**AI Virtual Assistant**

Vikhyat Jalota and Abhishek Verma

UG Student , CSE Department, Chandigarh University,

,Punjab, India

\_\_\_\_\_\_\_\_\_ Assistant Professor, CSE Department, Chandigarh University,

Punjab, India

*Abstract*— Abstract This report discusses ways in which new technology could be harnessed to create an intelligent Virtual Personal Assistant (VPA) with a focus on user-based information. It will look at examples of intelligent programs with natural language processing that are currently available, with different categories of support, and examine the potential usefulness of one specific piece of software as a VPA. This engages the ability to communicate socially through natural language processing, holding (and analysing) information within the context of the user. It is suggested that new technologies may soon make the idea of virtual personal assistants a reality.

Experiments conducted on this system, combined with user testing, have provided evidence that a basic program with natural language processing algorithms in the form of a VPA, with basic natural language processing and the ability to function without the need for other type of human input (or programming) may already be viable.

INTRODUCTION

A virtual assistant, also called AI assistant or digital assistant, is an application program that [understands natural language](https://www.techtarget.com/searchenterpriseai/feature/Natural-language-processing-drives-conversational-AI-trends) voice commands and completes tasks for the user.  
  
Such tasks, historically performed by a personal assistant or secretary, include taking dictation, reading text or email messages aloud, looking up phone numbers, scheduling, placing phone calls and reminding the end user about appointments. Popular virtual assistants currently include Amazon [Alexa](https://www.techtarget.com/whatis/definition/Alexa-Voice-Services-AVS), Apple's [Siri](https://www.techtarget.com/searchmobilecomputing/definition/Siri), Google Assistant and Microsoft's [Cortana](https://www.techtarget.com/searchenterprisedesktop/definition/Cortana) -- the digital assistant built into Windows Phone 8.1 and Windows 10.

Though this definition focuses on the digital form of virtual assistants, the term *virtual assistant*, or virtual personal assistant, is also commonly used to describe contract workers who work from home doing administrative tasks typically performed by executive assistants or secretaries.

Virtual assistants can also be contrasted with another type of consumer-facing AI programming, called smart advisers. Smart adviser programs are subject-oriented, while virtual assistants are task-oriented.

Virtual assistants are typically cloud-based programs that require internet-connected devices and/or applications to work. Three such applications are Siri on Apple devices, Cortana on Microsoft Devices and Google Assistant on Android devices.

There are also devices dedicated to providing virtual assistance. The most popular ones are available from Amazon, Google and Microsoft. To use the Amazon Echo virtual assistant, called Alexa, users call out the wake word, "Alexa." A light on the device signals to the user it is ready to receive a command, which typically involves simple language requests, such as "what is the weather today," or "play pop music." Those requests are processed and stored in Amazon's cloud.

The technologies that power virtual assistants require massive amounts of data, which feeds artificial intelligence (AI) platforms, including machine learning, [natural language processing](https://www.techtarget.com/searchenterpriseai/definition/natural-language-processing-NLP) and [speech recognition](https://www.techtarget.com/searchcustomerexperience/definition/speech-recognition) platforms. As the end user interacts with a virtual assistant, the AI programming uses sophisticated [algorithms](https://www.techtarget.com/whatis/definition/algorithm) to learn from data input and become better at predicting the end user's needs.

BACKGROUND

There already exist a number of desktop virtual assistants. A few examples of currentvirtual assistants available in market are discussed in this section along with the tasks they can provide and their drawbacks.

SIRI from Apple

SIRI is personal assistant software that interfaces with the user thru voice interface,recognizes commands and acts on them. It learns to adapt to user’s speech and thus improvesvoice recognition over time. It also tries to converse with the user when it does not identify theuser request.It integrates with calendar, contacts and music library applications on the device andalso integrates with GPS and camera on the device. It uses location, temporal, social and task based contexts, to personalize the agent behavior specifically to the user at a given point oftime.

Supported Tasks

•

Call someone from my contacts list

•

Launch an application on my iPhone

•

Send a text message to someone

•

Set up a meeting on my calendar for 9am tomorrow

•

Set an alarm for 5am tomorrow morning

•

Play a specific song in my iTunes library

•

Enter a new note

Drawback

SIRI does not maintain a knowledge database of its own and its understanding comesfrom the information captured in domain models and data models

3

ReQall

ReQall is personal assistant software that runs on smartphones running Apple iOS orGoogle Android operating system. It helps user to recall notes as well as tasks within alocation and time context. It records user inputs and converts them into commands, andmonitors current stack of user tasks to proactively suggest actions while considering anychanges in the environment. It also presents information based on the context of the user, aswell as filter information to the user based on its learned understanding of the priority of thatinformation.

Supported Tasks

•

Reminders

•

Email

•

Calendar, Google Calendar

•

Outlook

•

Evernote

•

Facebook, LinkedIn

•

 News Feeds

Drawback

 Will take some time to put all of the to-do items in – you could spend more time putting the entries in than actually doing the revision.

4

1.2 OBJECTIVES

Main objective of building personal assistant software (a virtual assistant) is usingsemantic data sources available on the web, user generated content and providing knowledgefrom knowledge databases.

The main purpose of an intelligent virtual assistant is to answerquestions that users may have. This may be done in a business environment, for example, onthe business website, with a chat interface. On the mobile platform, the intelligent virtualassistant is available as a call-button operated service where a voice asks the user “What can Ido for you?” and then responds to verbal input.Virtual assistants can tremendously save you time. We spend hours in online researchand then making the report in our terms of understanding. JIA can do that for you. Provide atopic for research and continue with your tasks while JIA does the research. Another difficulttask is to remember test dates, birthdates or anniversaries. It comes with a surprise when youenter the class and realize it is class test today. Just tell JIA in advance about your tests andshe reminds you well in advance so you can prepare for the test.One of the main advantages of voice searches is their rapidity. In fact, voice is reputedto be four times faster than a written search: whereas we can write about 40 words per minute,we are capable of speaking around 150 during the same period of time15. In this respect, theability of personal assistants to accurately recognize spoken words is a prerequisite for them to be adopted by consumers.

5

1.3

PURPOSE, SCOPE AND APPILCABILITY

Purpose

Purpose of virtual assistant is to being capable of voice interaction, music playback,making to-do lists, setting alarms, streaming podcasts, playing audiobooks, and providingweather, traffic, sports, and other real-time information, such as news. Virtual assistantsenable users to speak natural language voice commands in order to operate the device and itsapps.There is an increased overall awareness and a higher level of comfort demonstratedspecifically by millennial consumers. In this ever-evolving digital world where speed,efficiency, and convenience are constantly being optimized, it’s clear that we are movingtowards less screen interaction.

Scope

Voice assistants will continue to offer more

individualized

 experiences as they get better at differentiating between voices. However, it’s not just developers that need to addressthe complexity of developing for voice as brands also need to understand the capabilities ofeach device and integration and if it makes sense for their specific brand. They will also needto focus on maintaining a user experience that is consistent within the coming years ascomplexity becomes more of a concern. This is because the visual interface with voiceassistants is missing. Users simply cannot see or touch a voice interface.

Applicability

The mass adoption of artificial intelligence in users’ everyday lives is also fueling theshift towards voice. The number of IoT devices such as smart thermostats and speakers aregiving voice assistants more utility in a connected user’s life. Smart speakers are the numberone way we are seeing voice being used. Many industry experts even predict that nearly everyapplication will integrate voice technology in some way in the next 5 years.The use of virtual assistants can also enhance the system of IoT (Internet of Things).Twenty years from now, Microsoft and its competitors will be offering personal digitalassistants that will offer the services of a full-time employee usually reserved for the rich andfamous.

*Technologies*

1. Let’s say you want to create your own virtual assistant like Siri. How would you go about making it? Your first and possibly least difficult option would be to integrate Siri into your application directly. Siri, Cortana, and Google Assistant are three well-known examples of AI assistants that many developers integrate into their applications. In 2016, Apple Inc. announced SiriSDK, a development kit that allowed programmers to integrate functions of their own apps as “Tasks” that Siri could perform. SiriSDK uses “Intents” as labels for user intentions and associates Intents with custom classes and properties.
2. If your company doesn’t want to rely on existing AI assistant options, you’d need an expert team of AI engineers to build your own solution. Let’s dive into the key AI technologies behind intelligent virtual assistants.
3. SPEECH-TO-TEXT (STT) AND TEXT-TO-SPEECH (TTS)
4. If we’re talking about intelligent virtual assistants, they at the very least require Speech-to-text (STT) and Text-to-speech (TTS) capabilities.
5. Speech-to-text allows apps to convert human speech into digital signals. This is how it works. When you speak, you create a series of vibrations. Using an analog-to-digital converter (ACD) the software converts them into digital signals and extracts sounds, then segments them and matches them to existing phonemes. Phonemes are the smallest unit of a language capable of distinguishing the sound shells of different words. Based on complex mathematical models, the system compares these phonemes with individual words and phrases and creates a text version of what you said.
6. Text-to-speech does the opposite. This technology translates text into voice output. TTS is a computer simulation of human speech from text using machine learning. The system must go through three steps to convert text to voice. First, the system needs to convert text to words, then perform phonetic transcription and then convert transcription to speech.
7. Speech-to-text (STT) and Text-to-speech (TTS) are used in virtual assistant technology to ensure smooth and efficient communication between users and applications. To turn a basic voice assistant with static commands into a proper AI assistant, you also need to give the program the ability to interpret user requests with intelligent tagging and heuristics.
8. COMPUTER VISION (CV)
9. Computer vision is an AI technology that extracts meaningful information from visual inputs like digital images or videos. CV is an integral part of creating visual virtual assistants. These assistants can respond with creator-generated videos, not just sounds, which greatly enhances the user experience.
10. Computer vision allows the system to recognize body language which is a significant part of communication. Visual virtual assistants powered by this technology use a camera that stores data and utilizes real-time face detection to catch when someone is looking at the screen, this sends a signal to the rest of the system, which converts the user’s speech into text.
11. CV can also greatly increase the accuracy of speech recognition by comparing what the user has said verbally to the movement of the user’s face and mouth.

### NOISE CONTROL

Noise control is another critical feature for voice assistant accuracy. While many smartphones include software-based noise control and suppression features, you can’t count on this being the case for all of your customers. To compensate for a lack of onboard noise suppression software, top-shelf Bluetooth headsets also include hardware noise suppression, but once again there are no guarantees that your AI assistant is going to be able to detect what your customers are

.saying in a busy train car. By integrating in-house noise control packages, you minimize the risk of misunderstanding voice queries.

### SPEECH COMPRESSION

Your AI assistant will also need to at least temporarily store voice information for processing unless you’re going to fill up the customer’s hard drive locally with voice data. Speech compression is critical, but developers toe a fine line with compression. It’s possible to compress an audio file so much that substantial amounts of fidelity are lost, making it difficult or impossible to recover what was said during the processing. Compression technology is rapidly improving, but when developing your voice assistant, audio codecs and compression solutions merit a thorough investigation.

### NATURAL LANGUAGE PROCESSING (NLP)

Once you have the voice data, the AI assistant needs to process and interpret the data with Natural Language Processing (NLP) and then execute the requested command. NLP simplifies the speech recognition process. While many AI kits are pre-trained on countless hours of voice samples, you’d still need enough data from customers to adjust for precision for your use cases. If your AI assistant is going to respond verbally, you’ll need speech synthesis such as Google Cloud’s top-of-the-line solution, which produces realistic and clear voices.

However, speech processing is not enough to derive a person’s actual intent and maintain a normal conversation.

The request still needs to be interpreted right, and that’s when Natural Language Understanding comes into play.

### NATURAL LANGUAGE UNDERSTANDING (NLU)

Natural Language Understanding (NLU) is a different approach to Natural Language Processing and is considered by most computer and data scientists to be a subtopic of NLP. While NLP methods parse, tokenize, and standardize natural language into a standardized structure for command processing, NLU interprets the natural language without standardizing it and derives meaning from queries by identifying the context. In short terms, NLP processes grammar, structure, and compensates for the user’s spelling errors while NLU examines the actual intent behind the query.

### NATURAL LANGUAGE GENERATION (NLG)

Natural language generation produces natural language output. Thanks to this technology, users receive a human-like response from virtual assistants and chatbots. Models and techniques used for NLG can be different and depend on the goals of the project and development approaches. One of the simplest approaches is a template system that can be used for texts that have a predefined structure and require only a small amount of data to be filled in. This approach allows such gaps to be automatically filled in with data retrieved from a row in a spreadsheet, a record in a database table, and so on.

Another approach is dynamic NLG which does not require the developer to write code for each edge case and enables the system to react on its own. This is a more advanced type of natural language generation that relies on machine learning algorithms.

### DEEP LEARNING

Chatbots that utilize text-based responses only are substantially less complicated than voice assistants. Because you don’t have to then convert speech into text for interpretation, you remove a lot of tooling from the equation when constructing a chatbot. Next-gen text generation such as GPT-3 is capable of producing not only responses to basic queries, but entire news stories from a “seed”. Deep learning makes it happen.

Virtual assistants and chatbots powered by deep learning algorithms learn from their data and human-to-human dialogue. Chatbots that utilize deep learning examine existing

interactions between customers and support staff and create paired messages and responses and compensate for the user’s typos and grammatical errors.

### AUGMENTED REALITY (AR)

Augmented reality allows you to overlay 3D objects in the real world for an immersive experience. AR-based mobile chatbots and AR avatars are great examples of using this technology. For example, Arcade created a mobile AR Avatar Chatbot called Miss Perkins for the Ragged School Museum of East London. This assistant serves as a guide for museum visitors and quizzes them ensuring an interactive user experience.

Another example of an intelligent AR chatbot was developed for the Vienna Museum of Technology. The creators also used [mobile AR](https://mobidev.biz/services/augmented-reality). The functionality of the chatbot includes conducting tours and answering user questions about specific display items in the text, images, videos, and audio formats.

The rise of the Metaverse and VR technology leads to the logical conclusion of virtual assistants: 3D AI avatars. Combined with artificial intelligence, AR virtual assistants become more functional, bypassing the limitations of existing AR tools. For example, deep learning allows IVAs to capture user behavior in real-time to drive neural networks that automatically train and improve virtual assistant performance.

### GENERATIVE ADVERSARIAL NETWORKS (GANS)

Being algorithmic architectures that use neural networks, Generative Adversarial Networks create new instances of synthetic data. GANs consist of real image samples and generators fed into discriminators to generate a realistic 3D face for AI avatars and 3D assistants.

The technology has been utilized in many video games and other products to create true-to-life human figures. GANs can also be utilized to turn still images into full-depth 3D images. Perhaps the most advanced integration of AI avatars so far is Nvidia’s Omniverse Avatar Project Maxine, which creates a photorealistic real-time animation of a human face speaking a text-to-speech sample.

1. *Advantages*

1 ) They will stop single direction discussions and exhausting repetitive answers by answering FAQ. If you think these metadata questions are trivial chat bots reduce customer service costs by 30% when deployed in commercial sales marketing funnels. There are dozens of articles and academic research efforts validating the benefits to teachers when they use chat bots to support basic question answering – and they are right.

2) They can quickly devour a big chunk of information and take care of difficult issues so they can be understood—right away. Through natural language processing (NLP) algorithms Chat bots can answer complex questions on the first try that can take several minutes of explanation and repetition to disambiguate for a human. NLP is a complex process that yields simplicity in it’s outcome.

3 ) Easy to use, available 24/7 and fun. Call your instructor at 3 in the morning to find out your next due assignment – I dare you.

4) Students can ask questions privately that they might be embarrassed to ask face to face. Instructors understand the importance of impression management to today’s students who swim in social media presentation like water. How many bits of information are lost if a student thinks they have a “dumb” question?

5) Student impulse to ask questions may be increased. If a student can ask a question at any time without disrupting or interfering with anyone else – do you think they might ask more questions?

6) They give extensive criticism to coaches. Students can say whatever they want to their bot without fear of offending anyone.

7) Because the bots are always in development the discussion is always fresh and timely. Some teachers get into a routine – and appear dated or out of touch to their students as a result

8 ) They quickly can give visual answers – and a picture is worth a thousand words. Imagine being able to immediately show a picture related to the required answer.

9) They are intended to communicate like a human. In this way they can draw in students with cleverness and insight. Dialog design for a bot can be playful unique and compelling.

10) Students realize they can return and pose extra inquiries at anytime. The bot doesn’t become impatient with a topic simple becasue it has been previously discussed. They never say “Haven’t I already discussed with with you?” or “I have other students to attend to right now”

11) They can assist students with language difficulties as they can look at the responses and answers a few times, without feeling awkward. Even the best Instructors can become frustrated by 2nd language communication difficulties.

12) They deal with cell phone popularity by appearing on them. A bot never says “Students please put away your phones.”

RESULT

Usually, user needs to manually manage multiple sets of applications to complete onetask. For example, a user trying to make a travel plan needs to check for airport codes fornearby airports and then check travel sites for tickets between combinations of airports toreach the destination. There is need of a system that can manage tasks effortlessly.We already have multiple virtual assistants. But we hardly use it. There are number of people who have issues in voice recognition. These systems can understand English phrases but they fail to recognize in our accent. Our way of pronunciation is way distinct from theirs.Also, they are easy to use on mobile devices than desktop systems. There is need of a virtualassistant that can understand English in Indian accent and work on desktop system.When a virtual assistant is not able to answer questions accurately, it’s because it lacksthe proper context or doesn’t understand the intent of the question. Its ability to answerquestions relevantly only happens with rigorous optimization, involving both humans andmachine learning. Continuously ensuring solid quality control strategies will also help managethe risk of the virtual assistant learning undesired bad behaviors. They require large amount ofinformation to be fed in order for it to work efficiently.Virtual assistant should be able to model complex task dependencies and use thesemodels to recommend optimized plans for the user. It needs to be tested for finding optimum paths when a task has multiple sub-tasks and each sub-task can have its own sub-tasks. In sucha case there can be multiple solutions to paths, and the it should be able to consider user preferences, other active tasks, priorities in order to recommend a particular plan.

REQUIREMENT SPECIFICATION

Personal assistant software is required to act as an interface into the digital world byunderstanding user requests or commands and then translating into actions orrecommendations based on agent’s understanding of the world.JIA focuses on relieving the user of entering text input and using voice as primarymeans of user input. Agent then applies voice recognition algorithms to this input and recordsthe input. It then use this input to call one of the personal information managementapplications such as task list or calendar to record a new entry or to search about it on searchengines like Google, Bing or Yahoo etc. Focus is on capturing the user input through voice,recognizing the input and then executing the tasks if the agent understands the task. Softwaretakes this input in natural language, and so makes it easier for the user to input what he or shedesires to be done.Voice recognition software enables hands free use of the applications, lets users toquery or command the agent through voice interface. This helps users to have access to theagent while performing other tasks and thus enhances value of the system itself. JIA also haveubiquitous connectivity through Wi-Fi or LAN connection, enabling distributed applicationsthat can leverage other APIs exposed on the web without a need to store them locally.Virtual assistants must provide a wide variety of services. These include:

•

Providing information such as weather, facts from e.g. Wikipedia etc.

•

Set an alarm or make to-do lists and shopping lists.

•

Remind you of birthdays and meetings.

•

Play music from streaming services such as Saavn and Gaana.

•

Play videos, TV shows or movies on televisions, streaming from e.g. Netflix orHotstar.

•

Book tickets for shows, travel and movies.

Feasibility Study

Feasibility study can help you determine whether or not you should proceed withyour project. It is essential to evaluate cost and benefit. It is essential to evaluate cost and benefit of the proposed system. Five types of feasibility study are taken into consideration.1.

Technical feasibility:

 It includes finding out technologies for the project, bothhardware and software. For virtual assistant, user must have microphone to conveytheir message and a speaker to listen when system speaks. These are very cheap now adays and everyone generally possess them. Besides, system needs internet connection.While using JIA, make sure you have a steady internet connection. It is also not anissue in this era where almost every home or office has Wi-Fi.2.

Operational feasibility:

 It is the ease and simplicity of operation of proposed system.System does not require any special skill set for users to operate it. In fact, it isdesigned to be used by almost everyone. Kids who still don’t know to write can readout problems for system and get answers.3.

Economical feasibility:

 Here, we find the total cost and benefit of the proposedsystem over current system. For this project, the main cost is documentation cost. Useralso would have to pay for microphone and speakers. Again, they are cheap andavailable. As far as maintenance is concerned, JIA won’t cost too much.4.

Organizational feasibility:

 This shows the management and organizational structureof the project. This project is not built by a team. The management tasks are all to becarried out by a single person. That won’t create any management issues and willincrease the feasibility of the project.5.

Cultural feasibility:

 It deals with compatibility of the project with culturalenvironment. Virtual assistant is built in accordance with the general culture. The project is named JIA so as to represent Indian culture without undermining local beliefs.This project is technically feasible with no external hardware requirements. Also it issimple in operation and does not cost training or repairs. Overall feasibility study of the project reveals that the goals of the proposed system are achievable. Decision is taken to proceed with the project.

4

HARDWARE AND SOFTWARE REQUIREMENTS

The software is designed to be light-weighted so that it doesn’t be a burden on themachine running it. This system is being build keeping in mind the generally availablehardware and software compatibility. Here are the minimum hardware and softwarerequirement for virtual assistant.

Hardware:

•

Pentium-pro processor or later.

•

RAM 512MB or more.

Software:

•

Windows 7(32-bit) or above.

•

Python 2.7 or later

•

Chrome Driver

•

Selenium Web Automation

•

SQLite

Acknowledgement

In completing this project report on project titled

VIRTUAL ASSISTANT

, I had totake the help and guideline of a few respected people, who deserve my greatest gratitude.The completion of this project report gives me much Pleasure. I would like to showmy gratitude to

Prof. Miss Liza

for giving me a good guideline for project through out numerous consultations. I would also like to expand my deepest gratitude to all those who have directly and indirectly guided us in writing this project report. Many people, especially my classmates and friends themselves, have made valuable comments and suggestions on this proposal which gave me inspiration to improve my project. Here I thank all the people for their help directly and indirectly to complete this projectreport.The author

REFERENCE AND BIBLIOGRAPHY

Websites referred

www.stackoverflow.com

www.pythonprogramming.net

www.codecademy.com

www.tutorialspoint.com

www.google.co.in

Books referred

Python Programming - Kiran Gurbani

Learning Python - Mark Lutz

YouTube Channels referred

CS Dojo

edureka!

Documents referred

Designing Personal Assistant Software for Task Management using SemanticWeb Technologies and Knowledge Databases- Purushotham Botla

Python code for Artificial Intelligence: Foundations of Computational Agents- David L. Poole and Alan K. Mackworth