HO CHI MINH UNIVERSITY OF TECHNOLOGY

FALCUTY OF COMPUTER SCIENCE AND ENGINEERING

ASSIGNMENT 3

PROJECT 52

APPLICATION OF ARTIFICIAL INTELLIGENCE IN VIRTUAL ASSISTANTS

Le Minh Khoi	1952076
Nguyen Minh Tam	1952968
Nguyen Le Nhat Duong	1952638

Contents

1 Introduction

1.1 What is AI ?

Artificial intelligence (AI) is an area of computer science of making intelligent machines that emphasizes the creation of them that react like humans. Some of the activities computers with artificial intelligence are designed for include:

- Speech recognition
- Learning
- Planning
- Problem solving [1]

1.2 Recent feats of AI

Beating the world champion of Go

In 2016, Google DeepMind's AI software AlphaGo was the first computer program to defeat a professional human Go player and is arguably the strongest Go player in history.

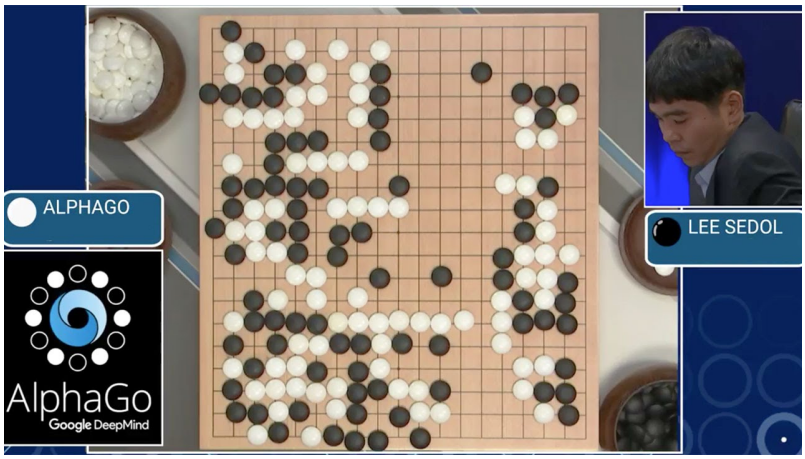


Figure 1: AlphaGo defeats Lee Sedol - the world Go champion

Dominating Atari games

Old Atari games used to be extremely fun for humans, but we never really managed to dominate them like the way AI has recently done. DeepMind – a London-based AI company acquired by Google

in 2014 – was recently presented with a few classic Atari games. At first, Google's DeepMind was not great but lately it started to play better than any human ever did.

Figure 2: Google Deepmind DQN plays Atari Breakout

2 Applications of AI in virtual assistant

2.1 History of Intelligent virtual assistant :

Radio Rex was the first voice activated toy released in 1911. Another early tool which was enabled to perform digital speech recognition was the IBM Shoebox. This early computer, developed almost 20 years before the introduction of the first IBM Personal Computer in 1981, which was able to recognize 16 spoken words and the digits 0 to 9.

Figure 3: Radio Rex

In the 1990s digital speech recognition technology quickly became a feature of the personal computer with Microsoft, IBM, Philips fighting for customers. Later on the market launch of the first smartphone IBM Simon in 1994 laid the foundation for smart virtual assistants as we know them today.

The first modern digital virtual assistant installed on a smartphone was Siri, which was first introduced as a feature of the iPhone 4S on October 4, 2011.

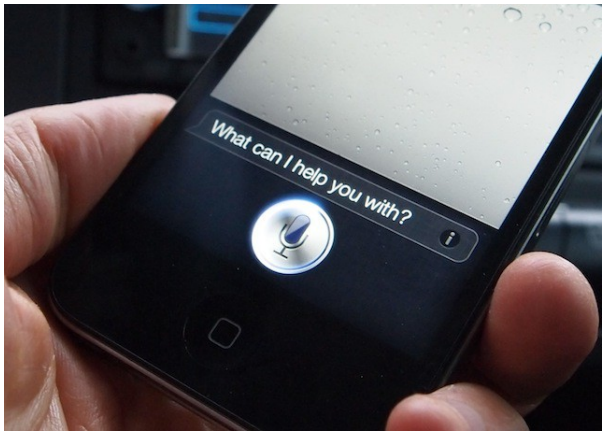


Figure 4: Siri on iPhone 4S

2.2 The concept working of virtual assistant :

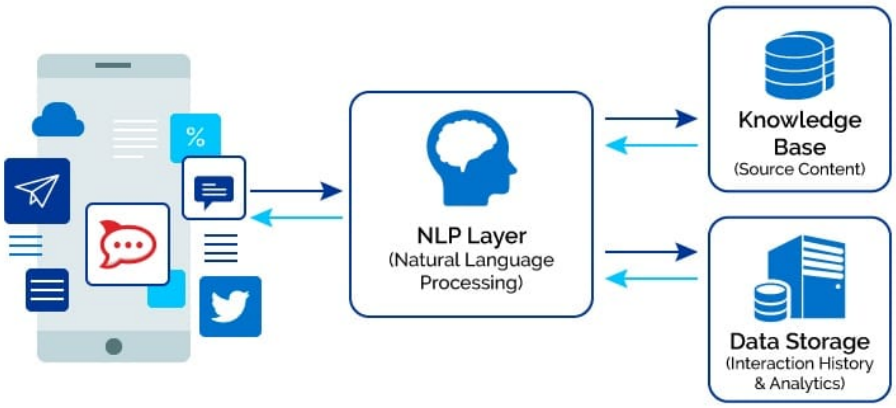


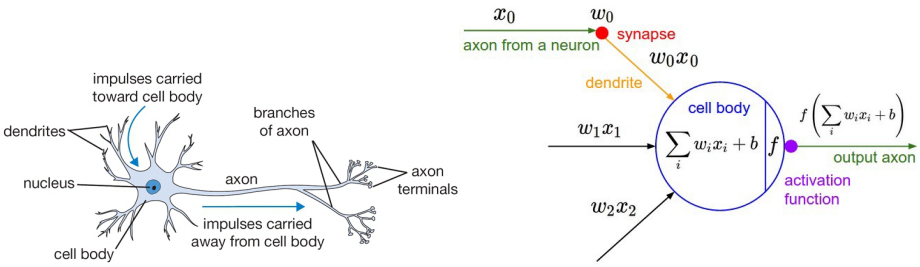
Figure 5: How Natural Language Processing help build better Virtual Assistant

Figure 5 has illustrated the concept of work behind most common virtual assistants. User will use device like smartphones, Alexa Echos, etc to use the interface. User can speak to virtual assistant (Siri, Alexa) or text to it. All the information the user has provided will be transferred to the NLP Part, which stands for Natural Language Processing. In this stage, the information will be processed, analysed by the machine. Then the machine will get the corresponding information from the Knowledge base and respond (in the case of user's asking question) or give out instructions to other part to perform a task (if user tell the virtual assistant to perform a job).

From the above explanation, NLP(Natural Language Processing) is considered to be the workhorse of the any common virtual assistant. In addition, NLP is built with the help of Artificial Neural Network(ANN).

2.2.1 Introduction to Neural Network

Artificial neural networks(ANN)[2] are computing system that are inspired by, but not identical to, biological neural networks that constitutes animal brains.



(a) Biological Neural Network (b) Biological Neural Network

Figure 6: Biological and Artificial Neural Network

Figure 6 illustrates some similarities of two types of neural network. From figure 6b, cell body (often referred as the node or unit, which is the basic unit of ANN) receives input data such as $x_0, x_1, ..., x_n$ and some weights $w_0, w_1, w_2, ..., w_n$. Those input data and weights are calculated, accumulated, summed up, applied function f on the data and yield out the result. Similar to biological neural network, from figure 6a, in which the input data here will be the impulses carried in and the result is the impulses carried out.

The idea that makes artificial neural network performs nearly like the biological neural network is that the weights w is learnable and control the strength of influence and its direction: excitory(positive weight) or inhibitory(negative weight).

From the above explanation, we can conclude that ANN is composed of the following components:

- **Input Nodes (input layer)**: No computation is done here within the layer, they just simply pass the information to the next layer
- **Hidden Nodes (hidden layer)**: In this layer, intermediate processing or computation is done, they perform computations and then transfer the weights from the input layer to next layer (which will the next hidden layer or output layer). There is sometimes possible that ANN does not have hidden layer.
- **Output Nodes (out layer)**: Use activation function that maps to the desired output format.

- **Connection and weights**: Connections transfer the output of 1 node to the input of another node. Each connection is assigned with the weight. For example, connection between node i and node j has the weight w_{ij}

2.2.2 Types of Neural Network

1. **Feedforward Neural Network**

   (a) **Single-layer Perceptron** No hidden layer. Weights and Inputs are directly put into node, which yield out the result

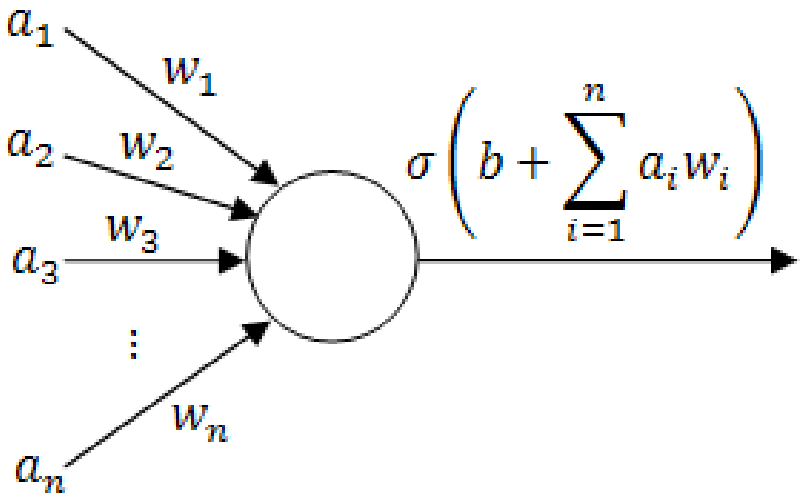


Figure 7: Single-layer Perceptron

   (b) **Multi-layer Perceptron(MLP)** This class of networks consists of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer.[3]

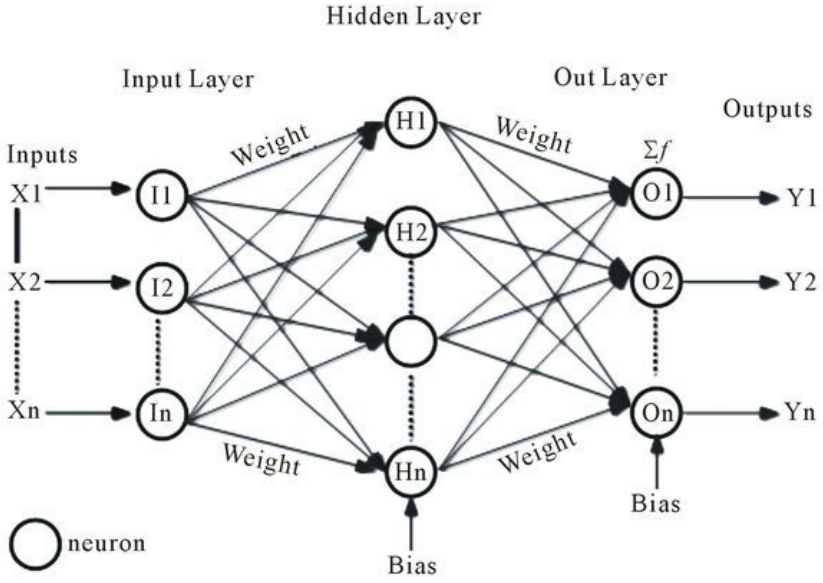


Figure 8: Multi-layer Perceptron

(c) **Convolutional Neural Network (CNN)** In neural networks, Convolutional neural network (CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used.

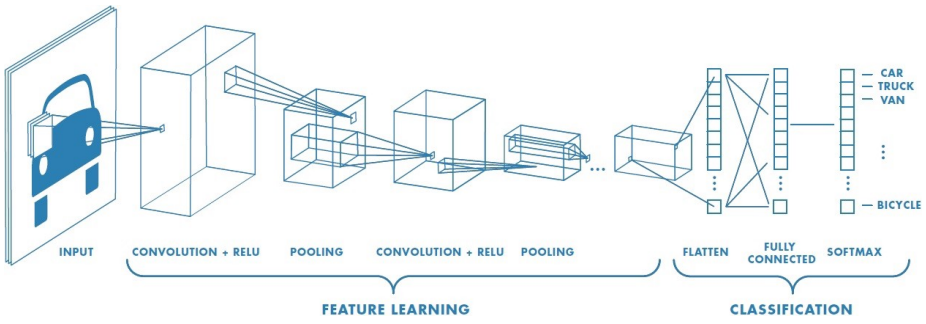


Figure 9: Convolutional Neural Network (CNN)

2. **Recurrent neural networks** In recurrent neural network (RNN), connections between units form a directed cycle (they propagate data forward, but also backwards, from later processing stages to earlier stages).

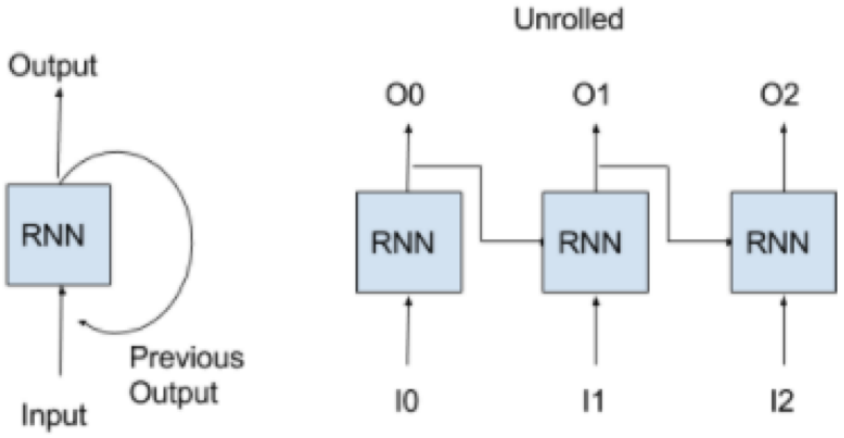


Figure 10: Recurrent neural networks

2.2.3 The process of NLP handle user's command

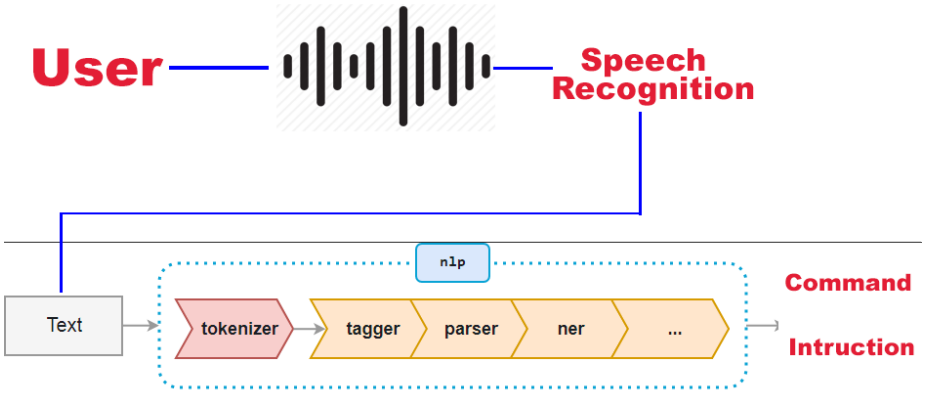


Figure 11: Process of handling user's command

1. **Speech Recognition**: As name suggested, this part will convert the audio file, which is the user's command in the form of speech into text. There are some famous libraries, API to implement the speech recognition such as: Google Cloud Speech-to-Text(Google), SpeechRecognition - Web APIS(Mozilla), DeepSpeech(Mozilla), Wav2Letter(Facebook)[4]

Figure 12: Speech Recognition

2. **Tokenization**: The process of chopping sentence(content) into piecess called tokens. A token is an instance of a sequence of characters in some particular document that are grouped together as a useful semantic unit for processing.[5]

```
Remind me to watch the SEA GAMES final game
1 Remind
2 me
3 to
4 watch
5 the
6 SEA
7 GAMES
8 final
9 game
```

Figure 13: Tokenization

3. **Part-of-Speech Tagging**: In corpus linguistics, part-of-speech tagging, also called grammatical tagging, is the process of marking up a word in a text (corpus) as corresponding to a particular part of speech, based on both its definition and its context.[6]

```
Remind me to watch the SEA GAMES final game
1 Remind VERB
2 me PRON
3 to PART
4 watch VERB
5 the DET
6 SEA PROPN
7 GAMES PROPN
8 final ADJ
9 game NOUN
```

Figure 14: Part-of-Speech Tagging

4. **Dependency Parser**: A dependency parser analyzes the grammatical structure of a sentence, establishing relationships between "head" words and words which modify those heads. The figure below shows a dependency parse of a short sentence.[7]

```
Remind me to watch the SEA GAMES final game
1 Remind ROOT
2 me dobj
3 to aux
4 watch xcomp
5 the det
6 SEA nmod
7 GAMES nmod
8 final amod
9 game dobj
```

Figure 15: Dependency Parser

5. **Named Entity Recognition (NER)**: The main task of named entity recognition (NER) is to classify named entities, such as Guido van Rossum, Microsoft, London, etc., into predefined categories like persons, organizations, locations, time, dates,

and so on. [8]

```
2019-12-07 15:03:26,978 loading file /root/.flair/models/en-ner-conll03-v0.4.pt
Sentence: "Remind me to watch the SEA GAMES final game" - 9 Tokens
The following NER tags are found:
MISC-span [6,7]: "SEA GAMES"
```

Figure 16: Named Entity Recognition

6. **Text classification and categorization**: Text classification or Text Categorization is the activity of labeling natural language texts with relevant categories from a predefined set.[9]

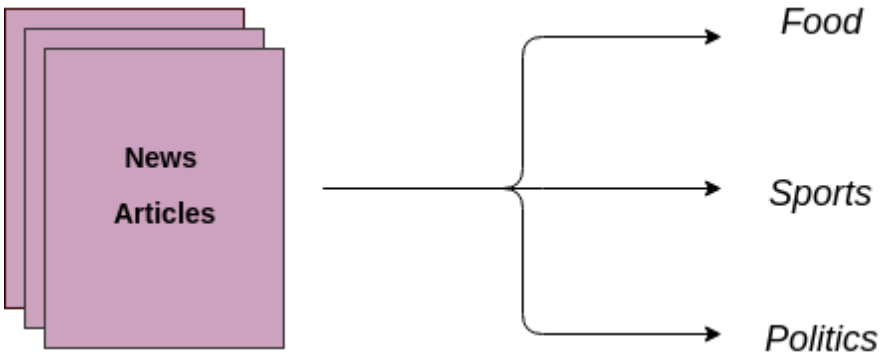


Figure 17: Text Classification

2.3 Some common virtual assistants we encounter in the daily life

2.3.1 Amazon Alexa

What is Amazon Alexa ?

The Amazon Echo is one of a range of hands-free speakers and devices from Amazon that can be controlled with your voice. The voice-controlled "personal assistant" on these devices is called Alexa, which will perform various tasks for you and control various systems. [10]

What devices offer Alexa ?

There are two sides to this question - devices that work with Alexa and devices that offer Amazon Voice Services, which is the platform that runs Alexa.

Firstly, Alexa is designed around Amazon's own Echo devices. The Amazon Echo range includes the standard Echo, Echo Plus, Echo Studio, and Echo Dot, which are all speakers, and then the Echo Show, Echo Show 5, Echo Show 8, and the Echo Spot, which also feature a display, so you can give visual feedback, like weather widgets, videos or song lyrics. There are several Amazon Alexa gadgets too, like the Echo Wall Clock and Echo Flex smart plug for example.

(a) Echo Show 5

(b) Echo Wall Clock

(c) Echo Flex

Figure 18: Amazon Echo devices

The cheapest Echo device is the Echo Dot, which is a good starting point for building an Echo system and getting started.

Figure 19: Echo Dot

There are plenty of other devices that offer Alexa voice control, such as the Sonos devices, Bose Home Speaker 500 or Polk Command Bar.

The way all these devices come to work is easy. When they hear the Alexa wake word from you, Alexa will swing into action and respond to your commands.

What can Alexa do ?

Alexa is able to play music, provide information, deliver news and sports scores, broadcast the weather forecast, control your smart home, etc. Alexa expands the information offered all the time and corrects the responses to give you more accurate information. No matter what Alexa device you ask, all can return with satisfactory answers at anytime you want.

One of the main functions of Alexa is playing music. Amazon Music is supported but beyond that there's support for many more services like Spotify, Deezer, Apple Music, TuneIn as well as apps from individual providers, like the BBC.

2.3.2 Google Assistant

What is Google Assistant ?

Google Assistant is an *artificial intelligence-powered virtual assistant* developed by Google that is primarily available on mobile and smart home devices.[11] The Google Assistant can engage in two-way conservation. It has made incredible progress since its 2016 launch and is probably the most advanced and dynamic of the assistants out there. Google Assistant supports both text or voice entry and it will follow the conversation whichever entry method you're using.

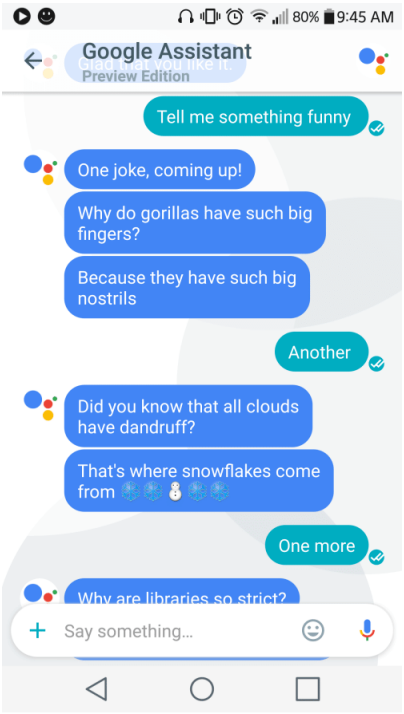


Figure 20: Google Assistant

What can Google Assistant do ?

Google Assistant offers voice commands, voice searching, and voice-activated device control, letting you complete a number of tasks after you've said the "OK Google" or "Hey, Google" wake words. It is designed to give you conversational interactions.[12] Specifically, Google Assistant is able to :

- Control your devices and your smart home
- Access information from your calendars and other personal information
- Find information online, from restaurant bookings to directions, weather and news
- Control your music
- Play content on your Chromecast or other compatible devices

- Run timers and reminders
- Make appointments and send messages
- Open apps on your phone
- Read your notifications to you
- Real-time spoken translations

Which devices offer Google Assistant ?

Google Assistant originally launched on the Google Pixel smartphones, Google Home and all modern Android devices, including Wear OS devices, Android TV, Nvidia Shield and in some cars if they offer support for Android Auto.

Smart home devices like Philips Hue, Nest products and Ikea's Home Smart range, for example, can be controlled by Google Assistant and not just through Google Home, but wherever you happen to interact with Assistant.

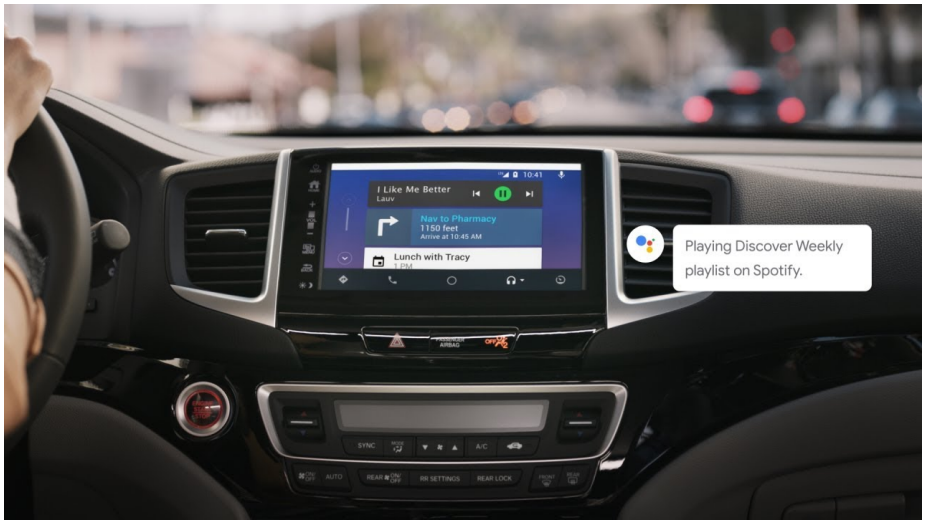
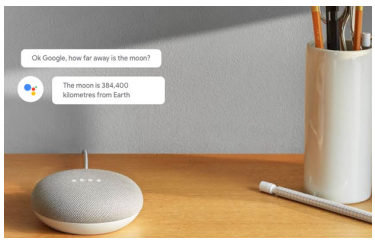


Figure 21: Android Auto



(a) Google Home

(b) Wear OS device

Figure 22: Google Assistant devices

2.3.3 Apple's Siri

What is Apple's Siri ?

Siri is a *built-in voice-controlled personal assistant* available for Apple users. She was first integrated on Iphone 4S in 2011.[13]

Siri is designed to offer you a seamless way of interacting with your iPhone, iPad, iPod Touch, Apple Watch, HomePod or Mac by you speaking to her and her speaking back to you to find or do what you need. You can ask her questions, tell her to show you something or issue her with commands for her to execute on your behalf, hands-free.

Siri has access to every other built-in application on your Apple device - Mail, Contacts, Messages, Maps, Safari and so on - and will call upon those apps to present data or search through their databases whenever she needs to. Ultimately, Siri literally does all the hard boring work for you.

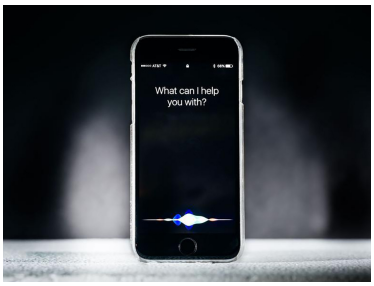


Figure 23: Siri

How Siri works ?

Siri is based on the based on the fields of Artificial Intelligence and Natural Language Processing, and it is comprised of three components - a conversational interface, personal context awareness and service delegation.[14]

The conversational interface is about how Siri understands you in the first place. The general workings of the straight word-for-word voice recognition have to be good in order to hear what you're saying but deciphering the meaning is all down to statistics and machine learning, which is where the personal context awareness system comes in.

The service delegation system is the unbridled access to all of the iPhones built-in apps and their inner workings. This access is how Siri works.

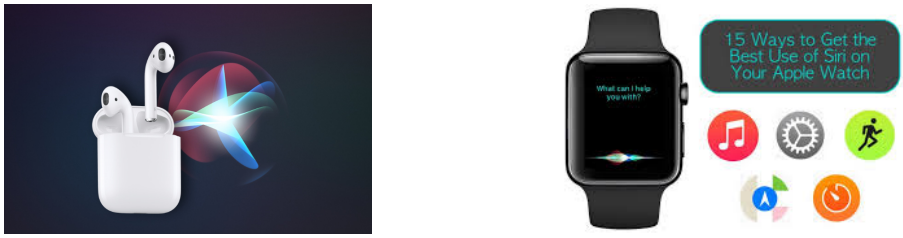


Figure 24: Siri is also available on Airpod and Apple's watch

What can Siri do ?

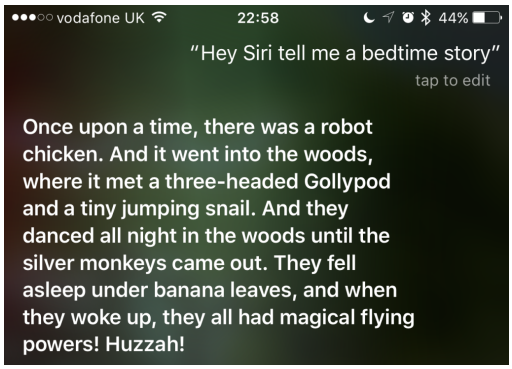
Like Amazon Alexa and Google Assistant, Siri can do common things such as setting a reminder, calling, playing music, etc. Here we will show you some more surprising things that Siri can do.

- Translation
  You only need to ask:"How do you say[word or phrase] in [language]?", Siri can help you translate English into other languages.
- Bedtime story
  "Tell me a bedtime story please" and Siri will lead you to an

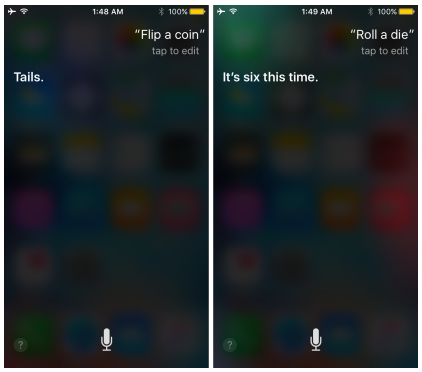
entertaining story, as told by Siri too. However you need to install Podcast app on your Iphone at first.

- Play dice
  Whether for fun or for decision-making, why don't you ask Siri for help, just by saying "Roll a dice".
- Flip a coin
  Or you can ask Siri to give you an answer by flipping a coin.
- Identify a song
  Sitting in the cafe and accidentally hear an interesting song at that time but do not know the name of it. Siri can help you identify a song by holding your phone close to the speaker and saying "Name that tune".

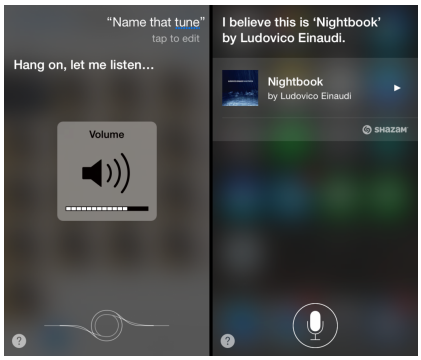
(a) Siri translation



(b) Bedtime story



(c) Dice and Coin



(d) Name of song

Figure 25: Siri devices

2.4 The future of virtual assistant

Virtual assistants like Alexa, Siri and Cortana are slowly becoming part of our everyday life. What prospects does the digital assistance technology have and how will it shape our lives in the future ?

The Future: A Time More Connected

Although we may be experiencing the growing pains of a technology not yet operating at full capacity, there is no doubt that this is fast changing.

According to the estimates, by 2020, 50% of all searches will be voice activated. With this comes a whole host of changes to the way the web operates as we know it. So how will these developments affect our everyday life and in what areas?

Search and Advertising

One of the most immediate changes that voice search will bring is in how questions will be posed to Google. This will, in turn, alter the way the search giant responds to commands. For the average person, this will have a gradual and possibly unnoticeable impact. However, this will completely change the way that online marketing works.

Currently, search optimization focuses primarily on a website's use of relevant keywords, which helps to determine the site's SERP ranking. Keywords are selected based upon terms posed to Google and make the process of testing websites using special SEO tools possible, and relatively simple.

With the advent of a voice search dominated field, however, optimization will need to adapt to a new form of command phraseology. Keywords will no longer consist of a few short instructional words but full sentences. People will no longer have to adapt to machine-processed language as the machines themselves adapt to natural human language.

Figure 26: Artificial Assistant will trigger a promising future

Teaching and Learning Tools

One of the productive outcomes of virtual assistant developments will come from their instructional value. Within a classroom setting, you can expect to see a smartboard that can be navigated and used solely from voice command wipe out the need for whiteboards and

clumsy projector gear.

Just as the chalkboard has taken on a relic-like status, so too shall the pen. The contents of textbooks can be digitized and made not only instantly accessible but searchable as well.

Connecting Home and Office

There is no doubt that technological developments will create seismic shifts in the way we occupy spaces. At home and at work we will see this happen most noticeably. For many people, these two spheres may even become more closely intertwined.

Broad guesses about the workplace of the future have already been made, yet the practical benefits are more measurable. Administration will be done hands-free with immediate time-saving effects for companies, while control of appliances in the home will be possible from any location.

The safety afforded by no longer having both your hands and sight tied to a device is projected to see a dramatic drop in traffic accidents and personal injuries sustained through mobile use.

Although for the time being, we can only project as to what the future holds, there is certainly little doubt that virtual assistants will be at the forefront.[15]

References

[1] Techopedia, "What does artificial intelligence mean ?." https://www.techopedia.com, Aug 2014. Accessed on 2019-12-7.

[2] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer*, vol. 29, no. 3, pp. 31–44, 1996.

[3] D. Fumo, "A gentle introduction to neural networks series—part 1," *Towards Data Science*, 2017.

[4] V. Pratap, A. Hannun, Q. Xu, J. Cai, J. Kahn, G. Synnaeve, V. Liptchinsky, and R. Collobert, "wav2letter++: The fastest open-source speech recognition system," *arXiv preprint arXiv:1812.07625*, 2018.

[5] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*, vol. 39. Cambridge University Press Cambridge, 2008.

[6] B. Plank, A. Søgaard, and Y. Goldberg, "Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss," *arXiv preprint arXiv:1604.05529*, 2016.

[7] M. M. Lopez and J. Kalita, "Deep learning applied to nlp," *arXiv preprint arXiv:1703.03091*, 2017.

[8] G. Lample, M. Ballesteros, S. Subramanian, K. Kawakami, and C. Dyer, "Neural architectures for named entity recognition," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (San Diego, California), pp. 260–270, Association for Computational Linguistics, June 2016.

[9] S. Gupta, "Text classification: Applications and use cases." https://towardsdatascience.com/text-classification-applications-and-use-cases-beab4bfe2e62, February 2018. Accessed on 2019-12-7.

[10] B. O'Boyle, "What is alexa and what can amazon echo do?." https://www.pocket-lint.com, Nov 2019. Accessed on 2019-12-7.

[11] Wikipedia, "Google assistant." https://en.wikipedia.org, Nov 2019. Accessed on 2019-12-7.

[12] M. Tillman and B. O'Boyle, "What is google assistant and what can it do?." https://www.pocket-lint.com, Aug 2019. Accessed on 2019-12-7.

[13] K. Long, "What is siri : Explained." https://cellularnews.com, Sep 2019. Accessed on 2019-12-7.

[14] B. O'Boyle, "What is siri and how does siri work?." https://www.pocket-lint.com, Sep 2019. Accessed on 2019-12-7.

[15] V. Soleil, "How will virtual assistants shape our lives in the near future." https://www.learning-mind.com, Jun 2018. Accessed on 2019-12-7.