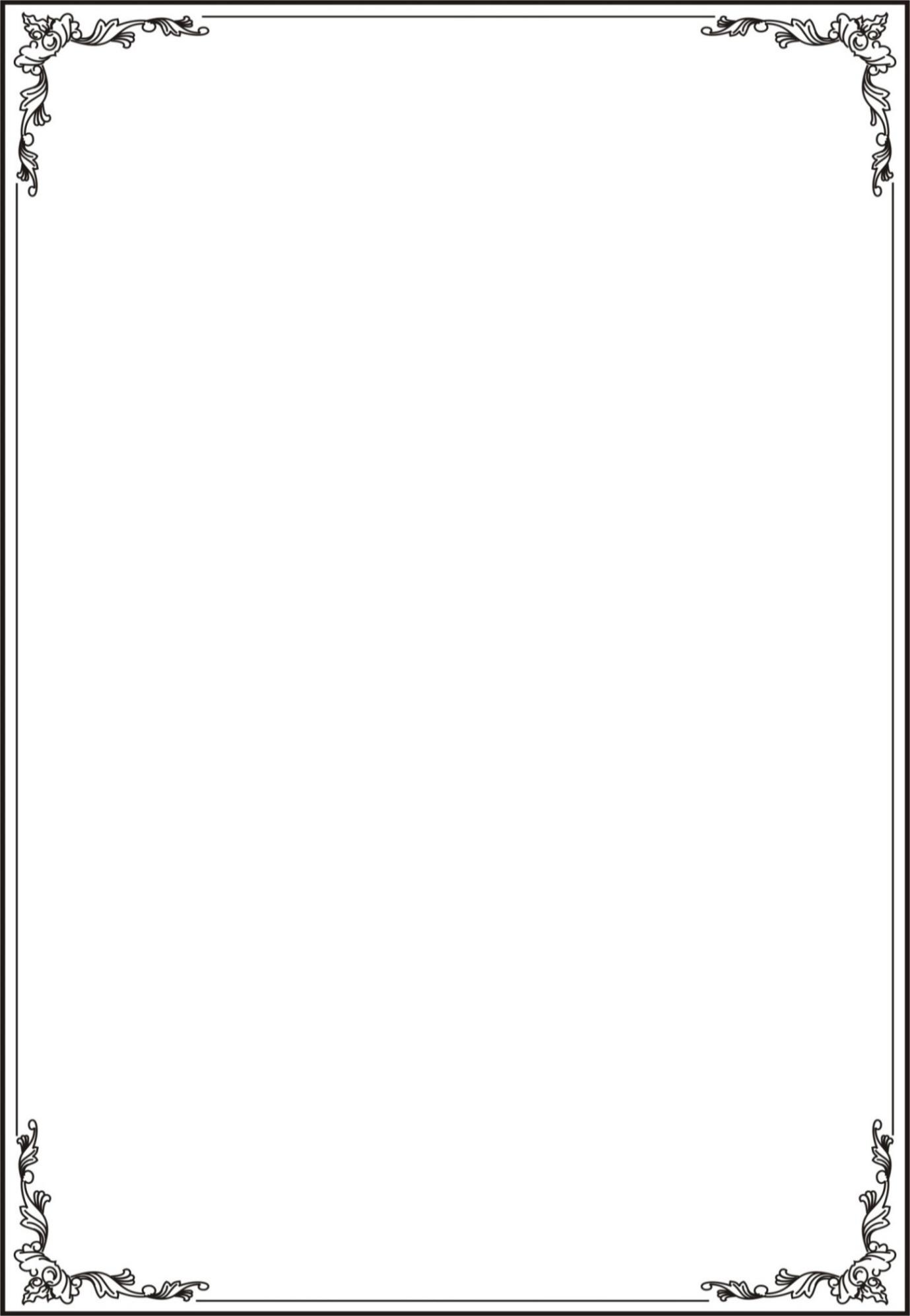
*** ĐẠI HỌC ĐÀ NẴNG***

***TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG VIỆT – HÀN***



***ĐỒ ÁN CƠ SỞ 4***

***Đề tài :***

***Phân tích tìm hiểu và xây dựng trợ lý ảo***

***Sinh viên : Nguyễn Trọng Tâm 18IT5***

***Mã sinh viên : 18IT294***

***Giáo viên hướng dẫn: TS.Nguyễn Sĩ Thìn***

***Đà nẵng tháng 12 năm 2020***

***ĐẠI HỌC ĐÀ NẴNG***

***TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG VIỆT – HÀN***



***ĐỒ ÁN CƠ SỞ 4***

***Đề tài :***

***Phân tích tìm hiểu và xây dựng***

***trợ lý ảo***

***Sinh viên : Nguyễn Trọng Tâm 18IT5***

***Mã sinh viên : 18IT294***

***Giáo viên hướng dẫn: TS.Nguyễn Sĩ Thìn***

***Đà nẵng tháng 12 năm 2020***

**LỜI CẢM ƠN**

Lời đầu tiên xin giữ lời cảm ơn tới thầy cô trong Trường Đại học công nghệ thông tin và truyền thông Việt – Hàn đã tạn tình giảng dạy, giúp đỡ và truyên đạt trong suốt kỳ học vừa qua. Em cũng cin cảm ơn các thầy cô bộ môn khoa học máy tính đã mang lại cho em những kiến thức vô cùng quý giá và bổ ích trong quá trình học tập.

Đặc biệt xin chân thành cảm ơn đến thấy giáo TS. Nguyễn Sĩ Thìn, người đã tận tâm, giúp đỡ, định hướng, trực tiếp hướng dẫn và tận tình chỉ bảo em trong suốt quá trình thực hiện đồ án cơ sở 4 này. Xin chân thành cảm ơn

**NHẬN XÉT**

**(Của giáo viên hướng dẫn)**

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………………………………………………………………………………………………………

………………………………………………………………………………………………

Xác nhận giáo viên hướng dẫn

*(Ký và ghi rõ họ tên)*

**MỤC LỤC**

[LỜI MỞ ĐẦU 4](#_Toc62195403)

[CHƯƠNG 1 : TỔNG QUAN 5](#_Toc62195404)

[1.1 Đặt vấn đề 5](#_Toc62195405)

[1.2 Mục tiêu 6](#_Toc62195406)

[1.3 Nội dung nghiên cứu 6](#_Toc62195407)

[1.4 Cấu trúc đồ án 6](#_Toc62195408)

[CHƯƠNG 2 : CƠ SỞ LÝ THUYẾT 8](#_Toc62195409)

[2.1 Trí tuệ nhân tạo là gì 8](#_Toc62195410)

[2.2 Giới thiệu về trợ lý ảo 8](#_Toc62195411)

[2.3 Nhận dạng giọng nói tự động (ASR – Automatic Speech Recogntion) 9](#_Toc62195412)

[2.3.1 Nhận dạng giọng nói là gì 9](#_Toc62195413)

[2.3.2 Nhận dạng giọng nói tiếng việt 11](#_Toc62195414)

[2.4 Xữ lý âm thanh 17](#_Toc62195415)

[2.5 Xữ lý hình ảnh IMM (Image Matching) 18](#_Toc62195416)

[2.6 Quản lý câu hỏi trả lời QA (Question Answering) 19](#_Toc62195417)

[2.6.1 Cấu tạo của quản lý câu hỏi 19](#_Toc62195418)

[2.6.2 Xử lý ngôn ngữ tự nhiên trong Quản lý câu hỏi trả lời 20](#_Toc62195419)

[*2.5.3* Các mức của NLP 20](#_Toc62195420)

[*2.6*  Ứng dụng NLP trong chatbot 22](#_Toc62195421)

[2.6.1 Cấu tạo hệ thống chatbot 22](#_Toc62195422)

[2.6.2 Cấu trúc hội thoại chatbot 22](#_Toc62195423)

[2.6.3 Xữ lý dữ liệu trong Chatbot 23](#_Toc62195424)

[CHƯƠNG 3 : TRIỂN KHAI THỰC HIỆN 25](#_Toc62195425)

[3.1 Cấu trúc trợ lý ảo 25](#_Toc62195426)

[3.2 Hoạt động của trợ lý ảo 26](#_Toc62195427)

[3.3 Xây dựng chương trình 28](#_Toc62195428)

[3.3.1 Các thư viện cần thiết 29](#_Toc62195429)

[3.3.2 Chức năng chuyển văn bản thành âm thanh 30](#_Toc62195430)

[3.3.3 Chức năng chuyển âm thanh thành văn bản 30](#_Toc62195431)

[3.3.4 Chức năng gửi Email 32](#_Toc62195432)

[3.3.5 Chức năng xem dự báo thời tiết 34](#_Toc62195433)

[3.3.6 Một số chức năng khác 35](#_Toc62195434)

[3.4 Ứng dựng thực tế của trợ lý ảo 36](#_Toc62195435)

[CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 37](#_Toc62195436)

[4.1 Kết luận 37](#_Toc62195437)

[4.2 Hạn chế trong đê tài 37](#_Toc62195438)

[4.3 Hướng phát triển 37](#_Toc62195439)

[TÀI LIỆU THAM KHẢO 37](#_Toc62195440)

**DANH MỤC HÌNH VẼ**

[Hình 2.1: Sơ đồ khối tổng quan của một hệ thống nhận dạng tiếng nó 11](#_Toc61600029)

[Hình 2.2: Các đường đặc tính của 6 thanh điệu tiếng Việt 13](#_Toc61600030)

[Hình 2.3: Đường pitch của câu nói “Nhận dạng tiếng Việt” 16](#_Toc61600031)

[Hình 2.4: Sơ đồ xữ lý âm thanh 18](#_Toc61600032)

[Hình 2.5: Sơ đồ xữ lý hình ảnh 19](#_Toc61600033)

[Hình 2.6: Quản lý câu hỏi trả lời 20](#_Toc61600034)

[Hình 2.7: Mô hình cơ bản của hệ thống chatbot 22](#_Toc61600035)

[Hình 2.8: Kiến trúc seq2seq 25](#_Toc61600036)

[Hình 3.1 : Sơ đồ hệ thống trợ lý ảo 26](#_Toc61600037)

[Hình 3.2: Sơ đồ hoạt động của trợ lý ảo 27](#_Toc61600038)

[Hình 3.3: Google Cloud Text To Speech 29](#_Toc61600039)

[Hình 3.4: Quy trình chung hệ thống nhận dạng giọng nói 30](#_Toc61600040)

[Hình 3.5: Quý trình hệ thống giữ Email 30](#_Toc61600041)

[Hình 3.6: Dự báo thời tiết OpenWeatherMap 31](#_Toc61600042)

**DANH MỤC BẢNG BIỂU**

[Bảng 2.1: Cấu trúc âm tiết tiếng Việt 12](#_Toc61600043)

[Bảng 2.2: Ví dụ cấu trúc ngữ âm của âm tiết “chuyển” 12](#_Toc61600044)

[Bảng 2.3: Kết quả nhận dạng của hệ thống cơ sở 15](#_Toc61600045)

# LỜI MỞ ĐẦU

Ngày nay, trí tuệ nhân tạo đang ngày càng phát triển mạnh mẽ. Các hãng lớn như Apple, Microsoft, Google đều đưa ra các công nghệ tương tác trực tiếp với người dùng. Trí tuệ nhân tạo đang ngày càng được nâng cấp hoàn thiện giúp người dùng dễ dàng tương tác, dễ sử dụng và giảm quá trình thực hiện. Trí tuệ nhân tạo đang là một lĩnh vực mới mẻ và được sự quan tâm rất lớn từ các hãng công nghệ hàng đầu. Với công nghệ đang ngày càng được áp dụng trong đời sống giúp con người làm việc hiệu quả hơn tiết kiệm thời gian và sức lực, trí tuệ nhân tạo như một hệ thống được xây dựng để phục vụ cho điều đó.

Hệ thống trợ lý ảo là một hệ thống giúp con người giao tiếp với máy thực hiện các yêu cầu ý muốn của người dùng. Hiện nay trên thế giới hệ thống trợ lý ảo đang được nghiên cứu và phát triển mạnh mẽ ở các hãng công nghệ hàng đầu thế giới. Với mong muốn hiểu sâu về trí tuệ nhân tạo, em quyết định chọn đề tài “Tìm hiểu xây dựng trợ lý ảo” làm đề tài đồ án cuối kỳ. Qua đề tài em muốn nâng cao sự hiểu biết về trí tuệ nhân tạo đồng thời nghiên cứu để có thể áp dụng thực tiễn tại Việt Nam.

Nội dung báo cáo được chia ra làm 5 phần như sau:

* Chương 1 : Tổng quan – Nêu lên vấn đề làm đồ án, những mục tiêu, các nội dung tìm hiểu của để tài
* Chương 2 : Cơ sở lý thuyết – Tổng hợp những nội dụng lý thuyết liên quan đến đề tài
* Chương 3 : Phân tích thiết kế hệ thống – Trình bày cách thức trợ lý ảo hoạt động.
* Chương 4 : Triển khai xây dựng – Nêu lên quy trình làm sản phẩm (Xây dựng trợ lý ảo)
* Chương 5 : Kết luận và hướng phát triển

# CHƯƠNG 1 : TỔNG QUAN

## 1.1 Đặt vấn đề

Với sự phát triển của khoa học công nghệ, sự phát triển nhanh chóng của AI và học máy đã giúp sự phát triển công nghệ nhận dạng giọng nói, công nghệ này ngày càng đi sâu vào mọi lĩnh vực trong cuộc sống.

Từ những điều trên, cũng với những phân tích bên dưới, ta sẽ chứng minh một cách sinh động rằng tại sao công nghệ điều khiển giọng nói nên được coi là xu thế công nghệ hiện tại không chỉ bới những ông lớn về IT hay những người có niềm đam mê với AI mà các doanh nghiệp cũng đang ráo riết nâng cấp, tối ưu hệ thống bằng công nghệ giọng nói hay còn gọi là trợ lý ảo.

Trợ lý ảo (có thể gọi là trợ lý kỷ thuật số, trợ lý giọng nói hay là trợ lý AI )là một ứng dụng lập trình hướng nhiện vụ, nhận dạng giọng nói của con người và thực hiện các lệnh được phát âm bỡi người dùng. Nền tảng của nó là AI và năng suất của nó dựa vào việc lưu trữ hàng triệu từ và hàng triệu cụm từ, Không giống như các thiết bị nhận dạng giọng nói đầu tiên mà các nhà khoa học đang nghiên cứu vào những năm 40-50 của thế kỷ trước, các trợ lý kỉ thuật số hiện đại không bị hạn chế bởi một mẫu ngôn ngữ hoặc từ vựng nhất định.Có hai lại phần mền trợ lý giọng nói phổ biến là trợ lý ảo tại nhà và trợ lý ảo tại công sở.

Loại đầu tiên thuộc về tự động hóa ngôi nhà (còn gọi là **nhà thông minh**), một hệ thống cho phép điều khiển thiết bị, ánh sáng, điện, thiết bị và những thứ khác lấp đầy trong nhà của chúng ta. Nhà thông minh đòi hỏi phải có kết nối internet và là một phần của Internet vạn vật (**IoT**). Loại thứ hai, trợ lý ảo cho nơi làm việc, có thể được sử dụng ngoài liên lạc với tự động hóa tòa nhà, như một ứng dụng độc lập được kết hợp như thói quen của mọi người trong nhóm.

Ba gã khỗng lồ đầu tiên trong sản xuất sản phẩm công nghê trợ lý ảo hàng đầu bao gồm Siri được phát triển bởi Apple Inc, công ty độc quyền cung cấp các thiêt bị của Apple, Google Assistant từ Google Inc, được phát triển cho hiệu điều hành Android Amazon Alexa, với một ứng dụng tương thích cho cả IOS và Android. Tất nhiên, có rất nhiều dụng trợ lý giọng nói nhỏ hơn nhưng vẫn được sự dụng trên thị trường.

Hiện nay với nền khoa học phát triển với các trợ lý ảo thông minh hỗ trợ AI như: Google Assistant (Google), Alexa (Amazon), Siri (Apple), Cortana (Microsoft), … Các thiết bị trong ngôi nhà thông minh ngoài được điều khiển qua các app trên điện thoại, web,… còn được điều khiển qua trợ lý ảo cũng dần được phát triển. Vì vậy, em đã tìm hiểu, nghiên cứu, chọn đề tài: “ Tìm hiểu xây dựng trợ lý ảo” để làm đồ án cuối kỳ của mình.

Kết quả nghiên cứu từ đề tài này sẽ giúp em có nhiều kinh nghiệm để sau khi tốt nghiệp chúng em có đủ khả năng nghiên cứu chế tạo hoàn chỉnh thiết bị điều khiển hệ thống điện cho ngôi nhà thông minh đáp ứng được sử dụng yêu cầu trên thi ̣trường với giá thành hợp ̣ lý, chất lượng đảm bảo, phù hợp ̣ với điều kiện sống tại Việt Nam

## 1.2 Mục tiêu

* Tìm hiểu về hệ thống điều khiểu thông minh bằng giọng nói của các trợ lý ảo đã có.
* Tìm hiểu và phân tích cách hệ thống trợ lý ảo hoạt động
* Tìm hiểu ứng dụng thực tế của trợ lý ảo
* Xây dựng và demo sản phẩm trợ lý ảo
* Kiểm tra, đánh giá tính ứng dụng của đề tài

## 1.3 Nội dung nghiên cứu

- Nghiêm cứu tài liệu qua sách báo về lĩnh vữa AI đặc biệt về xữ lý ngôn ngữ tự nhiên

- Tìm hiểu nguyên lý hoạt động của các trợ lý ảo đã có như : Google Assistant, Siri, Cortana, Bixby,…

- Tìm hiểu lý thuyết để xây dựng trợ lý ảo

## 1.4 Cấu trúc đồ án

Với đề tài " TÌM HIỀU XÂY DỰNG TRỢ LÝ ẢO" thì bố cục đồ án như sau:

* Chương 1 : Tổng quan – Nêu lên vấn đề làm đồ án, những mục tiêu, các nội dụng nghiên cứu và các hạn chế của để tài
* Chương 2 : Cơ sở lý thuyết – Tổng hợp những nội dụng lý thuyết liên quan đến đề tài
* Chương 3 : Phân tích thiết kế hệ thống – Trình bày cách thức trợ lý ảo hoạt động.
* Chương 4 : Triển khai xây dựng – Nêu lên quy trình làm sản phẩm (Xây dựng trợ lý ảo)
* Chương 5 : Kết luận và hướng phát triển

# CHƯƠNG 2 : CƠ SỞ LÝ THUYẾT

## 2.1 Trí tuệ nhân tạo là gì

Trí tuệ nhân tạo hay còn gọi là AI - [**Artificial Intelligence**](https://codelearn.io/sharing/tri-tue-nhan-tao-se-di-den-dau) là một ngành khoa học, kỹ thuật chế tạo máy móc thông minh, đặc biệt là các chương tình máy tính thông minh. AI được thực hiện bằng các nghiên cưu cách suy nghĩ của cong người cách con người học hỏi, giải quyết và làm việc trong khi giải quyết một vấn đề nào đó, và sữ dụng các kết quả nghiên cứu này như là một nền tảng  để phát triển các phần mềm và hệ thống thông minh, từ đó áp dụng vào các mục đích khác nhau trong cuộc sống. Nói một cách dễ hiểu thì AI là việc sử dụng, phân tích các dữ liệu đầu vào nhằm đưa ra sự dự đoán rồi đi đến quyết định cuối cùng.

Là một trí tuệ do con người lập trình tạo nên với mục tiêu giúp máy tính có thể tự động hóa các hành vi thông minh như con người hoặc thậm chí là hơn con người**.** Trí tuệnhân tạo khác với việc lập trình logic trong các ngôn ngữ lập trình là ở việc ứng dụng các hệ thống học máy (machine learning) để mô phỏng trí tuệ của con người trong các xử lý mà con người làm tốt hơn máy tính.

## 2.2 Giới thiệu về trợ lý ảo

Siri của Apple Google now của Google và Cortana của Microsoft đại diện cho một lớp dịch vụ web là các ứng dụng hỗ trợ cá nhân thông minh. IPA là một ứng dụng sữ dụng yếu tố đầu vào tiếng nói, hình ảnh thông tin theo ngữ cảnh của con người dùng để hỗ trợ bằng cách trả lời câu hỏi bằng ngôn ngữ tự nhiên, đưa ra khuyết nghị và hành động. Các IPA đang nỗi tiếng với dịch vụ Internet nhanh nhất vì gần đây họ đã phát triển trên các nền tảng nổi tiếng như IOS, Android và Windows, Phone, làm cho chúng phổ biến trên các thiết bị di động nỗi tiếng trên toàn thế giới.

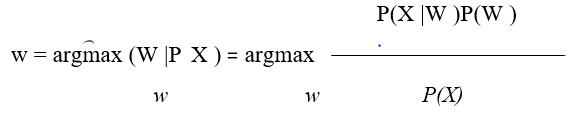
Sữ dụng IPA đang gia tăng nhanh chóng với các sản phẩm mới nhất trong các công nghệ có thể như đồng hồ thông minh và kính thông minh. Sự gia tăng thị trường phần này, cùng với việc thiết kế các sản phẩm đeo có thể phụ thuộc nhiều đầu vào hình ảnh và giọng nói, cho thấy nhu cầu sữ dụng IPA đang tăng lên nhanh chóng. IPA khác với nhiều khối lượng côgn việc dịch vụ web hiện có trong các máy tính có quy mô. Ngược lại với các truy vấn của các dịch vụ trung tâm duyệt web, luồng truy vân IPA thông qua các thành phần phần mền thúc đẩy những tiến bộ gần đây trong nhận dạng giọng nói, xữ lý ngôn ngữ tự nhiên và tầm nhìn máy tính để cung cấp người dùng một câu hỏi dựa trên câu hỏi và hệ thống trả lời. Do cường độ tính toán của các thành phần này và các mô hình dữ liệu dựa trên mô hình mà họ sữ dụng, các nhà cung cấp dịch vụ yêu cầu tính toán trong các nền tảng trung tâm dữ liệu dữ liệu lớn thay cho việc tính toán trên các tính bị di động. Cách tiếp cận giảm tải này được sữ dụng cả Siri của Appple lần Google Now của Google khi họ gữi các bản ghi nén lệnh truy vấn bằng giọng nói tới các trung tâm dữ liệu nhận dạng tiếng nói cà khai thác ngữ nghĩa.

Sự tương tác giữa thiết bị thông minh với con người không còn nằm tở việc gõ văn bản mà giờ là cả hình ảnh lẫn giọng nói, để đáp ứng được nhu cầu ngày càng cao của trải nghiệm người dùng, các hãng công nghệ thông tin lớn đang ngày càng đầu tư mạnh vào hệ thống trợ lý ảo.

## 2.3 Nhận dạng giọng nói tự động (ASR – Automatic Speech Recogntion)

### 2.3.1 Nhận dạng giọng nói là gì

Nhận dạng giọng nói là quá trình biến đổi tín hiệu âm thanh thu được của  
người nói thành một chuỗi các từ có nội dung tưng ứng dưới dạng văn bản. Nếu  
gọi tín hiệu giọng nói thu được trên miền thời gian là s(t) thì s(t) đầu tiên sẽ  
được rời rạc hóa để xử lý và trích chọn ra các thông tin quan trọng. Kết quả thu  
được là một chuỗi các vector đặc trưng tương ứng 𝑋 = {𝑥1, 𝑥2, 𝑥3, … , 𝑥𝑁}. Sau  
đó nhiệm vụ của hệ thống nhận dạng giọng nói là tìm ra một chuỗi các từ, 𝑊̂ =  
{𝑤1, 𝑤2, 𝑤3, … , 𝑤𝐿} có nội dung tương ứng với X về mặt ngữ nghĩa. Công thứ  
3.1 mô tả mô hình toán học của một hệ thống nhận dạng giọng nói theo nguyên  
lý xác suất của Bayes. Hầu hết các hệ thống nhận dạng giọng nói thống kê ngày  
nay đều dựa trên mô hình này

  
Trong đó *P(W)* là xác suất của chuỗi *W,* giá trị này có thể được tính toánthông qua một mô hình ngôn ngữ n-gram và nó hoàn toàn độc lập với tín hiệu  
tiếng nói X, P(X\W) là xác suất để X là W được xác định thông qua mô hình âmhọc (acoustic model). Giá trị *P(X)* có thể được bỏ qua do giá trị của nó khôngthay đổi trong một bộ dữ liệu cụ thể với tất cả các chuỗi dự đoán *W*

Các hệ thống nhận dạng giọng nói hiện nay có thể được phân loại theo các cách như sau:

* Nhận dạng các từ phát âm rời rạc hoặc liên tục.
* Nhận dạng giọng nói phụ thuộc hoặc không phụ thuộc người nói.
* Nhận dạng với hệ thống từ vựng nhỏ (vài trăm từ) hoặc từ vựng lớn (hàng nghìn từ).
* Nhận dạng giọng nói trong môi trường nhiễu cao hoặc thấp

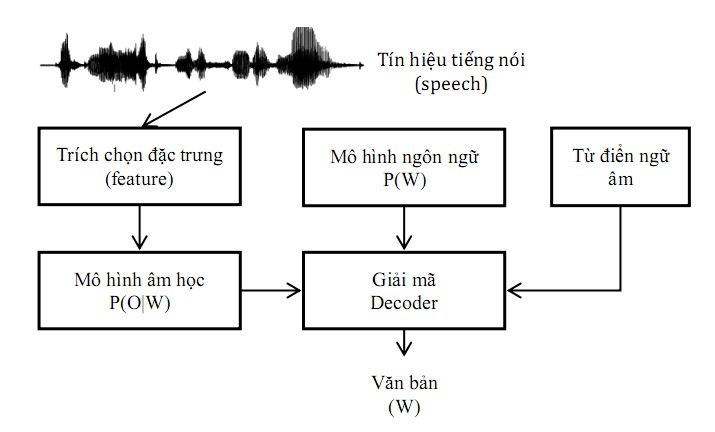
**Ứng dụng**

Cùng với sự phát triển nhanh chóng của các thiết bị tính toán tốc độ cao như máy tính, điện thoại thông minh, vi xử lý – vi điều khiển, ngày nay nhận dạng giọng nói được ứng dụng cho rất nhiều các lĩnh vực trong cuộc sống. Có thể kể đến một số ứng dụng thông dụng như hệ thống dịch tiếng nói tự động, giao tiếp robot, tóm tắt tiếng nói, …

**Các vấn đề trong nhận dạng giọng nói**

Nhận dạng giọng nói là một dạng bài toán trong lĩnh vực nhận dạng mẫu, vì vậy cũng tồn tại những khó khăn tương tự như các bài toán nhận dạng khác. Ngoài ra còn tồn tại một số vấn đề khác do đặc tính biến đổi ngẫu nhiên của tín hiệu tiếng nói. Các vấn đề chính ảnh hưởng đến độ chính xác và hiệu suất làm việc của một hệ thống nhận dạng giọng nói có thể kể đến như nhiễu và sự khác nhau trong tiếng nói: Vấn đề phụ thuộc người nói; Vấn đề về tốc độ phát âm, hiện tượng đồng phát âm; Vấn đề về kích thước của bộ từ vựng (từ điển); Vấn đề nhiễu; Vấn đề về ngôn ngữ;

**Các thành phần chính của một hệ thống nhận dạng tiếng nói**



##### Hình 2.1: Sơ đồ khối tổng quan của một hệ thống nhận dạng tiếng nó

### 2.3.2 Nhận dạng giọng nói tiếng việt

Nhìn chung tính đến hiện nay các nghiên cứu về nhận dạng giọng nói tiếng Việt vẫn còn rất hạn chế. Phần lớn các nghiên cứu mới chỉ dừng lại ở nhận dạng số hoặc nhận dạng các từ phát âm rời rạc. Tiếng Việt là một ngôn ngữ có thanh điệu, như vậy một hệ thống nhận dạng đầy đủ sẽ phải bao gồm 2 thành phần là nhận dạng âm vị và nhận dạng thanh điệu. Đã có một só nghiên cứu về nhận dạng thanh điệu cho tiếng Việt, tuy nhiên các nghiên cứu này mới chủ yếu tập trung vào việc phân tích đặc tính và tìm ra mô hình phù hợp trong việc mô hình hóa và nhận dạng thanh điệu đơn lẻ. Hầu hết chưa tích hợp việc nhận dạng thanh điệu với nhận dạng âm vị để tạo thành một hệ thống hoàn chỉnh

#### 2.3.2.1 Tổng quan về tiếng việt

Tiếng việt là một ngôn ngữ đơn âm tiết (Monosyllable), nghĩa là mỗi một âm tiết được thể hiện bởi một từ và cũng là đơn vị cơ bản trong phát âm. Các đặc tính chính của âm tiết tiếng Việt như sau:

* ***Âm tiết tiếng Việt có tính độc lập cao***

Âm tiết là đơn vị cơ bản trong hệ thống các đơn vị ngôn ngữ. Mỗi âm tiết đều có khả năng được thể hiện bởi một từ không biến hình, hay nói cách khác một âm tiết cũng đồng thời là một hình vị. Về mặt ý nghĩa và ngữ pháp trong tiếng Việt được thể hiện chủ yếu bằng trật tự giữa các từ. Như vậy tiếng Việt khác với một số ngôn ngữ khác như tiếng Anh, Pháp,… là luôn có ranh giới rõ ràng giữa hai âm tiết.

* ***Âm tiết tiếng Việt có khả năng biểu hiện ý nghĩa***

Hầu hết các âm tiết tiếng Việt khi đứng một mình đều có khả năng biểu hiện một ý nghĩa xác định. Như vậy âm tiết tiếng Việt ngoài vai trò là một đơn vị ngữ âm nó còn có vai trò về từ vựng và ngữ pháp.

* ***Âm tiết tiếng Việt có cấu trúc chặt chẽ***

Tất cả các âm tiết tiếng Việt đều có thể phân tích thành một cấu trúc gồm năm thành phần như sau:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Thanh điệu | |  |
| Âm đầu | Vần | |  |
| Âm đệm | Âm chính | Âm cuối |

###### Bảng 2.1: Cấu trúc âm tiết tiếng Việt

Ví dụ cấu trúc của âm tiết từ “chuyển” có thể được phân tích thành 5 thành phần như sau:

|  |  |  |  |
| --- | --- | --- | --- |
| Thanh điệu (Thanh hỏi) | | |  |
| Âm đầu (Ch) | Vần (uyên) | |  |
| Âm đệm  (u) | Âm chính  (yê) | Âm cuối (n) |

###### Bảng 2.2: Ví dụ cấu trúc ngữ âm của âm tiết “chuyển”

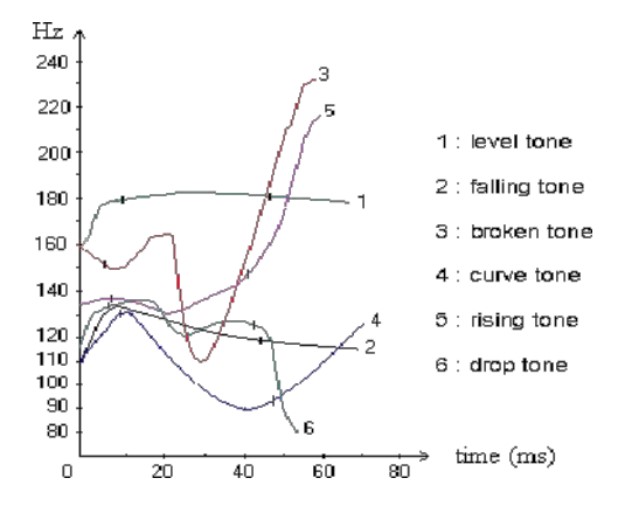
**a) Âm vị tiếng Việt**

Âm vị (phoneme) là đơn vị trừu tượng nhỏ nhất của một ngôn ngữ. Mọi âm tiết trong một ngôn ngữ đều được tạo ra bằng tổ hợp của các âm vị. Trong nhận dạng tiếng nói từ vựng lớn thì âm vị thường được chọn như là đơn vị của hệ thống nhận dạng với mục đích giảm số lượng đơn vị trong hệ thống. Trong khoảng thời gian ngắn từ 10ms đến 40ms có thể coi tín hiệu âm thanh là ổn định và đây cũng là khoảng thời gian phổ biển cho một âm vị. Vì vậy chọn âm vị là đơn vị nhận dạng còn là để giảm ảnh hưởng của sự biến đổi của tín hiệu tiếng nói. Dựa theo cấu trúc của âm tiết tiếng Việt thì hệ thống âm vị của tiếng

Việt bao gồm 21 âm đầu, 1 âm đệm, 16 âm chính và 8 âm cuối.

#### 2.3.2.2 Thanh điệu tiếng Việt

Về mặt hình thức nếu không xét đến sự biến đổi thanh điệu trên các phụ âm dừng ở cuối âm tiết thì tiếng Việt có 6 thanh điệu. Bao gồm thanh huyền, ngã, hỏi, sắc, nặng và thanh bằng (thể hiện trong chữ viết là không dấu).



##### Hình 2.2: Các đường đặc tính của 6 thanh điệu tiếng Việt

* **Thanh bằng:** Là thanh điệu cao, có đường đặc tính bằng phẳng như đường số 1 trong hình 2.2
* **Thanh huyền:** Là thanh điệu thấp, đường đặc tính có dạng bằng phẳng tương tự thanh bằng nhưng phần cuối có phần đi xuống thấp hơn như thể hiện ở đường số 2 trong hình 2.2.
* **Thanh ngã:** Đường số 3 hình 2.2, đường đặc tính của thanh ngã biến đổi từ ngang, thấp rồi cao;
* **Thanh hỏi:** Đường số 4 hình 2.2, là thanh thấp và có đường đặc tính gãy ở giữa;
* **Thanh sắc:** Đường số 5 hình 2.2, đường đặc tính của thanh sắc có hướng đi lên;
* **Thanh nặng:** Đường số 6 hình 2.2, là thanh thấp có đường đặc tính đi xuống.

Đối với các âm tiết kết thúc bởi các phụ âm đóng “p, k, t” thì các âm tiết này có xu hướng kết thúc nhanh hơn so với các âm tiết khác, chính vì thế hai thanh sắc và thanh nặng (trong tiếng Việt chỉ có hai thanh này tồn tại với các âm tiết kết thúc bằng các phụ âm đóng “p, k, t”) cũng có xu hướng kết thúc nhanh hơn khi cùng với các âm tiết khác. Trong trường hợp này có thể coi tiếng Việt có 8 thanh điệu.

#### 2.3.2.3 Mô hình nhận dạng tiếng Việt

Xét một hệ thống nhận dạng có bộ từ vựng *W =* *{W },i i* *= (1,...,N)*kích thước *N.*

Xét một ngôn ngữ *L* có tập từ vựng là W\* . Mục tiêu của đề tài là xây dựng mô hình hệ thống có thể nhận dạng từ trong *L* (trong phạm vi đề tài này *L* là tiếng Việt). Khi đó ta sẽ có *W\* Ξ W*. Nếu mô hình hóa mỗi một từ *Wi* bởi một mô hình λ*i* thì kích thước của hệ thống sẽ là *N.* Trong thực tế thì *N* thường rất lớn, và rất khó để có thể liệt kê hết tất cả các từ của W\*. Nguyên nhân là do hạn chế của người liệt kê, tính chất vùng miền, tính chất thế hệ hoặc theo sự phát triển của văn hóa, công nghệ thì vẫn luôn có những từ mới được bổ sung vào W\*. Như vậy nhược điểm của phương pháp này là kích thước hệ thống vô cùng lớn và hệ thống không có khả năng nhận dạng *Wj* nếu  *Wj thuộc* W\*nhưng  *Wj*  không thuộc W.

#### 2.3.2.4 Hệ thống nhận dạng cơ sở (Baseline)

Các nghiên cứu đã có về nhận dạng tiếng Việt hiện mới chỉ áp dụng mô hình HMM trên các loại đặc trưng phổ biến là MFCC hoặc PLP trên bộ từ vựng kích thước nhỏ cỡ vài trăm từ hoặc trên tiếng nói phát âm rời rạc. Tính đến hiện nay vẫn chưa có một công bố nào về bộ dữ liệu chuẩn cho huấn luyện và đánh giá chất lượng hệ thống chung cho cộng đồng nghiên cứu nhận dạng tiếng Việt. Nghĩa là các thử nghiệm của các nghiên cứu đã khó có thể so sánh với nhau do không cùng các điều kiện tiêu chuẩn như dữ liệu thử nghiệm, đầu ra của hệ thống nhận dạng. Vì thế để có thể so sánh và đánh giá chất lượng của các phương pháp mới trong đề tài thì một hệ thống nhận dạng cơ sở (Baseline) ban đầu cần được xây dựng. Hệ thống cơ sở này được xây dựng dựa trên mô hình không có thanh điệu. Tức là bộ âm vị không có thanh điệu sẽ được sử dụng làm đơn vị nhận dạng và được mô hình hóa bởi mô hình HMM truyền thống với các tham số chính sau:

Đặc trưng đầu vào: MFCC/PLP. Trong đó kích thước của mỗi vector đặc trưng MFCC/PLP là 39 bao gồm 13 thành phần MFCC/PLP, 13 thành phần Delta và 13 thành phần Acceleration của MFCC/PLP.

* Mô hình HMM: Được huấn luyện ở mức triphone với 2179 âm buộc (tiedstates).
* Từ điển: Từ điển sử dụng tập âm vị không có thông tin thanh điệu có 45 âm vị.
* Mô hình ngôn ngữ: 2-gram được xây dựng từ dữ liệu phiên âm của VOV.
* Kết quả đánh giá theo tham số độ chính xác theo từ ACC (Word Accuracy) đạt 77.7% với đặc trưng MFCC.

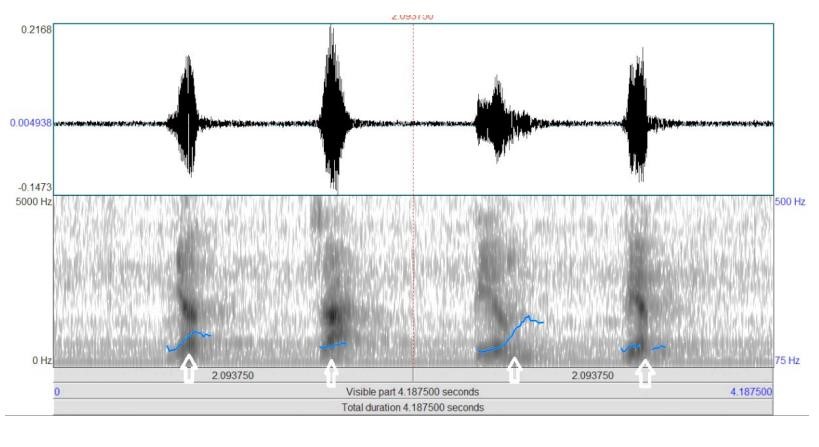
Kết quả nhận dạng trên dữ liệu kiểm thử VOV – test được đánh giá theo tham số độ chính xác theo từ ACC (Word Accuracy) trên hai loại đặc trưng PLP và MFCC được trình bày ở bảng sau:

|  |  |  |
| --- | --- | --- |
| Hệ thống | Đặc trưng | ACC (%) |
| Sys1 (Baseline) | MFCC | 77.70 |
| Sys2 | PLP | 76.77 |

###### Bảng 2.3: Kết quả nhận dạng của hệ thống cơ sở

Như vậy đặc trưng MFCC cho chất lượng nhận dạng tốt hơn PLP là 0.93% theo ACC

2.3.2.5 Đặc trưng thanh điệu và vấn đề không liên tục của dữ liệu

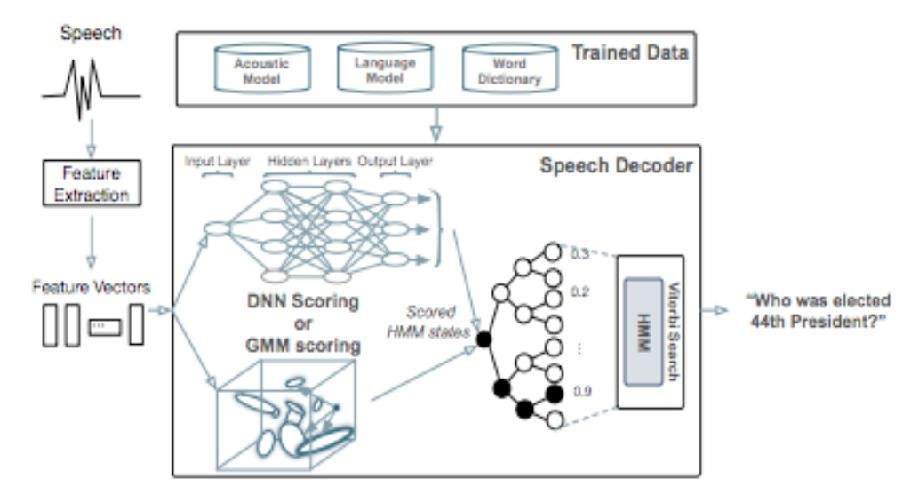
Hình 2.3: Đường pitch của câu nói “Nhận dạng tiếng Việt”

Thanh điệu được tạo ra do dao động của dây thanh. Tuy nhiên dây thanh chỉ dao động đối với các âm hữu thanh vì vậy mà trong vùng âm vô thanh không tồn tại thanh điệu. Nếu xét trong cả một câu phát âm thì đường đặc trưng của thanh điệu sẽ bị đứt gãy tại các vùng vô thanh. Hình 2.3 mô tả đường đặt trưng thanh điệu không liên tục của câu nói “nhận dạng tiếng Việt”. Để có thể mô hình hóa đặc trưng thanh điệu sử dụng mô hình HMM hoặc mạng nơron thì đặc trưng này cần phải đượcáp dụng một kỹ thuật tiền xử lý trước để bổ sung các giá trị cho các vùng đứt gãy. Biện pháp đơn giản nhất là thay thế các vùng đứt gãy bằng giá trị 0. Hoặc có thể áp dụng một số kỹ thuật là trơn khác. Tuy nhiên việc áp dụng các kỹ thuật khác để bổ sung giá trị vào vùng mà thanh điệu không tồn tại sẽ làm biến đổi đặc trưng này và đặc trưng mới không còn thể hiện đúng đắn đặc tính đó. Đối với các ngôn ngữ không có thanh điệu như tiếng Anh, Pháp đặc trưng thanh điệu chỉ làm tăng thông tin về ngữ điệu, người nói, giới tính,… do nó không làm thay đổi ngữ nghĩa âm tiết. Vì thế việc thay đổi đặc trưng thanh điệu bằng việc bổ sung các giá trị ”nhân tạo” cũng có thể chấp nhận được hoặc thậm chí có thể bỏ qua đặc trưng này khi xây dựng các hệ thống nhận dạng giọng nói. Đối với tiếng Việt do thanh điệu còn ảnh hưởng trực tiếp đến ngữ nghĩa của từ, vì vậy việc thay đổi nó có thể làm giảm chất lượng nhận dạng. Như vậy cần phải có một phương pháp mô hình hóa sao cho có thể mô hình hóa được đặc tính thanh điệu bị đứt gãy để mô tả đúng nhất đặc tính của nó trong việc góp phần thay đổi ngữ nghĩa trong tiếng Việt.

Tính đến hiện nay có rất nhiều cách nghiên cứu đề xuất các kỹ thuật để trích chọn đặc trưng thanh điệu thông qua việc tính toán tần số cơ bản (F0) từ tín hiệu tiếng nói. Mục tiêu là đề xuất một mô hình có khả năng mô hình hóa loại đặc tính đứt gãy hay nói cách khác là mô hình được loại đặc trưng đầu vào chứa cả giá trị liên tục và giá trị rời rạc. Sử dụng hai phương pháp trích chọn đặc trưng thanh điệu được sử dụng phổ biến là đặc trưng về độ lệch biên độ trung bình (AMDF- Average Magnitude Difference Function) và đặc trưng giá trị tương quan chéo đã chuẩn hóa (NCC – Normalized Cross Correlation). NCC tính toán đặc trưng thanh điệu bằng hương pháp tương quan chéo. Cả NCC và AMDF đều thay thế các giá trị ở vùng vô thanh bằng giá trị 0. Mục đích việc sử dụng 2 loại đặc trưng này là muốn kiểm chứng chất lượng của mô hình và xác định loại đặc trưng nào trong hai phương pháp NCC và AMDF phù hợp với loại mô hình này.

## 2.4 Xữ lý âm thanh

Khối nhận dạng giọng nói tự động là khối đầu vào giúp con người dùng tương tác với trợ lý ảo bằng giọng nói. Các đầu vào cho ASR là các vector đặc trưng đại diện cho đoạn nói, được tạo ra bở quá trình xữ lý nhanh vào truy xuất đặc tính của bài nói. Thành phần của ASR dựa vào sự kết nói của mô hình Hidden Markov, một mô hình hỗn hợp Gaussian hoặc một trạng thái nơ-ron sâu. Hidden Markov xây dựng một cây trạng thái cho khung lời nói hiện tại sữ dụng vector đầu vào. Gaussian hoặc nơ-ron đánh giá xuất sắc củasự chuyển đổi trạng thái trong cây, và thuật toán Viterbi sau đó tìm kiếm con đường có khả năng nhất dựa trên các điểm này. Đường dẫn có xác xuất cao nhất thể hiện kết xuất văn bản cuối cùng. Điểm Gaussian đánh giá Hidden Markov trạng thái chuyển tiếp bắng cách ánh xạ một vector đối tượng đầu vào mọt hệ tọa độ đa chiều và lặp lại điểm số các đặc tính chống lại mô hình âm thanh được đào tạo. Nơ- ron sâu điểm số sử dụng xác xuất từ một mảng thần kinh. Độ sâu của nơ-ron sâu được xác định bởi số lớp ẩn mà số điểm cho một chuyển tiếp đi qua mạng.

**

*Hình 2.4: Sơ đồ xữ lý âm thanh*

## 2.5 Xữ lý hình ảnh IMM (Image Matching)

Khối xữ lý hình ảnh là khối giúp con người tương tác với trợ lý ảo thong qua hình ảnh. Khi người dùng muốn hỏi trợ lý ảo về một hình ảnh nào đó chẳng hạn bông hoa hồng thì trợ lý ảo sẻ xữ lý hình ảnh đầu vào cung cấp đầu ra là thông tin về hình ảnh đó. Từ dữ liệu hình ảnh đưa vào hệ thống sẻ cố gắng tìm kiếm hình ảnh trong cơ sỡ dữ liệu trước khi chế biến và trả về thông tin và kết hợp với hình ảnh. Cơ sở dữ liệu được sữ dụng trong hệ thống là cơ sở dữ liệu Mobile Visual Search. Các điểm chính của hình ảnh được rút ra lần đầu tiên từ hình ảnh đầu vào sữ dụng thuật toán SURF. Trong tính năng chiếu xuất, hình ảnh được lấy mẫu vào xáo lộn nhiều lần để tìm kiếm điều thú vị ở các mô hình khác nhau. Khi ngưỡng đáp ứng lưu trữ các điểm chính của hình ảnh, các điểm then chốt được chuyển tới bộ phận mô tả tính năng, nơi chúng được chỉ định một vecto định hướng và các điểm chính được định hướng tương tự được nhóm thành các bộ mô tả tính năng. Quá trình này làm giảm biến đổi giữa các hình ảnh đầu vào, tăng cơ hội tìm kiếm sự kết hợp chính xác. Bộ mô tả hình ảnh từ đầu vào được đối sánh với bộ mô tả nhóm trước đại diện cho các hình ảnh cơ sở dữ liệu bằng cách sữ dụng tìm kiếm gần nhất, hình ảnh cơ sở dữ liệu với số lượng lớn nhất của cơ sở được trả lại.

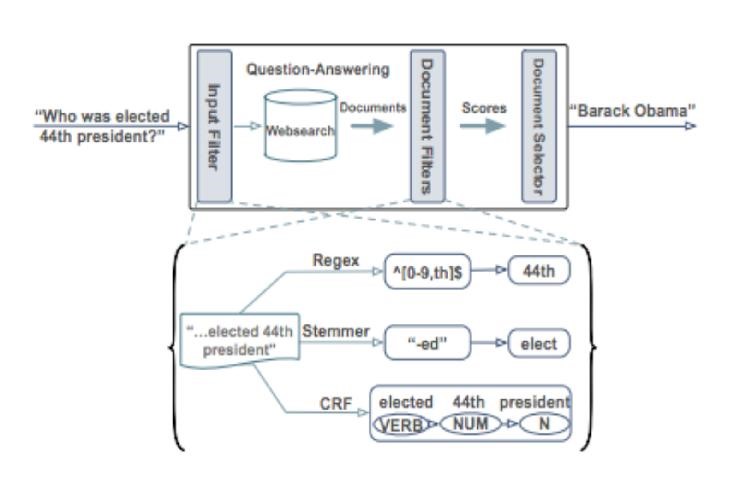


Hình 2.5: Sơ đồ xữ lý hình ảnh

## 2.6 Quản lý câu hỏi trả lời QA (Question Answering)

### 2.6.1 Cấu tạo của quản lý câu hỏi

Khối quản lý câu hỏi trả lời là khối xử lý dữ liệu tương tác ở trong hệ thống trợ lý ảo. khi người dùng hỏi hệ thống trợ lý ảo thì nó sẽ lấy dữ liệu từ khối QA để đưa ra câu trả lời phù hợp. Văn bản được xuất ra từ ASR hoặc nhập trực tiếp được chuyển đến OpenEphyra sử dụng ba quy trình cốt lõi đẻ trích xuất thông tin văn bản, bắt nguồn từ, kết hợp cụm từ thông dụng, và gắn thẻ theo từng thành phần của văn bản . Hình dưới mô tả sơ đồ OE kết hợp thành phần này, tạo ra các truy vấn tìm kiếm web và lọc các kết quả trả về. Thuật toán Porter (stemmer) phơi bày gốc của một từ bằng cách kết hợp và cắt ngắn các từ thông dụng, OpenEphyra cũng sử dụng một bộ mẫu biểu thức chính quy để đối sánh các từ truy vấn chung (như cái gì, ở đâu …) và lọc bất kỳ ký tự đặc biệt nào trong đầu vào. Trình phân loại ngẫu nhiên (CRF) lấy một câu, vị trí của mỗi từ trong câu và nhãn các từ hiện tại và trước đó làm đầu vào để đưa ra các dự đoán về phần nói của từng từ của một truy vấn đầu vào. Mỗi truy vấn đầu vào được phân tích cú pháp bằng cách sử dụng các thành phần nói trên để tạo truy vấn cho công cụ tìm kiếm. Tiếp theo, các bộ lọc sử dụng các kỹ thuật tương tự được sử dụng để lấy thông tin từ các tài liệu trả về; tài liệu với điểm tổng thể cao nhất sau khi tổng hợp số điểm được trả về là câu trả lời hay nhất.



##### Hình 2.6: Quản lý câu hỏi trả lời

### 2.6.2 Xử lý ngôn ngữ tự nhiên trong Quản lý câu hỏi trả lời

Theo Wikipedia, NLP (Natural Language Processing) là một nhánh của trí tuệ nhân tạo tập trung vào các ứng dụng trên ngôn ngữ của con người. Trong trí tuệ nhân tạo thì xử lý ngôn ngữ tự nhiên là một trong những phần khó nhất vì nó liên quan đến việc phải hiểu ý nghĩa ngôn ngữ - công cụ hoàn hảo nhất của tư duy và giao tiếp.

Theo cách hiểu thì có thể hiểu xử lý ngôn ngữ tự nhiên là một phạm vi lý thuyết các kỹ thuật tính toán để phân tích và mô tả các văn bản xảy ra tự nhiên ở một hoặc nhiều mức độ phân tích ngôn ngữ theo yêu cầu của con người mong muốn.

Mục tiêu của NLP là nhằm thể hiện ý nghĩa thực sự và ý định của người dùng khi thao tác dữ liệu. Điển hình ứng dụng NLP:

* Giải thích văn bản đầu vào
* Dịch văn bản sang một ngôn ngữ khác
* Trả lời các câu hỏi về nội dung của một văn bản
* Thu thập các suy luận từ văn bản

### Các mức của NLP

Ngữ âm học: mức này liên quan tới việc giải thích các âm thanh nói trong và giữa các từ. Có ba loại quy tắc được sử udnjg trong phân tích âm vị học: quy tắc âm thanh trong từ, quy tắt ngữ âm trong biến thể phát âm khi từ được nói với nhau, quy tắt biến động trong ngữ điệu của một câu. Một hệ thống NLP hỗ trợ đầu vào nói, song âm là phân tích và mã hóa tính hiệu thành tín hiệu số hóa để giải thích các quy tắc khác nhau hoặc bằng việc so sánh với mô hình ngôn ngữ cụ thể được sử dụng.

Hình thái học: mức độ này liên quan tới bản chất cấu thành của các từ bao gồm các đơn vị nhỏ nhất của ý nghĩa. Ví dụ từ preregistration có thể được phân tích thành trước tiền tố, gốc “registra” và hậu tố. Vì ý nghĩa của mỗi hình thái vẫn giữ nguyên qua các từ ngữ, con người có thể phân chia một từ không rõ thành các hình thái cấu thành để hiểu ý nghĩa của nó. Tương tự trong NLP có thể nhận ra ý nghĩa được truyền đạt bởi mỗi hình thái có thể đạt được và đại diện cho ý nghĩa. Ví dụ thêm hậu tố vào một động từ cho biết hành động của động từ diễn ra trong quá khứ.

Từ vựng học: con người hay hệ thống NLP diễn giải ý nghĩa của từng từ.

Thuật ngữ: tập trung vào việc phân tích các từ trong một câu để khám phá ra ngữ pháp cấu trúc của câu. Điều này đòi hỏi cả ngữ pháp và trình độ phân tích cú pháp. Kết quả của việc này là đại diện của một câu. Các mối quan hệ phụ thuộc cấu trúc giữa các từ. có nhiều ngữ pháp khác nhau có thể được sử dụng và do đó sẽ ảnh hưởng đến sự lựa chọn của một trình phân tích cú pháp. Ví dụ có câu “con chó đuổi con mèo” và “con mèo đuổi theo con chó” khác nhau về về ý nghĩa.

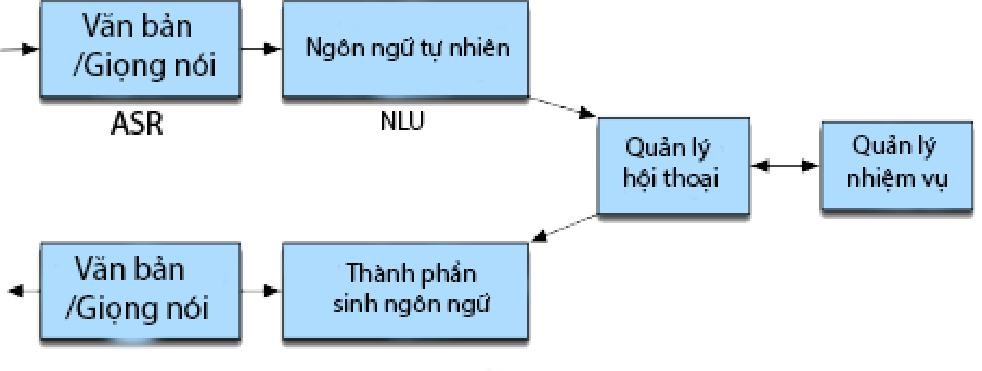
Ngữ nghĩa: mức độ mà hầu hết mọi người nghĩ rằng ý nghĩa được xác định tuy nhiên chúng ta có thể xem trong xác định ở trên của các cấp, đó là tất cả các cấp có ý nghĩa góp phần vào. Xử lý ngữ nghĩa xác định ý nghĩa của một câu bằng cách tập trung vào tương tác giữa các ý nghĩa cấp từ trong câu. Mức độ này bao gồm việc định hướng ngữ nghĩa của các từ với nhiều giác quan, theo cách tương tự để cách phân định cú pháp của các từ có thể hoạt động như nhiều phần cảu bài phát biểu là hoàn thành các cấp cú pháp. Ví dụ trong các nghĩa khác “file” là một danh từ có thể có nghĩa là một thư mục hoặc một công cụ để tạo. Đàm luận: mặc dù cú pháp và ngữ nghĩa làm việc với các đơn vị câu, mức độ diễn đạt của NLP làm việc với các đơn vị văn bản dài hơn một câu. Nghĩa là nó không giải thích văn bản như các câu ghép nối, mỗi câu có thể được giải thích đơn lẻ. Thay vào đó bài diễn thuyết tập trung vào tính chất của văn bản tập trung kết nối giữa các thành phần câu.

Thực dụng: liên quan đến việc sử dụng có mục đích ngôn ngữ trong các tình huống và sử dụng bối cảnh trên các nội dung của văn bản để hiểu mục đích là để giải thích làm thể nào thêm ý nghĩa được đọc vào văn bản. Điều này đòi hỏi nhiều kiến thức bao gồm sự hiểu về ý định, kế hoạch và những mục tiêu.

Hệ thống NLP hiện nay có khuynh hướng thực hiện thành các mô đun để đạt được mức độ yêu cầu. Mức độ thấp thì sử dụng mô đun thấp, mức độ cao thì dùng nhiều mô đun thấp kết hợp.

## *2.6* Ứng dụng NLP trong chatbot

### 2.6.1 Cấu tạo hệ thống chatbot

**

##### Hình 2.7: Mô hình cơ bản của hệ thống chatbot

Chatbot có thể hiểu là các hệ thống có thể thực hiện các cuộc hội thoại mở với mục đích bắt chước tính năng đàm thoại phi cấu trúc hoặc trò truyện của con người với máy. Hình 2.6.1 mô tả về cấu trúc mô hình của một chatbot . Trong mô hình chatbot kiểu truy xuất thông tin (retrieval-based), thành phần ngôn ngữ tự nhiên sử dụng rất nhiều kĩ thuật xử lý ngôn ngữ tự nhiên để phân tích câu hội thoại, xác định ý định của người dùng, trích xuất những thông tin của người dung cung cấp. Sau đó những thông tin này được chuyển tới cho thành phần quản lý hội thoại (Dialogue manager). Thành phần quản lý hội thoại sẽ xử lý thông tin và chuyển tới cho để sinh ra câu hội thoại.

### 2.6.2 Cấu trúc hội thoại chatbot

Bao gồm:

* Nội dung dành cho người dùng thao tác nhập liệu: ghi âm giọng nói, nhập văn bản, thêm file như hình ảnh, âm thanh, địa chỉ.
* Nơi hiển thị: giao diện tương tác là nơi hiển thị nội dung người dùng nói và nơi Chatbot trả lời.

### 2.6.3 Xữ lý dữ liệu trong Chatbot

***a) Tập văn bản***

Corpus-based chatbots dựa trên cơ sở thay vì dựa trên các quy tắc được xây dựng bằng tay, các cuộc hội thoại giữa con người với con người, hoặc đôi khi là phản ứng của con người từ cuộc trò chuyện. Serban et al đã tổng hợp một số tập văn có sẵn như vậy chẳng hạn như cuộc trò chuyện trên nền tảng trò truyện facebook, twitter hoặc trong hội thoại phim, có sẵn với số lượng lớn và đã được cho thấy giống với cuộc trò chuyện tự nhiên. Câu trả lời của Chatbot thậm trí có thể được trích ra từ các câu trong tập tin của văn bản không phải từ hộp thoại.

Có 2 loại chatbots dựa trên tập lệnh: các hệ thống dựa trên việc thu thập thông tin, và các hệ thống dựa trên việc học máy có giám sát dựa trên sự truyền tải trình tự.

Chatbots hầu hết đều dựa trên cơ sở có xu hướng làm mô hình rất ít của ngữ cảnh đối thoại. Thay thế điều đó có xu hướng tập trung vào việc tạo ra một biến đáp ứng duy nhất phù hợp với lời nói trước của người sử dụng, vì lý do này chúng thường được gọi là phản ứng hệ thống thế hệ và chỉ tập trung vào các phản hồi đơn lẻ trong khi bỏ qua ngữ cảnh hoặc mục tiêu đàm thoại.

***b) Thu thập thông tin (IR Chatbot - Information retrival chatbot)***

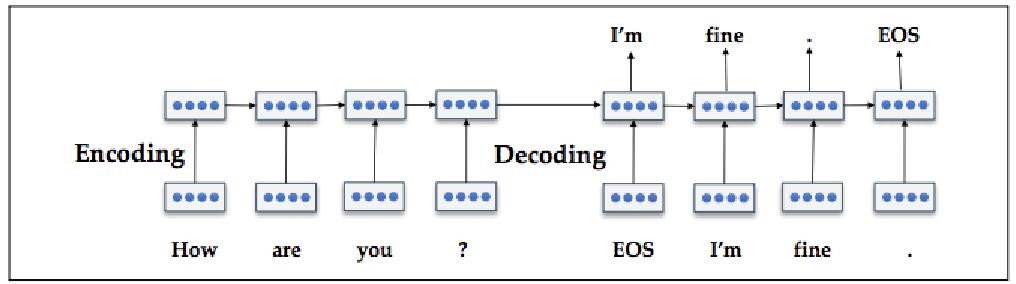
Nguyên tắc đằng sau việc thu thập thông tin dựa trên chatbot [8] là đáp ứng các yêu cầu của người sử dụng biến x bằng cách lặp lại một số biến y phù hợp từ một tập hợp của văn bản tự nhiên.

Sự khác biệt giữa các hệ thống như vậy nằm ở cách họ chọn tập văn bản và cách họ quyết định xem những gì được coi là sự biến đổi thích hợp của con người để sao chép. Sự lựa chọn chung của tập văn bản là thu thập các cơ sở dữ liệu về các cuộc đối thoại của con người. Cách tiếp cận khác là sử dụng tập tin trong hội thoại phim. Khi Chatbot được đưa vào thực tế các bước mà con người sử dụng để phản hồi lại chatbot có thể được sử dụng như dữ liệu cuộc hội thoại bổ sung cho việc đào tạo. Với câu của người dùng các hệ thống dựa trên IR có thể sử dụng bất kỳ truy xuất thuật toán để lựa chọn một phản ứng thích hợp từ hệ thống. Trong mỗi trường hợp, bất kỳ chức năng tương tự có thể được sử dụng, tính toán phổ biến nhất hoặc qua các từ hoặc trên các phần nhúng.

***c) Trình tự liên tiếp trong chatbot (Sequence to sequence chatbots)***

Mô hình seq2seq bao gồm hai mạng RNN : một cho bộ mã hoá, và một cho bộ giải mã. Bộ mã hoá nhận một chuỗi (câu) đầu vào và xử lý một phần tử (từ trong câu) tại mỗi bước. Mục tiêu của nó là chuyển đổi một chuỗi các phần tử vào một vector đặc trưng có kích thước cố định mà nó chỉ mã hoá thông tin quan trọng trong chuỗi và bỏ qua các thông tin không cần thiết. Có thể hình dung luồng dữ liệu trong bộ giải mã hoá dọc theo trục thời gian, giống như dòng chảy thông tin cục bộ từ một phần tử kết thúc của chuỗi sang chuỗi khác.

Mỗi trạng thái ẩn ảnh hưởng đến trạng thái ẩn tiếp theo và trạng thái ẩn cuối cùng được xem như tích luỹ tóm tắt về chuỗi. Trạng thái này được gọi là bối cảnh hay vector suy diễn, vì nó đại diện cho ý định của chuỗi. Từ bối cảnh đó, các bộ giải mãi tạo ra một chuỗi, một phần tử tại một thời điểm. Ở đây, tại mỗi bước các bộ giải mã bị ảnh hưởng bởi bối cảnh và các phần tử được sinh ra trước đó.



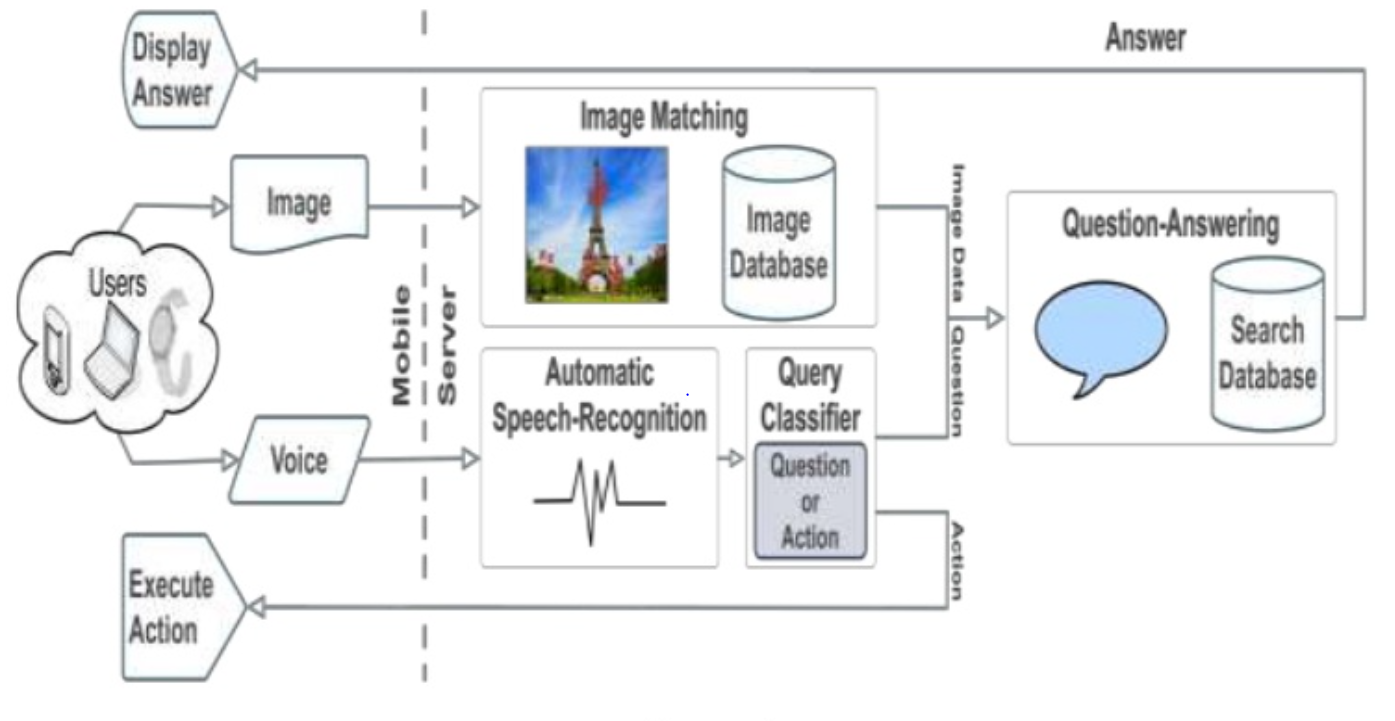
##### Hình 2.8: Kiến trúc seq2seq

Quá trình tìm hiểu về cấu trúc của một mô hình hệ thống trợ lý ảo, có thể thấy để xây dựng một hệ thống trợ lý ảo cần rất nhiều các mô đun khác nhau ghép lại, từ đầu vào cần xử lý dữ liệu, sau đó truyền vào hệ thống xử lý dữ liệu và cuối cùng là trả lại dữ liệu cho người dùng. Hệ thống trợ lý ảo thông minh là nhờ khâu xử lý dữ liệu cực kỳ quan trọng và sẽ được tìm hiểu sâu hơn ở chương tiếp theo.

# CHƯƠNG 3 : TRIỂN KHAI THỰC HIỆN

## 3.1 Cấu trúc trợ lý ảo

Để hiểu được một trợ lý ảo hoạt động như thế nào, cần phải biết cấu trúc và cấu tạo của hệ thống trợ lý ảo. Sợ đồ của một trợ lý ảo:



##### Hình 3.1 : Sơ đồ hệ thống trợ lý ảo

Dựa vào hình 3.1, có thể cấu tạo của một trợ lý ảo bao gồm 2 khối : Khối người dùng (User) có đầu vào là ảnh, giọng nói, văn bản và khối xữ lý dữ liệu (Server) gồm các khối xữ lý hình ảnh, văn bản, âm thanh, khối quản lý hội thoại.

Truy vấn bắt đầu bằng giọng nói, văn bản, hình ảnh của người dùng thiết bị thông minh như điện thoại, thiết bị đeo, kính thông minh. Các file hình ảnh, âm thanh, văn bản được giữ tới máy chủ để xữ lý.

Với giọng nói được xữ lý bằng giao diện ASR (Auto matic Speech Recogniton) chuyển câu hỏi nói của người dùng sang văn bảng tương ứng bằng mô hình thống kê. Sau đó văn bản đi qua quý trình phân loại truy vấn quyết định bài nói có phải là hành động hay câu hỏi không. Nếu là hành động, lệnh sẻ giữ lại cho thiết bị động để thực hiện. Nếu không thì là câu hỏi bằng văn bản thuần túy. Sữ dụng các kỷ thuật xữ lý ngôn ngữ tự nhiên (Natural Language Processing), dịch vụ câu hỏi trả lời (Question - Answering) sẻ trích xuất thông tin từ đầu vào, tìm kiếm cơ sở dữ liệu của nó và chọn câu trả lời tốt nhất để trả lại cho người dùng.

Nếu là một hình ảnh đi kèm với đầu vào giọng nói, hệ thống sẻ sử dụng kỷ thuật thị giác máy tính để cố gắng phù hợp với hình ảnh đầu vào cơ sở dữ liệu của nó.

## 3.2 Hoạt động của trợ lý ảo

Các ứng dụng trợ lý giọng nói hoạt động dựa trên hệ thống Nhận dạng giọng nói tự động (ASR). Các hệ thống ASR ghi lại lời nói và sau đó chia nhỏ thành các âm vị, sau này được xử lý thành văn bản. Một âm vị (không phải từ của âm tiết) là một đơn vị đo lường cơ bản để nhận dạng giọng nói của con người. Nhận dạng âm vị mang lại kết quả tốt hơn quá trình giải mã từ, vì người cuối cùng có xu hướng phân tích từ dưới dạng một đơn vị độc lập bỏ qua các giới hạn ngữ cảnh.



##### Hình 3.2: Sơ đồ hoạt động của trợ lý ảo

Có thể sử dụng loại phần mềm nhận dạng giọng nói nào, tất cả đều dựa trên ASR. Nếu bạn đã từng tự hỏi làm thế nào để tạo ra một phần mềm trợ lý ảo, điều quan trọng nhất cần làm là làm quen với cách ASR hoạt động. Tóm lại, quá trình bắt đầu với việc thiết bị thu thập âm thanh với micrô. Các dạng sóng giọng nói đã ghi được chuyển thẳng sang phân tích âm thanh, được thực hiện ở ba cấp độ khác nhau:

- Mô hình âm thanh, đại diện cho những âm vị được phát âm và những từ mà các âm vị này hoàn thành là gì;

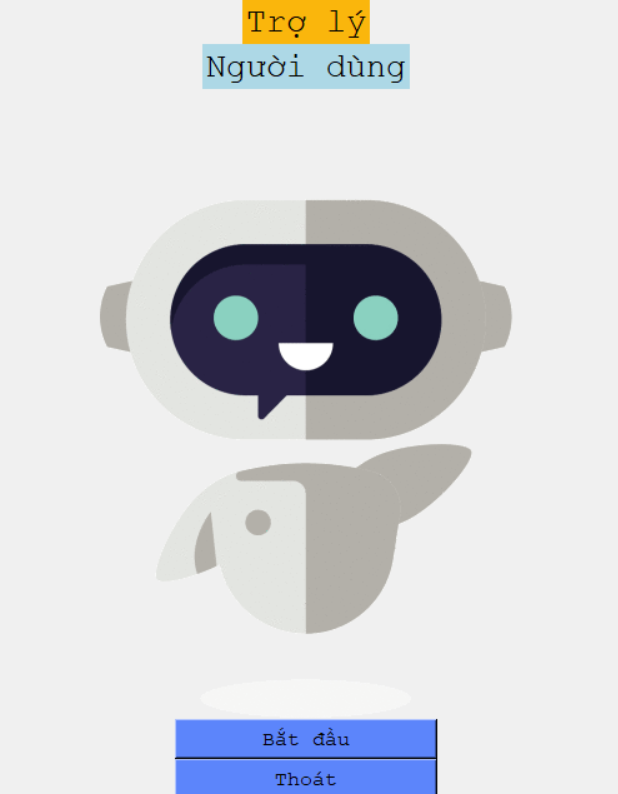
- Mô hình phát âm, phân tích cách phát âm của âm vị, có bất kỳ trọng âm hoặc đặc thù nào khác của bộ máy phát âm để nắm bắt sự biến đổi ngữ âm của lời nói;

- Mô hình hóa ngôn ngữ, nhằm mục đích tìm kiếm xác suất theo ngữ cảnh tùy thuộc vào âm vị nào được ghi lại.

Tất cả các dữ liệu được xử lý bởi AI mà không cần sự tương tác của con người. Chúng ta giảm tỷ lệ lỗi xuất hiện bằng cách sử dụng các thuật toán học máy. Dữ liệu dạng sóng giọng nói sau đó được truyền đến bộ giải mã, nơi cuối cùng nó chuyển thành văn bản để sử dụng thêm như lệnh hoặc chính tả.

Trí thông minh nhân tạo mang đến cho các ứng dụng trợ lý giọng nói hiện đại sự tự do không dựa vào vốn từ vựng hạn chế, mà sử dụng **lưu trữ đám mây** với hàng triệu từ và cụm từ thay thế. Nói cách khác, học máy làm cho các ứng dụng nghe toàn bộ bài phát biểu, không phải mỗi từ riêng biệt. Bằng cách đó, các ứng dụng giọng nói sẽ **phân tích bối cảnh** và xác suất để xác định những gì bạn đang cố gắng nói. Ví dụ: **Google**sử dụng một mạng nơ-ron kỹ thuật số tương tự như các mạng trong các tế bào não người. Về cơ bản, mạng lưới thần kinh nhân tạo này mô phỏng hoạt động thần kinh của bộ não của chúng ta và bắt chước các quá trình bắt đầu khi chúng ta nghe thấy dòng âm thanh liên tục. Tuy khác nhau về cấu trúc, ba trợ lý phổ biến nhất hiện nay Google Assistant, Conrtana và Siri đều dựa trên nền tảng là nghê công nghệ mạng nơ-ron sâu ở phần backend.

## 3.3 Xây dựng chương trình

Hình 3.3: Giao diện trợ lý ảo

### 3.3.1 Các thư viện cần thiết

Ngôn ngữ lập trình: Python 3.9

Các thư viện chính được áp dụng trong xây dựng trợ lý ảo và một só chắc năng chính để xây dựng trợ lý ảo

* **speech\_recognition**: Nhận dạng giọng nói
* **time, datetime**: Xử lý thời gian
* **wikipedia:** Tìm kiếm trên từ điển wikipedia
* **webbrowser, selenium, webdriver\_manager, urllib:** Truy cập web, trình duyệt (Chrome)
* **gTTS**: Chuyển văn bản thành âm thanh của Google (Chị Google)
* **requests:** Crawl thông tin từ web
* **smtplib:** Gửi Email bằng giao thức SMTP
* **re:** Biểu thức chính quy (Regular Expression)
* **os, sys, ctypes:** Truy cập, xử lý file hệ thống
* **playsound:** Phát âm thanh từ file mp3
* **json:** Xử lý kiểu dữ liệu JSON
* **youtube\_search:** Tìm kiếm video trên Youtube

Hiện tại, các nghiên cứu về trí tuệ nhân tạo ở Việt Nam đang đầu tư đẩy mạnh nghiên cứu mảng xử lý ngôn ngữ tự nhiên (Natural Language Processing) ở dạng tiếng việt và đã đạt được những kết quả đáng ghi nhận. Các mô hình nghiên cứu được có thể xử lý ngôn ngữ tiếng việt ở nhiều vùng miền, cả nam lẫn nữ, rất thực. Ở góc độ người sử dụng phổ thông thì khó có thể biết đâu là do máy tính, đâu là do người thực hiện. Các nghiên cứu đều có tính ứng dụng cao, đang được các công ty công nghệ áp dụng triển khai thực tế, nâng cao năng suất lao động một cách rõ rệt. Các thư viện được sử dụng ở trên đều phổ thông, tích hợp nhiều hàm xử lý. Tuy chất lượng xử lý tiếng việt chưa thực sự tuyệt vời nhưng củng rất là khá ổn

### 3.3.2 Chức năng chuyển văn bản thành âm thanh

##### Hình 3.3: Google Cloud Text To Speech

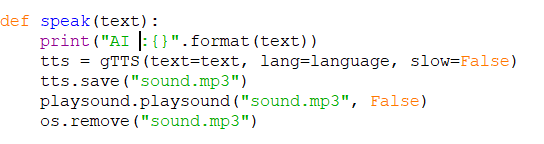
Nhờ thư viện **gTTS** (Chuyển văn bản thành âm thanh của Google ) mô hình có thể bắt chước giọng người với nhiều âm thanh khác nhau. Các yếu tố chính nhất trong công nghệ này là:

- Ngôn ngữ học bao gồm âm vị học, hình thái học và ngữ dụng học

- Vật lý học thì có âm học

- Công nghệ: Xử lý ngôn ngữ tự nhiên, học máy, xử lý tín hiệu số …

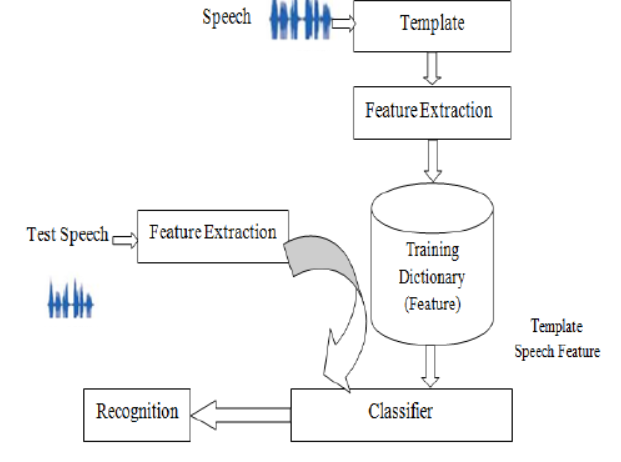
Code:

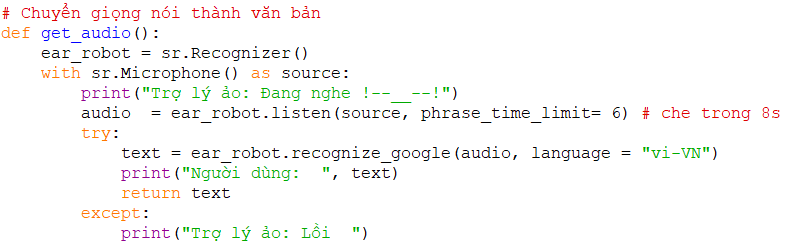


- Hàm *gTTS (google Text To Speech)* để chuyển văn bản thành âm thanh theo ngôn ngữ nhận dạng tiếng việt rồi lưu về máy tính dữ liệu âm thanh dưới file *sound.mp3*

*- playsound.playsound()* để đọc file *sound.mp3* trên máy tính. Sau khi đọc xong xóa file *sound.mp3* để tránh lỗi khi mình đọc một đoạn văn bản khác thì cũng được lưu lại dưới file *sound.mp3.*

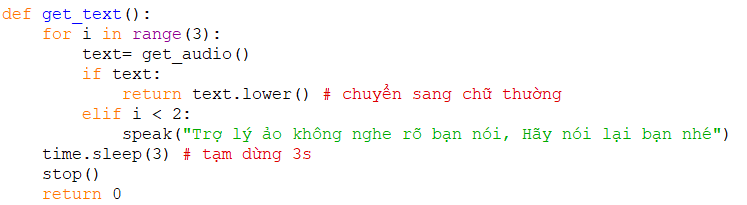
### 3.3.3 Chức năng chuyển âm thanh thành văn bản

 Hình 3.4: Quy trình chung hệ thống nhận dạng giọng nói



- Thư viện *speech\_recognition (sr)* có chức năng là nhận dạng giọng nói để chuyển âm thanh thành văn bản. Âm thanh được đọc vào *microphone*của máy tính sau đó được xử lý qua hàm *listen* của *sr.Recognition* rồi lưu dữ liệu âm thanh vào biến *audio.* Dữ liệu âm thanh audio thu được sẽ được nhận dạng ở ngôn ngữ tiếng việt trong hàm *r.recognize\_google* để chuyển thành dạng văn bản rồi lưu dữ liệu vào biến text.

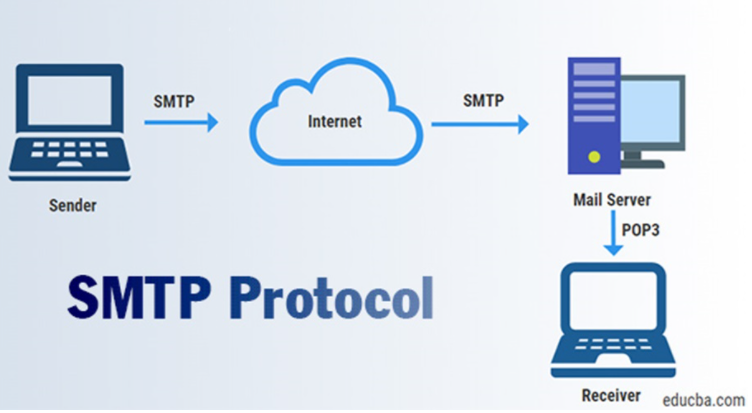
- Nếu dữ liệu âm thanh *audio* không lỗi tức là hàm r.recognize\_google có thể nhận dạng được *audio* để chuyên thành *text* thì hàm *get\_audio()* sẽ được trả về giá trị là *text* còn nếu dữ liệu *audio* bị lỗi mà hàm *r.recognition\_google* không nhận dạng được thì hàm *get\_audio()* sẽ được trả về giá trị là *lỗi* (Mục đích là khi máy tính không hiểu mình nói gì thì mình sẽ thực hiện lại hàm *get\_audio()* để đọc lại)



- Hàm get\_text() có chức năng là máy tính sẽ cố gắng nhận dạng âm thanh của người đọc tối đa 3 lần cho đến khi máy tính hiểu. Sử dụng hàm for lặp lại 3 lần, nếu đoạn *text* có giá trị khác *0* thì hàm *get\_text* sẽ được trả về giá trị *text.lower()* (Chuyển chữ cái in hoa thành in thường), còn nếu *text* nhận giá trị là *0* (Tức là minh đọc mà máy tính không hiểu) mà chưa đọc đến lần thứ 3 thì mình sẽ yêu cầu người sử dụng đọc lại. Nếu sau 3 lần mà máy tính vẫn không hiểu thì mình nói gì hay không nghe thấy gì thì mình sẽ cho dừng lại sử dụng hàm *stop* và hàm *get\_text* lúc này sẽ được trả về giá trị là*0* (Mục đích là khi máy tính không nghe thấy gì thì mình sẽ cho dừng chương trình luôn).

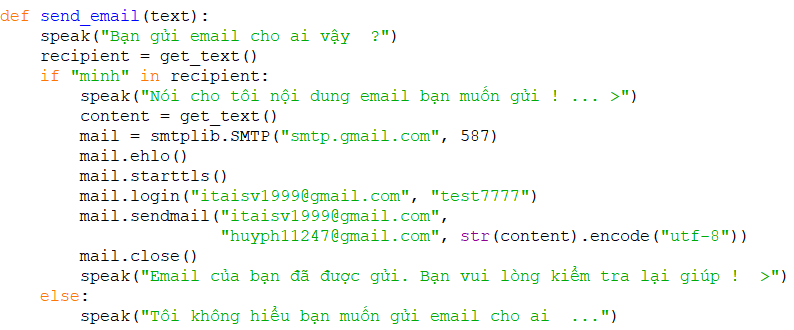


### 3.3.4 Chức năng gửi Email

Hình 3.5: Quy trình hệ thống giữ Email

**SMTP** là 3 chữ cái đầu viết tắt của Simple Mail Transfer Protocol dịch ra có nghĩa là giao thức truyền tải thư tín đơn giản hóa. Và giao thức này thực hiện nhiệm vụ chính là gửi mail còn việc nhận mail hay truy xuất dữ liệu mail server sẽ có giao thức IMAP hay POP3 đảm nhiệm.

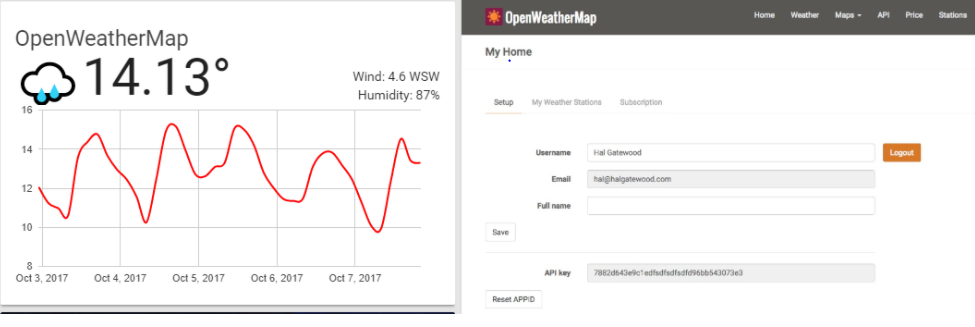
**SMTP Server** (server dùng để gửi mail) là một dịch vụ cho phép gửi email với số lượng lớn, tốc độ nhanh mà không bị giới hạn như các hòm mail miễn phí của Gmail hoặc mail đi kèm hosting. Nói cách khác các máy chỉ chủ giúp bạn thao tác gửi thư người ta thường gọi là SMTP server chúng thực hiện gửi thư qua giao thức TCP hoặc IP.



- Thư viện *smtplib* để thực hiện chức năng gửi Email bằng phương thức *smtp*. Lấy tên người gửi sử dụng bằng hàm *get\_text()* rồi lưu vào biến *recipient.*

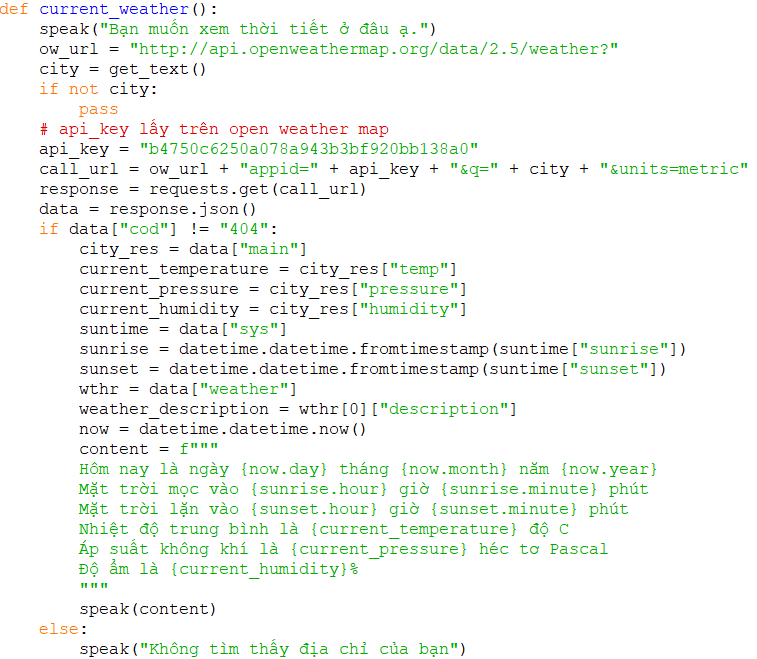
Ở hàm này, ví dụ đơn giản là nếu tìm thấy chữ *"tâm"* trong tên người cần gửi (*recipient*) thì mình sẽ thực hiện chức năng gửi *Email*. Tiếp tục, hàm *get\_text()* để lấy nội dung cần gửi rồi lưu vào biến*content*. Mở đường truyền gửi Email bằng *smtp* rồi đăng nhập lại vào tài khoản gmail của mình. Sau khi kết nối thành công, mình gửi email từ địa chỉ " *itaisv1999@gmail.com*"đến địa chỉ "huyph11247@gmail.com" với nội dung *content* đã lấy ở trên. Cuối cùng,đóng đường truyền lại.

### 3.3.5 Chức năng xem dự báo thời tiết



##### Hình 3.6: Dự báo thời tiết OpenWeatherMap

**- OpenWeatherMap** là một dịch vụ trực tuyến cung cấp dữ liệu thời tiết. Nó thuộc sở hữu của OpenWeather Ltd. Nó cung cấp dữ liệu thời tiết hiện tại, dự báo và dữ liệu lịch sử cho hơn 2 triệu khách hàng, bao gồm các công ty Fortune 500 và hàng ngàn doanh nghiệp khác trên toàn cầu.



Nguồn dữ liệu thời tiết của hầu hết các thành phố trên toàn thể thời cung cấp tại trang web ***openweathermap.org.***

Đầu tiên, sẽ sử dụng biến *ow\_url* để lưu đường đẫn đến *api* của trang web *openweathermap.org.*Thông tin các thẻ để thực hiện request, gọi hàm *get\_text()* để lấy thông tin thành phố cần truy vấn thời tiết rồi lưu vào biến *city.* Nếu máy tính không nghe được tên thành phố thì hàm sẽ*pass*.Tiếp theo là thực hiện kết nối với *api* trang web *openweathermap.org.* Phải đăng kí tài khoản mới ở đây và lấy *api\_key* để có thể thực hiện thao tác truy vấn. Biến *call\_url* sẽ lưu đường dẫn đầy đủ để truy vấn bao gồm thông tin tên thành phố *city* và *api\_key* mình đã lấy ở trên. Thư viện requests là một mô-đun Python có thể sử dụng để gửi tất cả các loại yêu cầu HTTP. Hàm *requests.get(call\_url)* để lấy thông tin truy vấn được từ trang web rồi lưu vào biến *response. response.json()* sẽ chuyển dữ liệu thuần về kiểu dữ liệu *json* rồi lưu vào biến*data*. Nếu *data["cod"]* không trả về *404*tức là requests không bị lỗi hay có nghĩa là thông tin tên thành phố *city* chuẩn xác (truy vấn được) thì mình sẽ thực hiện chức năng xem thời tiết. Từ dữ liệu kiểu json được lưu trong biến *data*lấy được thông tin thời tiết bao gồm *temperature, pressure, humidity,…* Cuối cùng mình kết nối những thông tin trên với nhau rồi lưu vào biến*content* và thực hiện đọc *content.*

### 3.3.6 Một số chức năng khác

* Hiển thị giời gian và ngày tháng
* Chụp ảnh và quay video từ webcam
* Tìm kiếm từ khóa trên google
* Nhận diện vật thể
* Thay đổi hình nền laptop
* Mở ứng dụng hệ thống, website
* Phát nhạc trên Youtube
* Đọc báo ngày hôm nay
* Tìm kiếm định nghĩa trên từ điển wikipedia
* Thực hiện phép tính

## 3.4 Ứng dựng thực tế của trợ lý ảo

Phát triển mạng nơ-ron và xử lýngôn ngữ tự nhiên hoàn toàn hoạt động theo hướng biến điều khiển giọng nói thành một tiêu chuẩn mới cho nhiều **sản phẩm** và **hành động** mà mọi người sử dụng/thực hiện hàng ngày. Chiến lược  bắt đầu một xu hướng mới của các công ty phát hành bộ công cụ phát triển thiết bị dịch vụ để tích hợp ứng dụng trợ lý giọng nói. Các nhà sản xuất thiết bị gia dụng như LG, Whirlpool và Phillips đã bắt đầu phát hành các mặt hàng hỗ trợ Alexa và phản ứng của người tiêu dùng nhanh chóng thu hút sự chú ý của các doanh nhân vào các ứng dụng trợ lý giọng nói. 72% chủ sở hữu trợ lý ảo thừa nhận rằng thiết bị của họ nhanh chóng trở thành một phần **thiết yếu** trong thói quen hàng ngày của họ và họ không muốn bỏ bộ điều khiển giọng nói này do thói quen bắt đầu hìnhthànhvàdễ sử dụng**.**

Vì vậy, các nghĩa vụ và nhiệm vụ thực tế mà ứng dụng trợ lý AI có thể làm cho con người là gì ? Bao nhiêu trong số chúng hoạt động tốt ?. Sau đây là một vài công việc mà trợ lý ảo có thể thực hiện cho bạn:

* Gửi thông tin cập nhật về các chủ đề mà bạn quan tâm mà không cần bạn tìm kiếm chúng
* Báo cáo thời tiết;
* Thêm các sự kiện và các cuộc họp vào lịch của một nhóm hoặc từng thành viên riêng biệt;
* Đặt báo thức và nhắc nhở để mọi thứ diễn ra theo đúng lịch trình;
* Trả lời câu hỏi chung bằng giọng nói (thay vì mở liên kết để bạn tìm kiếm câu trả lời);
* Tạo và điền vào danh sách việc cần làm;
* Thực hiện dịch thuật thời gian thực;
* Ngay lập tức cập nhật cho bạn về lưu lượng trên tuyến đường của bạn (đặc biệt hữu ích cho các hoạt động hậu cần);
* Theo dõi hàng tồn kho trong kho và tự động điền vào danh sách mua sắm với các mặt hàng sẽ được đưa ra ngoài;
* Điều khiển các thiết bị khác từ ánh sáng đến PC;
* Đọc email và các tài liệu khác thành tiếng thay vì bạn đọc chúng;
* Ghi lại lời nói chính tả và chuyển nó thành văn bản thay vì bạn gõ thủ công;
* Tìm các doanh nghiệp địa phương và nhà thầu cho nhu cầu hiện tại của bạn.

# CHƯƠNG 4: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 4.1 Kết luận

- Tìm hiểu được cơ chế nguyên lý hoạt động của trợ lý ảo

- Quý trình xây dựng được trợ lý ảo

- Biết được cấu trúc của trợ lý ảo

- Ứng dựng của trợ lý ảo trong đời sống như thế nào

- Sử dụng trợ lý ảo điểu khiển các hoạt động của máy tính

## 4.2 Hạn chế trong đê tài

- Code chưa được tối ưa

- Một số chức năng vẫn chưa được hoàn thiện như trong đề cương đã nêu

## 4.3 Hướng phát triển

Trong suốt chiều dài lịch sử phát triển máy tính, giao diện người dùng (User Interface) đang dần trở nên tự nhiên dễ sử dụng hơn. Đầu tiên đó là sự kết hợp của màn hình và bàn phím. Sau đó là chuột và đồ họa giao diện người dùng. Màn hình cảm ứng là bước phát triển vĩ đại trong cuộc cách mạng cải tiến máy tính. Chúng ta dự đoán bước tiếp theo sẽ có thể bao gồm sự kết hợp giữa thực tế ảo, hành vi con người và ra lệnh bằng giọng nói. Rốt cuộc thì việc giao tiếp luôn luôn là dễ dàng và hiệu quả để truyền đạt nội dung hơn là nhập tay một câu lệnh hay biểu mẫu nào đấy.

Càng nhiều người tương tác với các thiết bị kích hoạt bằng giọng nói thì hệ thống càng có nhiều dữ liệu về đặc điểm ngôn ngữ của người dùng. Hệ thống sẽ xác định được xu hướng và mô hình hóa thông tin nhận được. Sau đó, dữ liệu này có thể được sử dụng để xác định sở thích và thị hiều người dùng, đây là một trong những kế hoạch dài hạn để biến ngôi nhà trở nên thông minh hơn. Google và Amazon đang tìm cách tích hợp trí tuệ nhân tạo hỗ trợ giọng nói có khả năng phân tích và phản hồi lại cảm xúc của con người.

# TÀI LIỆU THAM KHẢO

[1] <https://github.com/marimeireles/virtual-assistant>

[2] <https://codelabs.developers.google.com/codelabs/cloud-text-speech-python3#7>

[3] https://becominghuman.ai/a-simple-introduction-to-natural-language-processing

[4] QuanTra - Trợ lý ảo tiếng việt

https://quantraai.com/xay-dung-tro-ly-ao-tieng-viet-bang-ngon-ngu-python-co-ban/