WEEK 4: Tìm hiểu về Voice Assistants

1. Khái niệm về Voice Assistants - VA

- Trợ lý giọng nói hay còn goi là trở lý ảo được định nghĩa là 1 phần mềm hay ứng dụng được tích hợp công nghệ nhận dạng giọng nói “speech recognition” và xử lý ngôn ngữ tự nhiên “natural language processing” để xử lí mệnh lệnh từ người dùng và cho phép điều khiển thiết bị khác, truy cập thông tin thông qua giọng nói.

“A voice assistant is intelligent software that responds to voice commands and runs on various devices like smartphones, speakers, computers, tablets, wearables, gaming consoles, TVs, VR headsets, cars, and internet of things devices. Examples include Amazon’s Alexa, Apple’s Siri, Google Assistant, and Microsoft’s Cortana.”

-Trích dẫn từ emarketer.com-

- Chúng ta có thể kể tên 1 vài trợ lý ảo nổi tiếng như Siri của Apple, Alexa của Amazon, Bixby của Samsung…

2. Lịch sử của VA

Lịch sử phát triển của VA

- 1922 – First Voice activated consumer product hits store shelves as “Radio Rex”
- 1952 – Audrey, or the Automatic Digit Recognition Machine, is announced
- 1962 – IBM Shoebox is shown for the first time at the State Fair
- 1971 – Darpa funds five years of speech recognition research and development
- 1976 – Harpy is shown at Carnegie Mellon
- 1984 – IBM releases “Tangora” the first voice activated typewriter
- 1990 – Dragon Dictate is released
- 1994 – Simon by IBM is the first modern voice assistant released

- 2010 – Siri is released as an app on the iOS app store
- 2011 – IBM Watson is released
- 2012 – Google Now is released
- 2014 – Amazon Alexa and Echo are released
- 2015 – Microsoft Cortana is released
- 2017 – Alan is developed and released with the Alan Platform

Sơ lược qua lịch sử phát triển của VA

1. **Radio Rex (1922)**
   – Con chó đồ chơi "Radio Rex" là sản phẩm kích hoạt bằng giọng nói đầu tiên, ra mắt năm 1922. Khi người dùng nói "Rex", một điện từ sẽ kích hoạt cơ cấu đẩy chú chó nhảy ra khỏi chuồng. Đây là ví dụ sớm nhất về thiết bị nhận diện tần số giọng nói, xuất hiện hơn 20 năm trước khi máy tính hiện đại ra đời.

2. **Audrey – Bell Labs (1952)**
   – Tại Hội chợ Thế giới 1952, Bell Labs giới thiệu "Audrey" (Automatic Digit Recognizer). Máy cao gần 1,8 m, chỉ để nhận dạng 10 chữ số từ giọng nói – minh chứng cho quy mô khổng lồ của công nghệ khi ấy.

3. **IBM Shoebox (1962)**
   – Năm 1962, tại Hội chợ Thế giới ở Seattle, IBM ra mắt "Shoebox": thiết bị có kích cỡ bằng hộp giày, có thể nhận dạng các chữ số 0–9 và 6 lệnh cơ bản như "plus", "minus". Hệ thống dùng ba bộ lọc âm thanh để so khớp tần số giọng nói với giá trị số đã định.

4. **Chương trình SUR của DARPA & Harpy (1971–1976)**
   – DARPA cấp vốn cho chương trình Speech Understanding Research (SUR). Kết quả nổi bật là "Harpy" của Carnegie Mellon, có khả năng hiểu hơn 1.000 từ, mở đường cho việc nhận diện ngôn ngữ tự nhiên quy mô rộng hơn.

5. **Dragon Dictate & Dragon NaturallySpeaking (1990 & 1997)**
   – **Dragon Dictate (1990):** Phần mềm đầu tiên dành cho PC gia đình, giá 9.000 USD, cho phép người dùng nói từng từ một và phải chờ hệ thống xử lý trước khi tiếp tục.
   – **Dragon NaturallySpeaking (1997):** Hiểu giọng nói liên tục, tốc độ lên đến 100 từ/phút, giá giảm xuống còn 695 USD.

6. **IBM Simon (1994)**
   – Là PDA đầu tiên tích hợp chức năng nhận diện giọng nói – được coi là "smartphone" đầu tiên trong lịch sử, ra mắt 1994, sớm hơn các điện thoại thông minh phổ biến gần 25

năm.

7. **Google Voice Search / Google Assistant (2008–2011)**
   – Khi Android ra mắt năm 2008, Google dần tích hợp tính năng tìm kiếm bằng giọng nói vào ứng dụng mobile. Đến 2011, Google phát hành app Google Voice Search độc lập, về sau phát triển thành Google Now và Google Assistant.

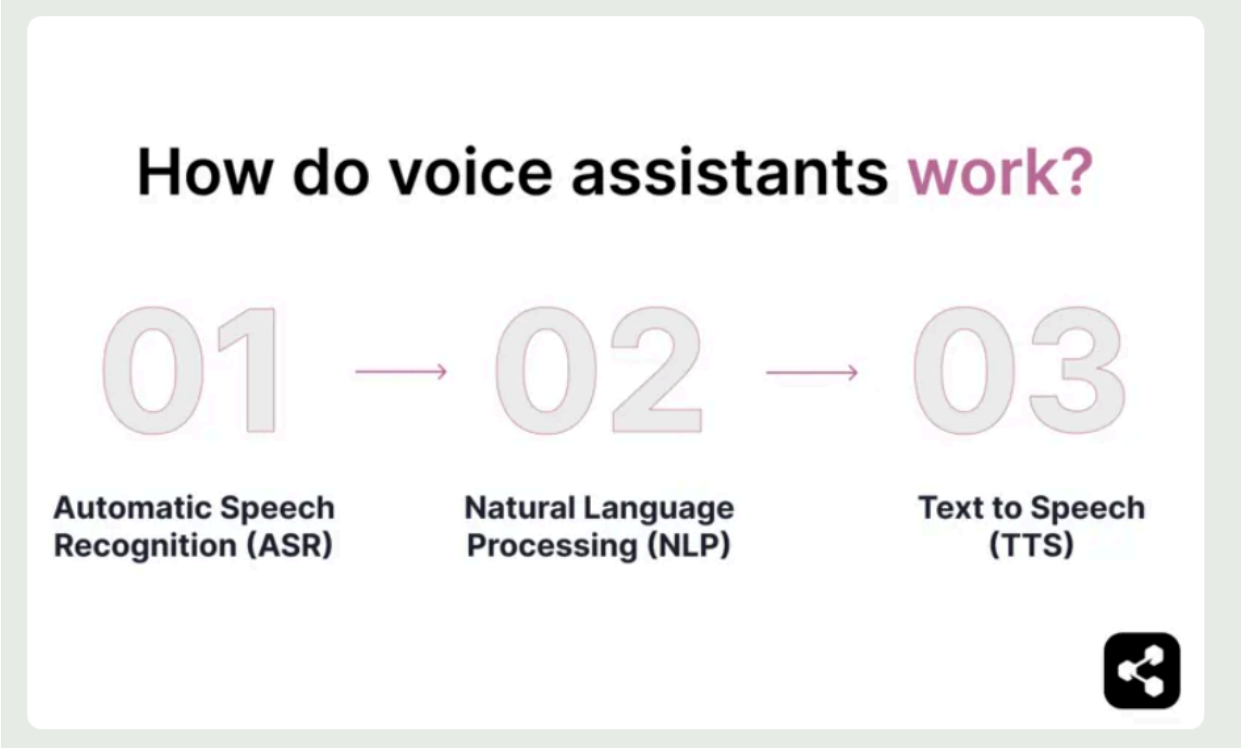
8. **Siri – Apple (2010–2011)**
   – Siri khởi nguồn từ dự án của SRI International, ra mắt dưới dạng app iOS năm 2010 với công nghệ nhận dạng của Nuance. Hai tháng sau, Apple mua lại và tích hợp Siri vào iPhone 4s phát hành cuối 2011. Từ đó, Siri có mặt trên mọi thiết bị Apple, kết nối trong hệ sinh thái chung.

9. **IBM Watson (2011)**
   – Watson, vốn được phát triển từ 2006 để thi đấu Jeopardy, công khai năm 2011. Đây là một trong những hệ thống AI có khả năng hiểu ngôn ngữ tự nhiên và xử lý dữ liệu tiên tiến nhất.

10. **Amazon Alexa (2015)**
    – Alexa ra mắt năm 2015, lấy cảm hứng từ thư viện Alexandria và âm “X” dễ nhận diện. Cùng với loa thông minh Echo, Alexa đem trợ lý giọng nói đến nhà người tiêu dùng với chi phí hợp lý.

11. **Alan (2017)**
    – Alan (ban đầu tên “Synqq”) công bố năm 2017, là nền tảng AI giọng nói hướng đến ứng dụng doanh nghiệp, giúp các công ty tích hợp nhanh chóng vào sản phẩm và dịch vụ.

3. Công nghệ cốt lõi và hướng tiếp cận của VA



Công nghệ lõi của VA tiêu biểu bao gồm:

1. **Nhận dạng giọng nói (ASR – Automatic Speech Recognition)**
   Trợ lý bắt đầu bằng cách sử dụng Nhận dạng giọng nói tự động (ASR) để ghi lại và chuyển giọng nói thành văn bản. Khi bạn nói "Thời tiết hôm nay thế nào?", hệ thống ASR của trợ lý sẽ chia sóng âm thanh của giọng nói thành các từ mà nó có thể xử lý, thậm chí tính đến cả giọng nói hoặc tiếng ồn xung quanh.

2. **Xử lý ngôn ngữ tự nhiên (NLU – Natural Language Understanding)**
   Tiếp theo, trợ lý sử dụng xử lý ngôn ngữ tự nhiên (NLP) để phân tích văn bản và xác định ý định của mình. Nó xác định yêu cầu chính – “thời tiết” – và hiểu rằng mình đang yêu cầu dự báo thời tiết cho hôm nay. Nó cũng có thể sử dụng các manh mối theo ngữ cảnh, như vị trí của bạn, để tinh chỉnh phản hồi của mình.

3. **Tổng hợp giọng nói (TTS – Text-to-Speech)**
   Biến văn bản phản hồi của hệ thống thành âm thanh tự nhiên, để nghe và tương tác giống như đang trò chuyện với con người.

   Sau khi trợ lý thu thập thông tin (ví dụ: kiểm tra API thời tiết để biết dự báo thời tiết địa phương), nó sẽ tạo ra phản hồi dưới dạng văn bản: "Thời tiết hôm nay nắng với nhiệt độ cao nhất là 75°F". Hệ thống Chuyển văn bản thành giọng nói sẽ chuyển đổi văn bản này

thành giọng nói rõ ràng, giống giọng người và phát lại cho mình.

Các trợ lý giọng nói (VA) hiện đại thường kết hợp hai hướng tiếp cận chính để phục vụ nhu cầu của người dùng:

1. **Hướng tiếp cận theo nhiệm vụ (Task-oriented)**
    - Tập trung vào việc hoàn thành các tác vụ cụ thể dựa trên "mục tiêu" do người dùng đưa ra.
    - VA sẽ tương tác trực tiếp với các ứng dụng có sẵn trên thiết bị để thực thi lệnh.
    - Ví dụ: Khi bảo "Đặt báo thức lúc 3 giờ chiều," trợ lý sẽ gọi API hoặc giao tiếp với ứng dụng Đồng hồ, tự động mở giao diện đặt báo thức, hỏi có muốn đặt tên cho báo thức hay thay đổi âm báo không, rồi sau đó xác nhận lại.
2. **Hướng tiếp cận theo kiến thức (Knowledge-oriented)**
    - Sử dụng dữ liệu, kiến thức đã lưu trữ hoặc truy vấn trực tuyến để trả lời các câu hỏi mang tính thông tin.
    - VA sẽ tra cứu cơ sở dữ liệu hoặc công cụ tìm kiếm để lấy thông tin chính xác nhất.
    - Ví dụ: Khi hỏi "Thủ đô của Việt Nam là gì?", trợ lý sẽ gửi truy vấn ra Internet hoặc cơ sở dữ liệu nội bộ, sau đó trả về đáp án "Hà Nội".

4. Lợi ích và tương lai của VA

Đối với người dùng, VA đem lại lợi ích như:

1. Tiện lợi
- cho phép làm việc từ xa, giúp dễ dàng đặt lời nhắc, điều khiển thiết bị thông minh hoặc nhận câu trả lời nhanh trong khi thực hiện nhiều tác vụ cùng lúc.
2. Sẵn sàng
- cung cấp giao diện thân thiện với người dùng cho những người khuyết tật hoặc những người gặp khó khăn với các tương tác công nghệ truyền thống, giúp truy cập thông tin và công cụ tốt hơn.
3. Hiệu quả
- hợp lý hóa các tác vụ như lên lịch, gửi tin nhắn hoặc truy xuất thông tin nhanh hơn so với các phương pháp nhập thủ công.
4. Cá nhân hóa

- tìm hiểu sở thích của người dùng theo thời gian, điều chỉnh phản hồi và gợi ý theo nhu cầu của từng cá nhân, chẳng hạn như đề xuất tuyến đường hoặc ghi nhớ các tác vụ thường xuyên.

5. Tích hợp vào thiết bị trong nhà

- có thể hoạt động như trung tâm cho các thiết bị nhà thông minh, cho phép người dùng điều khiển đèn, thiết bị hoặc hệ thống an ninh bằng các lệnh thoại đơn giản

Đối với các doanh nghiệp hay công ty thì:

1. **Tăng hiệu suất làm việc (Productivity)**

    - Cho phép nhân viên ra lệnh bằng giọng nói để tra cứu nhanh thông tin nội bộ (tài liệu, báo cáo, lịch họp…) mà không phải chuyển đổi qua lại giữa các ứng dụng.

    - Thực hiện các tác vụ lặp đi lặp lại (đặt lịch, gửi email mẫu, khởi tạo báo cáo đơn giản) chỉ bằng một câu nói, tiết kiệm thời gian so với thao tác thủ công.

2. **Cải thiện dịch vụ khách hàng (Customer Experience)**

    - VA có thể tích hợp vào kênh hotline hoặc chatbot để trả lời tự động 24/7 các câu hỏi thường gặp (FAQ), giảm tải cho bộ phận chăm sóc khách hàng.

    - Với khả năng nhận diện ngôn ngữ tự nhiên, trợ lý giọng nói giúp giao tiếp thân thiện, cá nhân hóa và nhanh chóng hơn, từ đó nâng cao mức độ hài lòng của khách hàng.

3. **Tối ưu quy trình nội bộ (Workflow Automation)**

    - Kết nối liền mạch với các hệ thống CRM, ERP, HRM… để tự động hóa quy trình: ví dụ mở ticket hỗ trợ, cập nhật tình trạng đơn hàng, ghi nhận chấm công hay đặt mua vật tư.

    - Giảm thiểu sai sót do con người và tăng tính đồng nhất trong các bước xử lý.

4. **Hỗ trợ ra quyết định thông minh (Data-Driven Insights)**

    - VA có thể truy vấn trực tiếp vào kho dữ liệu doanh nghiệp (BI, Data Warehouse) để cung cấp báo cáo ngắn gọn, số liệu KPI, biểu đồ hiệu suất… chỉ qua câu lệnh.

    - Giúp ban lãnh đạo và quản lý cấp trung đưa ra quyết định nhanh chóng dựa trên dữ liệu cập nhật tức thì.

5. **Tiết kiệm chi phí vận hành**

- Tự động hóa nhiều tác vụ thủ công sẽ giảm nhu cầu nhân lực cho các công việc lặp lại, từ đó cắt giảm chi phí lương và đào tạo.

- Hạ tầng VA đám mây (cloud-based) cho phép doanh nghiệp mở rộng hoặc thu hẹp dịch vụ linh hoạt, chỉ phải trả phí theo mức sử dụng thực tế.

6. **Hỗ trợ đa ngôn ngữ và tiếp cận toàn cầu**

    - Đối với công ty đa quốc gia, VA có thể được huấn luyện đa ngôn ngữ, giúp nhân viên và khách hàng ở nhiều vùng miền tương tác dễ dàng.

    - Nâng cao trải nghiệm người dùng khi họ được phục vụ bằng ngôn ngữ mẹ đẻ.

7. **Tăng tính linh hoạt và di động**

    - Nhân viên có thể tương tác với hệ thống khi đang di chuyển, họp bên ngoài hay thực hiện công việc tay chân (ví dụ: kiểm kho, bảo trì thiết bị) mà không cần dùng bàn phím.

8. **Nâng cao tính bảo mật và tuân thủ**

    - VA tích hợp xác thực đa yếu tố (voice biometrics), giúp kiểm soát truy cập vào các chức năng nhạy cảm.

    - Ghi lại lịch sử tương tác (log) để đảm bảo tính minh bạch, hỗ trợ audit và tuân thủ quy định.

Hướng phát triển của VA trong tương lai:

1. Tích hợp sâu hơn vào hệ sinh thái thiết bị

- **Mở rộng ra mọi thiết bị**: Từ đồng hồ treo tường, ô tô, thiết bị gia dụng cho tới không gian công cộng, VA sẽ luôn sẵn sàng lắng nghe và phản hồi ngay tức thì.

- **Tương tác "không chạm"**: Chỉ cần ra lệnh bằng giọng nói, người dùng có thể khởi động máy giặt, kiểm tra lịch trình xe buýt, hay đặt hẹn bác sĩ mà không cần chạm vào màn hình.

2. Hội thoại tự nhiên hơn

- **Xử lý song song**: Người dùng không còn phải nói từng cụm rồi chờ trợ lý "bắt kịp" nữa. VA sẽ hiểu mạch lời một cách liền mạch, như đang nói chuyện với một con người.

- **Giảm độ trễ**: Nhờ phần cứng mạnh hơn và thuật toán tối ưu, độ trễ giữa lệnh nói và phản hồi sẽ gần như bằng 0, mang lại trải nghiệm liền mạch, thư thái.

3. Luồng tác vụ phức tạp hơn

- **Kết nối ngữ cảnh liên tiếp**: Thay vì phải thực hiện từng bước riêng (hỏi thời gian di chuyển, rồi hỏi phương án khác), người dùng sẽ có thể hỏi phức hợp:

  "Nếu Uber nhanh hơn xe buýt tới công ty, đặt giúp tớ một chuyến Uber và cho biết mất bao lâu?"

- **Chủ động đề xuất**: VA có thể tự giám sát lịch làm việc, thời tiết, lưu lượng giao thông… rồi gợi ý phương án tối ưu mà không cần được nhắc.

4. Ứng dụng chuyên biệt mới

- **Chăm sóc sức khỏe cá nhân hóa**: VA hỗ trợ theo dõi thuốc men, nhắc lịch khám, thậm chí phân tích triệu chứng ban đầu dựa trên giọng nói và thông tin y tế cá nhân.

- **Giáo dục thông minh**: Giao diện giọng nói sẽ hỗ trợ bài giảng tương tác, trắc nghiệm bằng lời, giải thích khái niệm ngay lập tức.

- **Tiếp cận đa ngôn ngữ**: Hỗ trợ dịch song song, phiên dịch thời gian thực, giúp kết nối toàn cầu và xóa bỏ rào cản ngôn ngữ.

5. AI và Machine Learning làm "xương sống", tích hợp LLM vào Voice Assistants

- **Hiểu ngữ cảnh sâu (context-aware)**: Nhận diện thói quen, sở thích, hoàn cảnh (vị trí, thời gian, lịch làm việc) để tùy biến phản hồi.

- **Cá nhân hóa và chủ động**: Dựa trên lịch sử tương tác, VA sẽ gợi ý trước khi người dùng yêu cầu (ví dụ: nhắc nạp tiền điện thoại khi cước sắp hết).

- **Học không ngừng**: Hệ thống liên tục cập nhật mô hình từ phản hồi người dùng để ngày càng thông minh và chính xác.

Các tập đoàn , doanh nghiệp hàng đầu đang rục rịch để cải tiến VA của mình:

Công ty	Sản phẩm / Dịch vụ	Điểm nhấn tương lai
Amazon	“Alexa+” (2025)	- AI nâng cao từ Anthropic- Hội thoại tự nhiên hơn- Phân tích hình ảnh, tự động hóa routine, mua vé sự kiện, đặt hàng tạp hóa… (19.99 USD/tháng, miễn phí cho Prime)
Apple	Apple Intelligence (iOS 18)	- NLP cải tiến cho câu hỏi phức tạp- Giữ ngữ cảnh liên tiếp- “On-screen awareness” & hành động trong ứng dụng- Tích hợp ChatGPT để đàm thoại nâng cao
OpenAI	ChatGPT Advanced Voice Mode	- Hội thoại gần gũi hơn, đáp ứng nhanh- Dịch thuật trực tiếp (live translation)- Kết hợp cả tương tác âm thanh và hình ảnh
Google	Google Assistant (Gemini AI)	- Thay thế hoàn toàn Assistant cũ bằng AI-enabled- Nâng cao khả năng giao tiếp tự nhiên- Hoàn thành tác vụ phức tạp, tích hợp trong mọi thiết bị Android