Master of Science in Internetworking

Capstone Project

on

Analysis of currently open and closed-source software for the creation of an AI personal assistant.

Presented By:

Rubal Preet Singh

Under the Supervision of

Leonard Rogers And Dr. Mike McGregor

Abstract

In this modern age of the 21st century, personal assistants are fantastic for everybody. It has paved the way for modern technology where computers can be asked questions. As individuals do with humans and may communicate with Intelligent Personal Assistants.

In many ways, this new technology, including mobile phones, smartphones, computers, etc., attracted almost the entire world. Siri, Google Assistant, Cortana, Bixby, Mycroft AI and Alexa are just some big personal assistants. The concerns that have not yet been addressed in these IPAs are speech recognition, contextual comprehension and human interaction.

The implementation and use of Intelligent Personal Assistants (IPA) in operating systems, the Internet of Things (IoT), and various other fields such as Automobile, Business etc. There are several IPA implementations today, and companies such as Apple, Google, and Microsoft all have their implementations in their operating systems and devices as a significant feature.

Using Natural Language Processing (NLP), Artificial Intelligence (AI), Machine Learning (ML), and prediction models in Computer Science (CS), as well as theory and techniques from these areas, IPAs are becoming more intelligent and essential because of Human-Computer Interaction (HCI).

According to the findings, many resources were covered by these assistants, but there are still some improvements needed in voice recognition, contextual comprehension, and hands-free interaction.

This research paper's main aim would undoubtedly be to expand its use after discussing these changes in IPAs. Also, to examine and compare the existing significant implementations of IPAs to determine the specification of which is, at this moment, the most evolved and contributing to the sustainable AI's future.

Table of Contents

Table of Figures

1. Introduction

An AI Personal Assistant is a software package designed with oral and written commands to perform specific tasks for a user. It is an example of AI, that is, an AI machine trained for a specific job. Siri, Alexa, and Google Assistant include examples of AI assistants. [1]

In a world full of new technology, handy and timely information is accessed quickly via Intelligent Personal Assistants (IPAs). Using these Personal assistants built into mobile OS, a user's daily electronic tasks can be done 24/7.

IPAs like Apple's Siri, Google Assistant and Microsoft Cortana can complete such tasks as getting turn-by-turn directions, reminding daily appointments, taking dictation, setting reminders, vocalizing email messages, responding to any factual questions and invoking apps. The mentioned assistants programmed within Artificial Intelligence create an interaction between human and computer through a natural language process used in digital communication. [2]

The scope of this study is to thoroughly examine the potential use of IPAs that uses advanced cognitive computing technologies and Natural Language Processing (NLP) for learning. To achieve this goal, the working system of IPAs are thoroughly reviewed within the scope of AI that has become smarter to predict, comprehend and carry out multi-task and complex requests of users. [2]

2. Scope and objectives

The scope and objective of this Capstone project is to analyze and introduce the theory, concepts and branches of Artificial Intelligence (AI) that are building blocks of IPAs, followed by an introduction to some of the existing personal assistant's applications from six different companies such as Apple's Siri, Google Assistant, Mycroft AI, Samsung Bixby, Amazon Alexa and Microsoft Cortana, also to analyze and compare their behaviour and technology. [3]

This Project is also focused on the system required to run AI personal assistants and what role the system plays in improving the IPA's performance and behaviour.

After that, Natural Language Processing (NLP) and Human-Computer Interaction (HCI) are briefly explained with Machine Learning (ML) and Artificial Intelligence (AI). It also analyzes how close we are to achieving the fictional Jarvis.

After that, the current state and the future state of IPAs are discussed with predictions about their future expansion. [3] This project also focuses on the various concepts on what it would take to finally have an AI-based personal assistant playing more than music or turning on a light when we ask?

Finally, the conclusion deals with the current best implementation in combination with predictions about its future features.

3. Background

The concepts presented in this section, namely the background chapter, include the ideas and models important in Intelligent Personal Assistants.

The first section covers the main parts of AI, followed by the fundamentals of several statistical learning methods and, finally, the introduction and discussion of human-computer interaction and its importance in the sense of IPAs to Natural Language Processing.

3.1 Artificial Intelligence

Artificial intelligence (AI) is a science field that aims to create, study, and understand intelligent entities and different aspects, such as behaviour, processing and reasoning. In this area, the distinction between human and rational behaviour is sometimes discussed. [4] [5]

The two components of intelligence and instruments are needed to construct AI. These instruments have been developed in the field of computer science. This field is described as the study of intelligent agents, which means that any system which recognizes its surroundings and takes actions that maximize its chance of successfully achieving its objectives.

Typically known as AI, modern computer capabilities include efficient comprehension of a human voice, engaging in strategic game systems at the highest level, autonomously driving vehicles, intelligent routing in content distribution networks, and military simulations. [4]

While AI typically invokes pictures of science fiction's sentient computer master, this reality is entirely different. At its heart, AI uses similar essential recursive functions that drive traditional computer code. However, it applies them in an exceedingly completely different manner. A standard warehouse management system, for instance, shows the present levels of assorted merchandise, whereas an associate intelligent one may determine shortages, analyze the cause and its result on the supply chain, and also take steps to correct it. [4]

Artificial intelligence is often allowed by the user to switch a full system, creating all choices end-to-end or being accustomed to enhancing a selected method. For example, we analyze video footage to acknowledge gestures or exchange peripheral devices such as a keyboard, mouse and touchscreen, with a speech-to-text system, giving the impression that one is interacting with a sentient being. [5]

Several instruments are used in AI, including search and mathematical optimization methods, artificial neural networks, and statistical, probability and economic methods. The next segment involves Agents, which helps AI perform specific tasks assigned to it. [5]

3.1.1 Agent

An agent is something from the Latin verb, which means managing and performing tasks. Computer programs still behave to do tasks assigned to it, but to do more, an agent is exempted. It is assumed that an agent can behave independently, observe and respond to the climate and evolve, persist for a prolonged period and eventually develop, appreciate and follow objectives. [3]

3.1.1.1 Rational agent

In machine learning and artificial intelligence analysis, the rational agent could be a conception that guides the utilization of scientific theory and decision theory in applying AI to varied real-world situations. [3] [6]

The rational agent could be a theoretical entity supported by a practical model with preferences for advantageous outcomes and can obtain to realize them during a learning state of affairs. One of the simplest ways to know rational actors is to require an example of some form of business AI or machine learning project.

Suppose a business desires to know how individuals can use a fancy direction space like a drive-through with four lanes or a fancy edifice layout with multiple tables and chairs. The engineers and information scientists can construct profiles and properties for the rational actors that square measure sculptural on real-life customers. They'll then run the machine learning programs with these rational actors in mind and appearance at the outputs. [6]

Rational actors are applied in all told kinds of ways that AI comes. They assist individuals in learning. However, IPAs would possibly use these technologies and study human behaviour to help different humans build selections.

3.1.1.2 Intelligent Agent

An Intelligent Agent is a form of agent that's thought about to be rational, and with a collection of style principles, developers can produce successful agents that create the system to be reasonably intelligent to complete specific and intelligent tasks.

IPAs are code agents that act on behalf of the user to complete tasks and supply data. Therefore, the communication between the entity and the user is sometimes supported by voice inputs or commands in order words.

An intelligent agent could be a form of code application that searches, retrieves and presents data from the web. This application automates extracting information from the web, like data designated, supported a predefined criterion, keywords, or any such information to be searched. Intelligent agents square measure usually used as internet browsers, news retrieval services and on-line looking. [7]

An intelligent agent is primarily used to supplement data recovery operations which humans typically perform manually. Typically, an intelligent agent runs automatically at a scheduled time or when the user initiates it manually. To function on the primary search query, it then searches the entire internet or on user-defined websites. The intelligent agent copies, extract or lists the data when a significance or match is identified. The data gathered is then presented to the user in a raw or report-based format. [7]

Some advanced intelligent agent utilities use data inference matching and recovery techniques based on artificial intelligence, enabling them to collect higher quality and more applicable data. Shopping agents/bots, news feed/alert agents and Web crawlers are common intelligent agents. [8]

4. Statistical Learning

A paradigm for machine learning drawing from the fields of statistics and functional analysis is statistical learning theory. The theory of statistical learning deals with the issue of discovering a data-based predictive function. In fields such as speech recognition and bioinformatics, mathematical learning theory has contributed to successful applications. [9]

Understanding and prediction are the priorities of learning. There are many learning types, some of these are supervised learning, unsupervised learning, on-line learning, and reinforcement learning, falls into several categories.

Supervised learning is better viewed from the viewpoint of mathematical learning theory. Supervised learning requires learning from data from a training collection. An input-output pair is each point in the training, where the input maps to an output. The learning issue is that the function that maps between the input and the output is assumed so that the learned function can predict the output from future input. [9]

This session includes a brief analysis of specific fields relevant to AI Personal Assistants' creation. Machine Learning (ML), Natural Language Processing (NLP) and Sentiment Analysis (SA) are introduced together with their contexts and relevance to IPAs.

4.1 Machine Learning

Machine learning is a discipline of AI oriented towards human knowledge's technological growth via analysis, self-training, observation and practice, and it enables computers to navigate new situations. By exposure to new conditions, testing, and adaptation, machine learning promotes computation's continuous development, employing pattern and trend detection for better decisions in the following circumstances.

Alexa, Google Assistant, and Siri's accuracy, speed, and contextual abilities are all attributable to Machine Learning algorithms and servers owned by their developing businesses. The main distinction exists in their protocols and data protection intricacies, and they all function in a similar way. [10]

When a user makes a request, the request is instantly packed and submitted to respond to their respective businesses' server, so internet access is one of the fundamental criteria for the proper

functioning of Virtual Assistants'. A series of algorithms evaluate our request's terms and tone after the package is submitted to the server and are then matched with an order they assume we have requested. With the support of the server, not all the data are stored, just the complex ones.

The difficulty is about the speed of task completion and comprehension of what the consumer needs in the first attempt. That's a simple method of tapping into a server, third-party computer, or any other electronic device until it knows what it wants to do. [10]

Let's look at the example that helps us understand how machine learning works. Reviewing our past behaviour and feedback on different websites such as Trip Advisor and suggestions from other people takes us through the restaurant booking process and learns the kind of restaurants we want. It would then give us the options to book the restaurant from our past behaviour, and if we book an Uber through it, it ultimately picks up on the trend itself so that it offers from the beginning to the end of booking an Uber. It is just the most simple version. Before we have to ask, there is even more potential for the technology to step in, knowing all the wants and needs.

The same working principles apply to personal assistants such as Siri. Siri is Apple's digital assistant for voice recognition devices using iOS, macOS, tvOS, and watchOS. It is based on artificial intelligence, machine learning, and the usage of natural language processing. And it is based on three primary components: a conversational interface, knowledge of personal context, and delegation of service. It allows the mobile device and its applications to be controlled by the user using natural voice commands.

In addition, Siri is not all business; however, if users like it, users can have some fun and ask more cryptic questions. When the user is out and about, Siri helps us with sports and entertainment information, phone calls and messages, get organized, tips and tricks, read our last email, text our friends and family, playlist shuffle, even find a table for three in a restaurant and much more. [10]

One of the most significant factors currently holding back virtual assistants' acceptance is the public's frustration with talking to their assistants using their voice.

Amazon's Alexa works by incorporating 'hmms' and 'ums' into her responses to humanize her responses. Siri's assistant at Apple is known for making wry jokes, too. Using machine learning algorithms, the bot learns to generate accurate cue responses from the basis that the users generate.

Depending on the quality of the signal or feedback available to the learning system, machine learning approaches are generally split into three broad categories: Supervised Machine Learning, Unsupervised Machine Learning and Reinforcement Machine Learning. [10]

4.1.1 Supervised Machine Learning

Supervised machine learning is aimed at predicting sets of output data (y1, y2, ..., yn) from the input data sets given (x1, x2, ..., xn) for n observations. A general machine learning function is generated that was not a training package feature for predicting output from the data. [11] A training set of tuple data, such as ((y1, x1), (y2, x2), ..., (yn, xn)), from a known set of tuple data, shapes the output is predicted from the input. Example inputs and their desired outputs are introduced to the machine by a teacher or server, and the objective is to learn a general rule that maps inputs to outputs. [3] [11]

4.1.2 Unsupervised Machine Learning

Unsupervised machine learning is the data classification method without any links to branded data for preparation using data findings n (x1, x2..., xn). [3]

The primary objective of the unsupervised approach of machine learning is to gather knowledge with common characteristics and interactions in various classes. Because labelled information is not provided, unsupervised methods usually require more extensive methods. The learning algorithm is not given any marks, leaving it to find structure in its input on its own. Unsupervised learning may be an objective in itself (the detection of secret data patterns) or a means to an end (feature learning). [12]

4.1.3 Reinforcement Machine Learning

Reinforcement machine learning is a computer program that communicates with a lively world where a particular task must be accomplished (such as driving a vehicle or playing a game against an opponent).

The software is given feedback equivalent to incentives, which it seeks to optimize as it navigates its problem space. [13]

4.2 Natural Language Processing

The machine's interaction is no longer science fiction as intelligent personal assistants begin to play a more critical role in our everyday lives. But few have ever bothered to ask the question that

are all the Smart Personal Assistants such as Siri, Cortana, Alexa unavoidable? Or what helps us, in the end, to interact with a machine?

A viable intelligent personal assistant combines several tangible and intangible elements, from the software layer to the hardware layer. Any Intelligent Personal Assistant could be considered complex software. Though a working, intelligent personal assistant unit is the product of a more extensive system, the back-and-forth mechanism of human-machine interaction which is the most intuitive aspect from a user perspective. [14]

Most technology companies providing intelligent personal assistant services are trying to make their product more human-like at the current level. Again, this is an entire project consisting of big data, machine learning (deep learning), computational neural networks, and other artificial intelligence-related disciplines. But there is one subsystem on the front-facing end that we need to talk about, natural language processing (NLP). [14]

Natural Language Processing (NLP) is a technique for translating between human and machine languages. It is a method of getting a machine to read a text line understandably without hint or measurement being fed to the computer. NLP automates, in other words, the method of translation between computers and humans.

Feeding statistics and models have historically been the method of choice for phrase analysis. Recent advances in this field include speech recognition applications, translation of human language, information processing, and artificial intelligence. The development of human language translation software is difficult since the language is continuously evolving.

Natural language processing is also being developed to produce human-readable text and translate between one human language and another. NLP's ultimate goal is to create software that would naturally analyze, understand and produce human languages, allowing contact with a machine as though it were a human being.

A three-step method seems to be the standard procedure when breaking down the communication flow between people. [14]

The first step is to receive the data, and typically our ear picks up the sound wave produced and transmitted through the air by some kinds of vibration.

The second stage is to process the data. The auditory signal that was received must fit our brain's current pattern according to corresponding meanings.

The production of information is the third stage. By producing the acoustic signal through transducers, one would disseminate the message to be collected from the other end to preserve the conversation flow.

NLP follows a similar trend in human-machine interaction by imitating the three-step method of inter-human communication. By definition, NLP is a field of study that encompasses many different moving parts, culminating in the ten or so seconds that it takes to ask and receive an answer from IPA.

We may think of it as an approximately 3-stage process: listening, understanding, and reacting. Alexa was designed as a framework with several modules to control various phases of the procedure.

One of the front-end modules picks up the acoustic signal with a sensor on voice commands or activation phrases for the listening portion. It could send information to the back end for further processing as this module is connected to the internet with wireless technology. [15]

It could also be understood as the processing portion as the program for speech recognition takes over and helps the machine transcribe into corresponding texts the user's spoken English (or other supported languages). This approach is the tokenization of an acoustic wave, which is not a self-contained medium. Some waves were converted into tokens and strings that the computer could manage by transforming them. The ultimate objective of this phase of analysis is to translate the text into data.

Here's one of the most challenging aspects of NLP, understanding the natural language. It adds to the whole linguistic section of NLP, considering all the different and imprecise ways people speak and how meanings change with context. Natural language understanding teaches computers to understand semantics with methods such as part-of-speech tagging and classification of intent, how words make up sentences that express ideas and meaning. [15]

It comes to the final stage of reacting when a result has been obtained. As the details would be translated back into text, this would be an analogous Natural Language Comprehension method. Now that the machine has the result, two more attempts need to be made.

One is prioritizing, which means choosing the essential data to the user's demand, leading to the second effort: reasoning. It relates to the method of translating the reacting concept into a human-understandable way.

Finally, once the natural-language response is produced, speech synthesis technology turns the text back into speech.

4.2.1 Methods: Rules, Statistics and Neural Networks

4.2.1.1 Rules

Many language processing systems were constructed by symbolic methods in the early days, i.e. the hand-coding of a set of rules, combined with a search for a dictionary, such as writing grammars or creating stemming heuristic rules. [14]

There are several advantages to more current systems based on machine-learning algorithms over-rules generated by hand such as:

- The learning processes used during machine learning immediately concentrate on the most common situations, while it is often not evident where the effort should be guided when writing rules by hand.
- Automatic learning procedures may use statistical inference algorithms to generate models that are robust to unfamiliar input and erroneous input. Generally, it is challenging, error-prone and time-consuming to manage such feedback gracefully with handwritten rules, or more generally, to build systems of handwritten rules that make soft decisions.
- Systems based on learning can automatically be made more precise by merely providing more input data. However, only by increasing the rules' sophistication, which is a much more challenging task, can systems based on handwritten rules be made more effective. In particular, the complexity of systems based on handwritten laws, beyond which the systems are increasingly unmanageable, is minimal. However, it needs a corresponding increase in the number of working hours to generate more data to input machine-learning systems, usually without significant increases in the annotation process's complexity. [14]

4.2.1.2 Statistical methods

Much natural language processing research has focused heavily on machine learning since the so-called statistical revolution. Instead, the machine-learning paradigm calls for statistical inference to automatically learn these rules by analyzing many documents, likely with human or computer annotations. [16]

Natural-language-processing tasks have been extended to several different classes of machine-learning algorithms. These algorithms take a vast collection of features created from the input data as input.

However, research has increasingly focused on mathematical models that make soft, probabilistic decisions based on the attachment to each input function of real-valued weights. These models have the advantage that, when such a model is used as part of a larger scheme, they can convey several different possible answers' relative confidence rather than only one, providing more accurate results. [16]

Neural networks have increasingly replaced statistical approaches since the neural turn in NLP science. However, they remain essential in contexts where statistical interpretability and clarity are essential.

4.2.1.3 Neural networks

A significant downside of statistical methods is that they involve elaborate engineering of features. Thus, the field has mostly abandoned statistical approaches and moved to deep learning neural networks. Instead of depending on a pipeline of separate intermediate tasks to capture the semantic properties of words, common approaches are used, such as word embedding and an increase in end-to-end learning of a higher-level task (e.g., question answering). [17] This shift has led to significant changes in how NLP systems are built in some ways so that that deep neural network-based methods can be regarded as a new paradigm distinct from the processing of natural statistical language.

4.3 Artificial Neural networks

A theoretical model focused on biological neural network structure, and functions is an artificial neural network (ANN).

The ANN structure is influenced by information passing through the network since a neural network adjusts or learns based on input and output. Nonlinear statistical data modelling methods are known to be ANNs, where the complex relationships between inputs and outputs are modelled, or patterns are found. [17]

ANNs are models of deep learning capable of the identification of patterns and machine learning. They form part of the broader area of artificial intelligence (AI) technology. ANN is classified as a neural network as well.

There are three or more layers of an artificial neural network that are linked together.

Input neurons compressed the first layer. These neurons transmit data to deeper layers, sending the final data output to the last layer of output. All the inner layers are hidden and generated by units that adaptively modify the information obtained from layer to layer through a series of transformations. Each layer functions as a layer of input and output that enables the ANN to understand more complicated artifacts.

Collectively, the neural layer is the name of these inner layers. The neural layer units try to learn about the information obtained by measuring it according to the ANN's internal system. These guidelines allow units to produce a transformed outcome, supplied to the next layer as an output. [17]

An additional number of learning rules use backpropagation, a mechanism by which the ANN can change its performance results by taking errors into account. Through backpropagation, the information is sent backwards each time the performance is marked as an error during the supervised training stage. Each weight is updated to the amount of blame for the mistake. The error is then used to recalibrate the weight of the unit connections of the ANN to make the difference between the desired result and the real one into account.

The ANN learns how to mitigate the risk of mistakes and unexpected outcomes in due time. Training an ANN involves choosing from accepted models for which several similar algorithms are available. [17]

An ANN has many benefits learning from observing data sets is one of the most known of these. These instruments help estimate the most cost-effective and ideal methods to arrive at solutions

when specifying functions or computation distributions. Instead of whole data sets, ANN takes data samples to arrive at solutions, saving time and money. To improve current data analysis technologies, ANNs are considered simple mathematical models.

There are many practical applications in which an ANN is used, such as business intelligence predictive analysis, spam email identification and processing of natural language in chatbots. [18]

Since the 1940s, neural networking, a computer machine learning and pattern recognition technique inspired by the human brain's neurons, has been bandaged around. Still, the decades-old technology has been given new attention by implementing the technology in virtual personal assistants. After some preliminary research findings involving neural networking, Microsoft disclosed that the technique could improve speech recognition accuracy by about 25 percent.

Furthermore, researchers argue that neural networking has greater potential than the technology of existing speech recognition. It's no surprise that neural networking is one of the hottest areas of development today, acknowledges wireless industry analyst Jeff Kagan. [18]

To enhance its Siri personal assistant's efficiency, Apple has been drawing on scientists worldwide with neural network experience. But according to many experts, Apple is coming into the neural networking game a little late.

IBM, Microsoft, and Google have already implemented deep learning technology in some of their products. For instance, in the real-time translation function, Microsoft uses neural networks in Skype. Not only does neural networking increase the accuracy of the speech recognition involved, but the longer the device is used, the better the translations get. This is because the machine can recognize and learn from trends in how people interact through languages. [18]

For enhanced speech recognition, Microsoft's assistant, Cortana, also uses neural networking. A core component of IBM's Watson personal assistant technology is neural networking. Additionally, in some of the Google Now apps, Google has already introduced the technology. The technology is introduced in each case to learn about the needs of the customer.

Neural networking enables these assistant apps to learn and become increasingly personalized about their users, ultimately executing tasks before they ask. Instead of collecting disjointed

questions and responses, neural networking also facilitates multistep searches that create an informal atmosphere. [18]

Even Netflix, the on-demand Internet streaming video service, uses neural networking to compile subscriber taste profiles based on past viewing habits. This information can then be used to make informed recommendations.

4.4 Sentiment Analysis

There is the use of NLP in Sentiment Analysis (SA), or Opinion Mining (OM), analysis of text and computational linguistics to recognize and extract information (or Characteristics) from documents.

One of the things IPA's would have to learn to do sentiment analysis, is doing human analysis based on its emotions such as reading text and evaluating its emotional charge. There is a great deal of excitement and buzz about sentiment analysis today. It has its supporters and its critics, like any modern technology. And because technology is in the early stages of evolution, there can be plenty of setbacks for the skeptics and plenty of optimism for the supporters. [19]

Now let's start with the skeptics. Emotions are viewed as distinctly human, and even human beings are hard to read. So how does a computer predict the emotions in messages accurately?

The problem is further compounded by the subtleties of language and gestures, the unwritten but implicit insinuations in short messages, the symbols and abbreviations in social media posts that continually alter and develop, and the fact that more than one emotion can be conveyed by messages. These are all legitimate problems and issues that today do not have definitive solutions.

Early identification of warning signs and prospects is the engine of sentiment analysis. It can be used to avoid customer churn to predict that a customer is becoming upset with customer service and detecting early that negative posts have the potential to go viral can cause a counteraction to prevent significant reputational harm. [19]

We ask machines' to support this activity because the number of online communications is so high that organizations cannot put enough workers to sift through them all.

Naturally, individuals would look for solutions when there is a need and hype, and cynicism is the corollaries of searching for a solution. Industry-wide, we know that about 40 percent to 50 percent of messages can be assigned to sentiment ratings.

Sentiment analysis helps one reliably distinguish signals at the two ends of the bell curve, the highly positive and the extremely negative. Given the human condition, outraged messages have a very angry tone. And the emphasis of the messages is entirely aimed at the expression of powerful anger. The same applies to those happy posts. So, we can confidently argue that the program can allow us to process the messages at both ends of the bell curve, which is not a small advantage. [19]

Now let's figure out how to improve it further so that we can get more insight. What if we combine the study of sentiment with search technologies? We allocate sentiment to a post, in other words, index all the messages in a search index. Then the quest would show us similar messages until we find an angry message. Content would indeed be connected to those messages, but even a novice user can quickly sift through similar messages and pick up those of interest.

Sophisticated analyzers that can extract words from the text are already filled with searches. These terms and their frequencies can be displayed in different visualizations such as tag clouds, graphs of lexical diversity, and streaming graphs. With this enhancement, a user can filter the sentiment messages (detecting the angriest ones), display the related messages (using search), evaluate the aggregate word distributions in related messages (using text graphs) and drill down on the graphs to further filter the messages. They would read the few remaining messages at the end and know both the sentiment and the content precisely. [19]

In its current process, automated SA can not be as precise as human analysis. Methods for automated sentiment analysis do not account for the context, surroundings, irony, human body language or good subtleties.

Inter-rater reliability plays an essential role in human analysis, which is the degree of agreement between raters. Currently, no IPA in enterprise development is addressing these aspects today.

5. Speech Recognition

Technology for speech recognition is something that has been dreamt of and worked on for decades.

From the beep-bopping of R2-D2 in Star Wars to the disembodied but soulful voice of Samantha in Her, the voice of Jarvis in the Ironman movie, science fiction writers have played an enormous role in building expectations and predictions for what speech recognition in our world could look-like. However, for all the advances of modern technology, voice control has been a relatively unsophisticated affair. [20]

Speech recognition is interdisciplinary computer science and computer linguistics subfield that develops methodologies and technologies that make it possible for computers to recognize and translate spoken language into text using Automatic speech recognition (ASR), computer speech recognition speech-to-text recognition (STT). It includes expertise and research in the fields of computer science, linguistics, and computer engineering.

Specific speech recognition systems require training where text or isolated vocabulary is read into the system by an individual speaker. The system analyzes the individual's distinct voice and uses it to fine-tune the recognition of that person's speech, resulting in increased precision.

The independent speaker systems are those systems that do not use training. Systems using training are referred to as speaker dependent. [20]

Speech recognition implementations include voice user interfaces such as voice dialling, call routing, home appliance control, keyword search, basic data entry, formal document preparation, defining characteristics of speaker and speech to text processing. Instead of what they say, the word speech recognition or speaker identification refers to recognizing the speaker. Recognizing the speaker would simplify translating speech into a system trained on a particular person's voice or can be used as part of a protection mechanism to authenticate or verify a speaker's identity.

It is easy to take for granted how speech recognition technology works as we are surrounded by smartphones, smart vehicles, smart appliances, voice assistants and more. Since it is misleading to talk to personal assistants straightforwardly, recognition of speech is incredibly complicated. [20]

Think about how an infant learns a language.

They can hear words being used all around them from day one. Parents talk to their child, and while the child does not respond, they absorb all sorts of verbal signals such as intonation, inflection, and pronunciation so, their brain creates patterns and associations based on how their parents use language. We have been training our entire lives to acquire this so-called innate capacity, while it might seem as though human beings are hardwired to listen and understand.

Technology for speech recognition operates in the same way. We are still working out the best practices for computers, though people have perfected our process. We must educate them in the same way that our parents and teachers have taught us. And a lot of imaginative thinking, workforce, and research is involved in that training. [20]

It would take a lot more time and a lot more field data to perfect these speech recognition systems. There are, after all, thousands of languages, accents and dialects to consider. That is not to say we're not making progress. As of May 2019, Google's machine learning algorithms have now reached a 95 percent word precision score for the English language. The current rate also happens to be the human accuracy threshold.

In modern statistically based speech recognition algorithms, both acoustic modelling and language modelling are essential components. In many systems, Hidden Markov models (HMMs) are widely used in many other natural language processing applications, such as document classification or statistical machine translation, language modelling is also used.

5.1 Hidden Markov Models (HMMs)

Hidden Markov Models are the basis of modern general-purpose speech recognition systems. These are statistical models that produce symbols or quantities in a sequence.

In speech recognition, HMMs are used because a speech signal can be viewed as a stationary signal or a short-time stationary signal piecewise. In a short timescale, speech can be approximated as a stationary process. For many stochastic purposes, speech can be thought of as a Markov Model. [20]

5.2 Neural Networks

In ASR, neural networks emerged as an attractive acoustic modelling approach. Since then, neural networks, such as phoneme classification, phoneme classification through multi-objective

evolutionary algorithms, isolated word recognition, audiovisual speech recognition, audiovisual speaker recognition, and speaker adaptation, have been used in many aspects of speech recognition. [20]

Neural networks make less clear assumptions about statistical features than HMMs and have some characteristics that make them appealing speech recognition models. Neural networks allow discriminative training naturally and effectively when used to estimate a speech feature section's probability.

However, due to their limited ability to model temporal dependencies, early neural networks were rarely competitive in continuous recognition tasks, despite their effectiveness in classifying short-time units such as individual phonemes and isolated words. [20]

5.3 End-to-end Automatic Speech Recognition

There has been a great deal of research interest in ASR end-to-end since 2014. Traditional approaches based on phonetics (i.e., all HMM-based models) required separate components and training for the model of pronunciation, acoustics, and language. [20]

End-to-end models jointly learn all the elements of the speech recognizer. It is useful as the training process and deployment process is simplified. For example, for all HMM-based systems, an n-gram language model is required, and a typical n-gram language model often takes several gigabytes of memory to make it impractical to deploy on mobile devices. As a result, Google, and Apple's (as of 2017) new commercial ASR systems are installed on the cloud and require a network connection instead of the computer locally.

6. Human-Computer Interaction

Human-computer interaction (HCI) is an interdisciplinary research area that focuses on computer technology design and, in particular, human (user) interaction with computers. HCI has since grown to include almost all aspects of information technology design, although it was initially concerned with computers.

HCI systems are comfortable, safe, effective, and enjoyable. Software engineering focuses on developing software application solutions, while HCI focuses on exploring people-supporting approaches and techniques. For efficient user interaction, HCI designers often consider HCI usability and user interface objectives. As some combinations are incompatible, not all usability and user interface goals apply to any interactive computer device. HCI designers often consider potential contexts, activities at hand and users of computer systems. [21]

Humans interact with computers through a user interface. It includes software and hardware such as the mouse, keyboard, and other peripheral devices to show what is displayed on the computer monitor. Therefore, HCI's study focuses on user satisfaction. It is crucial to pay attention to interaction with human machines because a poor interface can make it difficult for users to benefit from even the most straightforward systems. A poor user interface could have more severe consequences in a corporate or factory setting. [21]

For the first time, sophisticated electronic systems for applications such as word processors, game units and accounting aids have been made available to general consumers. Consequently, since computers were no longer space-sized, costly tools built exclusively for experts in specialized environments, it became increasingly important for less experienced users to create human-computer interaction that was also easy and effective.

HCI would expand from its origins to incorporate multiple disciplines, such as computer science, cognitive science, and human factors engineering. [21]

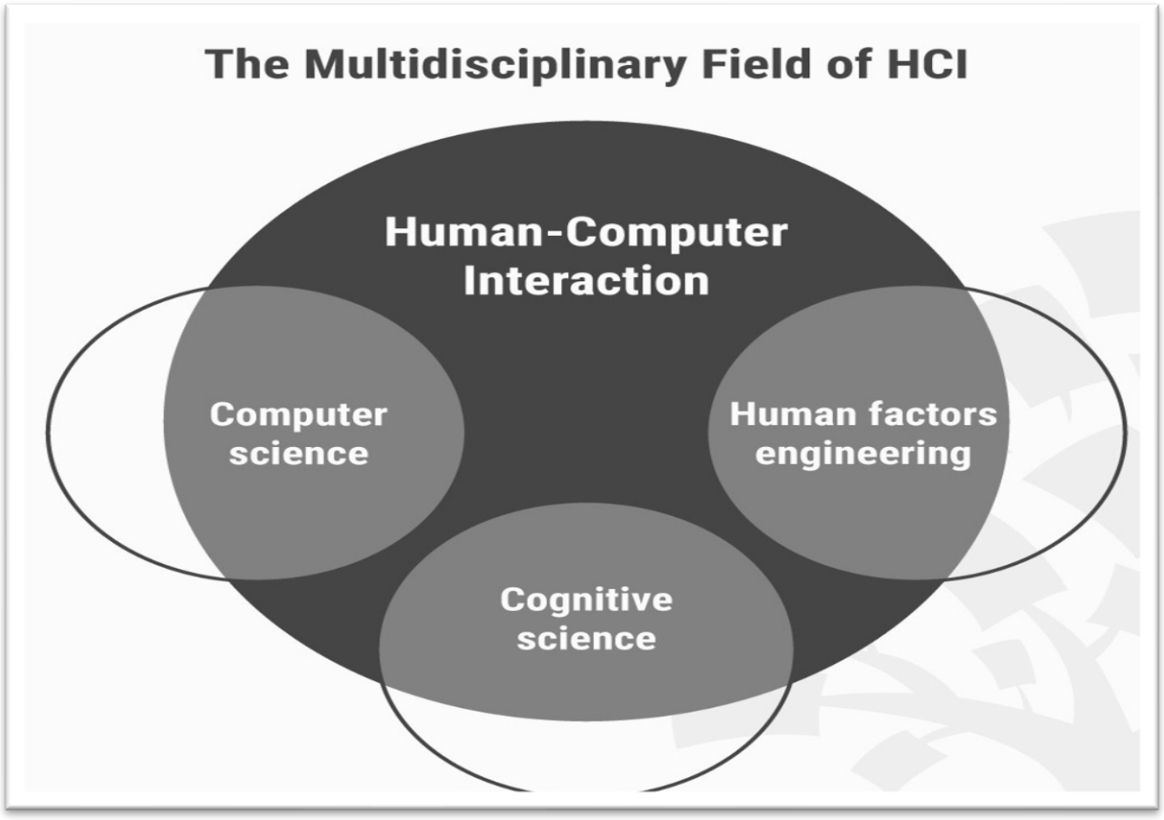


Figure 1: The Multidisciplinary Field of HCI [22]

6.1 Goals of Human-Computer Interface

Human-computer interaction studies how humans make use of computer artifacts, systems and infrastructures, or do not make use of them. Much of the field research aims to improve the interaction between humans and computers by improving computer interfaces usability. It is increasingly debated how usability is to be understood precisely. [21]

Much of the research in the human-computer interaction area is concerned with:

- Methods for designing new programming interfaces, optimizing a template for the desired property such as learning capacity, discoverability and usage efficiency.
- Methods for interface implementation, e.g. utilizing software libraries.
- Methods to determine and compare interfaces about their usability and other desirable characteristics.
- Methods for generally researching human-computer usage and its sociocultural ramifications.
- Methods to determine whether the user is a person or a machine or not.

- Human-machine usage models and hypotheses and philosophical structures for computer interface design, such as cognitivist user models, operation theory or ethnomethodological accounts of the use of human computers.
- Perspectives that focus critically on the principles underlying the practice of computational design, computer usage and HCI research. [21]

7. Analysis of closed-source AI Personal Assistant

This section introduces the analysis and results of the methods and background introduced in this text's previous sections. This text's scope is restricted to six different companies: Apple Inc., Google Inc., Samsung Corporation, Mycroft AI Inc., Amazon Inc., and Microsoft Corporation.

7.1 Siri

The Intelligent Personal Assistant from Apple is called Siri, and it is currently an integral part of their operating systems such as iOS, which is used by their smartphones and tablets, watchOS used by Apple Watch, tvOS is the operating system for their television hardware, namely Apple TV and macOS for their MacBooks.

7.1.1 Background, Research Development

A Conversational Interface, Personal Context Understanding, and Service Delegation are the key technical areas of Siri, all of which are essential pieces for an IPA. [23]

The engine for speech recognition that Siri has is provided by Nuance Communications, a corporation responsible for speech technology. For conversational and comic purposes, Siri also has hard-coded answers, or Easter eggs, such as What is the purpose of life? And Who is your creator? A great implementation of many technologies in Siri is provided by Nuance Communications, such as voice recognition and text-to-speech (TTS) technology. Artificial intelligence such as natural language processing engines and back-end services are also used in Siri.

A practical simplification is perhaps to say that Siri has three layers: voice processing, grammar analysis-context learning engine and services.

The Nuance speech-to-text engine converts the request into text when a voice file arrives at Apple's data center. Since 1994, Nuance has been in the voice technology market and was, interestingly, another spin-out from the same lab as Siri (SRI International STAR). It is safe to assume that the real interface of Siri is simply Nuance software, but it is Siri's back-end magic that makes things pop. [23]

7.1.2 Development of Specific Details and Speculations

Siri is most likely still under re-factoring and production of features and Quality analysis, is in continuous growth. In recent years, the Field of Machine Learning and Artificial Intelligence has become more of a focus for Apple Inc.

For years, contact between humans and computers has always been an essential part of Apple's goods and services. Developers from Apple and the Siri platform team announced that Siri is powered by Apache Mesos on 22 April 2015. And the next section focuses on the working of Siri using AI that was mentioned earlier. [23]

7.1.3 Working of Siri.

Phase 1: Voice Recognition

That's the simple part, but that's where it all starts, so it can't be trivial.

Our device captures our analog voice when we command Siri, transforms it into an audio file (translated into binary code), and sends it to Apple servers. The nuances of our accent, the noise around us and the local expressions make it hard to get it right. Compared to the standard Graphical User Interface we are used to, it is called a Human User Interface. It is significant here that Apple receives millions of queries every day from people who speak different languages, with different accents, living on different continents. In other words, people are contributing to their acts and errors to the biggest crowd-sourced speech recognition experiment ever attempted on earth. [23]

Today, the Siri app receives approximately a billion requests each week, and Apple notes that its ability to understand speech has just a 5% word error rate. To get to this point, Apple recently acquired the speech recognition company Novauris Technologies, a spinoff of Dragon Systems, and recruited several speech recognition specialists.

Phase 2: Send everything to Apple servers in the cloud.

Siri does not locally process our speech input on our device. If we're not linked to internet for whatever reason, this is an issue, but this is how Apple gets two significant advantages:

- Offload much of the work to powerful machines instead of consuming the mobile device is limited resources.
- Use the information it gathers to optimize the service continually.

The algorithm recognizes the keywords and begins to take us down the flowchart branches related to those keywords to get our reaction. In this exercise, if it fails because a part of the communication does not function, it goes down the flowchart's wrong branch. If it occurs just once, the whole question is destroyed and ends up in the, "*Would you like to check the internet for that?*", result. They is no difference from Google Now and Cortana. [23]

We know that this is far from the meaning of a human conversation. The Siri app is still designed with a pre-program logic to answer any possible questions and laws. This was even more apparent when Apple honoured Back to the Future Day in October 2015 by updating the Siri app with at least ten humorous responses related to the famous Back to the Future movie. [23]

Phase 3: Understand the meaning

The method of interpreting what the consumer asks for depends on science called the processing of natural languages.

People have hundreds of ways to pose the same question. Using infinite variations of phrases, we may articulate an idea. Is there an Italian restaurant nearby? I'm in the mood for a pizza. Today I would love a Margherita. Humans can easily understand what I do so, it's evident that Margherita is not an entity, but to draw the same conclusion, an algorithm must be sophisticated. It's sometimes only because words have a common sound or are mispronounced, the job is complicated by oyster and ostrich, school and skull, byte and bite, sheep and ship and many others. [23]

The Siri application software model principles are there to ease their lives. It analyzes how an object and a verb are related to the subject keyword. In other words, the syntactic structure of the text is studied.

Siri can make sense of inquiries and follow up commands on top of that. It is not necessarily what a person would term a conversation, but it means that the meaning is known, and it is the starting point for potential developments.

Phase 4: Transform the meaning into actionable instructions.

When the Siri app understands what we want, to make it happen it has to communicate with other apps, and each app is different and partly has its own language.

The device must have known domain knowledge, the subject area we are talking about must be known. This occurs every time we speak to specialists in a particular area in a human conversation, and they use technical terms that we barely understand. It's clear, for example, when we talk to a doctor, an architect, or a person in finance.

That's the same for the Siri app. It has to communicate with other apps and understand their meaning when providing a direction, book a flight or send a text. It is also essential. If the protocol does not work, Siri can give instructions to other apps to perform acts that we did not need and maybe potentially harmful. [23]

Finally, Siri must translate the outcome back into text that can be spoken to the user once a request has been processed. Though not as hard as processing a user's command, this task, known as the generation of natural language, still poses some challenges. Apple also plans to increase the number of languages that Siri understands, and that is another explanation of why the Siri app is not increasing as quickly as anticipated.

7.1.4 SDK

Apple, with the release of iOS 10, the Siri SDK, namely Sirikit for developers, was created, giving them the ability to develop their applications, to incorporate Siri use. Sirikit support is classified into Domains, each describing one or more tasks that can be carried out.

For the successful working of Sirikit, it must support one of the domains below:

- VoIP calls, texting, purchases, photography, workouts, scheduling a trip
- CarPlay and restaurant bookings

7.1.5 Future of Siri

It is evident with iOS 13 and Shortcuts that Apple has no intention of slowing down the assistant's growth. So, what is Siri's potential future? Several claims that a SiriOS is inevitable. This hypothetical operating system will break the silos and make the smart assistant of Apple a coherent, unified entity in the ecosystem of a customer. [24]

In order to alter outcomes, another rumour refers to Siri interpreting the emotions. The production of shortcuts would also lead to better assistant too. Siri still holds the crown for user customization

and third-party integration with Siri Shortcuts with superior responsiveness and natural language parsing and can keep the lead even tighter with its strong device integration.

From time to time, Apple also acquires new firms and technology, leading to changes in its services. In early April 2020, a business called Voysis was acquired that could improve the capacity of Siri to comprehend commands. [24]

7.2 Google Assistant

Google Assistant is a Google-developed virtual assistant powered by artificial intelligence primarily available on smartphones and smart home devices. Unlike the company's previous virtual assistant Google Now, the Google Assistant participates in two-way conversations.

Google Assistant was initially limited to the Pixel and Pixel XL smartphones for system-level integration outside of the Allo app and Google Home. Google revealed in February 2017 that it had started to allow Android smartphones running Android Marshmallow or Nougat to access the Assistant, starting in select English-speaking markets. [25]

The Assistant was not given to Android tablets as part of this rollout. The Assistant is now incorporated into Android Wear 2.0 and also included with Android TV and Android Auto.

Google Pixel book became the first laptop to integrate Google Assistant in October 2017. Google Assistant later accessed the Google Pixel Buds. Google revealed in December 2017 that the Assistant would be released via an update to Google Play Services for phones running Android Lollipop, as well as tablets running 6.0 Marshmallow and 7.0 Nougat.

In February 2019, Google officially started checking Google Assistant results for advertisements. In recent years, Google Assistant is updated to use in the latest android operating system, Android version 11, released in September 2020. [25]

7.2.1 Working of Google Assistant

If we want assistance with daily tasks, controlling smart home devices, enjoying music or games, interacting with friends and family, having instant answers or local information, or any other items.

Google Assistant wants to respond to our request in the most helpful way possible when we ask a question or tell it to do something. To do this, Google Assistant needs to understand what we are

asking for and why, to find the best way to help us do that task. This purpose is at the core of how an assistant works. [25]

Phase 1: Understanding the Request

Google Assistant speech recognition technology translates our request to text if we communicate with the Assistant by voice. Then the Assistant analyzes the text to distinguish potential meanings, in conjunction with valuable details such as recent requests or the type of question we are using.

For instance, if we say, "Hey Google stop", we may want to stop one of the two running timers, the music that's playing, or a running routine. We may just want to see the "Stop" search results or something completely different. The Assistant compiles a list of our request's various interpretations and how it reacts to each one to weigh the options. [25]

The next step is to find the best way to satisfy our request by rating these choices.

Phase 2: Ranking the available responses.

The Assistant ranks the available responses with several signals, including the following key factors:

- How clear it is that the Assistant heard what we asked.
- If there is an answer available for a specific understanding of our request.
- With an apparent response to similar requests, how happy previous users were.
- How the answer was created recently to help us get a selection of new, high-quality responses.
- How well a response works on the device that we are using? For example, responses designed for devices with screens on speakers are likely to be ranked lower. If we request something unique to that device, such as raising the volume or playing a video, on a partner device where the Assistant is built-in, the device manufacturer handles some or all of the response, according to what the partner determines is the best user experience.
- What else have we asked for lately? For instance, if we say, 'Hey Google, start a five-minute timer,' and then say, 'Hey Google, stop,' shortly after,' Assistant can use our previous request to understand what we meant. [25]
- What we are doing on our device at the moment? such as which app we have open when we ask for help from Assistant, or what Assistant is already helping us with. For instance, if we

listen to music and say, "Hey Google, skip," the Assistant jumps to the next song. Similarly, if we make a restaurant reservation using Google Assistant, it prioritizes completing the reservation based on the overall potential answers.

- In limited cases, some high-quality responses may be manually curated to rank higher to enhance the user experience. For example, we can curate information from reliable sources such as the World Health Organization and government health authorities to help users get timely information about COVID-19 and mitigate misinformation that may endanger public safety. [25]

Phase 3: Choosing a provider

By connecting us with responses given by other creators and companies, Google responses, the assistant responds to certain types of requests. For instance, we can ask our favourite creator for a game, "Hey Google, play Warzone," and the Assistant starts the game.

We may also make a general request, "Hey Google, play a game," which a variety of different providers that have told Assistant that they sell games could satisfy. [25] The Assistant selects a provider in circumstances where more than one provider satisfies the request by applying the following rules in this order:

- If we have selected a provider, the Assistant picks the provider. For example, we might have selected the preferred music provider by Assistant Settings or Setup Flows, or our request may specifically name the provider.
- The Assistant ranks the available choices if we have not selected a supplier, using the following key factors:
  (a) Information about our preferences: This information may include, depending on our Google Account settings, which providers we most frequently or most recently used, which applications we install or open on our phone or another computer, which providers we have connected to our Google Account, and other information about our Google Services activity.
  (b) Information about the provider: The consistency of the provider's customer experience, based on such things as overall popularity, average user rating, how much the provider effectively responds to user requests, and whether we have a subscription to that provider

and how well the provider responds to our requests, such as in-stock goods, select menu items, or unique flight times. [25]

- If no provider ranks the highest, the Assistant can ask us to choose a provider to respond to our request.

Phase 4: Providing the best response to fulfill our request.

After the ranking process is done, the Assistant then responds with what it feels is the best choice, a list of options, or lets us know if it does not understand our request.

If there are several highly rated answers, the Assistant asks us for more details to explain our purpose, provide us with follow-up suggestions (on-screen devices) or let us know about the related issues that we may ask for. [25]

Phase 5: Providing the best response from Google Search.

In some cases, the easiest way to assist with our request is to include Google Search results. For example, Assistant might display us Search results if it thinks we want to see an enormous collection of results or if no other answer is expected.

In general, when the Assistant provides Google Search results, those results are close to what we would find if we searched for them in Google Search. Small algorithmic improvements are introduced by Assistant to provide results that are acceptable and beneficial for Assistant users:

- The assistant can filter out inappropriate and explicit content, such as smart displays, on shared devices. [25]
- The context of our request, such as our previous inquiries, and the capabilities of our device, and typical use patterns for that type of device, may be considered by the assistant. For example, more video results can be seen on TVs than on phones.

7.2.2 SDK

A software development kit (SDK) was introduced in April 2017, enabling third-party developers to create their hardware to run Google Assistant. It has been incorporated into Raspberry Pi, Audi and Volvo vehicles, and smart home appliances from companies including iRobot, LG, General Electric, and D-Link, including fridges, washers, and ovens.

In December 2017, Google updated the SDK to include many features that had previously supported only the Google Home smart speakers and Google Assistant smartphone applications. [25]

The features include:

- Let third-party software manufacturers implement their own Actions on Google commands for their respective goods.
- Focusing more on the text-based interaction and more languages.
- Enabling users to set the device's precise geographic position to allow enhanced location-specific queries.

Google launched a new initiative on their blog on 2 May 2018, which focuses on investing through early-stage entrepreneurs in Google Assistant's future. Their focus was to create an ecosystem where developers could create for their users' richer experiences. This involves start-ups that expand the functionality of Assistant, create new hardware products, or simply distinguish between various industries. [25]

7.2.3 Future of Google Assistant

For the next step of the Google Assistant, Google shares our vision, as they make it more inherently conversational, visually assistive, and helpful in getting things done. [26]

One of the Assistant's most significant components is its voice, which needs to sound both personal and natural. Up until now, it took hundreds of hours in a recording studio to build a new sound. But Google can now create new voices in just a few weeks with developments in AI and WaveNet technology from DeepMind and can capture subtleties such as pitch, rhythm, and all the pauses that communicate meaning so that voices are natural-sounding and original.

Being able to ask about several things at once is a vital aspect of having a natural conversation. The Google Assistant can understand more complicated questions such as, "*What's the weather like in Edmonton and Calgary*"? with multiple acts, which is already beginning to roll out.

Google has also redesigned the on-screen Assistant experience that is with us all the time on our tablets. A brief snapshot of our day is provided by the Assistant, with feedback based on day, place and recent experiences with the Assistant. [26]

They showed a new technology called Google Duplex in May 2018 and the assistant can understand complicated sentences, short speech, and lengthy remarks so that in a phone conversation, it can respond naturally. The Assistant is explicit about the purpose of the call, even though the calls sound very typical so that companies understand the meaning. Duplex is a glimpse of things to come, potentially leading the world to an interface between bot-to-human contact and bot-to-bot.

7.3 Microsoft Cortana

Cortana is an intelligent personal assistant from Microsoft that can alter reminders and recognize our natural voice and answer questions using bing's details. Cortana is capable of working with its operating system as a personal assistant for stock applications but is dependent on Internet connectivity for Bing to obtain details. [27]

Cortana has been incorporated into many Microsoft products, such as Microsoft Edge, a browser bundled with Windows 10. The Cortana assistant from Microsoft is deeply incorporated into its Edge browser. Cortana can find opening hours by displaying retail coupons for websites on restaurant sites or showing weather details in the address bar.

Cortana can set reminders, recognize natural voice without the keyboard input requirement, and use Bing search engine knowledge to give answers. Windows 10 searches are only performed with the Microsoft Bing search engine, and all links are opened with Microsoft Edge, except when using a screen reader such as Narrator, where the links are opened in Internet Explorer. [27]

The universal Bing SmartSearch features of Windows Phone 8.1 are integrated into Cortana, which replaces the previous Bing Search app-enabled when users on their mobile pressed the "Search" button.

Cortana includes a service for music identification. Rolling dices and tossing a coin can be replicated by Cortana. Bing searches to decide the bands or musicians the consumer is interested in are tracked by Cortana's' Concert Watch. It can generate reminders and phone alerts connected via Bluetooth when it integrates with the Microsoft Band watch band for Windows Phone devices.

7.3.1 Working of Cortana.

By simply saying "Hey, Cortana" out loud, we can summon Cortana, and then she offers her assistance. We need a microphone on our computer for this feature, and we can change Cortana's settings to learn our own voice and respond only to our verbal commands. If we're not in a position where it would be practical to communicate, we can always type our Cortana commands as well. [27]

Cortana can also search directly from our desktop for files and directories, and this requires any files that we might have saved in the cloud, such as on OneDrive. The Notebook is another interesting integration-type feature. The Notebook pops up as a newsfeed tailored to our interests through smart suggestions based on our recent web searches by clicking Cortana's search bar.

Before we even ask for them, Cortana reviews our emails and calendar via the Notebook and pushes updates and reminders. Since Cortana has such an engaging personality, communicating with her on our computer comes naturally. Cortana is very funny, can tell us jokes, and is programmed with variability to answer our questions. [27]

Cortana is constructed based on natural language processing, translating speech into sounds, words and ideas.

- Microsoft records our speech first. The recording of our speech is sent to Microsoft's servers to be evaluated more effectively since decoding sounds take up many computing resources.
- What we said is broken down by Microsoft into individual sounds. In order to find the words most closely correspond to the combination of individual sounds, it then consults a database containing different words' pronunciations.
- To make sense of the tasks and execute related functions, it then defines critical terms. For example, if Cortana detects terms such as "weather" or "temperature," the weather app opens.
- The information is sent back to our computer by Microsoft servers, and Cortana can talk. It would go through the same process mentioned above, but Cortana needed to say something back to us in reverse order. [27]

7.3.2 SDK

For third-party developers, Microsoft opened their SDK for Cortana, and it offers developers much flexibility to incorporate Cortana into their technology and encourages new behaviour.

Microsoft enables third-party users to identify activities that provide users with features from their applications based on explicit user requests or user context. However, activities are limited to Desktop and Mobile Windows 10 and Android. Developers may describe their own acts from scratch or choose from two predefined steps, such as ordering food and sending messages. [27]

Examples of own activities include Get Nutrition Details or Switching on the Lights but not limited to them. This makes the Cortana SDK extraordinarily versatile and capable of handling various intentions and activities that the other IPAs coming from Google and Apple don't.

However, developers need to register their acts that can be done without any costs and be tested by a developer who works for Microsoft's Cortana Squad. The main objective of this section is to provide the features of the SDK in Cortana.

7.3.2.1 Speech

Microsoft provides the following speech platforms and services for our apps.

- Windows speech: Windows speech is a collection of UWP APIs that allow windows speech to be used in both speech recognition and synthesis of speech through multiple languages on all devices that are based on Windows-10, including IoT hardware, computers, tablets, and PCs. Cortana on Windows is using these speech APIs. [27]
- Speech Recognition: Recognize real-time audio from a built-in microphone, from a non-microphone source, such as a Bluetooth headset or a file.
- Speech Synthesis: Transforms text into audio.

Microsoft also lags Amazon's Alexa Skills ecosystem, which today boasts the most extensive collection of voice assistants. Meanwhile, as was evident at the Consumer Electronics Show (CES) 2018, Google is making a significant investment in promoting Google Assistant and its Google Home unit.

In my view, Cortana is only available on niche dedicated speaker hardware and does not have Amazon's Alexa mindshare. To better compete, Microsoft needs to invest more in building an ecosystem of alliances, hardware options and workplace use cases. [27]

Microsoft is on a level footing with Google and Amazon from a technological perspective, but it faces an uphill climb battling Siri, Alexa and Google, particularly for home devices. A new survey

by 451 Research found that Amazon is the leader among respondents planning to purchase a smart speaker (38%), followed by the just-released HomePod from Apple (28 %). I still consider Cortana to be the digital assistant for employees, considering the overall market difficulties.

7.3.3 Future of Cortana

The vision of Microsoft for Cortana continues to develop, and the next step for the digital assistant, according to a recent patent, could include new features for users on the go to better read and summarize emails, text messages and other forms of communications.

With its latest Play My Emails feature for Outlook, Microsoft has already made advances in Cortana's ability to read and relay messages. However, by giving Cortana new capabilities to pull essential points out of long messages and summarize them, the knowledge outlined in the patent brings that capacity to the next level. [28]

Microsoft is well placed to extend Cortana's work scenarios, but much of its growth seems to come from within Microsoft and its deep bench of productivity resources, databases, applications, and cloud-based services.

The scale of Cortana will also be decided in certain respects by how effective Microsoft is in persuading more substantial third-party suppliers to build skills and support the technology. The organization will have access to the full Office Graph and automatically complete advanced activities that employees would like to set up meetings, follow up with clients, file expense reports, log meetings, and send out action items. [28]

In the meantime, speech recognition in business applications is becoming a standard feature, and developments in understanding natural language and voice synthesis will allow enterprises even more flexibility to select the best for interactions between humans and computers. This year, we expect to see more enterprise applications with conversational interfaces coming out, significantly where mobility directly impacts employees' productivity and where what works best is data input through a voice interface.

7.4 Amazon Alexa

Amazon Alexa, also known as Alexa, is an Amazon-developed virtual assistant AI technology, first used in the Amazon Echo smart speakers designed by Amazon Lab126. It can interact with

voices, play music, make to-do lists, set alarms, stream podcasts, play audiobooks, and provide weather, traffic, sports, and other content, such as news, in real-time. Using itself as a home automation system, Alexa can also monitor many smart devices. Users can expand the capabilities of Alexa by installing skills or apps. [29]

Most Alexa devices allow users to use a wake-word (such as Alexa or Amazon) to activate the device, some devices (such as the Amazon mobile app on iOS or Android and Amazon Dash Wand) enable the user to press a button to activate the listening mode of Alexa.

However, some phones still allow a user to say a command, such as 'Alexa' or 'Alexa wake.' Currently, only English, German, French, Italian, Spanish, Portuguese, Japanese, and Hindi are eligible for Alexa interaction and communication.

Let's just take a quick look at one of the devices that use Alexa, like Amazon Echo. Amazon Echo is not only a smart speaker when Alexa-enabled but acts as an intelligent virtual assistant. Alexa communicates with different Alexa-enabled devices, such as Echo, as a cloud service, and can communicate with other compatible IoT devices and third-party applications by translating voice requests to other services' native communication protocol. [29]

Users can also access the cloud service using companion customers, such as PCs or mobile (Android and iOS) devices, to configure these Alexa-related environments. Therefore, it is complex and heterogeneous to build the environment generated by interconnected devices, third-party applications and companion clients.

7.4.1 Working of Alexa

We define the particular architectures associated with the target IoT Ambience. The Amazon Echo controls an interface for interacting with the cloud-based service, Alexa, as described above.

Cloud-based operations, such as Echo and Alexa, are a general IoT consumer product operating method since most are inseparable from cloud services for user convenience in offering interoperability with companion customers and compatible devices. Although the Echo cylinder has some capabilities, such as speakers, a microphone and a small computer that can awaken the device and blink its lights to let us know that it is enabled, its real capabilities happen once it sends whatever we have told Alexa to the cloud to be interpreted by Alexa Voice Services (AVS). [29]

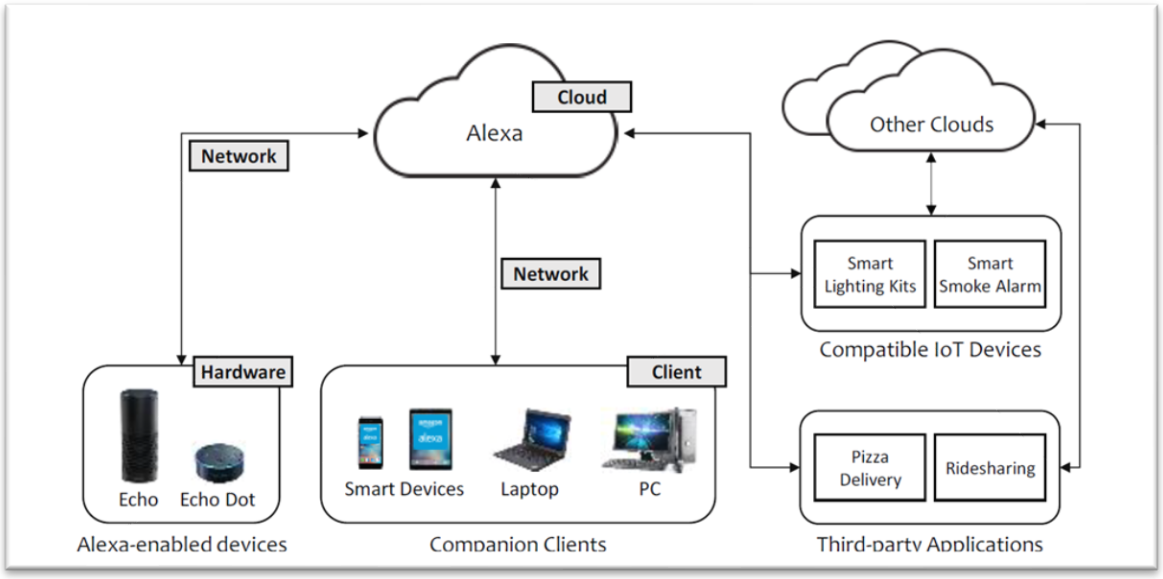


Figure 2: Working of Alexa. [30]

1. Hardware: It is essential to decompose each Alexa-enabled system to perform a hardware-level analysis. A category of AVS devices with a microphone array, speaker, and a core collection of Alexa features and functions are allowed by wake term. These things meet the specifications of the Alexa Built-in Badging Software.
2. Network: The devices and companion customers that are Alexa-enabled can interact with Alexa via the internet. It is established as a result of traffic analysis that most traffic associated with forensically significant artifacts is transferred over a cryptographic link after establishing a session with a valid user ID and password. [29]
3. Cloud: Alexa is a vital member of the Amazon AI ecosystem. Like other clouds, like Alexa, uses pre-defined APIs to transceive data services, but sadly the list of available APIs is not publicly open to the public.
4. Client: Finally, there is another level of Alexa companion customers that are linked to the work of Alexa. Interestingly, at least one companion client is essential for setting up and managing Alexa-enabled devices' service.

   Users can customize environment settings, review previous Alexa conversations, and enable/disable skills using a mobile phone or Web-browser, for example. A significant number of data associated with accessing Alexa can be stored naturally in companion customers in this process. This makes it possible to acquire and consolidate these client-centric objects with them and cloud-native objects.

So, when we ask Alexa, "*What will the weather be like today?*", the unit captures our voice. The recording is then sent to Amazon's Alexa Voice Services over the Internet, which parses the recording into commands that it understands. Then the machine sends back to our computer the correct output. An audio file is sent back when we inquire about the weather, and Alexa tells us the weather forecast, all without us having any idea that there was any back and forth between systems. Of course, that means that Alexa will no longer work if we lose the Internet connection. [29]

Cloud-based service enables computer manufacturers to incorporate into a linked product an ever-increasing collection of Alexa features and functions.

AVS takes the service to commercial computer manufacturers, while the Alexa service powers Amazon Echo products. To build Alexa into smart speakers, headphones, PCs, TVs, cars, and smart home devices, original engineering manufacturers (OEM), original design manufacturers (ODM), and systems integrators (SI) use AVS. AVS provides Cloud-based automatic speech recognition (ASR) and natural language processing.

7.4.2 SDK

To create an Alexa Built-in product, the Alexa Voice Service (AVS) Device SDK provides us with a collection of C ++ libraries. Our computer has direct access to cloud-based Alexa capabilities with these libraries to get voice responses instantly.

The SDK is feature-packed and conceptual. It provides separate components to handle the required Alexa features, including audio processing, continuous connection maintenance, and Alexa interaction management. To monitor interactions before integration, the SDK also includes a sample app. [31]

The following diagram illustrates the components of the SDK and how data flows between them.

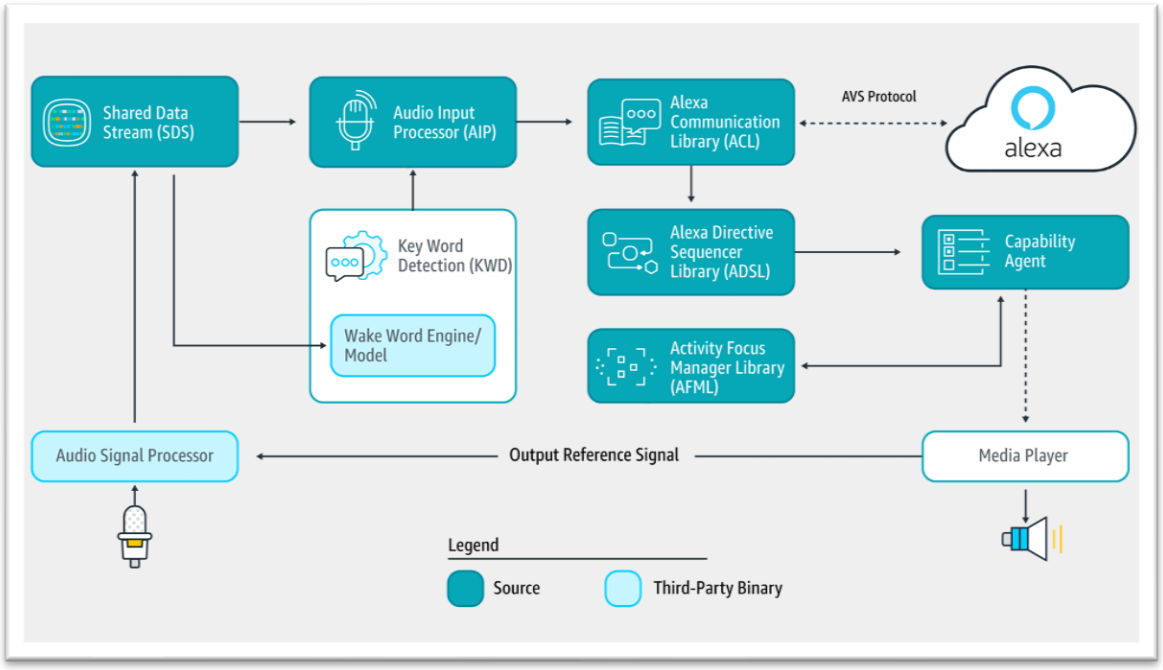


Figure 3: Components of Alexa's SDK [31]

These boxes are official components of the SDK – they include the following items:

1. Audio Input Processor (AIP): The AIP's duties include reading audio from the SDS and then forwarding it for processing to the AVS. The AIP also contains the logic for switching between various sources of audio input.
2. Shared Data Stream (SDS): The SDS is a multi-consumer, single-producer audio input buffer that transports information between a single writer and one or more readers. This ring buffer transfers data without replication in the SDK's various components. The memory footprint is minimized by this operation, as it overwrites itself continuously. SDS works on memory segments that are product-specific and user-specified, allowing for interprocess communication. [31]
3. Alexa Communication Library (ACL): The ACL handles the network connection between the SDK and AVS. It offers capabilities for message sending and receiving. Support for JSON-formatted text, binary audio content and forwarding of incoming ADSL directives are included in ACL functionality.
4. Alexa Directive Sequencer Library (ADSL): ADSL manages incoming directives, manages each directive's lifecycle, including, if necessary, queuing, reordering, or cancelling directives.

It also forwards directives by inspecting the directive header and reading the device namespace to the required Capability Agents.

5. Activity Focus Manager Library (AFML): The AFML makes sure the instructions are handled in the right order by the SDK. It decides which capability at any time has control over the device's input and output. If we are playing music, for example, and an alarm goes off on our computer, the alarm focuses on the music. The music will stop, and the alarm will ring. To control the prioritization of audiovisual inputs and outputs, Emphasis uses a term called channels. [31]
6. Capability Agent: What carries out the desired action on a system is a Capability Agent. They map to interfaces provided by AVS directly. For instance, if we ask Alexa to play a song, it's the Capability Agent that loads and plays the song into our media player.
7. Audio Signal Processor (ASP): On a dedicated digital signal processor, it is a Software On a Chip (SOC) or firmware (DSP). Even if our system uses a multi-microphone array, its task is to clean up the audio and create a single audio stream. Acoustic Echo Cancellation (AEC), noise suppression, beamforming, Voice Activity Detection (VAD), Dynamic Range Compression (DRC), and equalization are techniques used to clean the audio.
8. Wake Word Engine (WWE): The WWE is a program that continuously tracks the SDS, waiting for a preconfigured wake term. When WWE detects the correct word for the wake, the AIP starts reading the audio. The wake word is always "Alexa." when using the AVS System SDK. The SDK contains a Sensory wake word engine connector. [31]
9. Media Player: For the Gstreamer and Android Media Player, the SDK comes with a wrapper. We must create a wrapper for it with the Media Player GUI if we choose to use a different media player.

7.4.3 Future of Alexa

Amazon is reportedly taking measures to create new Alexa features and collect information to expand the assistant's capabilities. [32]

According to a patent, that could allow Alexa to take action without a "wake word" being needed to cause it. Rather than using the wake word "Alexa" before the command, a customer would say, "*play music, Alexa*." This small but essential change would allow customers to communicate with the assistant more naturally, which could help Amazon improve interaction and eliminate possible

barriers to making purchases or ordering voice services. But such initiatives would also pose privacy concerns, as it would suggest that not only would Amazon's always-listening devices begin processing after hearing a wake term. Instead, whenever they hear a wake phrase, they can still perceive any voices they pick up and act on previous commands.

Amazon is also looking to enhance camera-based body scanning, which for devices like the Echo Look could lead to a renaissance. [32]

In order to understand diversity among body shapes, Amazon's Body Labs unit performs a study in which it creates 3D models of participants based on scans, images and videos. Amazon may eventually incorporate body scanning technology to camera-equipped devices such as the Echo Look and probably even subsidiary home security systems such as Blink and Ring to provide personalized clothing to customers. Such a service could boost the ongoing movement of Amazon into apparel.

The company is considering an Alexa-enabled wearable capable of tracking the user's emotional state. While on this wrist-worn system, there are not many specifics readily accessible, it may theoretically allow Amazon to predict better how consumers feel and, therefore, what kinds of items they may like to buy or use. [32]

The business will need to balance these advantages against a mood-reading device's possible backlash and customer fears about the amount of data that the e-commerce giant can access and use to better target advertisements and marketing.

7.5 Samsung Bixby

Bixby is Samsung's digital assistant that allows us to monitor our smartphone via voice commands. We're able to open applications, check the weather, play music, turn Bluetooth on, and more.

In addition to supporting Google Assistant, Samsung's high-end Android phones come with their own voice assistant called Bixby. The effort by Samsung to take on Siri, Google Assistant and Amazon Alexa is Bixby. Although all of these three assistants' popularity has not been reached, it is still pre-installed on Samsung devices. It debuted on the Samsung Galaxy S8 but is designed to work across various Samsung products and is included in various other devices such as the Family Hub refrigerator and TVs from Samsung. [33]

We can use it to email, get personalized details (about the weather, meeting reminders, news stories, etc.), learn more about what the camera sees (such as where to purchase a couch in the camera view), and complete behaviour. Bixby learns individual voices, so depending on who is asking, it can personalize responses. Samsung notified that it "learns, evolves, and adapts."

For S Voice, Samsung's voice assistant app launched with the Galaxy S III in 2012, Bixby represents a significant redesign. On 1 June 2020, S Voice was discontinued. Samsung announced the arrival of Bixby 2.0 during its annual developer conference in San Francisco in October 2017. Bixby 2.0 is the newest edition of the digital assistant from Samsung. As demonstrated by the launch of the latest Galaxy Home Bixby-powered smart speaker, the new version will concentrate more on Samsung's IoT and related tech strategy. [33]

Samsung aims to build an ecosystem like the Amazon Alexa-powered one. It remains to be seen if it will compete, but Samsung's extensive range of appliances and products will give it an early advantage. As well as developers who want to incorporate their products with the Bixby ecosystem, Bixby 2.0 will impact Samsung clients.

With the Bixby SDK introduction, Samsung has opened the Bixby platform to developers, meaning it will likely impact hardware manufacturers contemplating which ecosystems to interact with.

7.5.1 Working of Bixby

Instead of launching an app, for example, or performing a single task, Bixby is designed to let us perform a full range of interactions. [33]

1. Bixby is contextually conscious, which means that, based on our requests, it can understand the state that the app is in and take the right actions, also enabling us to combine voice or contact.
2. Bixby should also interpret natural language, and this means we do not need to use fixed words, but we can include imperfect knowledge and Bixby can understand and take action. Recognition of natural language, for example, was crucial to the emergence of Alexa and is now a key feature of modern AI.
3. As with other AI solutions such as Google Assistant or Amazon Alexa, the service works in the same way in that it listens to our speech, interprets the details, and returns the resulting

action. The contextual interpretation suggests that we can get it to act without laboriously explaining what to do with what - it already knows where we are so that the next logical step can be taken.

7.5.2 Components of Bixby

In reality, it consists of three separate components that are all interrelated. Bixby comes with four parts known as "Bixby Voice," "Bixby Vision," "Bixby Home," "Bixby Routines," and Bixby Touch, which has recently been added. [33]

1. Bixby Voice: The most fascinating and useful aspect of the digital assistant is Bixby Voice, which helps us do things using voice commands. It works with all Samsung applications, including Instagram, Gmail, Facebook and YouTube, and a few third-party apps. We can send text messages, check sports scores, turn down the phone's brightness, check our calendar, launch applications, and more with Samsung Bixby Voice. Our latest messages can also be read out by the digital assistant and communicate in a male or female voice. More complex two-step acts, such as making an album with our vacation images and sharing it with a friend, can also be done by the feature.

   According to Samsung, more than 3,000 voice commands are assisted by the digital assistant, so anything we can do on our Samsung phone via touch can probably be done with our voice. Bixby Voice also supports simple commands, enabling us to execute a single phrase with multiple actions.
2. Bixby Vision: Samsung Bixby Vision, which is essentially Samsung's version of Google Lens, is the digital assistant's second component.

   It is an augmented reality camera that can, in real-time, recognize objects and then search for them on different services. Users can launch it from the camera app or Bixby Home directly. The execution of Samsung seems to be much better than the competition. It can interpret text, read QR codes and identify landmarks. Many applications can already do this, but Bixby excels in integrating all of them right there in our camera app for what seems to be reasonably good object recognition. To use it, we point it at something and wait for a tick (or several ticks, depending on the quality of our Internet connection). Flowers and a watch were recognized in my evaluation, and it can also recognize items like wine labels and books. [33]

Samsung has a few particular partners with whom it works, such as Amazon, Vivino, and Pinterest come to mind, and more are coming. I guess it's my favourite thing Bixby does.

3. Bixby Home: Bixby Home, which resides on our home screen, is the last part of Samsung's digital assistant. It's close to Google Feed and HTC Blinkfeed, displaying updates from social media, trending videos from YouTube, weather forecast, reminders, etc.

   We can configure it to display just the details we're interested in, and by pushing the ones that matter to us most to the top, alter the order of the Bixby Home cards. To a degree, Bixby Home can be personalized. Default apps from Samsung supply much of the content from Bixby Home.

   We can see a local weather forecast, activity stats from Samsung's Health app, and local files in the Music app. If we have connected our Google Calendar and Gmail to our Samsung computer, we will extract data from them.

4. Bixby Routines: Bixby Routines is a function that is supported by machine learning to adapt to our lives, suggesting ways to streamline our time on the phone.

   Context clues are activated by automatic actions: location, time, or event. For example, Bixby Routines can do a range of apps and settings, such as pulling up the Maps app, opening our music app, and keeping our phone unlocked for easy access, to add comfort to our drive Samsung devices attach to the Bluetooth in our car. Or if we fall asleep at night without putting our phone on the charger, unnecessary features would be shut down to conserve battery life. We also can create a custom routine with My Routine that fits our individual requirements.

5. Bixby Touch: Based on intelligent recognition, Bixby Touch makes recommendations. By pressing the screen, we can conveniently access translation, online shopping, and the media. It is the latest technology in the field of Bixby personal assistant, which is still in development, so not much data is available to describe this component. [33]

7.5.3 SDK

Using AI-powered tech, the Bixby Developer Center will allow developers to build Bixby voice apps to understand when to apply machine learning to automate tasks.

It provides a new way of helping users do things in a more profitable, personalized and natural way. To start developing Bixby, users now have access to all the software and documentation. They are using cases and templates to describe dialogue and layouts. To build conversational

experiences, we connect our capsule to our current APIs and add natural language training. Teach Bixby what users think and how it maps back to Capsule users' concepts and behaviour.

Bixby then extends that information to new terms and phrases to consistently apply the right interpretation to user requests. Production is a collaborative process with the AI for Bixby. [33]

Bixby is taught by developers the principles and behaviour their programs can perform. Within milliseconds, Bixby uses the models to create a new program to process user requests dynamically.

The Bixby Developer Studio is purpose-built with strong AI at its core to create rich conversational experiences. Users use this Bixby Developer Studio to create end-to-end Bixby Capsules, from concept and creation to testing and submission. When the customer has updated the Bixby Capsules, they publish them for discovery when the Bixby Marketplace opens. Samsung has its own app store for the Bixby marketplace where users can publish the capsule, and they have using the Bixby developer center.

7.5.4 Future of Bixby

In exchange for ditching both Bixby and the Galaxy Apps Store, Samsung is suspected of exploring a new global revenue sharing agreement proposed by Google. Future Samsung smartphones will instead use Google Assistant and the Play Store by default. [34]

On Samsung's mobile hardware, the deal is also expected to give Google's search service "greater prominence." Samsung is officially dismantling its developer relations team responsible for growing third-party voice apps for Bixby.

The exact terms being offered are not clear, but in terms of the percentage of ad revenue Samsung will earn from Google apps running on its phones, Google has increased its bid. It is also assumed that both businesses plan to have the terms of the deal finalized, indicating that Samsung is very close to being satisfied with the new bid from Google.

For now, Samsung does not seem to be making any attempts to give up entirely on Bixby. Cutting the developer relations team means we will not see any substantial improvement in Bixby Capsules' third-party support. The digital assistant continues to live. [34]

8 Analysis of Open-Source AI Personal Assistant

8.1 Mycroft AI

Mycroft is the first open-source voice assistant in the world. It can run anywhere, and it even runs on a Raspberry Pi on a desktop machine, inside an automobile. It is open so that it can be remixed, expanded, enhanced.

It can be seen in anything from a research experiment to an application for enterprise applications. It is possible to freely remix, expand, and deploy the Mycroft open-source voice stack anywhere. [35]

Mycroft may be used in everything from a research project to a global business environment. Mycroft is the name of a set of hardware and software tools that include an open-source voice assistant using natural language processing and machine learning. It is named after "The Moon Is a Harsh Mistress," a fictional computer from the 1966 science fiction book. [35]

8.1.1 Components and Working of Mycroft

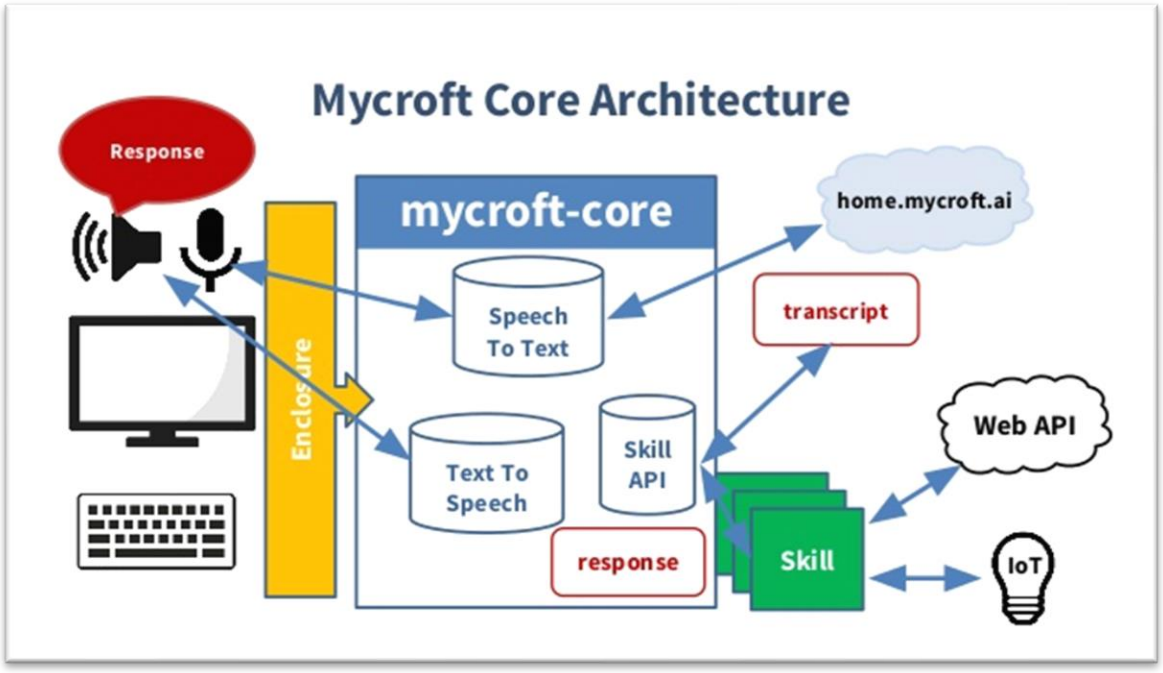


Figure 4: Components of Mycroft AI [36]

Mycroft is modular. Some components can be easily swapped out for others:

1. Wake Word detection: A Wake Word is a word we use to say we're about to give an order to Mycroft. This is "*Hey Mycroft*" by default, but we can customize our own Wake Term in our Mycroft Home account. [35]

   There are two technologies that Mycroft AI currently uses for Wake Word detection:

   - PocketSphinx: PocketSphinx, created by Carnegie Mellon University, is part of the more expansive CMUSphinx kit. PocketSphinx, tuned explicitly for handheld and mobile devices, is a lightweight speech recognition engine. The Wake Word currently needs to be an English word, such as Hello Mike, Hi there Mickey or Hey Mike, since PocketSphinx is educated on English speech. In other languages, including Spanish, French or German, Wake Words will not work as well.
   - Precise: Precise is a neural network trained on audio data, which is based on Speech-to-Text technology. What words we want to use for our Wake Word does not matter. Instead, we train it on sounds. The downside is that on our selected Wake Term, precision needs to be practiced. Precise is the default Wake Word Listener for the wake word "Hey Mycroft" if Precise is inaccessible, PocketSphinx offers a fallback to this.
2. Speech to Text (STT): Speech to Text (STT) software is used to transform them into text phrases that can then be acted on. The Mycroft team is collaborating to create Deep Speech with Mozilla. A fully open STT source engine based on the Deep Speech architecture of Baidu and implemented with the TensorFlow system of Google. Deep Speech is not yet ready for production use, and Google STT is currently used as the default STT engine by Mycroft. [35] Mycroft also supports other STT engines that can be configured using the Configuration Manager:
   - IBM Watson Speech to Text: IBM Watson Text to Speech provides a voice to Mycroft, allowing them to maximize consumer experience and interaction using any written text by communicating with users in their own languages. Increase user usability with various skills, including audio solutions to discourage distracted driving, or automate customer service interactions to increase efficiencies.
   - wit.ai Speech to Text: To achieve low latency and high robustness to both surrounding noise and paraphrastic variations, Wit incorporates multiple state-of-the-art Natural Language Processing methods and many speech recognition engines. Luckily, users don't

need to take care of all that equipment. From ice-cream delivery to space flights, Wit can adapt over time to our domain. Wit does not make any assumptions and remains 100% configurable.

3. Intent parser: Software that identifies what the intent of the user is based on their speech is an intent parser. The output of a Speech to Text engine is usually taken as an input by a purpose parser. This task intends to achieve what the consumer wants in speech recognition and voice assistance. In different ways, a user can accomplish the same mission. [35]

   The intent parser's job is to extract key data elements from the user's speech that define their intent in more detail. To help the user complete their desired goal, this knowledge can then be passed on to other programs, such as skills. For instance, Julie talks to Mycroft as follows: Hey Mycroft, tell me about the weather. Julie's goal is to learn about the weather (probably in her current location).

   To handle the intent, an intent parser should then fit the intent with relevant skill.

   - Adapt intent parser: For all Mycroft platforms, adapt is the default purpose parser. Adapt has been developed by Mycroft and is available under a license that is open source. It is lightweight and is designed to run on devices, such as embedded devices with limited computing resources. Adapt takes as an input the natural language and outputs a data structure that includes:
     - the Intent: What the Customer intends to do
     - a match probability: how sure Adapt is that the intent has been adequately established
     - a tagged list of entities: that can be used to execute functions through Skills [35]
   - Padatious: Padatious is an intent parser based on a neural network. Padatious is currently under Mycroft's active development and is accessible under an open-source license. Some Mycroft platforms are likely to turn in the future to using Padatious instead of Adapt. Padatious is an intent parser focused on machine-learning, neural-network. In comparison to Adapt, which uses small groups of specific words, Padatious is trained as a whole on the sentence. Over other deliberate parsing methods, Padatious has a range of main advantages such as: [35]
     - Intents are simple to build with Padatious.
     - A relatively small amount of data is needed for the machine learning model at Padatious

- It is essential to train models for machine learning. It is fast and simple to practice with the model used by Padatious.
- Intents run separately from each other. This allows new abilities to be built quickly without retraining any other abilities.
- We can quickly remove entities with Padatious and then use them in skills.

4. Text to Speech: Text-to-Speech software takes written text, such as a computer's text files, and uses a voice to speak the text. Depending on the TTS engine used, Text to Speech may have numerous voices. [35]
   - Mimic: Default local text to speech (TTS) engine from Mycroft, based on CMU's Flite (Festival Lite). Mimic is a fast, light-weight Text to Speech (TTS) engine based on FLITE software from Carnegie Mellon University.

     Mimic utilizes text as an input and uses the selected voice to output speech. Mimic has a minimal resource footprint and is low latency. It is also set apart from other open-source text-to-speech projects by its range of high-quality voices. Mimic's limited resource footprint, apart from being used as Mycroft's voice, makes it an appealing option for other embedded systems. Mimic is currently operating on Linux, Android and Windows, and other platforms could be supported in the future. [35]
   - Mimic2: Mycroft's own Tacotron-based cloud-based voice text to speech (TTS) engine offers a much better voice quality. Mimic2 is a useful TTS tool, but it can help solve other important issues as well.
5. Middleware: The Mycroft middleware has two components:
   - Mycroft Core: The core program that provides the 'glue' between other modules is written in Python. Mycroft Core is available under an open-source Apache 2.0 license.
   - Mycroft Home and Mycroft API: It is the portal where user and system information is stored. This framework offers abstraction services, such as storing API keys that are used to provide Skill features for third-party services to access. This platform code is available under an open-source AGPL 3.0 license. [35]
6. Mycroft Skills: Mycroft Skills are like 'add-ons' that provide extra features. Skills may be built and differ in their functionality and maturity by Mycroft Developers or Group Developers.

   Mycroft Skills Kit is a Python-based utility designed to promote the development, testing and submission of Skills to the Skills Marketplace by Skills Writers. Mycroft Skills Manager is a

command-line interface used by every installation of Mycroft to add, manage and remove skills.

7. Devices and Enclosures: Mycroft is designed for several different platforms to run on. Every dedicated platform is called a device, such as:
   - Mark 1 - uses a dedicated software image to be the first reference hardware unit.
   - Mark 2 - uses a dedicated software image to be the new reference hardware unit.
   - Picroft - any Raspberry Pi 3 or 4 that runs the software image of Picroft.

   The enclosure refers to the unique code for that system that is required. It could identify unique features, such as the Mark 1 eyes, or a particular way to communicate with the hardware, such as controlling the volume levels through i2c at the hardware level. [35]

8.1.2 SDK

Under the name Mycroft Skills Kit, Mycroft has an open SDK. Mycroft Skills Kit - MSK - is a Python-based utility built to make it much easier for Skill Writers to develop, test and submit skills to the Skills Marketplace. This usefulness will help shorten the development cycle and reduce some of the repetitive components of skills production. [35]

MSK currently provides the following features:

- Creating a new skill
- Create an Intent Test for a Skill
- Uploading a Skill
- Upgrade an established skill

8.1.3 Future of Mycroft AI

Mycroft is a big project that encompasses many critical technologies. Progress on the numerous technological fronts is mainly autonomous, enabling changes to occur within their own timetables. This might, however, make it hard to keep track of how things are progressing. For Mycroft, the best part is that they have an AI mission for everyone. [35]

Only this year, Mycroft open technology allowed customers and other businesses to easily combine Mycroft Holmes 1 electronics with some Texas Instruments sensors and GE's Green Bean hardware interface to make a dumb gadget a smart one.

Mycroft announced that via Kickstarter, a third hardware project, Mark III, would be offered. Mycroft has made several commercial partnerships. The organization has collaborated with WorkAround, an impact sourcing provider that brokers job opportunities for refugees, to conduct bulk machine learning training. [35]

9 Comparison of Apple Siri, Google Assistant, Amazon Alexa, Microsoft Cortana, and Samsung Bixby

I have made these comparisons based on results shown in some of the Internet's research and personal experience. I have used iPhone 6S plus and iPhone11 to test Siri, Pixel 4XL and Google nest to test Google Assistant, Windows 10 on my laptop to test Microsoft's Cortana, Alexa app available on Android play store to test Alexa, and Samsung Galaxy S10 to test Bixby.

I could not get all the expected Alexa results because I was not using Alexa Echo, the device Alexa works efficiently, but I was only using its android application. Unfortunately, I could not test Mycroft AI as Mycroft AI Mark I was all sold out, and Mark II is not available until next year.

I have divided this test into four different phases starting with fundamental questions, and in the last phase, I have given my final verdict based on these four tests.

9.1 Test 1: Answering Basic Questions

We first start with simple, straightforward questions to compare these AI Assistants that no AI assistant should have any trouble with. Tasks included weather questions, sports scores, telling a joke, and times for movie shows. [37]

As expected, all five assistants performed very well, but in the shortest time, Alexa gave the most useful details, which makes sense-it must be as concise as possible as a voice-only program, and we almost always found what we were looking for, but Bixby seemed to struggle with direct commands before the correct context was given.

However, if something doesn't work with Bixby, we can still ask it to conduct a Google search with the same query and piggyback off of the excellent answers from Google, which is undoubtedly an unusual workaround but still functional.

We also tried a variety of questions that we might put into a search engine: "*How old is Barack Obama?*", "*Who built the Taj Mahal?*", "*What is the population of Canada?*" and "*How big is Vancouver?*". The correct answers were returned by all the assistants, which means that each of them will make a good trivia night partner. [37]

9.2 Test 2: Performing Simple Tasks

There are occasions when, without actually calling them, we want to look up contact details. But we were trying to figure out an address for one of our contacts. Kudos to Siri and Bixby, who understood the command "*Where does [name] live?*" For a famous person of the same name, Cortana and Google Assistant returned a web result, while Alexa wasn't able to look up addresses.

All these digital assistants were asked to view our most recent emails. The most recent computer messages were all presented by Google Assistant, Bixby, and Siri, but Alexa and Cortana were unable to, hampered by not having strong integration with an Android or iOS operating system.

However, Cortana could send emails to particular contacts, like Bixby, Siri, and Google Assistant. Here, Alexa loses out again. [37]

Bixby, Google Assistant, and Siri all earn bonus points to perform various linked activities together, such as launching a web browser and opening a particular web page. Worth noting is that the Google Assistant is the only assistant that has not failed in any group to any degree here.

9.3 Test 3: Performing Daily tasks

Alarms, timers, and notifications are digital assistants' bread and butter jobs, and none of the contending applications can let us down in this category.

All assistants were able to "*set the alarm for 9 a.m. tomorrow*" They correctly defined the day and time and also displayed a screen confirmation. The "*cancel the alarm for tomorrow*" follow-up order only operated on Bixby, Google Assistant, and Alexa. Siri did not understand that word but replied to "*cancel the alarm*," perhaps an indication that the recognition of its natural language lags only a little behind the others. And until we asked for a list of alarms and manually toggled it off, we could not get Cortana to cancel the warning at all.

A few different answers were met with the necessary "*remind me to buy milk*" instruction for reminders. Before saving it, Alexa requested a day and time for the reminder, Google Assistant wanted either a time or a location where the reminder would kick in, and Bixby, Cortana, and Siri saved the note without asking for any further information in the default reminder app. If we provide a particular date and time in our original voice order, all apps can save the relevant information and time the reminder accordingly. [38]

Nobody wants all the work to be life and no fun. Therefore, we need to update our digital assistant on the news, have sporting scores, and play music and films.

Alexa reads out bulletins from a few local, national, and foreign services in response to "*What's the news?*". Some of the day's top headlines are seen by the other candidates, Cortana, Google Assistant, Bixby, and Siri, as taken from the internet, but only Cortana starts reading them aloud. However, if we have a built-in Google assistant in a Google Home speaker device, it can also read us the news.

For example, with Alexa, Google Assistant, and Cortana, we could call up and play Spotify playlists perfectly. But the assistants fell short when we tried other music apps, such as Apple Music and Google Play Music. Siri was only able to monitor Apple Music, and Bixby was even more limited; only audio files stored on the phone could be played. [38]

Likewise, movies have differed from app to app. As long as we have linked it to a Fire TV on the same wireless network, Alexa opens apps and plays specific movies when we say, "*show me an action movie*" or "*open Netflix*". Those images, and any YouTube clips, are pulled up by Google Assistant, Siri, and Bixby on the phone we are using. Siri can fire up everything we have saved in our iTunes library, and Google Assistant can beam any video material to a nearby Chromecast. We've left out Cortana because it doesn't have the same video-playing abilities that its rivals do, but with a Bing search, it can pull up a few YouTube videos.

9.4 Test 4: Navigating complex tasks.

"Hard mode" reflects the final collection of queries. These are dynamic tasks that would require some sort of integration of third-party applications. Tasks included monitoring fitness, booking an Uber, ordering food, and uploading an application.

A definitive response for which one is the best is difficult to find since many of these tasks require complex configurations or have other limitations. Siri can order pizza for us, for instance, but it is location-dependent, and Google Assistant can hail an Uber and order food. [38]

Overall, functionality is naturally volatile or incomplete, but things look promising to some degree with all five assistants. If the function rollout was more consistent across devices or regions, Siri and Google Assistant could have performed better, and although Cortana managed to perform a lot of the tasks, many attempts were needed. In this category, Alexa ended up being the best.

9.5 Final Verdict

This may sound like much research, but there are also several features that we haven't explored in these assistants, which is a testament to how capable they are becoming. Even we managed to rate these digital pals in various ways with our simple examination. Following figure shows the graph bar based on queries correctly answered by the categories.

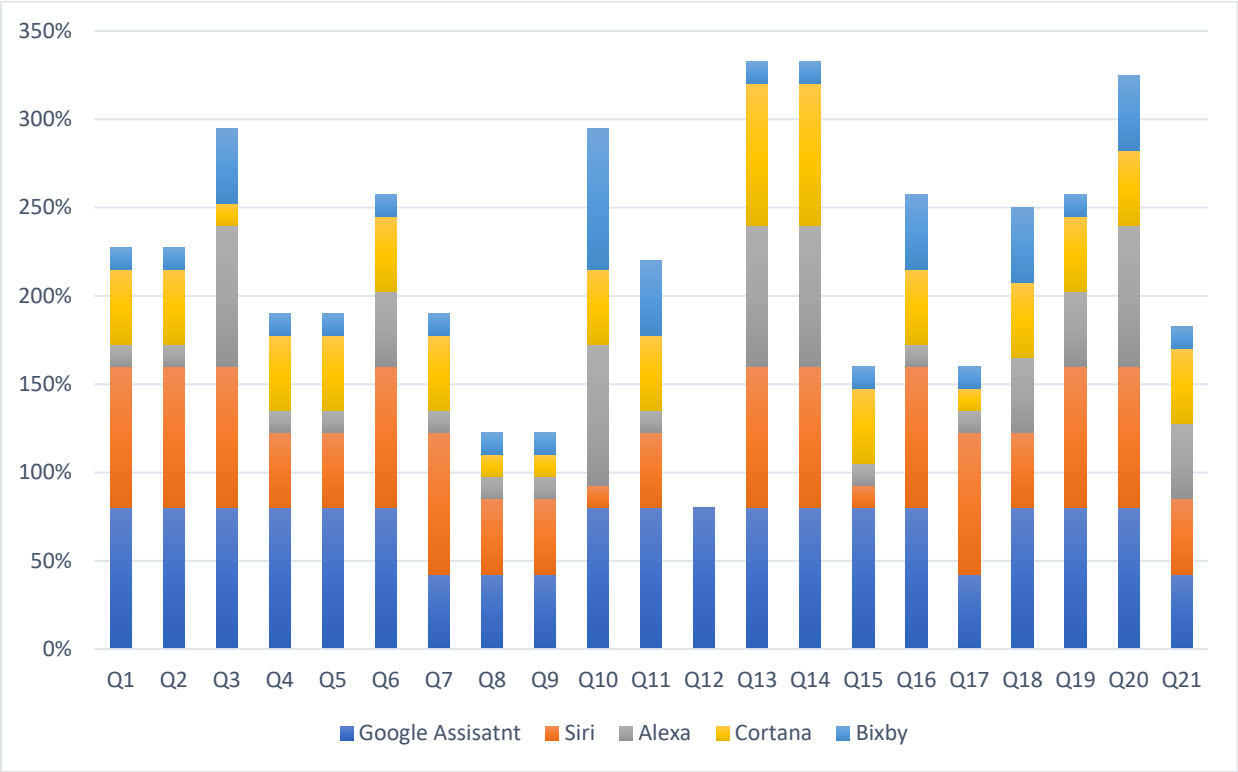
Questions Asked	Ranking based on accuracy, efficiency, and quality of the answers		
Range	Best (100% > 60%)	Better (60% > 25%)	Good (25% > 0%)
Q.1 How do I get to the Edmonton Airport?	Google Assistant, Siri	Cortana	Alexa, Bixby
Q.2 Book me a flight to America	Google Assistant, Siri	Cortana	Alexa, Bixby
Q.3 Call me an Uber	Google Assistant, Siri, Alexa	Bixby	Cortana
Q.4 Send an email to Harry	Google Assistant	Siri, Cortana	Alexa, Bixby
Q.5 Do I have any email from Sudhir?	Google Assistant	Siri, Cortana	Alexa, Bixby
Q.6 Send a text to Leonard	Google Assistant, Siri	Cortana, Alexa	Bixby
Q.7 Do I have any new text from Facebook?	Siri	Google Assistant, Cortana	Alexa, Bixby
Q.8 Who is the best player in IPL 2019?	N/A	Google Assistant, Siri	Alexa, Bixby, Cortana
Q.9 Last winner of Indian election?	N/A	Siri, Google Assistant	Alexa, Bixby, Cortana
Q.10 Play me some old music	Google Assistant, Alexa, Bixby	Cortana	Siri

Q.11 Play my favourite song on youtube?	Google Assistant	Siri, Bixby, Cortana	Alexa
Q.12 What's the weather going to be like the day after tomorrow?	Google Assistant	N/A	N/A
Q.13 Will I need an umbrella in the next two days?	Google Assistant, Siri, Alexa, Cortana	Bixby	N/A
Q.14 Do I have anything scheduled for the coming week?	Google Assistant, Alexa, Siri, Cortana	Bixby	N/A
Q.15 Asked each assistant to schedule dinner with Mom on my calendar Sunday night	Google Assistant	Cortana	Siri, Alexa, Bixby
Q.16 What are some Indian restaurants near me?	Google Assistant, Siri	Cortana, Bixby	Alexa
Q.17 I want to make a reservation at the Radisson hotel in Edmonton	Siri	Google Assistant	Alexa, Bixby, Cortana
Q.18 How do I say "where are my Clothes" in Hindi?	Google Assistant	Siri, Alexa, Bixby, Cortana	N/A
Q.19 Who is the current Prime Minister of Canada?	Google Assistant, Siri	Cortana, Alexa	Bixby

Q.20 Do you know Siri?	Google Assistant, Siri, Alexa	Cortana, Bixby	N/A
Q.21 What do you think who the best as a virtual assistant is?	N/A	Google Assistant, Siri, Cortana, Alexa	Bixby

Table 1: Best, better, and good personal assistants by feature.

Based on the table mentioned above I have made a graph comparing all the assistants based on the questions asked.



Graph 1: Questions Answered correctly by categories.

In my view, Google Assistant is the best to understand natural language and respond to follow-up questions. This is to be expected, as the app will rely on the hefty search and AI experience of Google. It also benefits from close integration with other Google services, such as Android, Gmail, and Google Maps. In terms of background, Google Assistant is unique, which is one of the drawbacks of most other personal assistant applications. We have to get the command precisely right for other assistants to complete the mission, but the Google Assistant is good at finding out what we need, even though we ask it to do the same thing in slightly different ways. [38]

Siri has a similar, strong mobile integration, but only of the iOS variety. Nevertheless, its lack of support for non-Apple applications and services still leaves this assistant down. If users prefer iPhones, there is no need to switch with Apple's Mail, Calendar, Contacts, and Maps apps; Siri will work well. During every turn of this text, Siri also performed surprisingly well. Apple has done a brilliant job of keeping Siri competitive with everyone else, and Siri certainly feels like a good alternative. Of course, for everyone in Apple's ecosystem, Siri is the obvious alternative, and few people will turn from Android to Siri. Nevertheless, the most remarkable aspect is its backwards compatibility.

The greatest strength of Cortana lies in the fact that, like iOS, Android, Windows 10, and even Xbox One, it is available everywhere. It is available on more computers than any other personal assistant, and in this respect, it's not even near competition. When it comes to third-party integration and more complex activities, though, it is increasingly falling behind.

Cortana is certainly not lacking in potential, but Microsoft needs to step up its game a lot to be a driving force in this space, and of the digital assistants we checked, Cortana, the Microsoft assistant hitching a ride on Android and iOS, seems to be the most disjointed. That said, it syncs neatly with Windows 10, runs through various devices, and makes some effort to understand the news stories, sports scores, and other interests that we follow. Sadly, it's not quite as polished as Google Assistant, Siri, or Alexa. [37]

Alexa excels in the number of third-party abilities; we can tap into the service from businesses such as Domino's and Uber and external applications such as iCloud and Spotify. It can also acknowledge natural patterns in language well. It's not available as a phone app on the downside, but as we speak, Amazon addresses the drawback so that it won't count as a disadvantage for much longer. Perhaps the most customizable assistant of the lot is Amazon Alexa. It is straight out of the box, but we can install additional "skills" that dramatically increase its functionality. It is up to us how many skills we want to add to get the assistant we want.

Finally, we have got a wild card, which is Bixby. During some of our research, Bixby did an excellent job, and where it excels the most is in feeling like an actual personal assistant. [38] Bixby can mimic screen taps like an actual individual when it does something, allowing it to communicate better than any other assistant with on-screen elements.

Bixby is a work in progress. It can't deliver as many features at the moment as its competitors do. However, it does monitor Samsung devices well (try commands such as "*close all recent apps*") and works well with the manufacturer's own mobile apps. Expect some significant upgrades to come. [38]

10 Discussion

An overview of the findings and a review of the results will be discussed in this chapter: Privacy Concerns, technology constraints, a conclusion and a future work of this report.

The drawbacks in terms of implementation and computation on time, data quantity and system limitations are addressed, emphasizing learning implementation. Ultimately, the conclusion of the findings reported is discussed with guidance in the fields of future studies. [39]

In our day-to-day lives, virtual assistants are increasingly prevalent. Many of us own at least one IPA, either Siri on iPhone or Google Assistant on Android phones, due to smartphones' assertive outreach. Due to Windows 10 and Alexa's large users as a home speaker, Cortana also has a substantial scope. It’s concluded from the entire scenario that recognizing voice required a range of significant distinct variations such as setting, voice modulation, frequency, etc.

The primary challenge of speech recognition is that people's voices vary, and they speak in different ways and different languages.

So, the discussion here is that we have to compare all of the personal assistants present today to the fictional AI assistant Jarvis. Who doesn’t know about Jarvis developed by Tony Stark? Jarvis is a personal assistant who can optimize Tony's tasks and follow instructions very well but is not great at independent decisions.

Alexa recollects things precisely as recorded and is unable to provide its perspectives, but on the other hand, Jarvis can provide its perspectives. When Jarvis communicates his findings, he uses the holographic display of Iron Man’s helmet and different graphical representations like the Amazon Echo.

Amazon’s Alexa aims to be a ubiquitous virtual assistant wherein the user can talk to devices & gadgets to provide information on a day-to-day basis. Jarvis is a high-end AI program developed by Tony Stark and was created to make decisions in extreme conditions. Like Google Now, Jarvis is hyper-charged to read all the conversations and emails of Tony Stark.

So, the real question here is how close are we to achieve the fictional Jarvis? Jarvis is knowledge-based, but all the personal assistants present today are search engine extensions. Jarvis is an AI-based assistant capable of making life easier by giving us the best answer to our questions.

So, do we need a high technological-based assistant for everyone?

If everyone has their own Jarvis, then many things can go wrong when learned from its human controller like if the human controller uses it in military application with wrong intentions than it can access nuke codes which leads to war between countries, or the controller wants to abuse it by using it in hospital and it can shut down the power system which can cause the death of many patients. So, it depends on the controller how they use this high-end assistant, whether for the benefit of humanity or to destroy it.

There is possible potential that when given to everyone hands for use than we are going to abuse it and to criminalize it. For example, an AI personal assistant developed by Facebook named "Jarvis" got shut down because AI was learning all the things from humans and became racially biased. So, the project got shut down.

To prevent this, we should a set of control walls for these assistants which contain restricts similar to the three basic laws of robots.

10.1 Privacy Concerns

Google, Amazon, Apple, and others had access to a fresh and useful stream of customer data as users opened their homes to data collection with smart home devices.

At a period when massive hacks, leaks, and misuse of personal information led to an environment of heightened customer interest and increased government oversight, this raised serious privacy concerns.

In March 2018, it was reported that Cambridge Analytica, a data analytics company, obtained unauthorized access to 87 million data from users, which further spurred the data privacy conversation. [39]

Devices such as Google Home and Amazon Echo responded to their respective wake words, "Hey, Google" or "Alexa". The smart assistant, once enabled, recorded what a user said and sent the recording to the backend servers, where it interpreted and processed the spoken input.

While only supposed be listening in response to their wake words, some Google Mini devices were found be listening to their owners and recording all the time. Although Google quickly solved the

problem, such events posed questions about how much data was collected, who had access to the data, and how it would be used in the future.

Alexa's privacy issues were heightened when Amazon admitted in May 2018 that one of its Echo systems accidentally captured parts of a couple's conversation and sent the recording of the conversation to a person on the contact list of the owner. The computer misinterpreted parts of the conversation, according to Amazon, as a series of commands instructing it to start recording and send the message to a contact. [39]

Another privacy issue was that any person who had access to the system could monitor voice-enabled assistants installed in smart speakers, allowing them to perform tasks that required confidential information. Theoretically, for example, if they were close enough to the computer, someone might make a transaction using stored payment information or retrieve personal information.

Like Apple had done for the iPhone, Amazon offered to set a 4-digit pin code for its speakers, but even this did not provide foolproof protection. By late 2017, the ability to differentiate between different voices and connect those voices to personalized accounts was introduced by both Alexa and Google Assistant. Neither, however, had the power to limit intelligent speakers' access to only approved voices, citing the need to preserve flexibility in responding to home guests.

In May 2018, the European Union prepared to introduce a sweeping new data protection law called the General Data Protection Regulation (GDPR) [39] as the voice wars were heating up. Under this new legislation, businesses will have to receive explicit and definitive consent from a customer to destroy and process personal information.

Besides, businesses will not be permitted to keep data for longer than is required, and at any time, a customer might ask for data to be removed. Fines for violations may be as high as 4 percent of the worldwide turnover of a company. Before recording and processing any personal information, voice-powered devices will soon need privacy disclosures and explicit opt-in consent with imminent legislation. [39]

10.2 Technology constraints

Accuracy has been a significant barrier to voice assistants achieving ubiquity in customer’s lives in multiple users and diverse environments.

Although machine learning developments helped improve voice as a UI, there were still technological challenges facing businesses. For instance, AI-powered voice assistants have struggled to discern user commands from background voices and ambient noise.

It was generally easy for a user to talk to their intelligent assistant with low intrusion levels in a home. However, a desired degree of precision has been a challenge in noisier environments such as urban areas. Machine learning researchers have discovered breakthroughs that would enable voices to be differentiated by AI assistants. [39]

Researchers at the Mitsubishi Electric Research Laboratory created one such breakthrough in May 2017. Their AI platform managed to differentiate multiple voices and recreate what was said individually by everyone. The device could correctly classify two individuals with up to 90 percent accuracy speaking through a single microphone; with three individuals speaking, the accuracy fell to 80 percent. However, this development was still an improvement, and previous technologies could only reliably discern two voices at 50 percent accuracy.

In its speech recognition technology, Google had achieved a 95 percent accuracy rate by 2017, a remarkable achievement, but not one without its own set of challenges. As customers and regulators entered the dialogue and shared their perspectives, these challenges will continue to develop. Such hurdles have produced questions about the potential success of Google, and how Google felt about its future in the voice industry were underpinned. [39]

11 Future Research

Intelligent personal assistants have been entities with many complex components: programming skills, statistics, Artificial Intelligence, Machine Learning and Ethics. This study's findings add to previous research and illustrate the significance of statistics, Machine Learning, Interaction and Ethics between humans and computers.

For the future, IPAs are of broader value than they used to be. A significant number of people often have cognitive disorders from the disabled population who may have trouble forming full sentences and communicating, and personal assistants may be a deal-breaker for those individuals. Although there is room for development in all tested devices, new research is continuously being carried out, and the abilities of IPAs are being tested and implemented in many devices. [40]

Combining these devices with different technologies and algorithms for machine learning also gives birth to numerous new possibilities such as education, finance, industry, advice, sales, etc.

Future research in this area could go deeper into the components that are necessary and essential for the development and creation of IPAs, the sustainability between the components, as well as building an entity with a rich SDK that allows for further extension in terms of the technical skills and competence required and what ethics to follow when doing so. It would be of great interest to compare these six IPAs and their respective SDKs to see what platform has come the furthest in the area, and by natural implication, the one that is contributing the most to the area.

12 Conclusion

This paper aimed to analyze the capabilities of some of the latest IPAs that exist in the industry today. It is possible to assume that today, the most versatile assistant is Google, which is open to all standard operating systems for mobile devices, including iOS and Android.

It is impossible to quantify intelligence and technologies at this time, primarily because the IPAs under consideration are being established securely. Google Assistant, which has the richest SDK and enables third-party developers to customize and build their own actions for their agents, is the most versatile agent.

Google is still the champion on the legal side of things. Google has created an ethical, well-formed agent available to developers worldwide, and without confirming the new behaviour, it does not blindly introduce fresh features for their agent. [40]

This is widely discussed in Superintelligence, where a secure, shared ground for the AI in question needs to be developed before the public can begin tweaking the intelligence source code. Google is now collaborating with Artificial Intelligence organizations like OpenAI with the most nuanced agent, which seeks to encourage and grow friendly AI in such a way as to help humanity rather than harm it.

In the end, I would like to conclude that if we compared all the AI assistants on a scale of one to ten where ten being the ultimate achievement in personal assistant then according to my opinion, Jarvis is a ten and Google assistant is a two. This indicates that it’s a long road ahead for all the assistants to grow and gain the abilities shown by fictional Jarvis. But it also indicates that there are many safeguards and privacy control that need to be put in place as well.

13 References

[1] N. Severt, "4 Amazing Ways AI Personal Assistants Can Impact Your Business," [Online]. Available: https://www.iteratorshq.com/blog/4-amazing-ways-ai-personal-assistants-impact-business/. [Accessed 08 01 2021].

[2] "Virtual Assistant," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Virtual_assistant. [Accessed 08 01 2021].

[3] O. Bahcecia, "Analysis and Comparison of Intelligent Personal," 16 11 2016. [Online]. Available: https://kth.instructure.com/files/1595276/download?download_frd=1. [Accessed 08 01 2021].

[4] "Artificial Intelligence," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Artificial_intelligence. [Accessed 08 01 2021].

[5] "Artificial Intelligence," Techopedia, [Online]. Available: https://www.techopedia.com/definition/190/artificial-intelligence-ai. [Accessed 08 01 2021].

[6] "Rational Agent," Techopedia, [Online]. Available: https://www.techopedia.com/definition/33315/rational-agent. [Accessed 08 01 2021].

[7] "Intelligent Agent," Techopedia, [Online]. Available: https://www.techopedia.com/definition/28055/intelligent-agent. [Accessed 08 01 2021].

[8] W. Murphy, "The 10 principles of intelligent agent design," [Online]. Available: https://towardsdatascience.com/10-principles-of-intelligent-agent-design-a2215c4ef0d6. [Accessed 08 01 2021].

[9] "Statistical learning theory," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Statistical_learning_theory. [Accessed 08 01 2021].

[10] R. Iriondo, "Machine Learning (ML) vs. Artificial Intelligence (AI) — Crucial Differences," 15 10 2018. [Online]. Available: https://medium.com/towards-artificial-

intelligence/differences-between-ai-and-machine-learning-and-why-it-matters-1255b182fc6. [Accessed 08 01 2021].

[11] D. Fumo, "Intro To Machine Learning (IML)," 14 01 2017. [Online]. Available: https://medium.com/simple-ai/intro-to-machine-learning-iml-1-c9ea966976b6. [Accessed 08 01 2021].

[12] M. Ryan and M. Talabis, "Supervised Learning," 2015. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/supervised-learning. [Accessed 08 01 2021].

[13] "What is Machine Learning?," talend, [Online]. Available: https://www.talend.com/resources/what-is-machine-learning/. [Accessed 08 01 2021].

[14] "Natural Language Processing," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Natural_language_processing. [Accessed 08 01 2021].

[15] D. M. J. Garbade, "A Simple Introduction to Natural Language Processing," 15 10 2018. [Online]. Available: https://becominghuman.ai/a-simple-introduction-to-natural-language-processing-ea66a1747b32. [Accessed 08 01 2021].

[16] "What is Natural Language Processing (NLP)," SAS, [Online]. Available: https://www.sas.com/en_ca/insights/analytics/what-is-natural-language-processing-nlp.html. [Accessed 08 01 2021].

[17] S. Walczak and N. Cerpa, "Artificial Neural Network," 2003. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/artificial-neural-network. [Accessed 08 01 2021].

[18] "Artificial Neural Networks (ANN) and Different Types," elprocus, [Online]. Available: https://www.elprocus.com/artificial-neural-networks-ann-and-their-types/. [Accessed 08 01 2021].

[19] R. Kotorov, "Sentiment analysis and artificial intelligence: Siri, should I open this email?," 01 04 2013. [Online]. Available:

https://venturebeat.com/2013/04/01/sentiment-analysis-and-artificial-intelligence-siri-should-i-open-this-email/. [Accessed 08 01 2021].

[20] "Speech Recognition," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Speech_recognition. [Accessed 08 01 2021].

[21] "Human-Computer Interaction (HCI)," Techopedia, [Online]. Available: https://www.techopedia.com/definition/3639/human-computer-interaction-hci. [Accessed 08 01 2021].

[22] "Human-Computer Interaction (HCI)," Interaction Design Foundation, [Online]. Available: https://www.interaction-design.org/literature/topics/human-computer-interaction.

[23] S. Reehal, "Siri –The Intelligent Personal Assistant," IJARCET, 2016. [Online]. Available: http://ijarcet.org/wp-content/uploads/IJARCET-VOL-5-ISSUE-6-2021-2024.pdf. [Accessed 08 01 2021].

[24] "Siri," appleinsider, [Online]. Available: http://web.archive.org/web/20210212162022/https://appleinsider.com/inside/siri. [Accessed 08 1 2021].

[25] "How Google Assistant helps you get things done," Google, [Online]. Available: https://developers.google.com/assistant/howassistantworks/responses. [Accessed 08 01 2021].

[26] "Bringing you the next-generation Google Assistant," Google, [Online]. Available: https://www.blog.google/products/assistant/next-generation-google-assistant-io/. [Accessed 08 01 2021].

[27] "Cortana," Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Cortana. [Accessed 08 01 2021].

[28] M. Kapko, "Cortana explained: How to use Microsoft's virtual assistant for business," 07 02 2017. [Online]. Available:

https://www.computerworld.com/article/3252218/cortana-explained-why-microsofts-virtual-assistant-is-wired-for-business.html. [Accessed 08 01 2018].

[29] "Amazon Alexa," Wikipedia, [Online]. Available: https://wiki2.org/en/Amazon_Alexa. [Accessed 08 01 2021].

[30] S. L. J. P. Hyunji Chung, "Digital Forensic Approaches for Amazon Alexa Ecosystem," ResearchGate, [Online]. Available: https://www.researchgate.net/publication/318729622_Digital_Forensic_Approaches_for_Amazon_Alexa_Ecosystem.

[31] "Overview of the Alexa Voice Service (AVS) Device SDK," Amazon, [Online]. Available: https://developer.amazon.com/en-US/docs/alexa/avs-device-sdk/overview.html. [Accessed 08 01 2021].

[32] P. Newman, "Amazon is hinting at Alexa's future functionality," Business Insider, 28 05 2019. [Online]. Available: https://www.businessinsider.com/amazon-gathering-data-developing-new-alexa-features-2019-5?r=US&IR=T. [Accessed 08 01 2021].

[33] M. Rutnik, "Bixby guide: Features, compatible devices, best commands," 01 07 2019. [Online]. Available: https://www.androidauthority.com/bixby-879091/. [Accessed 08 01 2021].

[34] M. Humphries, "Report: Samsung Considers Replacing Bixby With Google Assistant," PCMag, 29 06 2020. [Online]. Available: https://www.pcmag.com/news/report-samsung-considers-replacing-bixby-with-google-assistant-android. [Accessed 08 01 2020].

[35] K. Gesling, "Technology Overview," Mycroft AI, [Online]. Available: https://mycroft-ai.gitbook.io/docs/mycroft-technologies/overview. [Accessed 08 01 2021].

[36] "Ai, io t, and voice as a natural interface," SlideShare, [Online]. Available: https://www.slideshare.net/IntelSoftware/ai-io-t-and-voice-as-a-natural-interface.

[37] D. Nield, "We pitted digital assistants against each other to find the most useful AI," 28 01 2018. [Online]. Available: https://www.popsci.com/digital-assistant-showdown/. [Accessed 08 01 2021].

[38] J. Hindy, "Google Assistant vs Siri vs Bixby vs Amazon Alexa vs Cortana – Best virtual assistant showdown!," 29 08 2019. [Online]. Available: https://www.androidauthority.com/google-assistant-vs-siri-vs-bixby-vs-amazon-alexa-vs-cortana-best-virtual-assistant-showdown-796205/. [Accessed 08 01 2021].

[39] D. B. YOFFIE, L. WU, J. SWEITZER, D. EDEN and K. AHUJA, "Voice War: Hey Google vs. Alexa vs. Siri," 07 06 2018. [Online]. Available: https://wiwi.hs-duesseldorf.de/personen/peter.scheideler/PublishingImages/Seiten/downloads/Business%20Case%20Voice%20Commerce.pdf. [Accessed 08 01 2021].

[40] J. Vlahos and W. Shih, "Chapter 1. Voice Revolution," [Online]. Available: https://journals.ala.org/index.php/ltr/article/view/7361/10126. [Accessed 08 01 2021].