến trong giao tiếp trực tuyến sinh viên.

Từ đồng âm khác nghĩa rất phổ biến: "bàn" (furniture/discuss), "đạo" (đạo lý/đạo Phật/đảo ngược). Phân biệt đúng đòi hỏi hiểu sâu về ngữ cảnh.

b) Công cụ xử lý tiếng Việt

VnCoreNLP là thư viện toàn diện cung cấp tách từ, gán nhãn từ loại, NER và phân tích cú pháp với độ chính xác cao, được cộng đồng quốc tế công nhận.

PhoBERT áp dụng kiến trúc BERT cho tiếng Việt, cho phép tạo biểu diễn ngữ cảnh và có thể fine-tune cho ứng dụng cụ thể với dữ liệu tương đối nhỏ [5].

VnTokenizer, PyVi là công cụ chuyên biệt cho tách từ. UndertheSea là thư viện đa năng với giao diện thân thiện, phù hợp triển khai nhanh.

c) Thách thức trong xử lý tiếng Việt

Thách thức: Đa dạng phương ngôn (Bắc-Trung-Nam), từ viết tắt và tiếng lóng sinh viên (mik, vs, dc, ko, ntn), thiếu dữ liệu chuyên ngành giáo dục CNTT, và lỗi chính tả/thiếu dấu thanh điệu.

Giải pháp:

Xây dựng corpus chuyên ngành với thuật ngữ giáo dục và CNTT đặc thù PTIT (tên môn học, quy định, thủ tục)

Fine-tune PhoBERT với dữ liệu hội thoại thực tế từ sinh viên, bao gồm cách diễn đạt không chuẩn

Từ điển domain-specific chứa từ viết tắt phổ biến (PTIT, CTSV, ĐHCQ) và thuật ngữ chuyên ngành

Module chuẩn hóa văn bản chuyển đổi từ viết tắt về dạng chuẩn, phục hồi dấu thanh điệu, sửa lỗi chính tả

d) Giải pháp cho bối cảnh học viện

Để xây dựng Virtual Agent hiệu quả cho Học viện, cần triển khai bốn giải pháp chính:

Xây dựng corpus chuyên ngành: Thu thập và xây dựng cơ sở dữ liệu chứa thuật ngữ giáo dục và CNTT đặc thù PTIT, bao gồm tên môn học, khoa ngành, chương trình đào tạo, quy định học vụ và tài liệu hướng dẫn. Corpus chuyên ngành giúp hệ thống hiểu chính xác các khái niệm mà sinh viên thường hỏi.

Fine-tune mô hình PhoBERT: Thu thập và gán nhãn câu hỏi thực tế từ sinh viên, bao gồm cả cách diễn đạt không chuẩn, từ viết tắt và lỗi chính tả. Dữ liệu này dùng để fine-tune PhoBERT, giúp mô hình thích nghi với cách giao tiếp đặc trưng của sinh viên PTIT.

Xây dựng từ điển domain-specific: Tạo từ điển chứa thuật ngữ chính thức và từ viết tắt phổ biến (PTIT, CTSV, ĐHCQ, SV, GV). Hệ thống cần module chuẩn hóa văn bản tự động chuyển đổi từ viết tắt và tiếng lóng về dạng chuẩn trước khi xử lý.

Quy trình cải thiện liên tục: Ghi lại toàn bộ tương tác, đặc biệt các trường hợp hệ thống không hiểu hoặc phản hồi sai. Phân tích định kỳ để xác định pattern lỗi, bổ sung vào corpus và tinh chỉnh mô hình, đảm bảo hệ thống ngày càng hiểu rõ hơn về sinh viên.

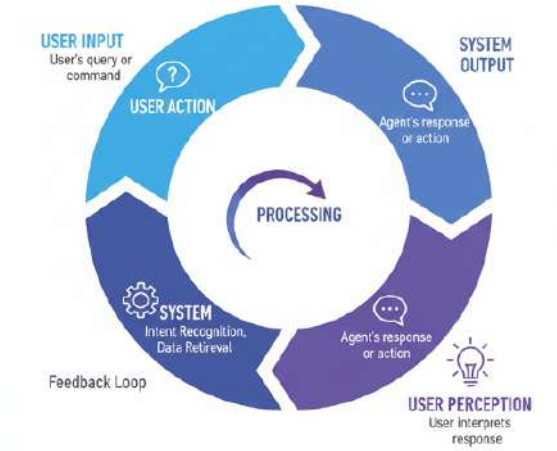
1.2.4. Lý thuyết tương tác người – máy và trải nghiệm người dùng

a) Khái niệm tương tác người – máy (Human Computer Interaction)

Tương tác người – máy là lĩnh vực nghiên cứu liên ngành tập trung vào thiết kế, đánh giá và triển khai các hệ thống máy tính tương tác để con người sử dụng. Theo định nghĩa của ACM SIGCHI, HCI là "một môn học liên quan đến việc thiết kế, đánh giá và triển khai các hệ thống máy tính tương tác dành cho con người sử dụng và nghiên cứu các hiện tượng chính xung quanh chúng" [6]. Lĩnh vực này kết

hợp kiến thức từ khoa học máy tính, tâm lý học nhận thức, thiết kế, nhân học và nhiều ngành khác để tạo ra các giao diện hiệu quả và dễ sử dụng.

Trong bối cảnh của Virtual Agent, HCI đặc biệt quan trọng vì hệ thống này đại diện cho một dạng tương tác tự nhiên hơn so với các giao diện truyền thống. Norman (2013) trong cuốn "The Design of Everyday Things" nhấn mạnh rằng thiết kế tốt phải tập trung vào người dùng, hiểu được mô hình tư duy của họ và tạo ra các tương tác phù hợp với kỳ vọng tự nhiên của con người [7]. Với Virtual Agent, mục tiêu là tạo ra trải nghiệm hội thoại gần giống với việc nói chuyện với một người thật, đồng thời vẫn đảm bảo hiệu quả và chính xác trong việc cung cấp thông tin.



Hình 1.9: Mô hình tương tác người – máy trong Virtual Agent

b) Các nguyên tắc thiết kế giao diện hội thoại

Thiết kế giao diện hội thoại (Conversational Interface Design) yêu cầu tuân thủ các nguyên tắc riêng biệt so với giao diện đồ họa truyền thống. Theo Nass và Brave (2005) trong nghiên cứu "Wired for Speech", con người có xu hướng tự nhiên áp dụng các quy tắc xã hội khi tương tác với máy tính như thể chúng là con người [8]. Hiện tượng này, được gọi là "Computers Are Social Actors" (CASA), có ý nghĩa quan trọng trong thiết kế Virtual Agent vì nó cho thấy người dùng kỳ vọng các tương tác tương tự như giao tiếp giữa người với người.

Grice (1975) đề xuất bốn nguyên tắc hợp tác trong giao tiếp (Cooperative Principle) mà các Virtual Agent nên tuân theo [9]: (1) Nguyên tắc số lượng - cung cấp đủ thông tin, không nhiều không ít; (2) Nguyên tắc chất lượng - thông tin phải

đúng sự thật và có căn cứ; (3) Nguyên tắc liên quan - chỉ nói những gì liên quan đến chủ đề; (4) Nguyên tắc cách thức - trình bày rõ ràng, tránh mơ hồ và lộn xộn. Việc tuân thủ các nguyên tắc này giúp Virtual Agent tạo ra các cuộc hội thoại tự nhiên và hiệu quả hơn.

Moore và Arar (2019) trong "Conversational AI" nhấn mạnh tầm quan trọng của việc thiết kế personality (tính cách) cho Virtual Agent [10]. Một agent có tính cách rõ ràng, phù hợp với thương hiệu tổ chức và nhóm người dùng mục tiêu sẽ tạo ra kết nối tốt hơn. Đối với Học viện, Virtual Agent nên được thiết kế với tính cách thân thiện, chuyên nghiệp, kiên nhẫn và hỗ trợ, phản ánh giá trị của một tổ chức giáo dục.



Hình 1.10: 4 nguyên tắc Grice trong thiết kế hội thoại

c) Trải nghiệm người dùng (User Experience - UX)

Trải nghiệm người dùng là một khái niệm toàn diện bao gồm tất cả các khía cạnh của tương tác giữa người dùng và sản phẩm, dịch vụ hoặc công ty. ISO 9241-210 định nghĩa UX như "nhận thức và phản ứng của người dùng khi sử dụng hoặc dự định sử dụng một sản phẩm, hệ thống hoặc dịch vụ" [11]. Hassenzahl và Tractinsky (2006) mở rộng khái niệm này bằng cách phân biệt giữa chất lượng thực dụng (pragmatic quality) - liên quan đến việc hoàn thành nhiệm vụ hiệu quả, và chất lượng khoái lạc (hedonic quality) - liên quan đến sự thỏa mãn, niềm vui và ý nghĩa cá nhân [12].

Nielsen (1993), một trong những chuyên gia hàng đầu về usability, đề xuất năm yếu tố cốt lõi của khả năng sử dụng [13]: (1) Learnability - hệ thống dễ học; (2)

Efficiency - người dùng thành thạo có thể đạt năng suất cao; (3) Memorability - dễ nhớ cách sử dụng sau thời gian không dùng; (4) Errors - tỷ lệ lỗi thấp và dễ phục hồi; (5) Satisfaction - người dùng cảm thấy hài lòng khi sử dụng. Các yếu tố này đặc biệt quan trọng cho Virtual Agent trong môi trường giáo dục, nơi người dùng có thể không sử dụng hệ thống thường xuyên nhưng cần có thể tìm được thông tin nhanh chóng khi cần.



Hình 1.11: Mô hình tổ ong UX (UX Honeycomb) của Peter Morville

Morville (2004) phát triển mô hình "UX Honeycomb" (Tổ ong UX) xác định bảy khía cạnh của trải nghiệm người dùng [14]: Useful (hữu ích), Usable (có thể sử dụng được), Desirable (mong muốn), Findable (có thể tìm thấy), Accessible (có thể truy cập), Credible (đáng tin cậy), và Valuable (có giá trị). Đối với Virtual Agent của Học viện, tất cả các khía cạnh này đều cần được chú ý: hệ thống phải thực sự hữu ích trong việc giải đáp thắc mắc sinh viên, dễ sử dụng ngay cả với người dùng lần đầu, có giao diện hấp dẫn, dễ tìm thấy trên website, truy cập được từ nhiều thiết bị, cung cấp thông tin chính xác đáng tin cậy, và mang lại giá trị thực sự cho cả sinh viên và Học viện.

d) Đo lường và đánh giá trải nghiệm người dùng

Việc đo lường hiệu quả của Virtual Agent đòi hỏi các phương pháp đánh giá phù hợp. Sauro và Lewis (2016) trong "Quantifying the User Experience" đề xuất nhiều thước đo khách quan và chủ quan để đánh giá UX [15]. Các thước đo khách

quan bao gồm: task completion rate (tỷ lệ hoàn thành nhiệm vụ), time on task (thời gian thực hiện nhiệm vụ), error rate (tỷ lệ lỗi), và number of steps (số bước cần thực hiện). Đối với Virtual Agent, các chỉ số này có thể được chuyển đổi thành: tỷ lệ câu hỏi được giải đáp thành công, thời gian trung bình để nhận được câu trả lời thỏa đáng, số lần người dùng phải đặt lại câu hỏi, và số lượt hội thoại trung bình để giải quyết một vấn đề.



Hình 1.12: Dashboard đo lường hiệu quả Virtual Agent

Các thước đo chủ quan tập trung vào cảm nhận của người dùng. System Usability Scale (SUS) được phát triển bởi Brooke (1996) là một bảng câu hỏi 10 mục đơn giản nhưng hiệu quả để đo lường khả năng sử dụng tổng thể [16]. Điểm SUS từ 0-100, với điểm trên 68 được coi là trên trung bình. Đối với Virtual Agent, có thể sử dụng các bảng câu hỏi đánh giá cụ thể hơn như Chatbot Usability Questionnaire (CUQ) do Holmes et al. (2019) phát triển, tập trung vào các khía cạnh đặc thù của giao diện hội thoại như naturalness (tính tự nhiên), understanding (khả năng hiểu), và helpfulness (tính hữu ích) [17].

Net Promoter Score (NPS) của Reichheld (2003) là một chỉ số đơn giản nhưng mạnh mẽ để đo lường sự hài lòng và lòng trung thành của người dùng thông qua câu hỏi [18]: "Trên thang điểm từ 0-10, bạn có khả năng giới thiệu hệ thống này cho bạn bè/đồng nghiệp đến mức nào?" Người trả lời 9-10 là promoters (người ủng hộ), 7-8

là passives (người thụ động), và 0-6 là detractors (người phê bình). NPS = % promoters - % detractors, với điểm số dương được coi là tốt.

e) Thiết kế đa kênh và đa thiết bị

Trong thời đại hiện nay, người dùng mong đợi có thể truy cập dịch vụ từ nhiều kênh và thiết bị khác nhau. Verhoef et al. (2015) định nghĩa omnichannel experience (trải nghiệm đa kênh) như "sự tích hợp liền mạch và nhất quán của các kênh khác nhau để người dùng có thể chuyển đổi tự do giữa các kênh trong cùng một hành trình" [19]. Đối với Virtual Agent của Học viện, điều này có nghĩa là sinh viên nên có thể bắt đầu một cuộc hội thoại trên website desktop, tiếp tục trên mobile app trong khi di chuyển, và có thể tương tác qua kiosk tại trường mà vẫn duy trì ngữ cảnh và lịch sử hội thoại.

Wroblewski (2011) trong "Mobile First" lập luận rằng việc thiết kế cho mobile trước (mobile-first design) buộc nhóm thiết kế phải ưu tiên các tính năng quan trọng nhất do hạn chế về không gian màn hình [20]. Cách tiếp cận này đặc biệt phù hợp với Virtual Agent vì giao diện hội thoại vốn dĩ đã tối giản, tập trung vào nội dung hơn là các yếu tố trang trí. Khi thiết kế cho mobile, cần chú ý đến kích thước nút bấm (tối thiểu 44x44 pixels theo khuyến nghị của Apple), khả năng đọc văn bản (font size tối thiểu 16px), và tối ưu hóa cho cả nhập liệu bằng giọng nói lẫn văn bản.



Hình 1.13: Responsive design của Virtual Agent trên các thiết bị

f) Tính năng tiếp cận (Accessibility) và thiết kế toàn diện (Inclusive Design)

Tính năng tiếp cận trong HCI đảm bảo rằng hệ thống có thể được sử dụng bởi tất cả mọi người, bao gồm cả những người khuyết tật. Web Content Accessibility

Guidelines (WCAG) 2.1 của W3C (2018) đưa ra bốn nguyên tắc chính [21]: Perceivable (có thể nhận thức được), Operable (có thể vận hành được), Understandable (có thể hiểu được), và Robust (bền vững). Đối với Virtual Agent, điều này có nghĩa là cần hỗ trợ screen readers cho người khiếm thị, cung cấp lựa chọn nhập liệu bằng giọng nói cho người khó gõ phím, đảm bảo độ tương phản màu sắc phù hợp cho người yếu thị lực, và hỗ trợ phím tắt bàn phím cho người không thể sử dụng chuột.

Microsoft (2016) trong "Inclusive Design Toolkit" mở rộng khái niệm accessibility sang inclusive design, không chỉ giải quyết các nhu cầu của người khuyết tật mà còn nhận ra rằng tất cả mọi người đều có thể gặp phải các tình huống khuyết tật tạm thời (như bị thương tay) hoặc khuyết tật tình huống (như ở trong môi trường ồn ào không thể nghe) [22]. Thiết kế toàn diện tạo ra các giải pháp linh hoạt mang lại lợi ích cho tất cả người dùng. Ví dụ, tính năng voice input ban đầu được thiết kế cho người khuyết tật nhưng giờ được nhiều người sử dụng vì tiện lợi.

g) Niềm tin và tính minh bạch trong AI

Khi làm việc với các hệ thống AI như Virtual Agent, vấn đề niềm tin (trust) trở nên đặc biệt quan trọng. Ribeiro et al. (2016) trong nghiên cứu về "Why Should I Trust You?" nhấn mạnh rằng người dùng cần hiểu tại sao hệ thống đưa ra quyết định hoặc câu trả lời cụ thể [23]. Explainable AI (XAI) là xu hướng ngày càng quan trọng, yêu cầu các hệ thống AI phải có khả năng giải thích logic đằng sau các quyết định của mình.

Đối với Virtual Agent, tính minh bạch có thể được thể hiện qua việc: (1) Rõ ràng ngay từ đầu rằng đây là hệ thống tự động, không phải người thật; (2) Cung cấp nguồn thông tin khi đưa ra câu trả lời (ví dụ: "Theo quy chế đào tạo năm 2024, điều 15..."); (3) Thừa nhận khi không chắc chắn về câu trả lời và đề xuất chuyển sang tư vấn viên người; (4) Giải thích tại sao yêu cầu một số thông tin cá nhân. Lee và See (2004) chỉ ra rằng appropriate trust (niềm tin phù hợp) - không quá tin cũng không

quá nghi ngờ - là mục tiêu thiết kế lý tưởng, và điều này đạt được thông qua tính minh bạch và khả năng dự đoán của hệ thống [24].



Hình 1.14: Các yếu tố xây dựng niềm tin trong Virtual Agent

h) Ứng dụng lý thuyết và thiết kế Virtual Agent cho Học viện

Việc áp dụng lý thuyết HCI và UX vào thiết kế Virtual Agent cần tuân theo quy trình có hệ thống. Dựa trên User-Centered Design process của ISO 9241-210, thiết kế nên trải qua bốn giai đoạn [6]: (1) Nghiên cứu ngữ cảnh và hành vi người dùng; (2) Xác định yêu cầu thông tin và dịch vụ sinh viên cần; (3) Phát triển prototype; (4) Thử nghiệm và cải thiện với người dùng thực.

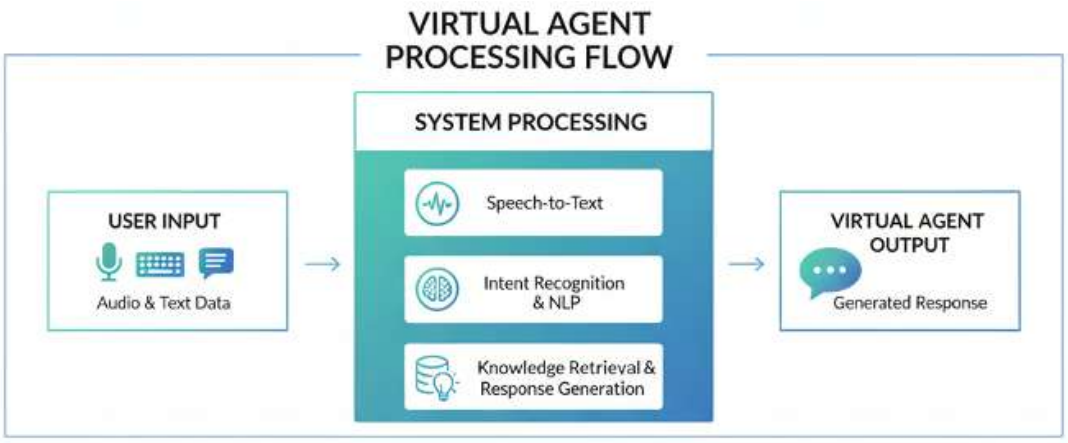
Garrett (2011) đề xuất năm lớp thiết kế từ trừu tượng đến cụ thể [25]: Strategy (mục tiêu và nhu cầu), Scope (tính năng và nội dung), Structure (luồng hội thoại), Skeleton (giao diện và điều hướng), và Surface (thiết kế trực quan). Mô hình này đảm bảo Virtual Agent vừa đáp ứng nhu cầu chiến lược, vừa thân thiện với người dùng.

Nielsen Norman Group (2020) khuyến nghị thử nghiệm usability sớm và thường xuyên với tối thiểu 5 người dùng, giúp phát hiện ~85% vấn đề [26]. Đối với Virtual Agent của Học viện, nên testing với các nhóm sinh viên đại diện từ nhiều khóa và ngành khác nhau.

1.3. Ứng dụng trong hoạt động tư vấn lễ tân

1.3.1. Tổng quan hệ thống

Đề án phát triển hệ thống Virtual Agent thông minh hỗ trợ tư vấn và cung cấp thông tin cho sinh viên Học viện PTIT, hoạt động 24/7 nhằm bổ sung và dần thay thế các hình thức tư vấn truyền thống. Hệ thống hướng đến bốn mục tiêu chính: (1) Giảm 70-80% khối lượng câu hỏi lặp lại cho nhân viên lễ tân và phòng CTSV; (2) Đảm bảo thời gian phản hồi dưới 2 giây; (3) Nâng cao trải nghiệm sinh viên qua thông tin nhanh, chính xác và nhất quán; (4) Thu thập dữ liệu insights về nhu cầu và hành vi người dùng phục vụ cải tiến dịch vụ. Với đầu vào (Input) là audio/text với nội dung là các câu hỏi của sinh viên, học viên,.. Và đầu ra (Output) là thông tin hệ thống đã xử lý gửi tới sinh viên, học viên,… qua phương thức audio/text



Hình 1.15: Sơ đồ chung của hệ thống

1.3.2. Công nghệ và kiến trúc hệ thống

a) Kiến trúc hệ thống

Hệ thống theo mô hình Client-Server với 3 thành phần chính:

Frontend: Xây dựng bằng **ReactJS** và **JavaScript** cho giao diện web tương tác cao. Sử dụng **Electron.js** để đóng gói thành desktop app chạy trên Windows, macOS và Linux. Giao diện responsive, hoạt động mượt mà trên mọi thiết bị.

Backend: Phát triển hoàn toàn bằng **Python** với các thư viện AI/ML:

- PhoBERT (transformers, PyTorch): Nhận diện ý định và xử lý tiếng Việt

- spaCy, PyVi, UndertheSea: Tách từ, NER, xử lý ngôn ngữ tự nhiên
- SQLAlchemy: Kết nối database PostgreSQL
- Redis: Lưu trữ context và session
- Elasticsearch: Tìm kiếm knowledge base

Communication**:** Frontend và Backend giao tiếp real-time qua **WebSocket** sử dụng thư viện **Socket.IO** (cả JavaScript client và Python server). Đảm bảo tương tác tức thì, bidirectional communication và auto-reconnect.

b) Luồng xử lý

Khi sinh viên hỏi "Học phí ngành CNTT là bao nhiêu?":

- Frontend (ReactJS) gửi message qua Socket.IO
- Backend (Python) nhận và xử lý:
    - Text Preprocessing (PyVi): Chuẩn hóa, tách từ
    - Intent Recognition (PhoBERT): Nhận diện "hỏi_học_phí"
    - Entity Extraction: Trích xuất "ngành CNTT"
    - Query Database (PostgreSQL): Lấy thông tin học phí
    - Response Generation: Tạo câu trả lời
- Frontend nhận response qua Socket.IO và hiển thị
- Tổng thời gian: < 1 giây

c) Tech Stack

Bảng 1.2: Tech Stack Overview

Component	**Công nghệ**	**Mục đích**
Frontend	ReactJS + Electron.js + JavaScript	Giao diện web & desktop
Backend	Python + PhoBERT + spaCy + PyVi	AI/NLP processing
Communication	WebSocket (Socket.IO)	Real-time messaging
Database	PostgreSQL + MongoDB + Redis	Dữ liệu + Logs + Cache
Search	Elasticsearch	Knowledge base

1.3.3. Tính năng và khả năng của hệ thống

a) Chức năng cốt lõi

Hệ thống Virtual Agent được thiết kế với năm nhóm chức năng chính phục vụ sinh viên PTIT:

Tư vấn thông tin cơ bản: Cung cấp thông tin về lộ trình đào tạo các khoa/ngành, học phí và học bổng, lịch học/thi, các quy chế, cơ sở vật chất và dịch vụ.

Hướng dẫn thủ tục hành chính: Hỗ trợ phương thức đăng ký học phần trực tuyến, bảo lưu/chuyển trường, và đăng ký ký túc xá - hoạt động 24/7 như một nhân viên lễ tân ảo.

Tra cứu thông tin cá nhân: Cho phép sinh viên sau khi xác thực truy cập thông tin cá nhân cơ bản.

Tư vấn định hướng: Cung cấp thông tin về các nganh, hướng nghiệp dựa trên xu hướng thị trường CNTT-Viễn thông, và cơ hội thực tập/việc làm.

Hỗ trợ kỹ thuật: Hướng dẫn về các vấn đề về khôi phục mật khẩu, hướng dẫn sử dụng hệ thống quản lý đào tạo, và khắc phục sự cố đăng nhập/truy cập dịch vụ số.

b) Chức năng mở rộng

Hỗ trợ đa ngôn ngữ: Hỗ trợ một số ngôn ngữ phổ biến để hỗ trợ sinh viên trong nước và quốc tế.

Đa phương thức tương tác: Ngoài chat văn bản, hệ thống hỗ trợ tương tác hình ảnh, biểu cảm để tăng tính trực quan.

Cá nhân hóa: Hệ thống hỗ trợ lưu lịch sử tương tác.

1.4. Kết chương

Chương 1 đã trình bày lý thuyết cho việc phát triển hệ thống Virtual Agent tư vấn lễ tân tại Học viện PTIT với các nội dung chính:

Về người đại diện ảo: Phân tích khái niệm, phân loại theo hình thức (2D, 3D, Digital Human), khả năng tương tác (Static, Reactive, Intelligent Agent) và chức năng. Xác định hệ thống phù hợp là Intelligent Agent có khả năng xử lý ngôn ngữ tự nhiên tiếng Việt, tích hợp đa nền tảng và cá nhân hóa trải nghiệm.

Về cơ sở công nghệ: Hệ thống hóa kiến thức về AI, Machine Learning và đặc biệt là NLP. Phân tích thách thức xử lý tiếng Việt (tách từ, dấu thanh điệu, phương ngôn) và giải pháp (PhoBERT, VnCoreNLP, PyVi, UndertheSea).

Về tương tác người-máy: Trình bày các nguyên lý thiết kế giao diện hội thoại, yếu tố tâm lý học nhận thức, quản lý đối thoại và phương pháp đánh giá UX - tạo cơ sở cho thiết kế hệ thống tối ưu.

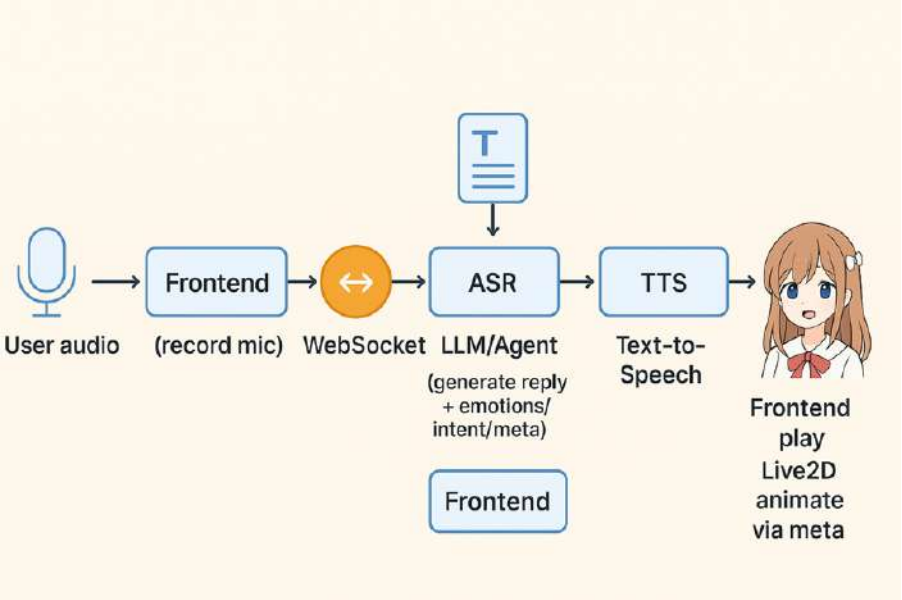
Về ứng dụng thực tiễn: Đề xuất giải pháp dựa trên framework mã nguồn mở Open-LLM-VTuber với các cải tiến quan trọng: Frontend (ReactJS + Electron.js), Backend (Python + PhoBERT + Vietnamese NLP), giao tiếp real-time qua WebSocket. Hệ thống tích hợp database PTIT, custom dialogue management và domain-specific features cho giáo dục.

Chương 1 chứng minh tính khả thi công nghệ của hệ thống, với yêu cầu cốt lõi là xử lý tiếng Việt chính xác thông qua kết hợp công cụ NLP chuyên biệt (PhoBERT, PyVi) và xây dựng corpus chuyên ngành PTIT. Việc sử dụng Open-LLM-VTuber làm base framework giúp tiết kiệm thời gian phát triển và tận dụng kiến trúc đã được kiểm chứng.

Chương 2 sẽ phân tích yêu cầu hệ thống chi tiết thông qua khảo sát nhu cầu sinh viên và cán bộ, xác định use cases cụ thể, yêu cầu chức năng và phi chức năng - tạo cầu nối giữa nền tảng lý thuyết và giai đoạn thiết kế, triển khai.

Chương 2. PHÂN TÍCH THIẾT KẾ VÀ XÂY DỰNG HỆ THỐNG

2.1. Phân tích yêu cầu hệ thống



Hình 2.1: Luồng xử lý cơ bản của hệ thống

2.1.1. Phân tích các bên liên quan

Việc xác định đầy đủ các bên liên quan (stakeholders) và nhu cầu của họ là bước đầu tiên trong phát triển hệ thống. Phân tích này giúp đảm bảo hệ thống Virtual Agent được thiết kế phù hợp với mọi đối tượng sử dụng tại Học viện PTIT. Các đối tượng có thể sử dụng hệ thống:

a) Sinh viên đang học

Nhóm sinh viên đang học là đối tượng chính với khoảng 15,000+ sinh viên thuộc thế hệ Gen Z (18-23 tuổi), có kỹ năng công nghệ tốt và quen thuộc với smartphone. Nhu cầu chính của họ là tra cứu nhanh lịch học, lịch thi, kết quả học tập; hướng dẫn thủ tục hành chính (đăng ký học phần, xin giấy tờ); thông tin học phí, công nợ, học bổng; và giải đáp về quy chế. Họ mong đợi phản hồi tức thì (< 3 giây), thông tin chính xác, giao diện mobile-friendly, hỗ trợ 24/7 đặc biệt trong cao điểm đăng ký học phần.

b) Sinh viên mới và thí sinh tuyển sinh

Nhóm này chưa quen với Học viện và cần thông tin định hướng ban đầu. Nhu cầu tập trung vào thông tin tuyển sinh, hướng dẫn thủ tục nhập học, thông tin ký túc xá và cơ sở vật chất, cùng tư vấn chọn ngành định hướng nghề nghiệp.

c) Sinh viên quốc tế

Số lượng nhỏ nhưng đang tăng, đối mặt với rào cản ngôn ngữ và văn hóa. Họ cần hỗ trợ tiếng Anh, thông tin về visa và lưu trú, cùng hướng dẫn thích nghi văn hóa Việt Nam.

2.1.2. Phân tích yêu cầu chức năng

Yêu cầu chức năng mô tả những gì hệ thống phải làm được. Dựa trên phân tích stakeholders, ta xác định các yêu cầu chức năng sau:

a) Chức năng cốt lõi

CF01: Hỏi đáp thông tin tổng quát

Hệ thống cung cấp thông tin cơ bản về Học viện, chương trình đào tạo và các quy định. Chức năng này bao gồm tra cứu thông tin ngành học/chuyên ngành, cơ sở vật chất (thư viện, phòng lab, ký túc xá), câu lạc bộ và hoạt động ngoại khóa, lịch học kỳ và lịch nghỉ lễ, cũng như quy chế và quy định đào tạo. Sinh viên đưa ra câu hỏi bằng ngôn ngữ tự nhiên (tiếng Việt hoặc tiếng Anh), hệ thống phản hồi bằng text hoặc audio kèm link tham khảo, hình ảnh/video minh họa nếu có.

CF02: Hướng dẫn thủ tục hành chính

Hệ thống hướng dẫn sinh viên thực hiện các thủ tục theo quy trình, bao gồm đăng ký học phần (các bước, thời hạn, lưu ý), xin giấy tờ (bảng điểm, xác nhận sinh viên), quy trình khiếu nại điểm, thủ tục bảo lưu/nghỉ học/chuyển trường, và đăng ký ký túc xá. Khi sinh viên yêu cầu về thủ tục cụ thể, hệ thống cung cấp hướng dẫn từng bước chi tiết, danh sách giấy tờ cần chuẩn bị, thời hạn và địa điểm nộp, form mẫu có thể tải xuống và link đến hệ thống online (nếu có).

CF03: Tư vấn học tập và định hướng

Hệ thống hỗ trợ sinh viên đưa ra quyết định về học tập thông qua tư vấn chọn học phần tự chọn, chọn chuyên ngành, gợi ý lộ trình học tập, cung cấp thông tin về

cơ hội thực tập và việc làm, cũng như định hướng nghề nghiệp theo ngành. Chức năng này yêu cầu thông tin ngữ cảnh cá nhân (ngành, năm học, sở thích) nhưng không đưa ra quyết định thay sinh viên mà chỉ cung cấp thông tin để sinh viên tự quyết định.

CF04: Hỗ trợ kỹ thuật cơ bản

Hệ thống giải quyết các vấn đề kỹ thuật phổ biến như hướng dẫn reset password, khắc phục lỗi đăng nhập, hướng dẫn sử dụng các hệ thống (portal, email), troubleshooting kết nối WiFi, và hướng dẫn cài đặt phần mềm học tập.

b) Nhóm chức năng hỗ trợ

CF06: Quản lý ngữ cảnh hội thoại

Hệ thống theo dõi và duy trì ngữ cảnh qua nhiều lượt hội thoại, lưu trữ lịch sử 10 lượt gần nhất, hiểu đại từ tham chiếu ("nó", "đó", "thế"), xử lý nhiều ý định trong một câu, và yêu cầu làm rõ khi câu hỏi mơ hồ. Ví dụ, khi sinh viên hỏi "Học phí năm nay là bao nhiêu?" và hệ thống trả lời "Học phí năm học 2024-2025 là 12 triệu đồng/năm", câu hỏi tiếp theo "Còn năm sau?" sẽ được hiểu là hỏi về năm 2025-2026 dựa trên ngữ cảnh trước đó.

2.1.3. Phân tích yêu cầu phi chức năng

Yêu cầu phi chức năng mô tả cách thức hệ thống hoạt động, bao gồm hiệu suất, bảo mật, khả năng sử dụng và bảo trì.

a) Hiệu suất

Hệ thống yêu cầu thời gian trung bình phản hồi không quá 2 giây, với 95% yêu cầu được xử lý trong 3 giây và 99% trong 5 giây. Thời gian này được đo từ đầu đến cuối (end-to-end latency) bao gồm cả network và processing, có giám sát realtime với cảnh báo khi vượt ngưỡng. Để đạt được yêu cầu này, hệ thống áp dụng cache cho câu hỏi thường gặp, load balancing, tối ưu database queries và sử dụng CDN cho static content.

Về độ sẵn sàng, hệ thống cần đảm bảo uptime 99.5% (tương đương khoảng 3.6 giờ downtime mỗi tháng). Bảo trì có kế hoạch được thông báo trước 24 giờ và thực hiện ngoài giờ cao điểm, với Mean Time To Recovery (MTTR) không quá 30

phút. Các biện pháp đảm bảo bao gồm redundancy với nhiều server, health checks tự động, cơ chế failover và backup hàng ngày.

b) Bảo mật

Dữ liệu nhạy cảm như điểm thi và thông tin cá nhân được mã hóa AES-256 trong database. Toàn bộ communication sử dụng HTTPS bắt buộc. Password không lưu dạng plaintext mà được hash bằng bcrypt hoặc Argon2. Thông tin cá nhân (PII) được ẩn danh trong logs. Hệ thống tuân thủ Nghị định 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân và OWASP Top 10.

c) Khả năng sử dụng

Giao diện responsive hoạt động tốt trên desktop, tablet và mobile với font size tối thiểu 14px, contrast ratio đạt 4.5:1 theo WCAG AA, hỗ trợ dark mode và không yêu cầu plugin/extension. Thiết kế tuân thủ nguyên tắc đơn giản, rõ ràng, trực quan với tối đa 3 cú nhấp chuột để đến thông tin quan trọng.

Hệ thống hỗ trợ đa nền tảng: web app chạy trên Chrome, Firefox, Safari, Edge (2 phiên bản gần nhất); mobile app tùy chọn cho iOS 14+ và Android 8+; tích hợp Facebook Messenger và Zalo trong Phase 2.

d) Khả năng bảo trì

Code quality đảm bảo coverage tối thiểu 70%, tuân thủ coding standards (PEP 8 cho Python, ESLint cho JavaScript), bắt buộc code review trước khi merge và có tài liệu đầy đủ (docstrings, README, architecture docs).

Khả năng debug được hỗ trợ qua error messages chi tiết cho developer (không expose cho user), stack trace đầy đủ trong logs, monitoring dashboard theo dõi CPU, RAM, disk, network và alerting khi có bất thường.

Deployment sử dụng CI/CD pipeline với automated testing trước khi deploy, blue-green deployment để đạt zero downtime và cơ chế rollback trong vòng 5 phút.

Kiến trúc hệ thống được thiết kế theo mô hình microservices hoặc modular monolith với loose coupling, high cohesion, API-first design, cho phép thay thế module mà không ảnh hưởng toàn hệ thống.

2.2. Thiết kế kiến trúc hệ thống

2.2.1. Kiến trúc tổng quan của hệ thống

Kiến trúc Client – Server kết hợp 3 lớp mô-đun AI chính STT (ASR), LLM (ngôn ngữ/agent), và TTS

a) Thành phần chính

- Frontend (web/desktop): giao diện người dùng, phát âm thanh, hiển thị Live2D, gửi/nhận sự kiện qua WebSocket.

- WebSocketServer (backend ASGI): nhận kết nối client, luồng dữ liệu thời gian thực, tạo session/chat context.

- Config Manager: đọc/validate conf.yaml, đồng bộ/merge cấu hình (merge_configs.py, upgrade.py).

- Model Manager / Adapter Layer: interface chung để tải và chạy các mô-đun ASR, LLM, TTS. Các adapter này ẩn chi tiết triển khai (local model, API).

- ASR (Speech-to-Text): chuyển audio -> text.

- LLM / Agent: xử lý ngữ cảnh, prompt, trả về text (có thể kèm "internal thoughts"/actions).

- TTS (Text-to-Speech): chuyển text -> audio, hỗ trợ voice cloning, chuyển ngôn ngữ cho giọng.

- Live2D Controller: ánh xạ cảm xúc/intent -> biểu cảm/animation Live2D (dựa trên model_dict.json cho mapping).

- Visual Perception (tùy chọn): camera / screen capture -> feature extraction -> đưa vào context LLM.

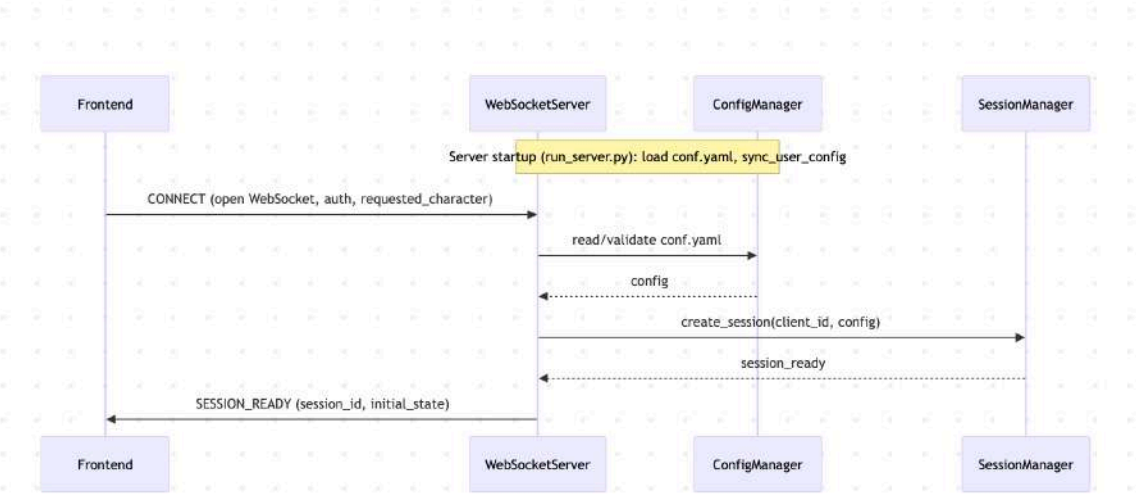
- Logging & Persistence: lưu chat logs, debug logs (log files), và health logs.

b) Vai trò của tác giả

Trong hệ thống tác giả đóng vai trò chính là thiết kế Frontend (web/desktop) của hệ thống, gửi audio/text tới hệ thống qua socket và nhận thông tin và hiển thị tới người dùng qua text/audio và cử chỉ Live2D

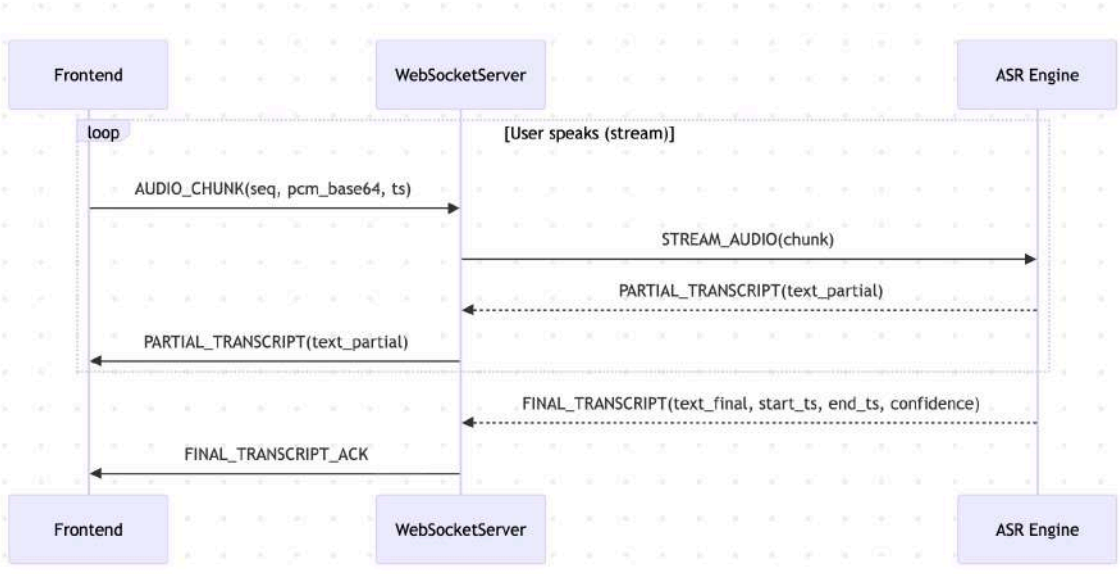
2.2.2. Luồng chính trong hệ thống

a) Luồng khởi tạo kết nối & Session



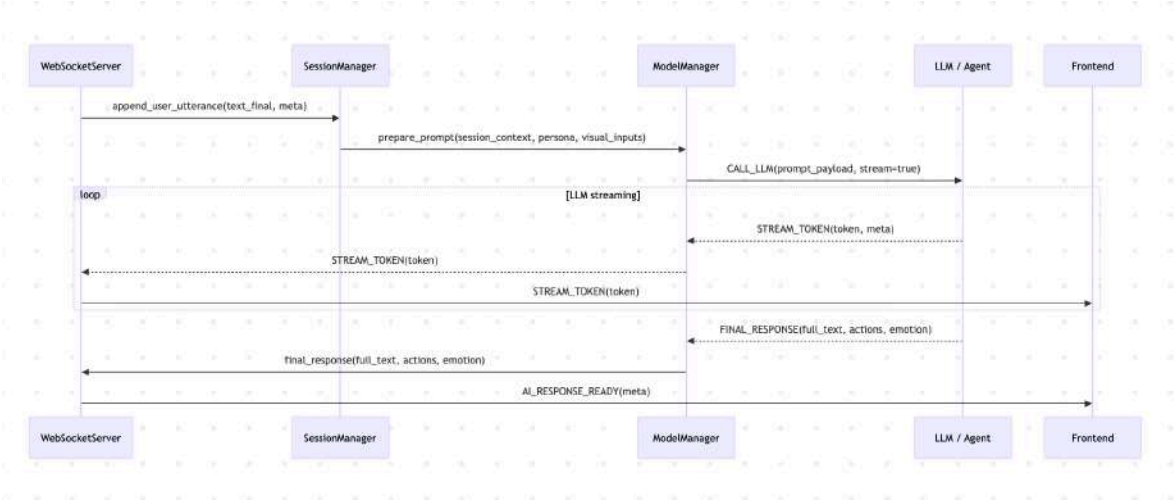
Hình 2.2: Luồng khởi tạo và kết nối session

b) Luồng: Ghi âm ở client -> ASR (streaming) -> partial + final transcript



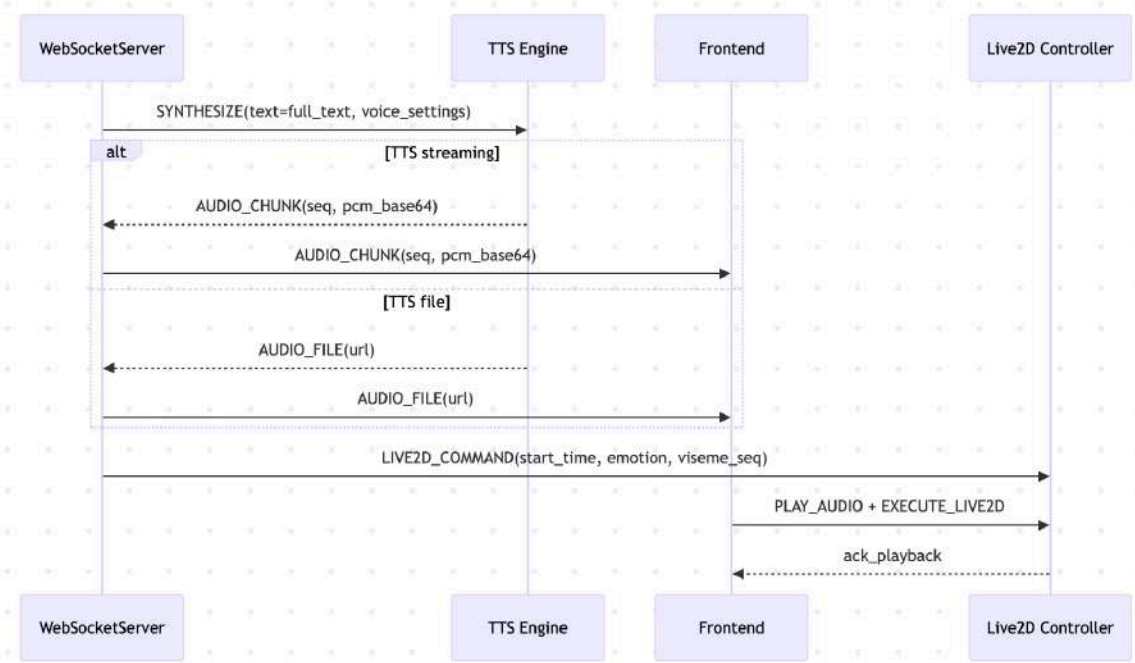
Hình 2.3: Luồng ghi âm ở client

c) Luồng: From final transcript -> build prompt -> gọi LLM (streaming tokens)



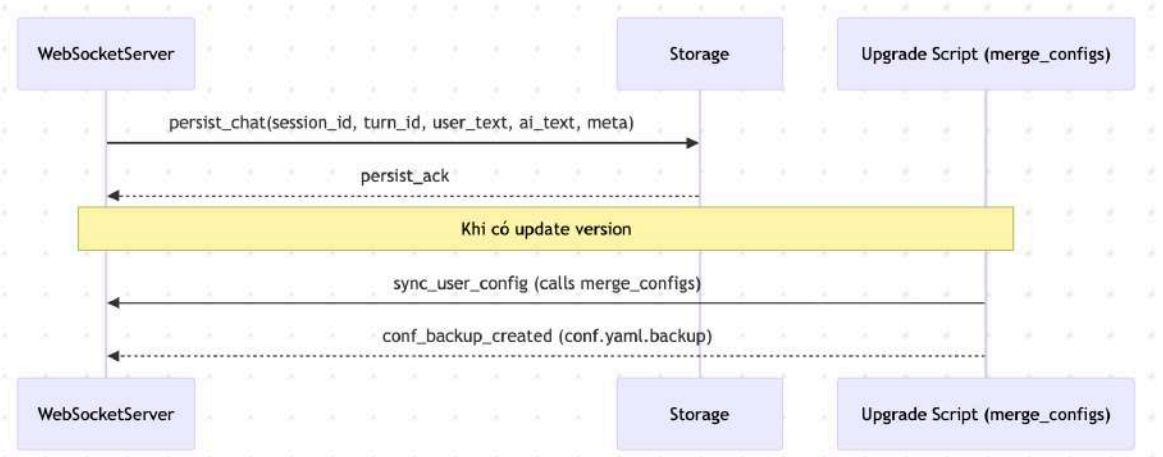
Hình 2.4: Luồng gọi LLM

d) Luồng: TTS synthesis -> gửi audio tới client -> Live2D sync



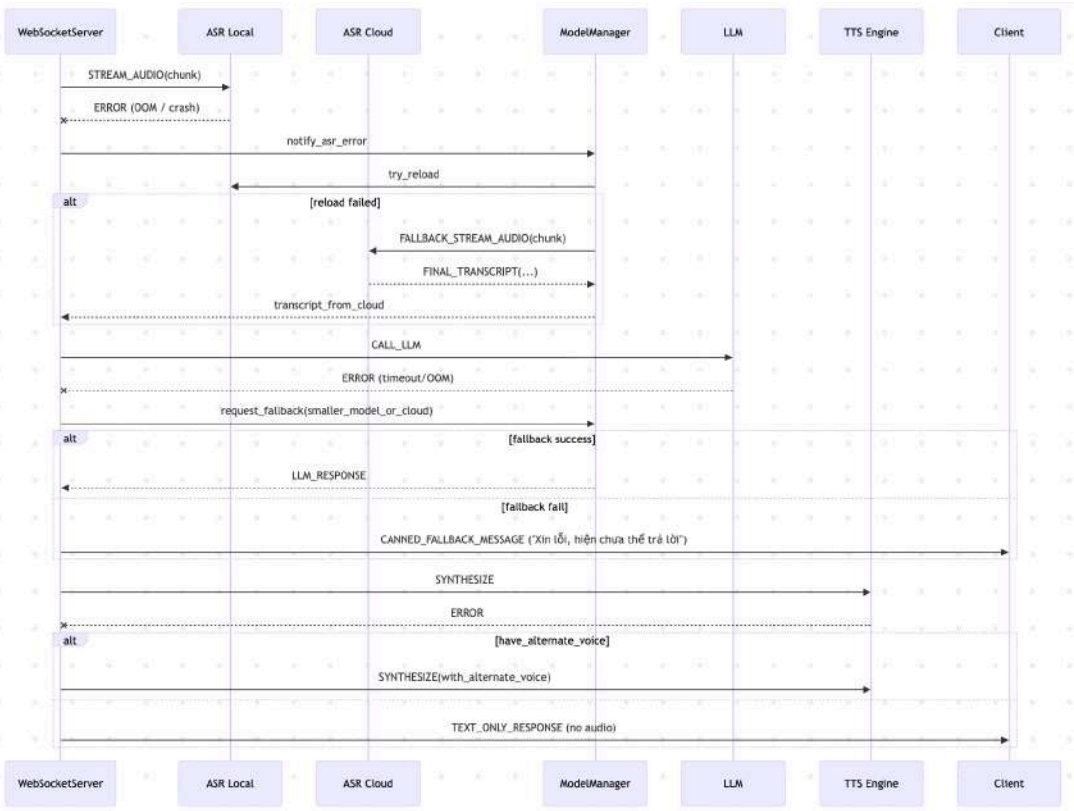
Hình 2.5: Luồng gửi audio tới client

e) Luồng: Lưu trữ hội thoại & logging & backup config



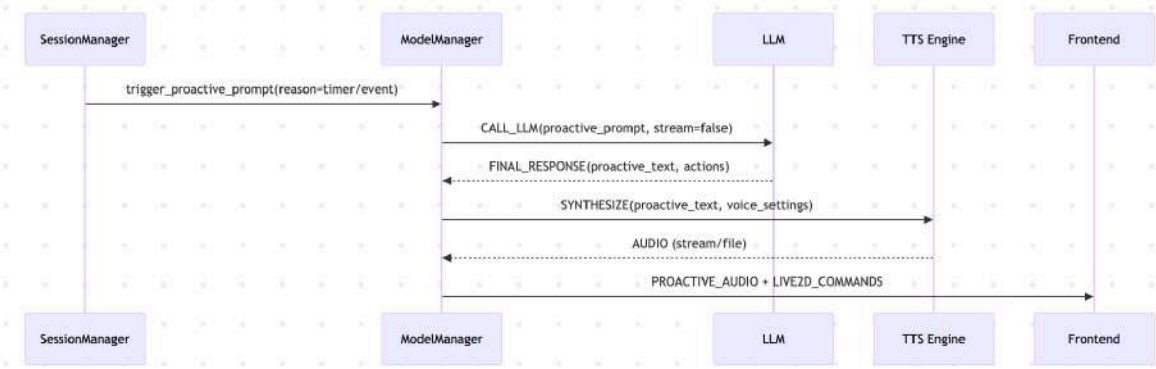
Hình 2.6: Luồng lưu trữ hội thoại

f) Luồng: Xử lý lỗi / fallback (ASR, LLM, TTS)



Hình 2.7: Luồng xử lý lỗi

g) Luồng: Proactive speaking (agent-triggered)



Hình 2.8: Luồng proactive speaking

2.2.3. Thiết kế chi tiết các mô-đun

a) Mô-đun STT

1. Tác dụng và vai trò

Module ASR đóng vai trò "tai" của hệ thống, chuyển đổi giọng nói sinh viên thành văn bản để xử lý:

Giao tiếp tự nhiên

- Nhận diện giọng nói tiếng Việt, chuyển câu hỏi từ âm thanh sang văn bản
- Hỗ trợ đa giọng địa phương
- Tạo trải nghiệm giao tiếp tự nhiên như nói chuyện với lễ tân thật

Tiếp nhận thông tin realtime

- Xử lý và nhận diện ngay lập tức khi sinh viên đang nói
- Hiển thị live transcript giúp sinh viên biết hệ thống đang "nghe"
- Giảm thời gian chờ, tăng tính tương tác

2. Tính năng chính

Real-time Streaming

- Nhận diện liên tục trong khi sinh viên đang nói
- Phản hồi dưới 500ms, đảm bảo trải nghiệm mượt mà
- Tự động điều chỉnh theo tốc độ nói
- Tự động kích hoạt khi backend hỗ trợ streaming

Phát hiện điểm kết thúc và im lặng

- Phát hiện thông minh khi sinh viên nói xong (không cắt giữa câu)
- Điều chỉnh độ dài im lặng cần thiết để kết thúc phát ngôn
- Tự động phân đoạn phát ngôn tự nhiên
- Hiểu ngữ cảnh để tránh cắt sai (ví dụ: đọc số điện thoại)

Điểm tin cậy và kết quả từng phần

- Hiển thị live transcript: Từng từ đang được nhận diện xuất hiện trên màn hình, tạo cảm giác tin tưởng
- Đánh giá độ tin cậy: Chỉ chấp nhận kết quả có confidence > 80%, tự động hỏi lại nếu không chắc chắn
- Tinh chỉnh dần: Kết quả được cải thiện khi có thêm ngữ cảnh
- Phản hồi trực quan: Màu sắc/style khác nhau cho các mức độ tin cậy

Chế độ "AI không nghe TTS"

- Loại bỏ tiếng vọng của chính hệ thống TTS
- Ngăn AI nhận diện lại giọng nói của chính mình (khi AI đang trả lời, ASR tự động "tắt tai")
- Chỉ kích hoạt lại khi phát hiện giọng nói thật của sinh viên
- Chuyển đổi mượt mà giữa chế độ nghe và nói

3. Tối ưu cho bối cảnh lễ tân PTIT

Xử lý thuật ngữ chuyên ngành

- Từ điển tùy chỉnh: Tên khoa, ngành, môn học, giảng viên
- Nhận diện viết tắt: ĐHBKHN, CNTT, ĐTVT, v.v.

Đa dạng giọng địa phương

- Hỗ trợ sinh viên từ Bắc - Trung - Nam
- Tự động điều chỉnh model theo giọng người nói

Khả năng chống nhiễu

- Lọc tiếng ồn môi trường (quầy lễ tân thường ồn ào)
- Phân biệt giọng nói chính trong môi trường có nhiều người

b) Mô-đun LLM / Agent

1. Tổng quan

Module LLM (Large Language Model) là "bộ não" trung tâm của hệ thống người đại diện ảo, đảm nhận vai trò chính trong việc tư vấn và cung cấp thông tin cho sinh viên.

2. Chức năng và vai trò chính

Hiểu và xử lý câu hỏi của sinh viên

- Phân tích ý định (NLU): Hiểu đúng yêu cầu (ví dụ: "Em muốn đăng ký học bổ sung" → cần thông tin quy trình đăng ký môn học)
- Nắm bắt ngữ cảnh: Nhớ thông tin đã trao đổi để hiểu câu hỏi liên quan ("Còn điều kiện gì nữa không?" → điều kiện của vấn đề vừa thảo luận)
- Phân loại mục đích: Xác định sinh viên muốn hỏi thông tin, khiếu nại hay đăng ký dịch vụ
- Nhận diện cảm xúc: Điều chỉnh cách phản hồi theo trạng thái (lo lắng, vội vàng, bình thường)

Tư vấn và cung cấp thông tin

- Tra cứu knowledge base: Trích xuất quy chế đào tạo, lịch học, thời khóa biểu, thông tin khoa/ngành, dịch vụ sinh viên (học bổng, KTX, thư viện)
- Tạo câu trả lời: Đưa ra thông tin tự nhiên, dễ hiểu, phù hợp từng đối tượng
- Hướng dẫn từng bước: Chỉ dẫn chi tiết quy trình phức tạp (đăng ký môn học, xin giấy xác nhận sinh viên)
- Tổng hợp thông tin: Kết hợp nhiều nguồn để đưa ra câu trả lời toàn diện

Ra quyết định và điều phối hệ thống

- Xác định hành động: Trả lời trực tiếp, tra cứu thêm, chuyển nhân viên thật hoặc gợi ý liên hệ bộ phận khác
- Điều phối module: Kích hoạt TTS (đọc câu trả lời), RAG (tra cứu tài liệu), hiển thị thông tin (bản đồ, form, link)
- Quản lý luồng hội thoại: Dẫn dắt cuộc trao đổi đạt mục tiêu tư vấn

- Chuyển tiếp khi cần: Nhận biết lúc cần nhân viên thật xử lý (vấn đề phức tạp, sinh viên không hài lòng)

Cá nhân hóa trải nghiệm

- Phong cách lễ tân: Giao tiếp lịch sự, thân thiện, nhiệt tình, phù hợp văn hóa Việt Nam và PTIT
- Điều chỉnh theo đối tượng: Phân biệt sinh viên năm 1 (giải thích chi tiết) và năm cuối; sinh viên vội (ngắn gọn) và cần tư vấn kỹ (đầy đủ)
- Tích hợp thông tin cá nhân: Sử dụng dữ liệu (tên, ngành, năm học) để cá nhân hóa
- Ghi nhớ lịch sử: Lưu tương tác để phục vụ tốt hơn lần sau

Suy luận và giải quyết vấn đề

- Suy luận logic: Phân tích quy chế, dữ liệu để trả lời câu hỏi phức tạp (ví dụ: học 2 môn trùng lịch)
- Giải thích từng bước: Trình bày quá trình lý luận khi cần
- Chia nhỏ vấn đề: Phân tách vấn đề phức tạp thành các bước đơn giản
- Đáp ứng ràng buộc: Tìm giải pháp thỏa mãn nhiều điều kiện đồng thời

3. Backend LLM được hỗ trợ

Local Inference

- GGUF Models: Chạy trên server học viện (Vistral, Qwen2, Llama-3 fine-tuned tiếng Việt), đảm bảo bảo mật, không phụ thuộc internet, chi phí thấp
- LM Studio: Giao diện quản lý local models cho giai đoạn nghiên cứu
- vLLM: Engine hiệu suất cao, phục vụ nhiều sinh viên đồng thời với tốc độ nhanh

Cloud API

- Google Gemini: AI đa phương thức, xử lý tốt text/image, hỗ trợ tiếng Việt

Lựa chọn đề xuất

- Giai đoạn nghiên cứu và production: **Google Gemini** (chất lượng cao, dễ triển khai)

4. Kiến trúc và khả năng mở rộng

Hệ thống Prompt Template

- Cá nhân hóa Persona: Định nghĩa tính cách người đại diện ảo (lễ tân chuyên nghiệp PTIT, lịch sự, nhiệt tình, chính xác)
- Điều chỉnh động: Tự động thay đổi prompt theo ngữ cảnh (thời gian → chào hỏi phù hợp; giờ cao điểm → trả lời ngắn gọn)
- Hỗ trợ đa ngôn ngữ: Template tiếng Việt và tiếng Anh cho sinh viên quốc tế

Tích hợp ngữ cảnh

- Xử lý hình ảnh: OCR đọc giấy tờ, thẻ sinh viên, form đơn từ
- Lịch sử hội thoại: Duy trì ngữ cảnh liên tục (sinh viên hỏi "Cái đầu ạ" → LLM hiểu là đề cập đến lựa chọn đầu tiên đã nói trước đó)
- Dữ liệu sinh viên: Cá nhân hóa response dựa trên mã sinh viên, họ tên, khoa, ngành, năm học, lịch sử tương tác
- Tri thức bên ngoài:
  - RAG: Tra cứu quy chế đào tạo, văn bản hướng dẫn, thông báo mới nhất
  - Truy vấn database: Kết nối hệ thống quản lý (thời khóa biểu, điểm thi, học phí)
  - Dữ liệu real-time: Lịch thi, lịch nghỉ, sự kiện sắp diễn ra

5. Tối ưu cho bối cảnh lễ tân PTIT

Tri thức chuyên ngành

- Thuật ngữ PTIT: Tên khoa (Viễn thông 1, 2; Công nghệ thông tin), ngành (ĐTVT, CNTT, KHMT, ATTT), khái niệm (học bổ sung, học cải thiện, học phần, tín chỉ)
- Quy chế và quy định: Nắm vững quy chế đào tạo tín chỉ, quy định học vụ, thi cử, xét tốt nghiệp
- Hướng dẫn quy trình: Chi tiết các thủ tục (đăng ký học, rút học phần, hoãn thi, xin giấy xác nhận, xin bảng điểm)

Thích ứng văn hóa và ngôn ngữ

- Xưng hô đúng chuẩn: "Anh/chị" (AI tự xưng) - "Em" (gọi sinh viên), điều chỉnh theo độ tuổi

- Giọng điệu phù hợp: Vừa chuyên nghiệp vừa thân thiện, không quá cứng nhắc cũng không quá thoải mái
- Ngữ cảnh địa phương: Hiểu địa điểm trong campus, tên tòa nhà, sự kiện và hoạt động đặc trưng PTIT

c) Mô-đun TTS

Tác dụng & Vai trò của Module:

Module TTS đóng vai trò **"giọng nói"** của hệ thống người đại diện ảo, chuyển đổi câu trả lời dạng văn bản thành giọng nói tự nhiên để giao tiếp với sinh viên

d) Live 2D & Visual Perception

Tác dụng & Vai trò của Module:

Live2D - Giao diện Hình ảnh

- Tạo avatar lễ tân ảo: Hình ảnh sinh động, chuyên nghiệp của lễ tân PTIT
- Biểu cảm cảm xúc: Mỉm cười, nghiêm túc, quan tâm theo tình huống
- Tăng tính thân thiện: Sinh viên cảm thấy dễ tiếp cận hơn so với giao diện text thuần

Visual Perception - Nhận diện Hình ảnh

- Nhận diện sinh viên: Phát hiện khi sinh viên đang chờ tại quầy
- Đọc tài liệu: OCR để đọc thẻ sinh viên, đơn từ, giấy tờ
- Cung cấp ngữ cảnh: Giúp LLM hiểu tình huống để tư vấn chính xác hơn

Live2D System

Quản lý Model

File model_dict.json lưu thông tin cấu hình:

- model3.json path: Đường dẫn đến file model Live2D
- scale: Tỷ lệ hiển thị avatar (1.0 = 100%)
- position/offset: Vị trí hiển thị trên màn hình (x, y)
- emotionMap: Ánh xạ cảm xúc → animation tương ứng

Emotion Map - Ánh xạ Cảm xúc

Bảng 2.1: Ánh xạ cảm xúc

Tình huống	**Emotion**	**Animation**
Sinh viên vừa đến	greeting	Vẫy tay chào
Đang lắng nghe	listening	Gật đầu
Đang suy nghĩ	thinking	Nhìn lên trên
Giải thích	explaining	Chỉ tay
Sinh viên khó khăn	empathy	Quan tâm
Idle (rảnh)	neutral	Đứng chờ tự nhiên

Motion Groups - Nhóm Động tác

Idle animations: Khi không tương tác

- Thở nhẹ, chớp mắt tự nhiên
- Random đổi tư thế (3-5 giây/lần)
- Tạo cảm giác "sống", không đứng yên như tượng

Tap motions: Khi user click vào avatar

- Click vào đầu → Nghiêng đầu mỉm cười
- Click vào thân → Giật mình nhẹ
- Click nhiều lần → Animation đặc biệt (easter egg)

Gesture animations: Động tác khi nói

- "Em điền vào form này" → Chỉ tay vào màn hình
- "Hạn cuối ngày 15" → Giơ tay nhấn mạnh
- Đồng bộ với giọng nói (TTS)

e) Vai trò của học viên

Thiết kế hệ thông ghi nhận audio/text gửi đến các mô-đun STT, LLM/agent, TTS, và hiển thị kết quả nhận được từ hệ thống đến sinh viên qua audio và text.

2.3. Thiết kế thuật toán và mô hình AI

- Streaming + partial results: hỗ trợ trả transcript/tokens từng phần để UX realtime.

- Interrupt handling: khi TTS đang phát, microphone không gửi input tới ASR, hoặc hệ thống lọc audio do TTS để tránh vòng lặp.

- Prompt management & context window: LLM adapter chịu trách nhiệm quản lý history, cắt giảm context theo token budget của backend.

- Emotion mapping: LLM hoặc sentiment detector xuất thông tin cảm xúc -> ánh xạ motion/expr Live2D.

- Config merge & upgrade: merge_configs.py đảm nhiệm merge an toàn user config với template new version, ghi backup, log các key mới/hết hạn.

2.4. Cài đặt và phát triển hệ thống

Hệ thống được xây dựng dựa trên open source **Open-LLM-Vtuber**, từ đó thiết kế hệ thống sao cho phù hợp với sinh viên Học viện Công nghệ Bưu chính Viễn thông. Hệ thống sử dụng các ngôn ngữ phát triển phổ biến như **Python** bởi vì Python là 1 trong những ngôn ngữ khá phổ biến và có thể kết hợp được nhiều thư viện AI hiện có. Ngoài ngôn ngữ Python để phát triển Backend thì tôi có sự dụng ngôn ngữ **Javascript** và các framework như **Reactjs**, **electronjs** để phát triển Frontend mang đến giao diện chuyên nghiệp và trực quan tới người dùng.

Ngoài ra để kiểm thử hệ thống trước khi tới người dùng tôi còn sử dụng phầm mềm kiểm thử **Jmeter** để thực hiện mô phỏng khoảng 1000 user cùng vào trang web để kiểm tra khả năng chịu tải của hệ thống.

2.5. Kết chương

Chương 2 trình bày quá trình phân tích, thiết kế và xây dựng hệ thống Virtual Agent qua ba giai đoạn chính:

Phân tích yêu cầu: Xác định các bên liên quan và nhu cầu cụ thể, bao gồm yêu cầu chức năng (chào hỏi, cung cấp thông tin, tư vấn) và phi chức năng (hiệu suất, độ tin cậy, bảo mật), tạo nền tảng cho thiết kế hệ thống.

Thiết kế kiến trúc: Triển khai mô hình 3-tier với các module chức năng rõ ràng (xử lý đầu vào, nhận diện ý định, quản lý hội thoại, sinh phản hồi, giao diện). Cơ sở dữ liệu được tối ưu hóa cho khả năng truy vấn và mở rộng.

Thiết kế AI và thuật toán: So sánh lựa chọn mô hình học máy, xây dựng giải pháp NLP cho tiếng Việt, và phát triển thuật toán quản lý hội thoại thông minh làm cơ sở cho cài đặt hệ thống.

Kết quả chương này cung cấp cơ sở kỹ thuật hoàn chỉnh cho triển khai thực tế. Chương 3 sẽ đánh giá hiệu năng và kết quả thử nghiệm hệ thống.

Chương 3. THỬ NGHIỆM VÀ ĐÁNH GIÁ

3.1. Tập dữ liệu

Tôi sử dụng tập dữ liệu là các câu hỏi và câu trả lời trong các thủ tục hành chính,…. Với tổng dữ liệu sấp sỉ 1600 dữ liệu bao gồm 424 dữ liệu về thông tin, 1120 dữ liệu về câu hỏi/trả lời, với các chủ đề chính như: Thông tin về học viện, các ngành học, tuyển sinh, các câu lạc bộ, các mẫu đơn,.. Với cách thức thu thập từ các trang chính thống của trưởng, sổ tay sinh viên.

A	B	C	D	E	F	G
STT	Dữ Liệu	Câu Hỏi	Câu Trả Lời	Nguồn(Nếu Có)	Thông Tin Thêm	Đánh giá
1	1	Cổng thông tin Học viện Công nghệ Bưu chính Viễn th	Cổng thông tin Học viện Công nghệ Bưu chính			
2	2	Link phòng giáo vụ của Học viện Công nghệ Bưu chín	Link phòng giáo vụ của Học viện Công nghệ Bư			
3	3	Facebook của Học viện Công nghệ Bưu chính Viễn thô	Học viện Công nghệ Bưu chính Viễn thông - PT			
4	4	Trang quản lý đào tạo của Học viện Công nghệ Bưu ch	Trang quản lý đào tạo của Học viện Công nghệ			
5	5	Web slink Học viện Công nghệ Bưu chính Viễn thông	Web slink Học viện Công nghệ Bưu chính Viễn			
6	6	Trụ sở chính của Học viện Công nghệ Bưu chính Viễn	Trụ sở chính của Học viện Công nghệ Bưu chín 122, Hoàng Quốc Việt, Cầu Giấy, Hà Nội Điện thoại: 024.37562186 Fax: 024.37562036 Website: www.ptit.edu.vn Email: vanphonghv@ptit.edu.vn			5
7	6	**Học viện Công nghệ Bưu chính Viễn thông - PTIT**	Học viện có 3 cơ sở đào tạo với 2 cơ sở tại Hà N			5
8	6	Em nên liên hệ với ai nếu học tại Học viện Công nghệ	Nếu bạn học tại Hà Nội, bạn có thể liên hệ với Điện thoại: 024.38544451 Fax: 024.33829236 Email: ctsv@ptit.edu.vn			5
9	6	Cơ sở đào tạo tại TP.HCM của Học viện Công nghệ Bư	Thông tin cơ sở đào tạo tại TP.HCM của Học vi 97 Man Thiện, P. Hiệp Phú, TP. Thủ Đức Điện thoại: 028.38295258 Fax: 028.39105510 Email: hvbcvthcm@ptithcm.edu.vn Website: https://ptithcm.edu.vn			5
10	7	Tôi muốn tìm hiểu về chương trình đào tạo và các ngà	Bạn có thể liên hệ trực tiếp với Phòng Đào tạo t Thông tin liên hệ: Số điện thoại: (024) 3351 2252 Email: daotao@ptit.edu.vn Địa chỉ: Tầng 2, nhà A1, Km10, Nguyễn Trãi, H Số điện thoại chung: 024 3756 2186 Email: ctsv@ptit.edu.vn Hoặc liên hệ Phòng Đào tạo cơ sở Hà Nội để đư Số điện thoại: (024) 3351 2252 Email: daotao@ptit.edu.vn Địa chỉ cơ sở TP. Hồ Chí Minh của Học viện: Số 11 Nguyễn Đình Chiểu, Phường Đa Kao, Qu			5
			Bạn có thể liên hệ với Phòng Chính trị & Công Thông tin liên hệ:			

Hình 3.1: Tập dữ liệu câu hỏi

3.2. Môi trường thử nghiệm

Môi trường thử nghiệm được hiển thị trong bảng sau:

Bảng 3.1: Thông số môi trường thử nghiệm

Thông số	Giá trị
CPU	Xeon Silver 4309
GPU	Nvidia Quadro P4000
Bộ nhớ RAM	32 GB
Hệ điều hành	Ubuntu

Python	3.9

3.3. Xây dựng ứng dụng thử nghiệm

3.3.1. Giới thiệu hệ thống

Ứng dụng được xây dựng bước đầu là một trang web, với ngôn ngữ lập trình là Python xử lý Backend và Reactjs và electronjs để xử lý FrontEnd. Tôi đã lựa chọn Python bởi ngôn ngữ này có nhiều thư viện hỗ trợ tốt cho việc xử lý âm thanh, hình ảnh hỗ trợ tốt cho ứng dụng.

3.3.2. Xây dựng hệ thống

Ứng dụng sử dụng Python bao gồm nhiều thư mục với các mục đích sử dụng khác nhau. Phần chính bao gồm các thư mục data bên trong chứa các tệp dữ liệu âm thanh mẫu. Ngoài ra, các thư mục khác chứa các tệp giao diện .html, javascript, view.py, url.py…Ngoài ra tôi cũng xây dựng giao diện trên trang web để có thể tương thích với desktop và mobile.

3.4. Kết quả thử nghiệm và đánh giá

3.4.1. Thiết lập thử nghiệm

a) Mục tiêu thử nghiệm:

- Đánh giá độ chính xác của hệ thống trong việc nhận diện ý định và trả lời câu hỏi của sinh viên
- Kiểm tra hiệu suất và thời gian phản hồi của hệ thống trong các tình huống khác nhau
- Đo lường mức độ hài lòng của người dùng đối với trải nghiệm tương tác
- Xác định các lỗi, hạn chế và điểm cần cải thiện của hệ thống
- Đánh giá khả năng xử lý đồng thời nhiều yêu cầu và tính ổn định của hệ thống

b) Phương pháp thử nghiệm:

Thử nghiệm chức năng (Functional Testing):

- Kiểm tra từng chức năng theo use case đã thiết kế
- Sử dụng phương pháp black-box testing

- Ghi nhận các lỗi, bug và hành vi bất thường

Thử nghiệm hiệu năng (Performance Testing):

- Load testing: Mô phỏng [X] người dùng đồng thời
- Stress testing: Kiểm tra giới hạn chịu tải của hệ thống
- Đo lường thời gian phản hồi trung bình, thời gian xử lý tối đa

c) Kịch bản thử nghiệm:

Kịch bản 1: Tương tác cơ bản

- Người dùng đặt câu hỏi đơn giản về thông tin cơ bản (giờ làm việc, địa chỉ phòng ban)
- Hệ thống phản hồi thông tin chính xác
- Đánh giá: Thời gian phản hồi, độ chính xác câu trả lời

Kịch bản 2: Tương tác phức tạp

- Người dùng đặt câu hỏi yêu cầu kết hợp nhiều thông tin (ví dụ: "Em muốn đăng ký học lại môn X, phải làm thế nào và mất bao nhiêu tiền?")
- Hệ thống cần hiểu đa ý định và trả lời đầy đủ
- Đánh giá: Khả năng hiểu ngữ cảnh, tính đầy đủ của câu trả lời

Kịch bản 3: Xử lý hội thoại nhiều lượt

- Người dùng đặt câu hỏi ban đầu, sau đó hỏi thêm các câu liên quan
- Hệ thống cần duy trì ngữ cảnh hội thoại
- Đánh giá: Context management, tính mạch lạc của cuộc hội thoại

Kịch bản 4: Xử lý tình huống ngoại lệ

- Người dùng đặt câu hỏi ngoài phạm vi kiến thức
- Người dùng nhập input không hợp lệ hoặc spam
- Đánh giá: Khả năng xử lý lỗi, thông báo phù hợp cho người dùng

Kịch bản 5: Kiểm tra tải đồng thời

- Mô phỏng nhiều người dùng truy cập đồng thời trong giờ cao điểm
- Đánh giá: Tốc độ phản hồi, tỷ lệ lỗi, độ ổn định hệ thống

d) Công cụ và phương tiện thu thập dữ liệu:

Công cụ ghi log và monitor:

- Hệ thống logging: Ghi nhận toàn bộ hội thoại, thời gian xử lý

- Sử dụng google analytics để kiểm tra số lương truy cập, các hành vi nổi bật

3.4.2. Giao diện hệ thống

a) Giao diện trang chủ



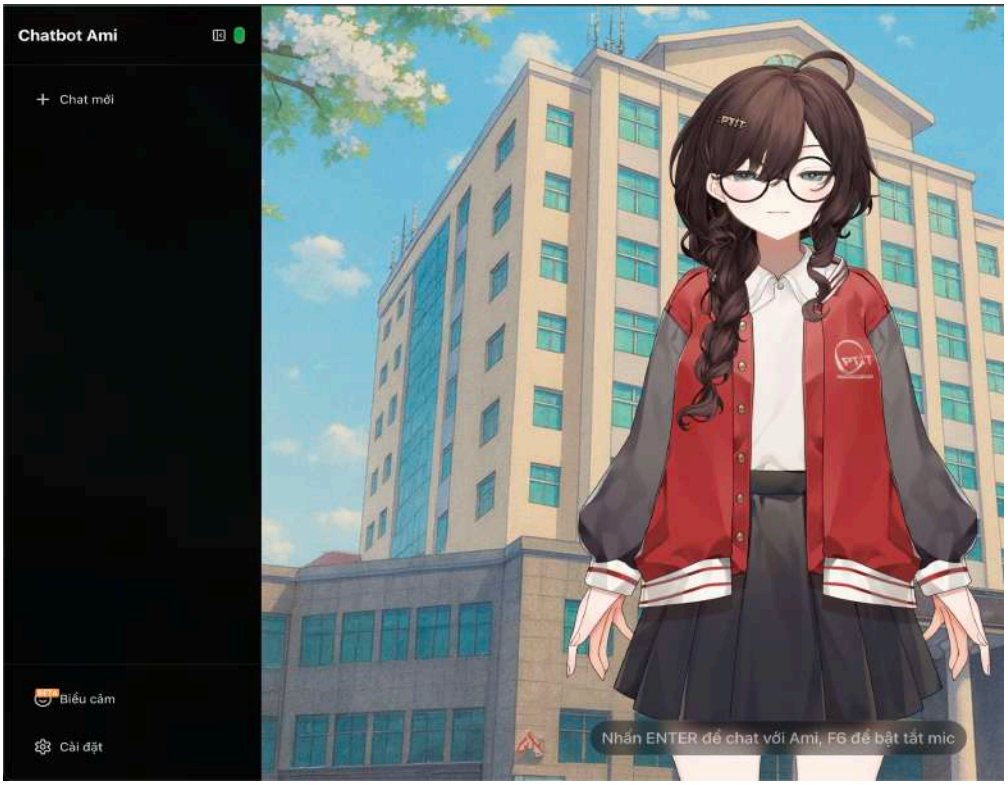
Hình 3.2: Trang chủ

b) Giao diện khung chat



Hình 3.3: Khung chat

c) Giao diện thanh công cụ



Hình 3.4: Thanh công cụ

d) Giao diện cài đặt



Hình 3.5: Cài đặt

e) Giao diện thanh công cụ biểu cảm



Hình 3.6: Thanh công cụ biểu cảm

3.5. Kết chương

Chương 3 đã đi vào chi tiết thực nghiệm đề tài, việc lựa chọn tệp dữ liệu đào tạo sẽ quyết định sự thành công của mỗi mô hình. Chính vì vậy tôi đã lựa chọn kỹ lưỡng mô hình và bộ dữ liệu thử nghiệm. Trong chương này cũng đã đi sâu vào phân tích và đánh giá mô hình, các thức thiết lập thử nghiệm, kết quả đạt được và so sánh kết quả đó với các nghiên cứu khác. Thông qua đó sẽ định hướng tiếp theo cho sự mở rộng của đề án.

KẾT LUẬN

Đề án "***Nghiên cứu phát triển hệ thống người đại diện ảo trong hoạt động tư vấn lễ tân, cung cấp thông tin cho sinh viên Học viện Công nghệ Bưu chính Viễn thông***" đã hoàn thành các mục tiêu đề ra, bao gồm nghiên cứu cơ sở lý thuyết, phân tích thực trạng và thiết kế thành công hệ thống với các chức năng đáp ứng yêu cầu nghiệp vụ.

Hệ thống đã được xây dựng với kiến trúc hợp lý, tích hợp công nghệ AI hiện đại và giao diện thân thiện. Qua triển khai thử nghiệm, hệ thống đã hỗ trợ sinh viên tra cứu thông tin hiệu quả, giảm tải công việc cho bộ phận lễ tân và nâng cao chất lượng dịch vụ hỗ trợ sinh viên 24/7.

Về ý nghĩa, đề án góp phần làm phong phú nghiên cứu về ứng dụng AI trong giáo dục đại học Việt Nam, đồng thời chứng minh hiệu quả trong việc hiện đại hóa công tác quản lý sinh viên, cải thiện trải nghiệm người học và tạo tiền đề phát triển các hệ thống thông minh khác.

Hạn chế của hệ thống bao gồm cơ sở dữ liệu cần cập nhật thường xuyên, khả năng xử lý câu hỏi phức tạp còn hạn chế và chưa tích hợp đầy đủ với các hệ thống thông tin khác của Học viện.

Hướng phát triển tiếp theo bao gồm mở rộng cơ sở dữ liệu, bổ sung nhận dạng giọng nói và đa ngôn ngữ (ngắn hạn); kết nối với các nền tảng quản lý đào tạo và phát triển ứng dụng di động (trung hạn); ứng dụng công nghệ metaverse và phát triển khả năng học tự động (dài hạn).

Đề án đã chứng minh tính khả thi của việc ứng dụng công nghệ người đại diện ảo vào hoạt động tư vấn sinh viên, hứa hẹn trở thành công cụ quan trọng góp phần xây dựng môi trường giáo dục thông minh và hiện đại.

TÀI LIỆU THAM KHẢO

[1] T. T. Hewett and e. al., ACM SIGCHI Curricula for Human-Computer Interaction, ACM, 1992.

[2] D. Norman, The Design of Everyday Things: Revised and Expanded Edition, Basic Books, 2013.

[3] R. J. Moore and R. Arar, Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework, ACM Books, 2019.

[4] H. P. Grice, "Logic and conversation," in *Syntax and Semantics, Vol. 3: Speech Acts*, Academic Press, 1975, p. 41–58.

[5] D. Q. Nguyen and A. T. Nguyen, "PhoBERT: Pre-trained language models for Vietnamese," in *Findings of EMNLP 2020*, 2020.

[6] T. T. e. a. Hewett, ACM SIGCHI Curricula for Human-Computer Interaction, 1992.

[7] D. Norman, The Design of Everyday Things: Revised and Expanded Edition, Basic Books, 2013.

[8] C. &. B. S. Nass, Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship, MIT Press, 2005.

[9] H. P. Grice, Logic and conversation, In P. Cole & J. Morgan (Eds.), Syntax and Semantics, Vol. 3, Speech Acts, pp. 41-58. Academic Press., 1975.

[10] R. J. &. A. R. Moore, Conversational UX Design: A Practitioner's Guide to the Natural Conversation Framework, ACM Books, 2019.

[11] ISO, ISO 9241-210: Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems, ISO, 2019.

[12] M. &. T. N. Hassenzahl, User experience – A research agenda, Behaviour & Information Technology, 25(2), 91-97, 2006.

[13] J. Nielsen, Usability Engineering, Morgan Kaufmann, 1993.

[14] P. Morville, User experience design, Semantic Studios, 2004.

[15] J. &. L. J. R. Sauro, Quantifying the User Experience: Practical Statistics for User Research (2nd ed.), Morgan Kaufmann, 2016.

[16] J. Brooke, SUS: A "quick and dirty" usability scale, Usability Evaluation in Industry, pp. 189-194, 1996.

[17] S. e. a. Holmes, Chatbot Usability Questionnaire: Development and validation, International Journal of Human-Computer Interaction, 35(11), 1-12, 2019.

[18] F. F. Reichheld, The one number you need to grow, Harvard Business Review, 81(12), 46-54, 2003.

[19] P. C. e. a. Verhoef, From multi-channel retailing to omni-channel retailing, ournal of Retailing, 91(2), 174-181, 2015.

[20] L. Wroblewski, Mobile First, A Book Apart, 2011.

[21] W3C, Web Content Accessibility Guidelines (WCAG) 2.1, W3C Recommendation, 2018.

[22] Microsoft, Inclusive Design Toolkit, Microsoft Design, 2016.

[23] M. T. S. S. &. G. C. Ribeiro, “Why should I trust you?” Explaining the predictions of any classifier, Proceedings of KDD 2016, pp. 1135–1144, 2016.

[24] J. D. &. S. K. A. Lee, Trust in automation: Designing for appropriate reliance, Human Factors, 46(1), 50–80, 2004.

[25] J. J. Garrett, The Elements of User Experience: User-Centered Design for the Web and Beyond (2nd ed.), New Riders, 2011.

[26] N. N. Group, How many test users in a usability study?, NN/g Research, 2020.