**HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG**

**CHI NHÁNH**

**THÀNH PHỐ HỒ CHÍ MINH**

****

**BÁO CÁO MÔN HỌC**

**Đề tài: Xây dựng chương trình nhận dạng giọng nói một người dựa trên các giải thuật học máy**

*Giảng viên hướng dẫn:*

**Nguyễn Ngọc Duy**

*Sinh viên thực hiện:*

**Võ Nguyễn Duy Long N18DCCN116**

**NHẬN XÉT CỦA GIẢNG VIÊN**

MỤC LỤC

[A. PHẦN MỞ ĐẦU 4](#_Toc79237636)

[**1.** **Mục đích, nhiệm vụ nghiên cứu** 4](#_Toc79237637)

[**2.** **Ý nghĩa đề tài** 4](#_Toc79237638)

[**3.** **Kết cấu đề tài** 4](#_Toc79237639)

[B. PHẦN NỘI DUNG 5](#_Toc79237640)

[**Chương 1** 5](#_Toc79237641)

[**1.** **Giới thiệu về ngôn ngữ lập trình Python** 5](#_Toc79237642)

[**2.** **Các đặc trưng của ngôn ngữ lập trình Python** 6](#_Toc79237643)

[**3.** **Các lĩnh vực sử dụng ngôn ngữ Python** 8](#_Toc79237644)

[**4.** **Lý do tại sao lại sử dụng Python cho các dự án về AI và Machine Learing?** 9](#_Toc79237645)

[**Chương 2** 11](#_Toc79237646)

[**1.** **Giới thiệu giản lược về học máy (Machine Learning - ML)** 11](#_Toc79237647)

[**2.** **Giới thiệu về nhận dạng giọng nói** 12](#_Toc79237648)

[**3.** **Lịch sử công nghệ nhận diện giọng nói** 13](#_Toc79237649)

[**4.** **Các dạng nhận diện giọng nói** 13](#_Toc79237650)

[**5.** **Ứng dụng của phần mềm nhận dạng giọng nói** 14](#_Toc79237651)

[**6.** **Ưu điểm và nhược điểm của phần mềm nhận dạng giọng nói** 14](#_Toc79237652)

[**7.** **Mô hình triển khai công nghệ giọng nói** 15](#_Toc79237653)

[**8.** **Cách thức xây dựng công nghệ Nhận dạng, giả lập giọng nói** 16](#_Toc79237654)

[C. PHẦN KẾT LUẬN 17](#_Toc79237655)

[D. PHẦN TÀI LIỆU THAM KHẢO 17](#_Toc79237656)

1. PHẦN MỞ ĐẦU
2. **Mục đích, nhiệm vụ nghiên cứu**
3. Mục đích nghiên cứu

Tìm hiểu về bài toán nhận dạng giọng nói và xây dựng một chương trình nhận dạng giọng nói dựa trên học máy.

1. Nhiệm vụ nghiên cứu

Nêu rõ về bài toán nhận dạng giọng nói, cách xây dựng một ứng dụng nhận dạng giọng nói bằng ngôn ngữ lập trình và ngôn ngữ lập trình sẽ sử dụng.

1. **Ý nghĩa đề tài**

Qua đề tài giúp ta hiểu rõ được:

* Ngôn ngữ lập trình Python
* Nhận dạng giọng nói
* Cách xây dựng một chương trình nhận dạng giọng nói bằng Python

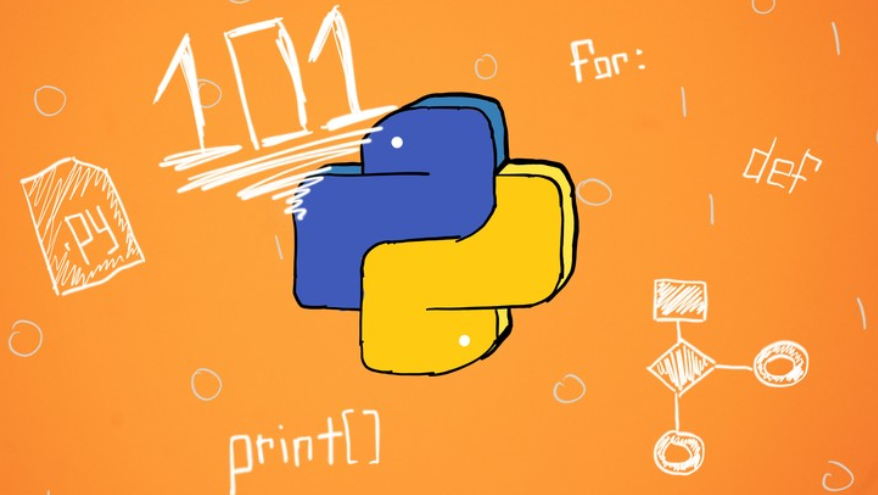
1. **Kết cấu đề tài**

* Đề tài gồm 4 phần:
* Phần mở đầu
* Phần nội dung
* Phần kết bài
* Phần tài liệu
* Phần nội dung gồm 3 chương:
* Chương 1: Giới thiệu về ngôn ngữ lập trình Python
* Chương 2: Giới thiệu về bài toán nhận dạng giọng nói
* Chương 3: Giới thiệu và hướng dẫn sử dụng chương trình nhận dạng giọng nói

1. PHẦN NỘI DUNG

**Chương 1**

**GIỚI THIỆU VỀ NGÔN NGỮ LẬP TRÌNH PYTHON**

1. **Giới thiệu về ngôn ngữ lập trình Python**

Python là ngôn ngữ lập trình cấp cao, đa mục đích được sử dụng rộng rãi. Ban đầu được thiết kế bởi Guido van Rossum vào năm 1991 và được duy trì, phát triển bởi tổ chức Python Software Foundation.

Python là một ngôn ngữ lập trình tuyệt vời. Không chỉ gây thích thú với việc dễ học, dễ tiếp cận với người mới mà nó còn có một hệ thống cú pháp vô cùng đơn giản và dễ nhớ. Nhưng đừng vì vậy mà bạn lầm tưởng rằng Python là thứ ngôn ngữ chỉ dùng để “học cho biết”, các ứng dụng cũng như sản phẩm từ những công ty như Google, Mozilla, Cisco, Microsoft hay Instagram đều được viết bằng Python[1].

Ngoài ra, sự bùng nổ của kỷ nguyên số về các lĩnh vực như Trí Tuệ Nhân Tạo (**Artificial Intelligence**) và Dữ Liệu Lớn (**Big Data**) đã góp phần gia tăng không nhỏ nhu cầu sử dụng Python trong những năm gần đây và còn tiến xa hơn nữa trong tương lai[1].

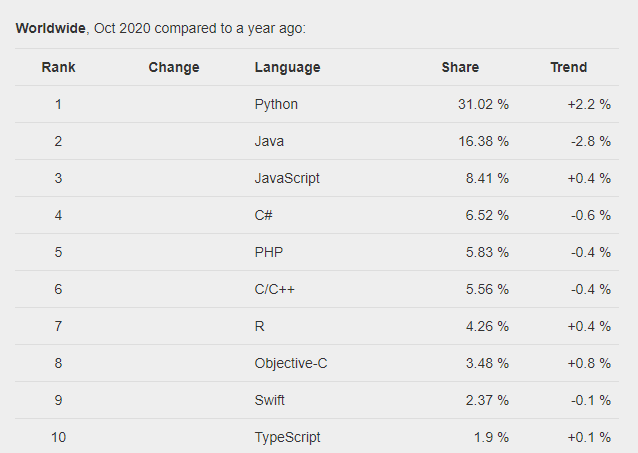
1. **Các đặc trưng của ngôn ngữ lập trình Python**

Đặc trưng của ngôn ngữ lập trình là yếu tố được quan tâm khi chúng ta tìm hiểu về nó. Python là một ngôn ngữ cấp cao, có quá trình phát triển dài và sau đây là một số đặc trưng nổi bật của nó:

1. Tính phổ biến.

Python có sự phát triển rất nhanh trong những năm gần đây. Theo một khảo sát vào năm 2018 trên trang web Stack Overflow, Python được xếp hạng thứ 7 trong những ngôn ngữ lập trình phổ biến nhất và xếp hạng thứ nhất trong những công nghệ được ưa thích nhất năm. Không những vậy, những công ty phần mềm hàng đầu thế giới sử dụng Python trong các dự án của họ mỗi ngày.

Theo một báo cáo của Dice, Python tính thời điểm hiện tại được xem như là một ngôn ngữ mà lập trình viên nào cũng nên biết và là ngôn ngữ phổ biến nhất trong bảng xếp hạng của Popularity of Programming Language Index.



Bảng xếp hạng mức độ phổ biến của các ngôn ngữ lập trình theo Popularity of Programming Language Index.

Chính vì mức độ phổ biến rộng rãi, tính ứng dụng cao nên Python là một trong những ngôn ngữ lập trình với nhu cầu tuyển dụng và mức lương hậu đãi rất cao.

1. Dễ học

Khi so sánh giữ Python và các ngôn ngữ lập trình cấp cao khác như C++, C#, Javascript, Java, … thì việc học ngôn ngữ Python dễ hơn.

Python có cú pháp rất đơn giản, rõ ràng.

Nó dễ đọc và viết hơn rất nhiều khi so sánh với những ngôn ngữ lập trình khác. Python làm cho việc lập trình trở nên thú vị, cho phép bạn tập trung vào những giải pháp chứ không phải cú pháp.

Nhờ vào tính đơn giản và dễ học của Python, nhiều người từ các ngành nghề khác khi muốn thử sức ở lĩnh vực Công Nghệ Thông Tin, có thể học Python như một bước đệm đầu để chinh phục như thứ khó khăn hơn ở phía sau.

1. Không tính phí và mã nguồn mở

Trình thông dịch của Python được phát triển dưới giấy phép mã nguồn mở (**OSI-Approved Open-Source License**), chính vì vậy ta có thể tải và cài đặt nó hoàn toàn miễn phí, kể cả dùng cho các mục đích thương mại khác.

1. Hướng đối tượng

Mọi thứ trong Python đều là hướng đối tượng. Lập trình hướng đối tượng (**Object Oriented Programming** hay **OOP**) giúp giải quyết những vấn đề phức tạp một cách trực quan. Với OOP, bạn có thể phân chia những vấn đề phức tạp thành những tập nhỏ hơn bằng cách tạo ra các đối tượng. Python hỗ trợ cả lập trình hướng đối tượng, một trong những tính năng chính của nó.

1. Ngôn ngữ thông dịch

Hầu hết các ngôn ngữ lập trình đều chạy bằng trình biên dịch, đều này có nghĩa là source code mà bạn viết ra cần phải được dịch qua mã máy (**Machine Code**) trước khi chạy được chương trình. Tuy nhiên, những chương trình được viết bằng ngôn ngữ thông dịch sẽ bỏ qua bước đó mà chạy thẳng qua trình thông dịch.

Vì vậy, tất cả những gì bạn cần làm là chạy đoạn code Python của bạn mà không phải lo lắng về việc liên kết với các thư viện và những thứ khác.

1. Dễ dàng mở rộng

Tính năng mở rộng của Python đó là Python có để được viết trong một ngôn ngữ khác và có thể dùng trình biên dịch/thông dịch để thực thi chương trình. Ví dụ, chúng ta có thể viết mã nguồn Python và biên dịch nó bên trong ngôn ngữ C/C++.

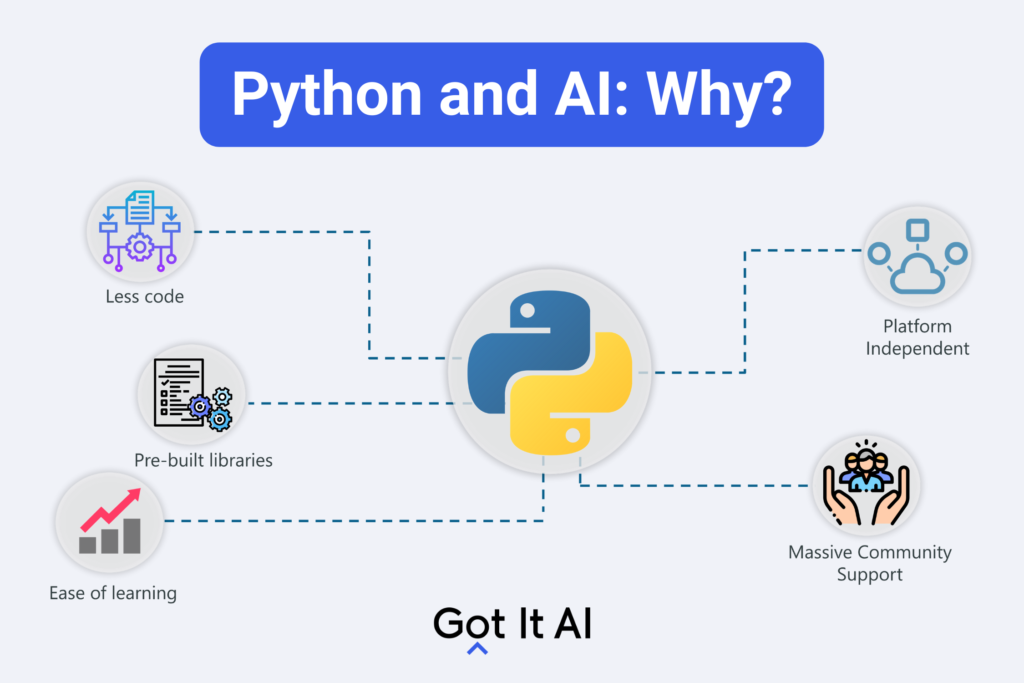
1. Ngôn ngữ không cần định nghĩa kiểu dữ liệu

Đây là một đặc trưng khác biệt hoàn toàn so với các ngôn ngữ lập trình C, C++, Java,… Các biến trong chương trình Python khi khai báo không cần chỉ rõ kiểu dữ liệu cho nó và một biến có thể nắm giữ các đối tượng có kiểu dữ liệu khác nhau.

1. **Các lĩnh vực sử dụng ngôn ngữ Python**

Các đặc trưng của Python giúp cho Python được sử dụng rộng rãi, phổ biến trong cộng đồng lập trình. Có đến hàng hàng ngàn thư viện hỗ trợ ngôn ngữ lập trình Python với mục tiêu nhất định. Dưới đây là những ứng dụng quan trọng của ngôn ngữ lập trình Python có thể tạo ra.

* + Phát triển Web và Internet (**Web and Internet Development**).
  + Games và Đồ Họa 3D (**Games and 3D Graphics**).
  + Phục vụ cho khoa học và tính toán (**Scientific and Numeric**).
  + Ứng dụng giao diện người dùng (**Desktop GUIs**).
  + Sử dụng trong lĩnh vực Trí tuệ nhân tạo (**Artificial Intelligence**).

1. **Lý do tại sao lại sử dụng Python cho các dự án về AI và Machine Learing?**

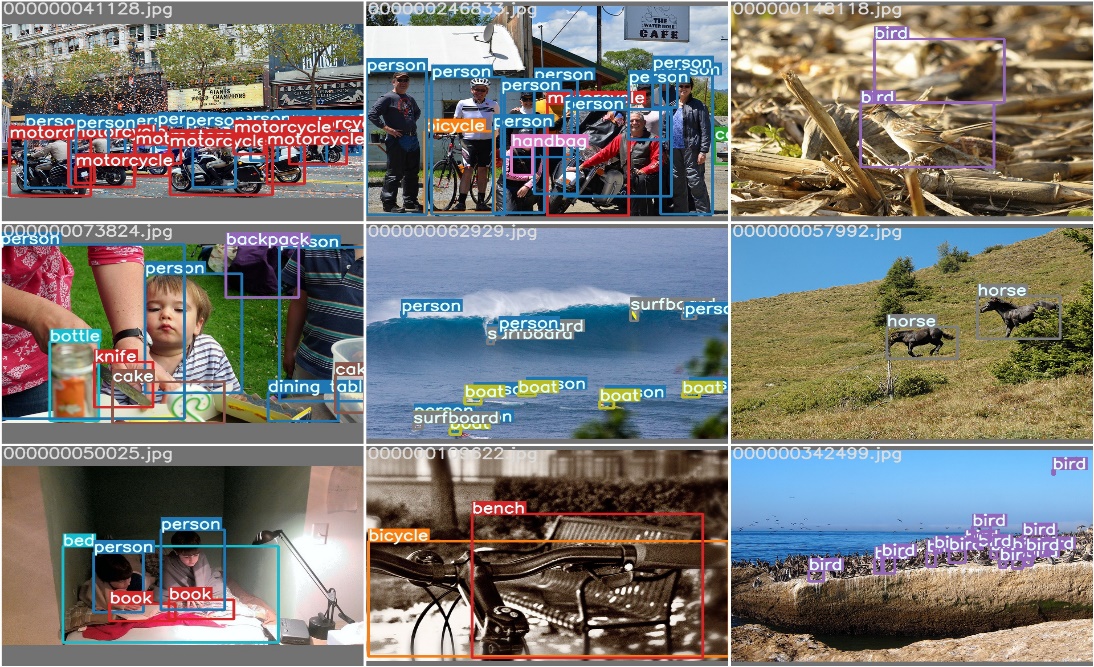
Điều đầu tiên bạn nên ghi nhớ là các dự án về AI khác với các dự án phần mềm truyền thống.

Sự khác biệt nằm ở Tech Stack, các kỹ năng cần thiết cho một dự án AI và sự cần thiết của những nghiên cứu chuyên sâu không phải ai cũng có đủ thời gian thể nắm giữ tất cả.

*Tech Stack được định nghĩa là tập hợp các công nghệ mà một tổ chức sử dụng để xây dựng một ứng dụng web hoặc thiết bị di động. Nó là sự kết hợp của các ngôn ngữ lập trình, ... [2]*

Do đó, để thực hiện tham vọng với dự án AI, bạn nên sử dụng ngôn ngữ lập trình ổn định, linh hoạt và có sẵn các công cụ / thư viện hỗ trợ.

May mắn là Python cung cấp tất cả những thứ này, đây cũng chính là lý do tại sao ngày nay chúng ta thấy RẤT NHIỀU dự án AI làm bằng Python.

Object Detection sử dụng YOLOv3

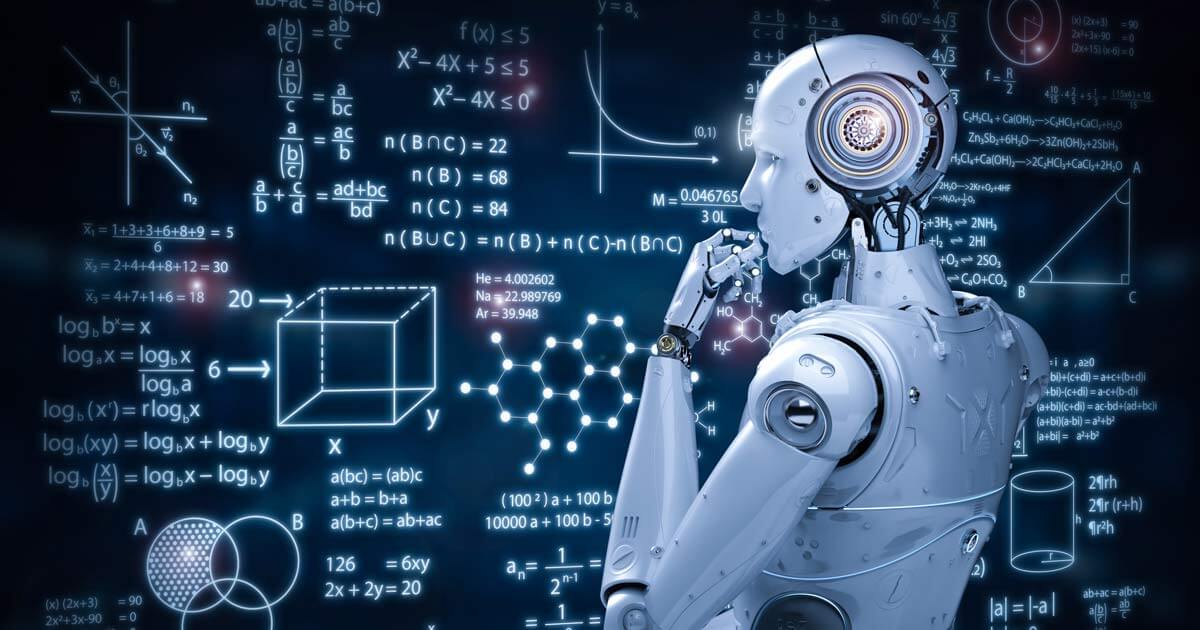
Từ giai đoạn phát triển đến triển khai và bảo trì, Python giúp các lập trình viên làm việc hiệu quả và tự tin về phần mềm mà họ làm ra.

*Bởi vì phần mềm của họ làm ra có sự đóng góp của hàng trăm bộ não thiên tài trong đó (thông qua các công cụ, thư viện, framework họ sử dụng) [2]*

## **Chương 2**

**GIỚI THIỆU VỀ BÀI TOÁN NHẬN DẠNG GIỌNG NÓI**

1. **Giới thiệu giản lược về học máy (Machine Learning - ML)**

Trước khi đi vào chủ đề của chương là nhận dạng giọng nói ta cần hiểu sơ về học máy\_một nhân tố không thể thiếu trong bài toán nhận dạng giọng nói.

Machine Learning là một tập con của AI. Theo định nghĩa của Wikipedia, Machine learning is the subfield of computer science that “gives computers the ability to learn without being explicitly programmed” [3]. Nói đơn giản, Machine Learning là một lĩnh vực nhỏ của Khoa Học Máy Tính, nó có khả năng tự học hỏi dựa trên dữ liệu đưa vào mà không cần phải được lập trình cụ thể, đồng thời là cách để cải thiện hiệu suất theo thời gian vô cùng hiệu quả.

Thuật toán học máy được chia thành 2 loại chính bao gồm: học có giám sát và học không giám sát.

* + **Học có giám sát (Supervised Learning)**

Học có giám sát là phương pháp sử dụng những dữ liệu được gán nhãn sẵn để suy luận ra quan hệ giữa đầu vào và đầu ra. Sau khi tìm hiểu cách tốt nhất để mô hình hóa các mối quan hệ cho dữ liệu được gán nhãn, thuật toán huấn luyện sẽ được sử dụng cho các bộ dữ liệu mới.

* + **Học không có giám sát (Unsupervised Learning)**

Học không giám sát sử dụng những dữ liệu chưa được gán nhãn sẵn để suy luận và tìm cách để mô tả dữ liệu cùng cấu trúc của chúng.

1. **Giới thiệu về nhận dạng giọng nói**

Trong thời đại công nghệ ngày nay, giọng nói là một thứ rất “quyền lực”. Các thiết bị công nghệ hiện đại đều có thể nhận diện giọng nói của bạn để thực hiện nhiều tính năng và hành động khác nhau. Vậy có bao giờ bạn thắc mắc rằng công nghệ nhận diện giọng nói là gì và chúng hoạt động như thế nào không? Ta cùng tìm hiểu nào

Công nghệ nhận diện giọng nói là gì?

Nhận dạng giọng nói là một lĩnh vực phụ liên ngành của khoa học máy tính (**Computer Science**) và ngôn ngữ học máy tính (**Computational Linguistics**) là công nghệ cho phép máy tính phát triển các phương pháp luận để nhận dạng và dịch ngôn ngữ nói thành văn bản. Nó còn được gọi là nhận dạng giọng nói tự động (**Automatic Speech Recognition -** ASR), nhận dạng giọng nói máy tính(**Computer Speech Recognition)** hoặc chuyển lời nói thành văn bản (**Speech to Text -** STT).

**🡺** Nói đơn giản công nghệ nhận dạng giọng nói là một bộ máy hoặc hệ thống có khả năng nhận và viết (hoặc hiểu và thực hiện) các lệnh từ giọng nói con người [4].

1. **Lịch sử công nghệ nhận diện giọng nói**

Nhận diện giọng nói lần đầu xuất hiện trên các thiết bị điện toán cá nhân từ khoảng 20 năm trước, khi Windows 98 được giới thiệu. Song, có thể bạn sẽ ngạc nhiên khi biết rằng công nghệ này đã bắt đầu được nghiên cứu từ năm 1936. Hệ thống nhận diện giọng nói ban đầu chỉ hiểu chữ số vì các kỹ sư cho rằng ngôn ngữ con người quá phức tạp.

**1961 — IBM Shoebox**

Năm 1952, Bell Laboratories thiết kế ra “Audrey”, hệ thống nhận diện chữ số từ một giọng nói nhất định. 10 năm sau tại triển lãm World’s Fair, IBM trình diễn hệ thống “Shoebox” có thể nhận diện 16 từ tiếng Anh khác nhau. Lúc này kẻ tấn công có thể thấy gói tin của các máy gửi cho nhau trong mạng nhờ vào các công cụ Sniffer (nghe lén).

1. **Các dạng nhận diện giọng nói**

Nhận dạng giọng nói có 2 thuật ngữ khác nhau: Voice Recognition và Speech Recognition

* **Voice Recognition**liên quan đến việc xác định một giọng nói chính xác của một cá nhân nào đó. Cụ thể là xác thực người nói bằng cách phân tích các mẫu và trình tự giọng nói của một người từ đó có thể định danh người nói chính xác, tương tự các phương pháp nhận diện sinh trắc học.
* **Speech Recognition**là việc xác định những từ ngữ trong câu nói rồi dịch chúng sang ngôn ngữ máy tính.

1. **Ứng dụng của phần mềm nhận dạng giọng nói**

Chuyển giọng nói thành văn bản được xem là ứng dụng phổ biến nhất của công nghệ nhận dạng giọng nói hiện nay. Chúng ta có thể dễ dàng nhìn thấy những phần mềm ứng dụng công nghệ này ngay trên smartphone.

Người dùng không chỉ xài giọng nói của mình như một phương thức nhập liệu mà chính bản thân thiết bị cũng có thể dùng giọng nói để đọc ra những thông tin cần thiết. Nó mang lại lợi ích to lớn cho những người khiếm thị khi mà họ có thể tận hưởng những tiến bộ công nghệ tương tự như những gì mà một người bình thường có thể làm.

1. **Ưu điểm và nhược điểm của phần mềm nhận dạng giọng nói**

**Ưu điểm của phần mềm nhận dạng giọng nói**

- *Khả năng truy cập*: Đây là một thuận lợi đối với người khuyết tất khi họ không thể dùng chuột hay bàn phím, nhưng có thể dùng giọng nói để hệ thống chuyển thành văn bản, giúp nhập liệu hay điều khiển một cách dễ dàng.

- *Kiểm tra chính tả*: Người dùng có thể truy cập vào các công cụ chỉnh sửa tương tự một giải pháp xử lý văn bản chuẩn. Đương nhiên mọi thứ sẽ không chính xác 100% nhưng phần mềm có thể nhận diện và xử lý phần lớn lỗi chính tả, ngữ pháp.

- *Tốc độ nhanh*: Phần mềm nhận dạng giọng nói có thể nắm bắt giọng nói của người dùng với tốc độ nhanh hơn so với khi nhập liệu bằng bàn phím, vì vậy tốc độ khi nhập liệu bằng giọng nói sẽ cải thiện đáng kể.

**Nhược điểm của phần mềm nhận dạng giọng nói**

-    *Thiết lập và "dạy"*: Mặc dù tất cả phần mềm nhận dạng giọng nói hiện nay đều hứa hẹn có thể hoạt động sau vài phút thiết lập, nhưng thực sự quá trình ghi nhận, làm quen với giọng nói, âm điệu và tốc độ nói của người dùng có đôi chút phức tạp và tốn thời gian. Một số phần mềm nhận dạng giọng nói còn bắt người dùng nói lại, thậm chí không thể nhận diện được bạn đang nói gì.

-    *Chưa thực sự ổn định*: Việc đang nói mà bị ngắt giữa chừng có thể khiến người dùng cảm thấy khó chịu. Đặc biệt, một số người không thích phần mềm nhận dạng giọng nói vì nó gây bối rối cho người dùng khi lên xuống giọng hay bỗng dưng nói nhỏ lại.

-    *Kho từ vựng hạn chế*: Người dùng phải sẵn sàng chấp nhận trường hợp phần mềm xử lý quá lâu vì những từ vừa nói không nằm trong từ điển có sẵn. Đó là điều các nhà nghiên cứu đang cố gắng cải tiến ở phần mềm nhận dạng giọng nói hiện nay.

1. **Mô hình triển khai công nghệ giọng nói**

Có nhiều cách thức mà các công ty hiện nay đang triển khai voice technology, có thể kể đến 2 phương pháp phổ biến như sau:

* *Điện toán đám mây*: Trong trường hợp này, việc nhận dạng, xử lý ngôn ngữ sẽ diễn ra trên máy chủ của các công ty cung cấp dịch vụ. Phương pháp đám mây giúp việc nhận dạng được chính xác hơn, ứng dụng thì có dung lượng nhỏ, nhưng bù lại thì thiết bị ở phía người dùng phải luôn kết nối với Internet. Độ trễ trong quá trình gửi giọng nói từ máy lên server rồi trả kết quả từ server về lại máy cũng là những thứ đáng cân nhắc.
* *Tích hợp thẳng vào app*: Với phương thức này, quá trình xử lý giọng nói sẽ diễn ra trong nội bộ ứng dụng, không cần giao tiếp với bên ngoài, chính vì thế tốc độ sẽ nhanh hơn. Người dùng cũng không bắt buộc phải kết nối vào mạng thường trực. Tuy nhiên, giải pháp này gặp nhược điểm đó là khi có cập nhật hoặc thay đổi gì đó về bộ máy nhận dạng, nhà sản xuất sẽ phải cập nhật lại cả một app, trong khi với phương thức đám mây thì những thay đổi đó chỉ cần làm ở phía server. Kích thước ứng dụng cũng sẽ tăng lên, có thể lên tới cả vài trăm MB.

1. **Cách thức xây dựng công nghệ Nhận dạng, Giả lập giọng nói**

Để AI thông minh thì cần phải có dữ liệu để huấn luyện cho nó, cả về nhận diện hình ảnh, văn bản, giọng nói. Google có hàng tỷ người dùng với công cụ tìm kiếm, nó có thể biết được trong khoảng thời gian nào, trong từng thời điểm người dùng quan tâm từ khóa nào, lĩnh vực nào. Đó là một cách người dùng tự tạo dữ liệu cho AI. Cũng còn một cách là người dùng trực tiếp cung cấp dữ liệu cho AI.

Vậy người ta áp dụng công nghệ giọng nói vào phần mềm như thế nào? Thông thường một bộ máy giọng nói sẽ có hai phần:

* Phần thứ nhất gọi là **Speech Synthesizer** (còn gọi là Text to Speech hay TTS). Đây là một trình tổng hợp giọng nói và thiết bị hoặc ứng dụng xài để tương tác, giao tiếp với người dùng, ví dụ: đọc văn bản, thông báo về tiến độ một tác vụ nào đó.
* Phần thứ hai là một công nghệ **Recognition** cho phép app biết được người dùng đang nói gì, từ đó chuyển thể thành lệnh để thiết bị thực thi hoặc chuyển đổi thành các kí tự nhập liệu. Nói cách khác, đây là thứ thay thế cho bàn phím của chúng ta.

Thoạt nhìn thì việc triển khai công nghệ nhận dạng giọng nói khá đơn giản, nhưng thực chất thì không phải như thế.

* Thứ nhất, các nhà phát triển phải xây dựng nên một công nghệ có thể lắng nghe, phân tích và phiên dịch một cách chính xác giọng nói của người dùng. Nếu không thì làm sao app biết bạn đang nói gì, còn nếu độ chính xác không cao thì cũng như không.
* Thứ hai, vấn đề bản địa hóa (**Localization**) cũng là một chuyện làm đau đầu các lập trình viên. Mỗi quốc gia sẽ có ngôn ngữ của riêng mình, vấn đề đó là làm thế nào để có thể hỗ trợ càng nhiều ngôn ngữ càng tốt.

Có một kĩ thuật được nhắc đến nhiều trong thời gian gần đây, đó là Xử lý ngôn ngữ tự nhiên (**Natural Language Processing – NLP**). Nó là tập hợp của nhiều thuận toán phức tạp nhằm phân tích mệnh lệnh của người dùng nhưng không bắt buộc họ phải nói theo một cấu trúc câu định sẵn.

1. PHẦN KẾT LUẬN

Tiềm năng của công nghệ Nhận dạng giọng nói là rất lớn. Hãy nghĩ đến một tương lai nơi bạn có thể hoàn toàn để điện thoại trong túi quần và thực hiện tất cả chỉ bằng cách nói vào tai nghe Bluetooth. Tính năng nhận dạng giọng nói không chỉ dừng lại ở việc nhập liệu mà nó còn mở ra cả một chân trời để chúng ta khai thác và đơn giản hóa cuộc sống của mình. Công nghệ được sinh ra là để làm cho cuộc sống dễ dàng hơn và việc nhận dạng giọng nói chắc chắn sẽ không phải là ngoại lệ!

1. PHẦN TÀI LIỆU THAM KHẢO

[1]. Anh Quan Nguyen, “[Python] Bài 1: Giới thiệu ngôn ngữ lập trình Python”, <http://tutorials.aiclub.cs.uit.edu.vn/index.php/2020/10/27/introductionpython/>, 27/10/2020

[2]. “TẠI SAO lại sử dụng PYTHON cho AI và Machine Learning?”, <https://niithanoi.edu.vn/tai-sao-lai-su-dung-python-cho-ai-va-machine-learning.html>, 04/11/2019

[3]. Wikipedia, “Machine learning”, <https://en.wikipedia.org/wiki/Machine_learning>

[4]. “TỔNG QUAN VỀ CÔNG NGHỆ NHẬN DIỆN GIỌNG NÓI”, <https://learn.qts.edu.vn/qts-cong-nghe/tong-quan-ve-cong-nghe-nhan-dien-giong-noi-884>