

**AN ARTIFICIAL INTELLIGENCE BASED BENGALI VOICE ASSISTANT  
'SAATHI'**

**BY**

**ALEX SARKER**

**ID: 183-15-11910**

This Report Presented in Partial Fulfillment of the Requirements for the  
Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

**Shah Md Tanvir Siddiquee**

Assistant Professor

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**JANUARY 2023**

## APPROVAL

This Project/internship titled “**An Artificial Intelligence Based Bengali Voice Assistant 'Saathi'**”, submitted by Alex Sarker, ID No: 183-15-11910 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on *date*.

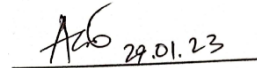
## BOARD OF EXAMINERS



**Dr. Touhid Bhuiyan**  
**Professor and Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

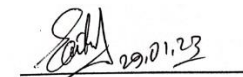
**Chairman**



**Arif Mahmud**  
**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

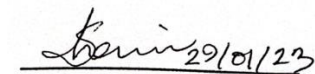
**Internal Examiner**



**Saiful Islam**  
**Assistant Professor**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



**Dr. Shamim H Ripon**  
**Professor**

Department of Computer Science and Engineering  
East West University

**External Examiner**

## DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Shah Md Tanvir Siddiquee, Assistant Professor, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

### Supervised by:



**Shah Md Tanvir Siddiquee**

Assistant Professor

Department of CSE

Daffodil International University

### Submitted by:



---

**Alex Sarker**

ID: 183-15-11910

Department of CSE

Daffodil International University

## ACKNOWLEDGEMENT

First of all, we would like to give our robust thanks and gratitude to almighty Allah for the blessings to make it possible to finished the final year thesis successfully.

I am really thankful to **Shah Md Tanvir Siddiquee, Assistant Professor**, Department of Computer Science & Engineering, Daffodil International University, Dhaka. He helped me spontaneously in the field of Machine Learning (ML), Internet of Things (IoT)-based research with his deep knowledge. Finally, I have finished my work on ‘Voice Assistant’. His energetic supervision, worthy instruction, unlimited patience help me to correct at all stage have made it easier to finished this project.

At last, I give thanks to all the good wishers surrounding me who helped me and inspired me to make the task complete.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

## **ABSTRACT**

Voice assistants are becoming more and more commonplace in real-world applications. Voice recognition technology is currently being incorporated into a diverse assortment of products, such as mobile applications and smart speakers found in consumers' homes. Additionally, voice assistants are rapidly turning into an important component of our day-to-day lives. A significant number of people who speak Bengali as their native language are illiterate, and thus have trouble using computers because the controls are in English. People who have trouble communicating in English may have an easier time using a computer or smartphone if they are able to give instructions in their native language of Bengali. A Bengali virtual assistant may be the solution to this problem. In this article, a Bengali virtual assistant known as "Saathi" is constructed. In order to understand the commands given in Bengali, "Saathi" makes use of the CNN model. The CNN employs a spectrogram to determine the nature of the orders and then carries out the corresponding responses.

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	i
Declaration	ii
Acknowledgement	iii
Abstract	iv
List of Tables	vii
List of Figures	viii
<b>CHAPTER</b>	<b>PAGE</b>
<b>CHAPTER 1: Introduction</b>	<b>1-6</b>
1.1 Introduction	1
1.2 Motivation	1-2
1.3 Rationale of the Study	2-3
1.4 Research Questions	3
1.5 Expected Output	4
1.6 Project Management and Finance	4
1.7 Report Layout	4-6
<b>Chapter 2: Background</b>	<b>7-12</b>
2.1 Preliminaries/Terminologies	7-8
2.2 Related Works	8-9
2.3 Comparative Analysis and Summary	10
2.4 Scope of the Problem	11
2.5 Challenges	11-12

<b>Chapter 3: Research Methodology</b>	<b>13-22</b>
3.1 Research Subject and Instrumentation	13-14
3.2 Data Collection Procedure/Dataset Utilized	14-15
3.3 Statistical Analysis	16
3.4 Proposed Methodology/Applied Mechanism	17-22
3.4.1 Audio Pre-Processing	17-18
3.4.2 Prepare the Train and Test dataset	18
3.4.3 Build CNN model	18-19
3.4.4 CNN architecture	19-20
3.4.5 Proposed Model	20-22
3.5 Implementation Requirements	22
<b>Chapter 4: Experimental Results and Discussion</b>	<b>23-35</b>
4.1 Experimental Setup	23
4.2 Experimental Results & Analysis	24-33
4.3 Discussion	34-35
<b>Chapter 5: Impact on Society, Environment and Sustainability</b>	<b>36-42</b>
5.1 Impact on Society	36-38
5.2 Impact on Environment	39-40
5.3 Ethical Aspects	40-41
5.4 Sustainability Plan	42
<b>Chapter 6: Summary, Conclusion, Recommendation and Implication for Future Research</b>	<b>43-45</b>
6.1 Summary of the Study	43-44
6.2 Conclusions	45
6.3 Implication for Further Study	45
<b>References</b>	<b>46</b>
<b>PLAGIARISM REPORT</b>	<b>47</b>

## **LIST OF TABLES**

### **TABLE NO**

### **PAGE NO**

Table 3.2.1: there are presented 3 variations of 13 commands.

15



## LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.3.1: Graphical representation of User's intend to use Saathi	16
Figure 3.4.4.1: Overall Process of Bengali Voice Assistant -Saathi	20
Figure 3.4.5.1: Working Procedure of Voice Assistant - Saathi	21
Figure 4.2.1: Output of Welcome command using Saathi-(VA)	24
Figure 4.2.2: Output of time asking command using Saathi-(VA)	25
Figure 4.2.3: Output of date asking command using Saathi-(VA)	25
Figure 4.2.4: Output of playing music using Saathi-(VA)	26
Figure 4.2.5: Output of playing music through youtube using Saathi	26
Figure 4.2.6: Output of Information from wikipedia using Saathi	27
Figure 4.2.7: Output of recent Address using Saathi-(VA)	28
Figure 4.2.8: Output of Open Microsoft using Saathi-(VA)	28
Figure 4.2.9: Output of Open Chrome using Saathi-(VA).	29
Figure 4.2.10: Output of Open YouTube using Saathi-(VA)	30
Figure 4.2.11: Output of Open Google Search Engine using Saathi	30
Figure 4.2.12: Output of Captured Photo from Camera using Saathi	31
Figure 4.2.13: Output of saved photo in Folder.	31
Figure 4.2.14: Output of Praising Saathi using Saathi-(VA)	32
Figure 4.2.15: Output of search command which is not included using Saathi-(VA)	32
Figure 4.2.16: Output of search command by Saathi-(VA)	33
Figure 4.2.17: Output command of Saathi-(VA) for turn off.	33

# CHAPTER 1

## INTRODUCTION

### 1.1 Introduction

Only in recent years we have been able to observe major changes in the manner in which users connect with one another and the experiences that users have. This new technique of communicating with technological gadgets makes lexical communication a new ally to this technology. We already use them for various jobs, such as turning lights on and off and playing music through streaming apps like Music and Spotify, among other things. Historically, the term "virtual assistant" was used to refer to web-based professionals who offered additional services. [1] The role of a voice can be broken down into three stages, which are as follows: Voice assistant will have all of its features completely developed, including text-to-speech, text-to-intention, intention-to-action, and voice assistant, which will improve the present range. [2] Voice assistants are not to be confused with virtual assistants, which are individuals that work on a part-time basis and are therefore able to fulfill a wide variety of responsibilities. Because they are powered by artificial intelligence, voice assistants can now predict our needs and respond accordingly. Voice-based search is going to be the future for the next generation of people, where users are going to be most dependent on voice assistants for everything they need. AI-based voice assistants can be useful in a variety of fields, including IT helpdesks, home automation, HR-related tasks, and voice-based searches, among other things. In this proposition, we have constructed the AI-based voice assistant named "Saathi" in Bengali "সাথী" that is capable of performing all of these duties without causing any inconvenience.

### 1.2 Motivation

The primary Motivation of our Research is to help people in need in general. In addition to this, we planned to make use of technologies that involve AI. As a consequence of this, we were discussing the various ways in which AI technology such as Voice Assistant could be used to assist regular people. When compared to other languages with a smaller population, such as German or French, the amount of study that has been

done on Bengali speech recognition is minimal [3]. Because the majority of people in Bangladesh live in the countryside and are not particularly tech- specialist, it is much simpler for them to provide voice commands in Bengali in order to use a mobile phone or a computer. People who are disabled, elderly, or illiterate can have an easier time operating a mobile device or computer if it supports voice command. Therefore, using a virtual assistant makes it significantly simpler to control a smart phone or computer by merely issuing voice commands, and this can be done without the need for any additional understanding. Users of a smart phone or computer can interact with a virtual environment created by virtual help in the same way that they would interact with another person.

As a direct consequence of this, there are an aggregate of three hundred characters speaking Bengali. There is not a suitable dataset available in the Bengali language for research, despite the fact that Bengali is both one of the most commonly spoken languages and a rich language. In most instances, researchers will collect their corpus in order to conduct study on it. In addition to this, the Bengali corpus that already exists does not have very good labeling. Because of this, research on the Bengali virtual assistant is made more difficult due to a lack of a benchmark corpus, morphological analysis, and tagged data [4].

### **1.3 Rationale of the Study**

Some of the most dreaded aspects of working in an office are the mundane tasks of figure crunching, sorting through files, and answering calls. Most redundant tasks in the office have been removed thanks to technological advancements and new ideas, but a few remain. This effort relied heavily on automation, with AI serving as the primary motivating factor. AI has been progressively reducing redundancy in previously cumbersome processes across the board, from customer service to daily administration. Now, a new partnership between AI and Voice Technology promises to skyrocket productivity. The goal of our project is to help both normal and blind people to live their life more comfortably.

The main objectives of our projects are,

- To find out about anything, such the current weather, the latest news, or the traffic situation in your area.
- Helping the blind people by searching anything on google or any other search engine anything just by commanding.
- Helping the people in their every work by simply saying some command.
- To find anything anywhere anytime on internet by commanding some word.

## **1.4 Research Questions**

The research questions for a study or research project are the questions that the study or project strives to find answers to in order to construct a research study. The findings of the review indicate that the answer to this problem is most frequently offered by the interpretation of data. As a direct result of this, the following is a list of significant questions that have served as the driving force behind this investigation:

- What is AI voice Assistant?
- How can we active AI voice assistant?
- What is the main purpose of Voice Assistant?
- What happens when people command something to voice assistant?
- What percentage of people use AI base voice assistant in Bangladesh?
- How can we collect data?
- What method can be applied to collect the data?
- What kind algorithms be used in order to execute our research project?
- How much data is needed to complete our research?
- What kind of device is needed to complete the model?
- How much training data is need to train the model?
- How we can command to the Voice Assistant?
- Are there any limitations of this project?
- How can we overcome the limitations this project?

## **1.5 Expected Outcome**

Even though speech technology has been around for quite some time, its adoption rate has only very recently witnessed a large jump in the most recent few years. This is despite the fact that voice technology has been around for quite some time. Because there are so many positive aspects associated with the application of voice technology, it is nearly impossible to list all of them in this context. The expected outcomes of our project are:

- The Internet is a wealth of knowledge, and we can learn just about anything there, from the local weather to the breaking news to the traffic situation.
- At any time, we can simply request the music that best complements our current state of mind.
- By issuing commands like "turn on the television," they can communicate with other smart devices and carry out predetermined actions.
- You may check your account balance, view recent transactions, and even order banking items like debit cards from some institutions' online portals.
- It's a simple method of keeping track of time that doesn't require a phone, a watch, or a planner.

## **1.6 Project Management and Finance**

In order to complete our research, We Collect data from different ages people with different accent of their local languages too. For doing this we need a recorder and a Microphone to collect the best data. It cost almost 2000 BDT. We Also go to the University to collect the data, and it takes total around 200 BDT. We have built our projects in our computer. And it's need Internet On all the time. It cost total around 1000 BDT. So total amount we spend to build our project is almost 3200 BDT.

## **1.7 Report Layout**

This research-based project is divided into six distinct phases. Different perspectives are offered in separate chapters throughout the book. Each chapter is then further divided into multiple specific subsections, all of which are presented in an approachable manner. Here's a rundown of the report's contents:

## **Chapter 1**

It presented the initiative and talked about its motivations, research questions, and anticipated outcomes. What we've covered so far: 1.1 Introduction, 1.2 Objective, 1.3 Motivation, 1.4 Expected outcome, 1.5 Research Questions and 1.6 Research Layout

## **Chapter 2**

It provides a summary of prior efforts in this area. The ramifications of the authors' decision to limit the scope of this investigation are presented later in Chapter 2. Among the many things we've talked about are: 2.1 Preliminaries/Terminologies, 2.2 Related Works, 2.3 Comparative Analysis and Summary, 2.4 Scope of the Problem and 2.5 Challenges.

## **Chapter 3**

It is important to the study's theoretical debate. To examine the theoretical part of the investigation, this work included enhancing the existing statistical methodologies. This chapter also demonstrates the inner workings of deep learning. Here are some of the issues that will be discussed: 3.1 Research Subject and Instrumentation, 3.2 Data Collection Procedure/Dataset Utilized, 3.3 Statistical Analysis, 3.4 Proposed Methodology/Applied Mechanism, 3.4.1 Audio Pre-processing, 3.4.2 Prepare the train and test dataset, 3.4.3 Build CNN Model, 3.4.4 CNN architecture, 3.4.5 Proposed Model, and 3.5 Implementation Requirements.

## **Chapter 4**

It Provides experimental data together with an analysis and discussion of the outcomes. Here, we showcase some of the experimental photographs we captured as the project evolved. 4.1 Experimental Setup, 4.2 Experimental Results & Analysis and 4.3 Discussion has been discussed.

## **Chapter 5**

Provides information that should be included in the whole report on the enterprise, not just the proposal. Limitations of our work are discussed to provide a foundation for

additional investigation. Following is some of the points we've covered thus far: 5.1 Impact on Society, 5.2 Impact on Environment, 5.3 Ethical Aspects and 5.4 Sustainability Plan.

## **Chapter 6**

Discussion of the study's findings and directions for further research are included. Here is a summary of what we've covered so far: 6.1 Summary of the Study, 6.2 Conclusions and 6.3 Implication for Further Study.

## **CHAPTER 2**

### **BACKGROUND**

#### **2.1 Preliminaries**

Voice assistants are a new category of product that is being promoted by tech giants like Apple, Amazon, and Google. These devices are also known as intelligent personal assistants or linked speakers. Speech recognition technology that is able to comprehend natural language forms the basis of these assistants. In addition to making it possible to access information via the use of voice synthesis, they make it possible to conduct a search through the use of a spoken command supplied by the user. It is already plainly clear that in the not-too-distant future, each and every appliance in a home will be connected to the internet and will listen to voice instructions issued by Google and Amazon. This will happen in a relatively short amount of time. The fact that the appliances in your home are connected to the internet makes it possible for you to control that equipment even when you aren't in the same room as them. Additionally, digital voice assistants are becoming more mobile and regularly accompanies us wherever we go. This trend is expected to continue. Both the Google Assistant and the Amazon Alexa virtual assistants have already been integrated into a variety of vehicles that have been manufactured by their respective companies. For instance, Amazon has developed in-ear headphones that enable users to converse with Alexa by merely speaking to her. These headphones may be purchased on Amazon.com. If you have a linked house and a mobile digital assistant, you will be able to control the appliances in your home no matter where you are in the globe. This is because you will be able to access the internet through your mobile device.

Imagine for a second that your shift is over and you are about to go home. You need to get back to your place, but getting there will take another half an hour. You are starving. You are suddenly overcome with the desire to eat pizza. It seems to come out of nowhere. Simply addressing the digital assistant that is built into your vehicle will cause it to raise the temperature of the oven to 200 degrees without requiring any more action from you. If you do things in this manner, you won't have to wait about for as long after



you get back to your house. Any electronic equipment that is connected to the internet will be under the command of a plain verbal command in the not-too-distant future.

These days, a lot of people place their orders for things online. We are all familiar with the process: first, you access the site or app, then you choose your items, and last, the order is placed with only a few clicks. In the meantime, conducting business via the internet has developed into a channel that is not only convenient but also trustworthy when doing so. In the subsequent phase, digital assistants have the potential to initiate us into the realm of automated commerce, also known as a-commerce.

The majority of households adhere to a consistent consumption pattern. The majority of the foods, beverages, and items that we consume on a weekly basis are the same ones that we always use. After recognizing that pattern for the first time, the computer can then construct an algorithm to make these kinds of purchases automatically. After that, all that will be required is a single command to be given to your virtual assistant in order to order a new supply of these products whenever the virtual assistant determines that it is necessary to do so. In this approach, the computer handles all purchases that are considered to be routine. A-commerce is so totally automated that you don't even have to give it a second thought, which frees up more time for you to focus on the things that are genuinely important to you in life.

## **2.2 Related Works**

A voice assistant, which is a subset of artificial intelligence that is enabled with voice recognition, is no longer merely a character found in science fiction films. At the moment, voice recognition technology is being incorporated into a wide range of goods, including mobile applications and smart speakers found in consumers' homes. In addition, voice assistants are quickly becoming an essential part of our day-to-day life. Voice assistant personalities can have an effect on everyday interactions with our surroundings, much like human personalities do. Human personalities define the way we interact with the world around us.

According to the findings of Poushneh and Atieh's research, consumers are given the ability to take control of their voice interactions with a VA, concentrate on their voice

interactions, and engage in exploratory behavior when the voice interaction involves a virtual assistant (VA) that incorporates functional intelligence, sincerity, and creativity. Exploratory activity on the part of customers ultimately results in consumer happiness and a willingness on the part of consumers to continue using voice assistants.[5]

AI base Voice assistant interact with people more easily. It can be a personal assistant who can do everything what the people need with a short time. As part of the research, Nasirian, Farzaneh, Mohsen Ahmadian, and One-Ki Daniel Lee constructed a conceptual model that incorporates a newly proposed system quality construct, which we refer to as interaction quality. We believe this model is able to more accurately represent the adoption of AI-based technologies.[6]

Voice assistants are becoming increasingly common in everyday family activities, such as listening to music, searching for information, asking for jokes, and participating in games together. However, very little study has been done to show how such technology affects the dynamic family interactions that occur within the house throughout time. Over the course of three weeks, a field study was carried out in six different family homes. Every family's Alexa was equipped with a variety of abilities, and they were encouraged to make use of them. The findings indicated that there were shifts in usage throughout time. To begin, the behaviors that contributed most to family cohesion were family rituals and behaviors. At the conclusion of the research project, it was discovered that the skills motivate distinct patterns of family interaction. These patterns include increased collaboration to manage Alexa as well as scaffolding of children's interactions with Alexa, which is necessary due to the fact that users' understanding of Alexa's capabilities and limitations varies depending on their developmental stage.

Beirl, Diana, Y. Rogers, and Nicola Yuill explore the many engagement patterns that can be supported by voice assistants, as well as the possible learning possibilities that these present[7]. In 2018 The researcher Tuzovic, Sven, and Stefanie Paluch said that, Voice-controlled digital assistants and other devices are expected to increase in popularity over the next few years.[8] According to Tractica, Forecasts for the global market for voice and speech recognition software range from about \$2 billion in 2020 to nearly \$7 billion in 2025.[9]

## 2.3 Comparative Analysis and Summary

Attacks that fool a voice assistant into executing dangerous behaviors are becoming increasingly concerning since they represent a threat to users' security, privacy, and even physical safety as firms add new features to voice assistants. However, designers face difficulties in identifying, understanding, and mitigating security threats against voice assistants due to the wide variety of attacks and isolated responses in the literature.

Yan, Chen, et al. offers a comprehensive look at the threats facing voice assistants and the solutions put out to defend against them. they classify known countermeasures based on defensive techniques from the standpoint of a system designer, and they organize a broad category of relevant but seemingly unrelated attacks by the vulnerable system components and attack methods. they give a qualitative evaluation of available countermeasures in terms of implementation cost, usability, and security, and they propose practical solutions to help designers develop defense based on their needs. Their hope is that this will encourage further study in this exciting, rapidly developing field and lead to more trustworthy voice assistants.[10]

As a result of imitating human behavior by answering in full phrases, virtual assistants are unable to reach their full potential as a practical instrument. This limits the design possibilities available to developers. Haas, G., Rietzler, M., Jones, M., & Rukzio, E. developed a virtual assistant that can respond to questions using one of three formats: a full-sentence baseline, one of two other short keyword-based response formats, or both. In a user research, 72 individuals communicated with their virtual assistant (VA) by making eight separate requests. According to the findings, the shorter responses were rated as having a similar level of usefulness and favorability, while also being rated as being more efficient, particularly for instructions, and occasionally being easier to comprehend than the baseline. Instead of constantly responding in full phrases, they believe that virtual assistants should be configurable and adapt to users' preferences in order to gain universal adoption.[11].

## **2.4 Scope of the problem**

The most difficult aspect was maintaining a flow of dialogue throughout. The idea for responding to individual requests such as "এটা করো" and "এটা খোঁজো" is straightforward and uncomplicated. However, constant discourse allows for an excessive number of paths that the person could take. The most challenging aspect was coming up with a way for the assistant to carry in a discussion with its own operators, outside from predetermined or manually prepared scenarios.

In addition to this, it is an imperative requirement to have access to high-quality microphone equipment at your disposal. This cannot be negotiated in any way. The microphone will have a difficult time picking up the actual words or instructions that you are speaking if you move further than two or three steps away from the person who is serving you. Either a significant amount of engineering work would have to be done in order to accomplish this objective, or a microphone that was crafted by a skilled expert and had its design meticulously mapped out would be necessary.

## **2.5 Challenges**

The work that we were tasked with completing presented us with a number of challenges, the most difficult of which was undoubtedly the collection and processing of data. We do so because the various age groups each provide us with their very own specific collection of information, which requires us to collect both sets of data. As a result, not all of the data that were gathered possessed an adequate level of clarity to be utilized for subsequent processing, such as having them translated into text. However, some of those data were of a quality that allowed them to be utilized. In today's globalized world, it's easy to forget that English is not spoken everywhere. As a result, it is unreasonable to assume that consumers in different regions will all have the same degree of expertise. The limited number of languages supported by AI is a major deterrent for 38% of consumers who are on the fence about using voice technology.

If the ASR has not been trained on regional language models, it is unlikely that the voice assistants will perform well when deployed there. Even after being taught the

language, ASR still faces the difficulty of distinguishing between regional accents and dialects, which can lead to inaccuracies in translation.

Noise is a major obstacle to accurate speech recognition. Everywhere you look, there it is. In this case, the ASR solution is responsible for correctly identifying the speech input despite the presence of background noise. Even in a noisy environment with a great distance between microphones, an ASR should be able to pick up the sound waves of the input. Imprecision is exacerbated by factors like echo. The receptor's capacity to unerringly process the actual input is distorted by sound waves that have been reflected from surfaces in the space.

Putting in place an ASR system requires foresight. This is a marathon, not a sprint; it won't happen overnight. Keep in mind that developing, testing, and releasing a system into the market requires significant investment of time, money, and other resources. Building a user interface (UI) for a chatbot is simpler than designing a VUI for a voice-activated system because of the presence of visual elements.

This technique might have some potential drawbacks, one of which is that the process of training language models might be challenging and time intensive. This might be one of the potential negatives. This is just one of the numerous potential downsides that could occur. This is only one of the many difficulties that can appear in the future. This is just one example of the many additional challenges that might appear in the years to come. There might be much more. You might find that either amassing a sufficient quantity of linguistic resources or making efficient use of those that are already available to you is a laborious task that requires a significant investment of both time and material. If this is the case, you should prepare yourself for this possibility. If this is the case, then you need to get yourself ready for the likelihood of this happening. If you follow either of these two courses of action, there is a chance that you will wind up having to pay a higher total amount. In the end, the manual development method would result in serious issues that pertain to the finances.

## Chapter 3

### Research Methodology

#### 3.1 Research Subject and Instrumentation

As a result of the wide variety of morphemes that are found in the Bengali language, it is currently ranked ninth on the list of the languages that are spoken the most frequently throughout the entire world. On the other hand, research into Bengali Natural Language Processing is nowhere near as sophisticated as it is in other languages. This is especially true when compared to research into other languages spoken by populations that are smaller than Bengali's. In addition to this, a sizeable portion of the population who speak Bengali as their first language are illiterate and have difficulty using computers because the instructions are written in English. This is a problem in many parts of the world where Bengali is spoken as a first language. If a person has difficulty speaking in English, learning how to give directions in their native language of Bengali may make it easier for them to use a computer or smartphone. There is a possibility that the issue that we are having can be remedied by a virtual assistant that is proficient in Bengali. This is the possibility. This article gives information regarding the construction of a Bengali virtual assistant known as, which uses the CNN model in order to interpret the user's instructions even when they are given in Bengali. The article is available here. Simply clicking on this link will take you to the article. Our language is not only utilized in its native tongue of Bengali, but it is also comprehensible to individuals whose primary language is English. In order to determine which of the instructions or the replies came first, CNN uses a spectrogram to examine the characteristics of both the commands and the answers.

The Python programming language has been applied to a number of different libraries and packages, some of which include speech recognition, pyttsx3, pyaudio, gTTS, and so on. In order to evaluate a wide variety of recorded voices and texts, we made use of a powerful computer that was equipped with a graphical processing unit (GPU), a significant amount of random-access memory (RAM), and a variety of other components. This allowed us to perform the evaluations quickly and accurately.

Because of this, we were able to translate not just between spoken languages like Bengali and written languages, but also between spoken languages and written languages. In order to train and test the data, as well as our proposed model for an AI-based voice assistant, we make use of the software package PyCharm, which is installed on a high-performance computer. This allows us to execute a number of Python routines simultaneously. This affords us the opportunity to evaluate our model in addition to testing and training the data. We are able to train and analyze not only our model but also the data as a result of this.

### **3.2 Data Collection Procedure/Dataset Utilized**

At first, thirteen of the most fundamental and frequently employed commands are selected. The topic of the command is offered to other people so that they might come up with other variations of that command. For instance, the command " গানটা বাজাও" translates to "Play the Music" in English.

It is possible to execute this command in a variety of different ways. Examples include "Give me the Time update" and "Tell me the current date," amongst many others. Because different people have different ways of expressing command, the goal is to collect all of the different ways that command can be expressed.

After the selection of thirteen distinct categories of different commands, we requested that every individual carry out one instruction from each of the thirteen groups. This information is compiled digitally and stored in a disk that is hosted by Google. The audio file is little under 840 kilobytes in size, and it uses up a total of 20 gigabytes of storage space on the computer. It is assumed, for the sake of clarity, that there is no background noise to be heard among the gathering voices. In the Table 3.2.1 there are 13 types of command with three variations are given below.

Table 3.2.1: there are presented 3 variations of 13 commands.

Serial No.	Task List	Command Variation 0.0.1	Command Variation 0.0.2	Command Variation 0.0.3
1	Hi	হাই	কি খবর	কি অবস্থা
2	Asking the time	সময় কত	সময় বল	সময় জানাও
3	Asking the date	আজকে কত তারিখ	আজকের তারিখ বল	আজকের তারিখ টা জানাও
4	Playing Music	বাজাও	গানটা বাজাও	গানটা বাজাও তো
5	Getting Information from Wikipedia	খুঁজে বের করো	অনুসন্ধান করো	সন্ধান করো
6	Find Address by Google Map	ঠিকানা	লোকেশন বল	অবস্থান বল
7	Opening Microsoft	মাইক্রোসফট ওপেন করো	মাইক্রোসফট চালু করো	মাইক্রোসফট খোলো
8	Opening Google Chrome	ক্রোম ওপেন করো	ক্রোম চালু করো	ক্রোম খোলো
9	Opening YouTube	ইউটিউব ওপেন করো	ইউটিউব চালু করো	ইউটিউব খোলো
10	Opening Google Search Engine	গুগোল ওপেন করো	গুগোল চালু করো	গুগোল খোলো
11	Captured Photo	ছবি তোলো	ছবি তোল	ছবি উঠাও
12	Praise Saathi	সাবাশ	ওয়েল ডান	ভালো করেছো
13	Turn Off	বন্ধ করো	থামো	ওফ করো



### 3.3 Statistical Analysis:

We polled people of all ages to find out their thoughts on the most compelling case for using Bangla Voice Assistant. The time and date will be useful information for many of them, since they have told us they intend to use it. They'll use it for things like Facebook, Twitter, YouTube, etc., according to several of them. However, the vast majority of them were interested in using it to conduct internet searches on google. About 13.4% people wanted to use Google, 14.6% people asked about time and 5.3% people asked about date, also 9.9% Chrome users and 6.2% Firefox users are recommended both beside in social media like Facebook wanted by 13.4% people, Telegram wanted by 9.7% people, Whatsapp wanted by 6.9% people, Instagram wanted by 4.9% people, Youtube wanted by 12.6% people, Twitter wanted by 3.4% people. Other 12.6% people also wanted to focus on Youtube.

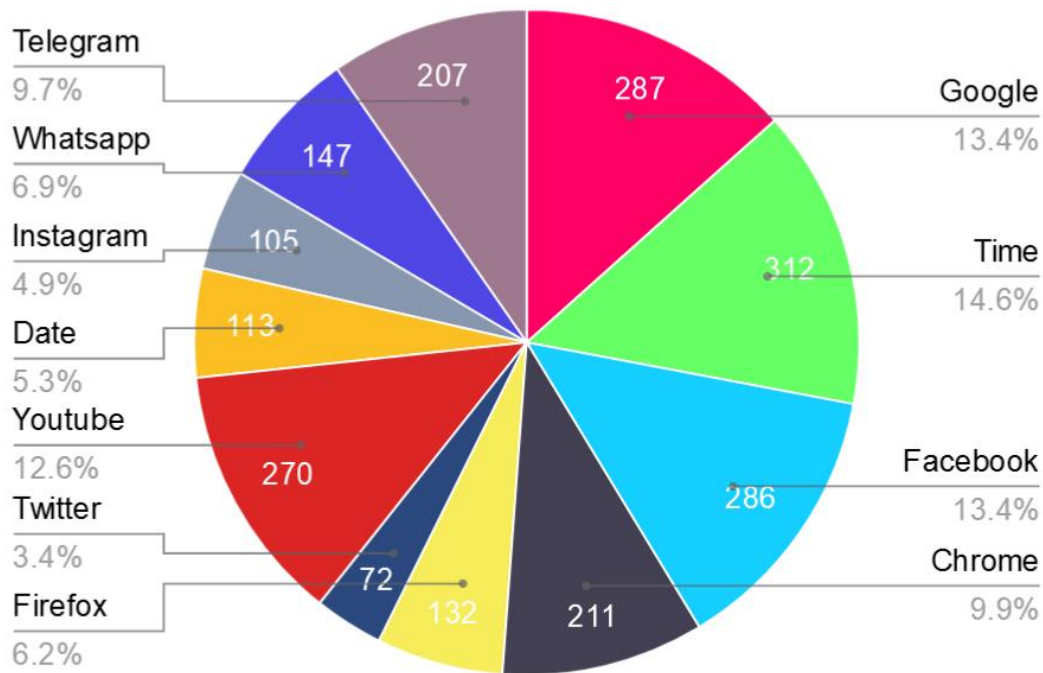


Figure 3.3.1: Graphical representation of User's intend to use Saathi

## **3.4 Proposed Methodology/Applied Mechanism**

### **3.4.1 Audio Pre-processing**

After the recording of an audio file, it is then saved using the .wav file type, and a spectrogram is constructed utilizing the audio files. A spectrogram is a visual representation of a waveform that shows the many frequencies that are present in the waveform as well as the strength of a signal and how it varies over time. A signal's non-stationary or nonlinear features can be accurately represented by a spectrogram. This is because the spectrogram is capable of doing so. As a consequence of this, the spectrogram is an efficient instrument for conducting analyses of signals that comprise a number of frequency components and/or noise that is brought about by a combination of electrical and mechanical sources. The spectrogram is constructed with the help of the short-time Fourier transform (STFT), which is applied to the audio data. With the assistance of the Fourier transform, it is possible to move a signal from its original location in the time domain to its new location in the frequency domain. This allows one to identify the degree to which the signal is fluctuating, which is very useful information. Following the convolution of a small window with the signal, the discrete Fourier transform (DFT) is performed within of the window that has been convolved with the signal. This is a modification of the traditional discrete Fourier transform (DFT). The computational complexity of DFT techniques is not reduced nearly enough to meet industry standards. There is a method known as the quick Fourier transform that can be utilized to bring about a reduction in it (FFT). The Fast Fourier Transform, also known as the FFT, is an algorithm that allows for the discrete Fourier transforms to be computed in a manner that is algorithmic, with the complexity of the computational element being able to be reduced. After the spectrogram has been created, the audio samples and the labels that correspond to them are then stored as an image. This is done after the spectrogram has been displayed. The class of the command that is being sent can be determined with the help of these spectrograms. When it is necessary, these image files that have been maintained are imported in order to make comparisons so that spectrograms do not need to be produced continuously. By examining the spectrograms of a person's speech, it is possible to distinguish specific words, which can then be used to detect orders. This process is called word recognition.

A spectrogram is a comprehensive representation of an audio signal that displays time, frequency, and amplitude on a single graph. Spectrograms are commonly used in the audio industry.

### **3.4.2 Prepare the train and test dataset**

In order to get the model ready, a command dataset consisting of relevant activities is developed. The categorical and nominal data are taken from the spectrogram image so that they can be processed further.

For example, the keyword "time," the keyword "date" etc. However, there are machine learning algorithms that are incapable of directly operating on label data. For the majority of the machine learning process, it is necessary for all of the input data variables to be represented as numbers. As a direct consequence of this, it is essential that our category data be converted into numerical form.

Because the data do not have an ordinal relationship with one another, a one-hot coding technique is utilized to accomplish this task. A method known as one hot encoding is a way of converting categorical data into a form that can be given to machine learning algorithms so that they can perform better when it comes to prediction. After that, mix the data up and separate it into a training set and a testing set. To train the model, sixty percent of the entire data is utilized, while the remaining forty percent is used for testing reasons.

### **3.4.3 Build CNN Model**

In this section, spectrograms are analyzed as if they were photographs, and an attempt is made to extract features from these spectrograms that can be used to determine the category of the audio sample. In order to classify data, a Convolutional Neural Network, or CNN, is utilized. The photos themselves would serve as input for CNN, which would then learn the spatial characteristics of the images and forecast their class. The CNN was constructed with the help of TensorFlow. The process begins by loading the model, and then it moves on to reading the spectrogram images. These pictures are symbolic representations of the words that we have uttered. A spectrogram is created for each and every syllable that is spoken. It is to be anticipated that spectrograms of the sound

produced by the word "time" will be comparable across a variety of speakers of both genders. It is to be assumed that the word "Time" when pronounced by anyone, should have similarities with other time sounds. These similarities should exist despite differences in volume, pitch, timbre, and so on. It should be noted that this also applies to the other commands. Therefore, if there are specific similarities between the sounds of the same word across all of these factors, then CNN will be able to identify those similarities in the spectrogram.

### **3.4.4 CNN architecture**

In order to identify different commands based on their individual spectrograms, a convolutional neural network (CNN) is used. The most important selling point is the fact that it is feasible to extract visual attributes in an efficient manner by making use of learnable convolution operators. In order to do classification, the output of the convolution operators is fed into a neural network. The rectified linear unit comes after the convolutional layer, which consists of two 3x3 kernels and is the next layer in the neural network. The first network layer has now reached its conclusion (shortened to ReLu). The second layer consists of four kernels that are of a size that is 3x3x2 in dimensions. After that, a max-pooling operation is implemented into the two convolutional layers as well as the ReLu layer in order to reduce the quantity of data by fifty percent. This is done in order to improve accuracy. In the second layer, there are eight different 3x3x4 kernels that make use of ReLu. In order to avoid overfitting, this layer makes use of a dropout strategy to discard forty percent of its output. In the end, a neural network that has all of its connections established is used so that an output result may be derived from the categorization. During the process of evaluating the model, the cross-entropy loss function is applied, whereas the RMSProp optimizer is utilized during the process of optimizing the model.

The sound is continuously captured by the microphone of the laptop and then fed into the CNN model so that predictions can be made using the data. If the CNN model is able to categorize the command with a high level of matching, the system will carry out the corresponding task. Figure 3.4.4.1 shows the Overall process of our project Saathi

the Bengali Voice Assistant. Including Data collection. Audio Processing then obtaining the Speech and construct spectrogram and apply CNN to classify. After Identifying the classification, it will train the model for further process.

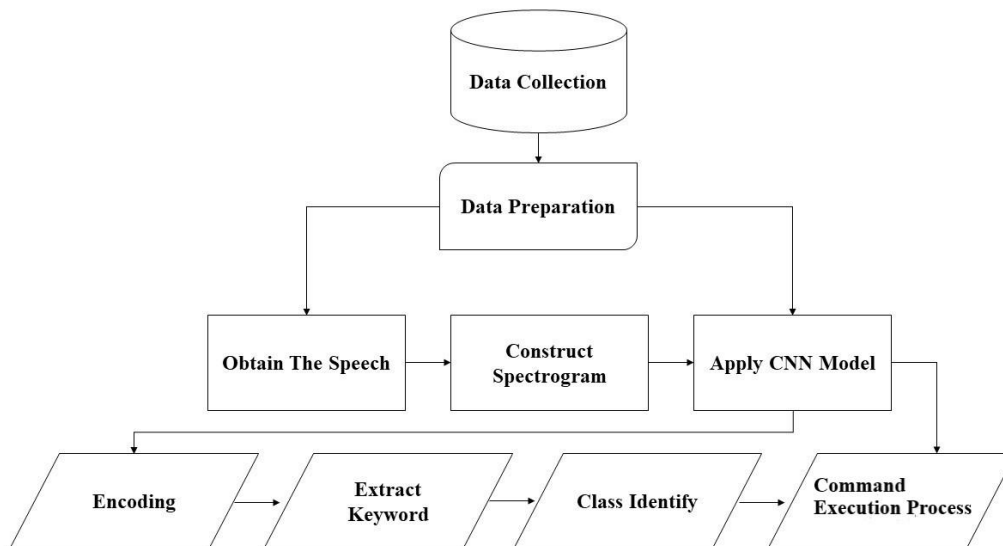


Figure 3.4.4.1: Overall Process of Bengali Voice Assistant -Saathi

### 3.4.5 Proposed Model

The goal of the system that has been proposed is to produce a useful virtual assistant that is able to comprehend Bengali voice instructions that are delivered via the use of natural language processing. After the instruction has been translated into Bengali and understood, the virtual assistant will put it into action in order to create the result that has been requested. We were given the instruction in Bengali to understand. It is vitally necessary to train the system so that it can recognize the command in order to put virtual help into action. Only then will the system be able to function properly. When this requirement is satisfied, and only then, will the system be able to provide assistance to those in need. Because of this, it is necessary for the training of the system to make use of a benchmark database that is exclusively devoted to the Bengali language. On the other hand, it is not possible to access the benchmark database in the Bengali language at this time. This restriction is in place for the time being. This restriction will continue to apply for the foreseeable future. The Bengali language has been put to use in the

construction of a number of instructions, each of which presents a novel perspective on the many categories in an effort to make the process of finding a solution to this problem more manageable. The goal of these instructions is to make the process of finding a solution to this problem more manageable. Following the completion of this stage and the preprocessing of the audio command, the data extracted from the audio will be utilized in the creation of a picture. When CNN is finished with the preparations for the model, it will use the image that was just produced as the input. This will take place once the preparations have been completed. Following the completion of the training, the model is implemented into the issue of recognizing Bengali voice instructions in order to find a solution to the problem.

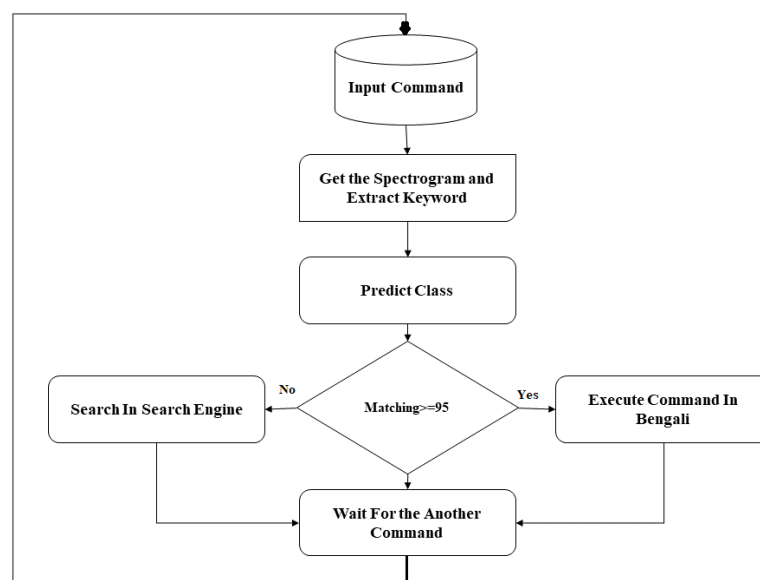


Figure 3.4.5.1: Working Procedure of Voice Assistant - Saathi

In Figure 3.4.5.1 shows that the operational methodology of our project, which is referred to as Saathi. The first thing that will happen is that it will wait for commands from the user. After that, it will retrieve the spectrogram and derive the keyword from that. Then, based on that information, it will determine their class. Following that, it will begin to match if it finds an equal or greater value than 95. The instruction will be carried out in Bengali. The user will be able to hear it right away. If it does not satisfy the condition, the saathi will inform the user that it will do a search of the search

engine for anything relevant to the query. And then saathi once more, wait for the Another Command to appear.

### **3.5 Implementation Requirements**

In our study, we used a powerful computer with a graphical processing unit (GPU), a large amount of random-access memory (RAM), and other components to evaluate a variety of recorded voices and texts, as well as to translate between spoken languages like Bengali and written ones. To train and test the data and our suggested model for an AI-based voice assistant, we utilize the software PyCharm to run the various Python codes on a high-performance computer. For your convenience, we've included all of our research's prerequisites below. Used Requirements are:

#### Hardware Requirements

- ✓ CPU having 4.10 GHz
- ✓ Core i5 6-core Processor
- ✓ 8 GB DDR4 RAM
- ✓ 500GB SSD
- ✓ 500GB HDD
- ✓ I/O device
- ✓ Internet Connection

#### Software Requirements

- ✓ Operating System – Windows 10
- ✓ PyCharm Software
- ✓ Python software
- ✓ Image Editor
- ✓ Lucid Chart

#### Developing Tools

- ✓ Python Environment
- ✓ PyCharm Software

## Chapter 4

### Experimental Results and Discussion

#### 4.1 Experimental Setup

We evaluated many recorded voices and sentences on a powerful computer with a GPU, a lot of RAM, and other components. We wanted to find the most accurate copies. We accomplished much because of this. This computer wrote Bengali. PyCharm runs Python procedures on a powerful computer, training and testing our AI-based voice assistant model and the data. This permits model testing and training. Examine, practice, and grade our model. This verifies and enhances data quality. Thus, we can practice and assess data. Microsoft Windows 10, an 8-gigabyte CPU, and a GPU enabled this (GPU). This ensured the model was correctly executed. Our computer has an input/output (I/O) device, a 4.10 GHz Core i5 6-core processor, 8 GB DDR4 RAM, a 500 GB SSD, and an Internet connection. Our computer's HDD holds. The computer also has I/O. Our inquiry used Python, the computer language. The language used was Python. Our research is named after computer programming. PyCharm was our main tool for designing and training all our models, including the machine learning model, which required Python code to function. We designed and trained this model in Python. The models needed this code to learn from each other. This code let models teach each other.



## 4.2 Experimental Results & Analysis

Our Project makes it simple for anyone to engage with cutting-edge research. The present day and time are simply accessible to them. Not only that, but with a simple command users can search for anything in any web browser. Saathi will deliver this information in an approachable and straight forward manner. This is all being discussed in a brief right now.

```
File Edit View Navigate Code Refactor Run Tools VCS Window Help Saathi - main.py
AlexaSiri main.py
Project
main.py x
60
61 if 'হাই' in command or 'কি খবর' in command or 'কি অবস্থা' in command:
62     print('হ্যালো!')
63     talk('হ্যালো!')
64
65 elif 'সময়' in command or 'সময় বল' in command or 'সময় জানাও' in command:
66     time = datetime.datetime.now().strftime('%I:%M %p')
67     print('এখন সময়: ' + " " + time)
68     talk('এখন সময়' + time)
69
while True
Run: main x
D:\AlexaSiri\venv\Scripts\python.exe D:\AlexaSiri\main.py
আমি শুনছি...
হ্যালো!
```

Figure 4.2.1: Output of Welcome command using Saathi-(VA)

Figure 4.2.1 shown that the output result of welcome which are 'হাই'. When the user command like this he/she will gets the response from the Bengali voice assistant Saathi which is 'Hello' in Bengali that is 'হ্যালো'.

```

61 if 'হাট্টে' in command or 'কি খবর' in command or 'কি অবস্থা' in command:
62     print('হ্যালো!')
63     talk('হ্যালো!')
64
65 elif 'সময়' in command or 'সময় বল' in command or 'সময় জানাও' in command:
66     time = datetime.datetime.now().strftime('%I:%M %p')
67     print('এখন সময়: ' + " " + time)
68     talk('এখন সময়' + time)
69
while True
Run: main
D:\AlexaSiri\venv\Scripts\python.exe D:\AlexaSiri\main.py
আমি শুনছি...
এখন সময়: 09:18 PM

```

Figure 4.2.2: Output of time asking command using Saathi-(VA)

When the user need to know the current time, he/she will ask Saathi by commanding 'সময় বল '. From there they will get reply from Saathi in Bengali that is 'এখন সময়' with current time and show in terminal.

```

66     time = datetime.datetime.now().strftime('%I:%M %p')
67     print('এখন সময়: ' + " " + time)
68     talk('এখন সময়' + time)
69
70 elif 'তারিখ' in command or 'আজকে কত তারিখ' in command or 'আজকের তারিখ বল' in command:
71     date = datetime.datetime.now().strftime('%m/%d/%Y')
72     print('আজকের তারিখ: ' + " " + date)
73     talk('আজকের তারিখ' + date)
74
75 elif 'বাজাও' in command or 'গানটো বাজাও' in command or 'গানটো বাজাও তো' in command:
76
while True
Run: main
D:\AlexaSiri\venv\Scripts\python.exe D:\AlexaSiri\main.py
আমি শুনছি...
আজকের তারিখ: 01/02/2023

```

Figure 4.2.3: Output of date asking command using Saathi-(VA)

When the user need to know the current date, he/she will ask Saathi by commanding 'আজকে কত তারিখ'. From there they will get reply from Saathi in Bengali that is 'আজকের তারিখ' with current date and show in terminal.

```

28 audios = 'audio.mp3'
29 ttss.save(audios)
30
31 playsound.playsound(audios)
32
33 os.remove(audios)
34 saathi.say(ttss) # (13)saathi will ask me
35 saathi.runAndWait() # (14)then run and wait for my reply
36
37 def take_command(): # (12b)
38     try: # (3)
39         with sr.Microphone() as micro: # (4)Microphone call korlam, nickname
40             print('আমি শুনছি...') # (6)micro ready kina ta check kora
41             voice = listener.listen(micro) # (5)voice read koralam ami ja bolchi ta shonar janno micro input niye

```

Run: main x

D:\AlexaSiri\venv\Scripts\python.exe D:\AlexaSiri\main.py

আমি শুনছি...

তুমি বলেছো: জেমসের মা

Figure 4.2.4: Output of playing music using Saathi-(VA)

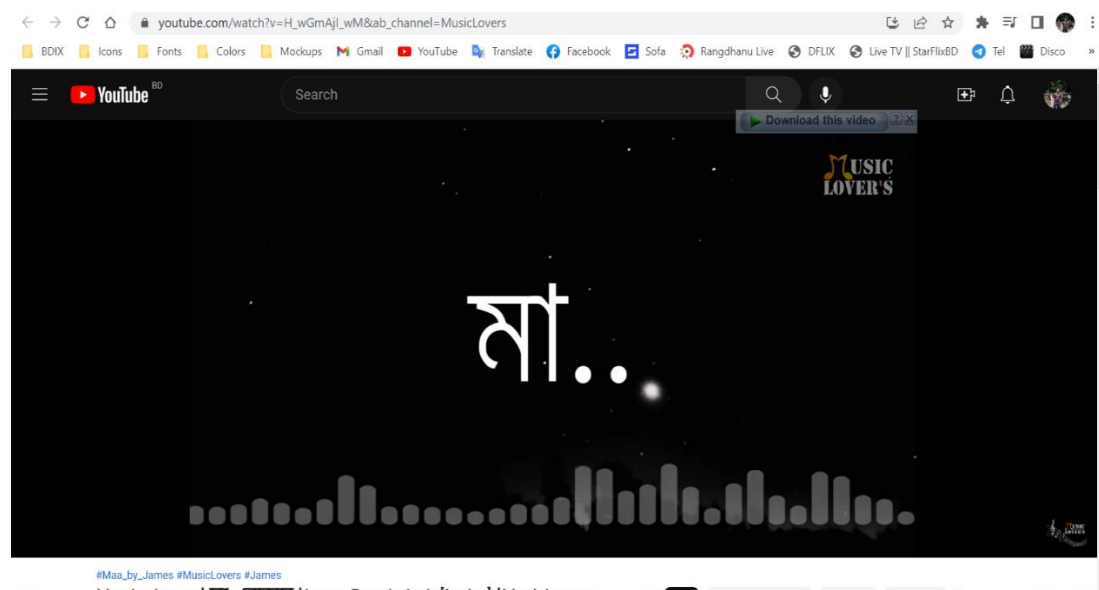


Figure 4.2.5: Output of playing music through youtube using Saathi-(VA)

In this section, the person will ask to Saathi to play music by commanding 'গানটা বাজাও' with song name and Saathi open and play the song In youtube via web browser directly.



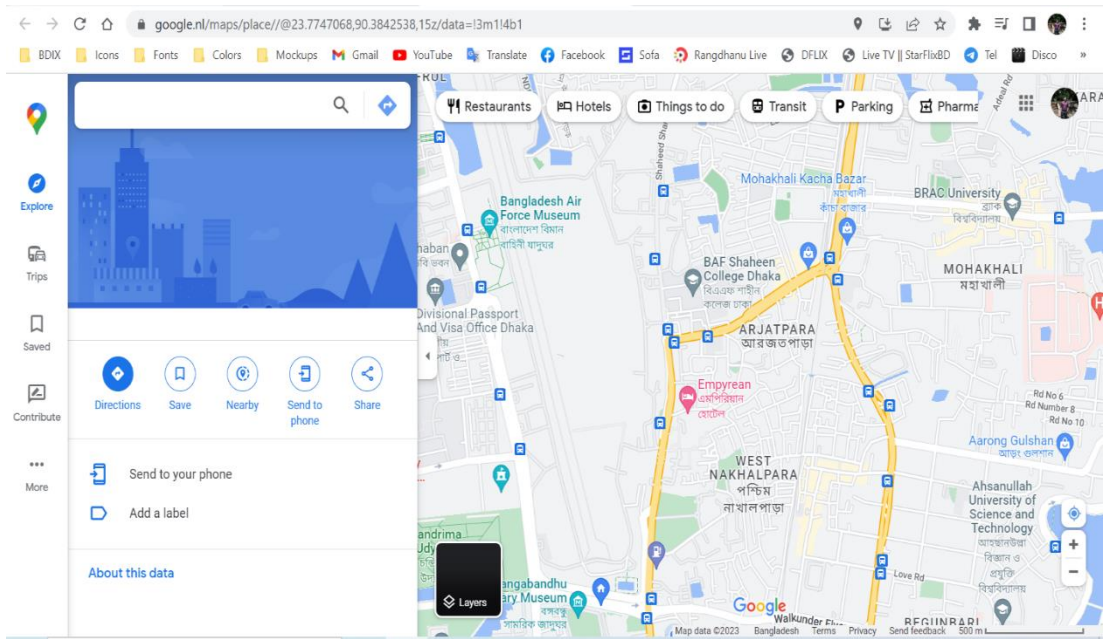


Figure 4.2.7: Output of recent Address using Saathi-(VA)

Furthermore, After asking Saathi to find my address through in Bengali 'অবস্থান বল' and Saathi reply 'এখানে অবস্থান আপনার' and locate them where they are now via Google Map in web browser.

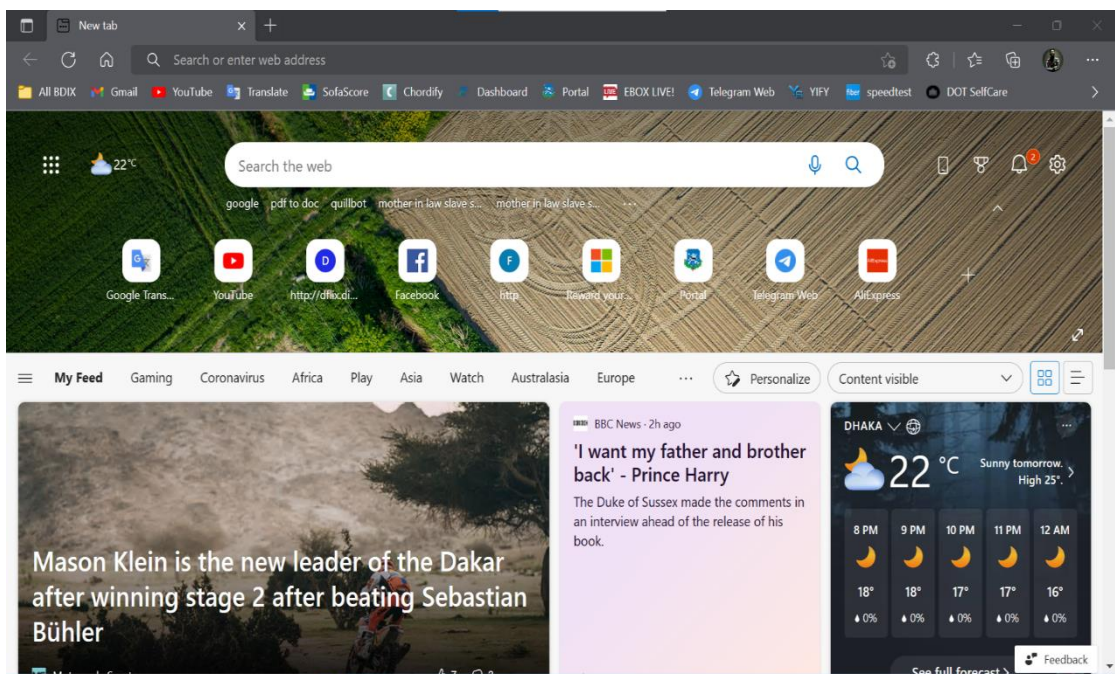


Figure 4.2.8: Output of Open Microsoft using Saathi-(VA)

Here people will ask to Saathi ' মাইক্রোসফট ওপেন করো ' in Bengali and Saathi do open Microsoft Edge directly as new tab.

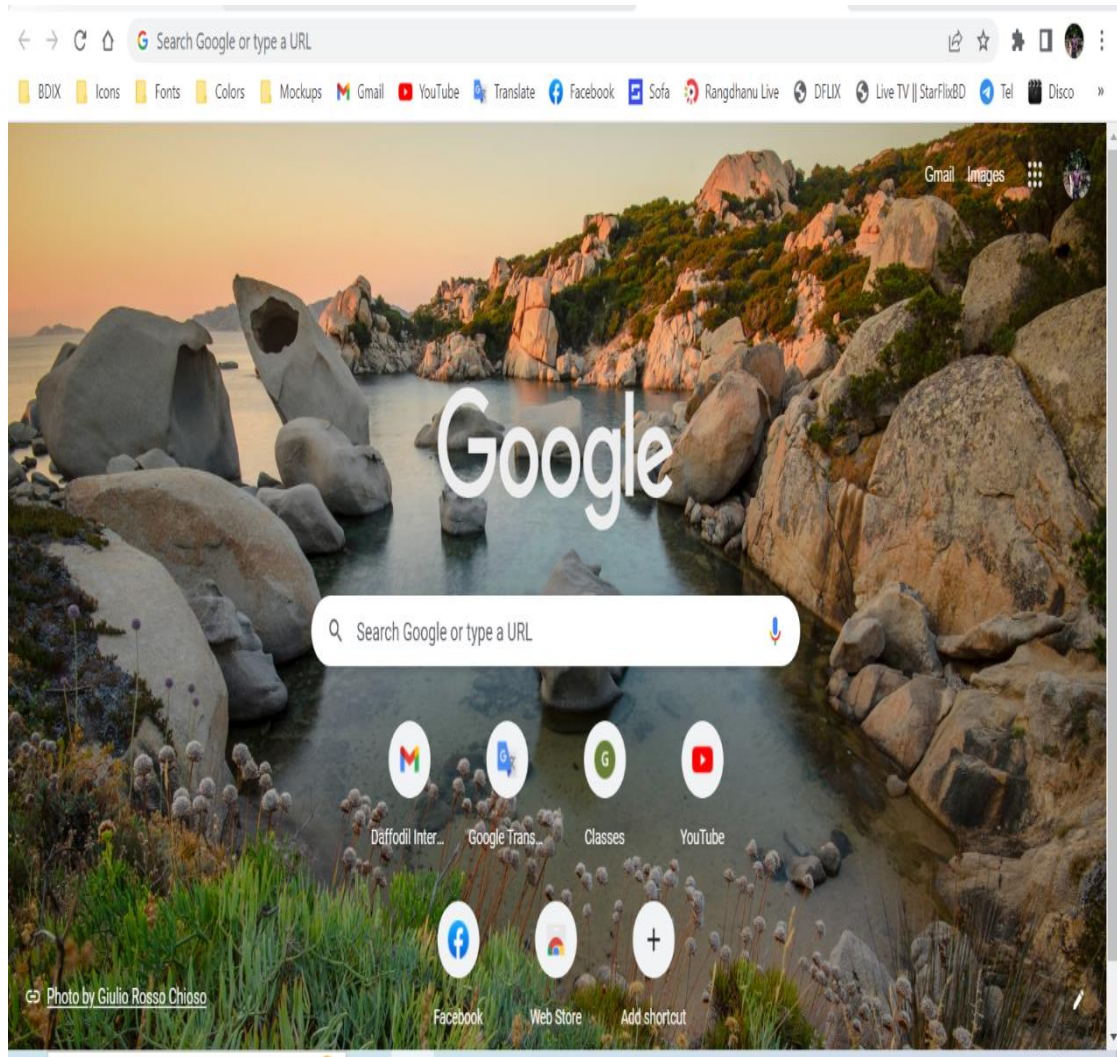


Figure 4.2.9: Output of Open Chrome using Saathi-(VA)

Also here people will ask to Saathi 'ক্রোম ওপেন করো' or 'ক্রোম চালু করো' in Bengali and Saathi do open into Google Chrome directly as new tab.

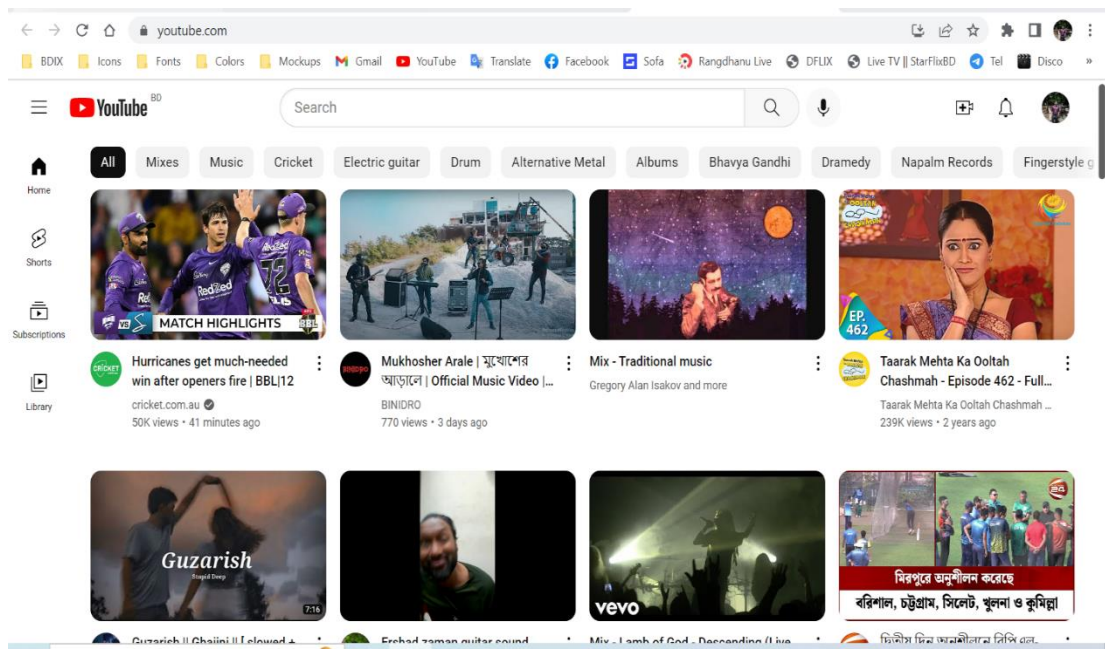


Figure 4.2.10: Output of Open Youtube using Saathi-(VA)

Also here people will ask to Saathi 'ইউটিউব চালু করো' or 'ইউটিউব ওপেন করো' in Bengali and Saathi do open into Google Chrome directly as new tab for youtube.

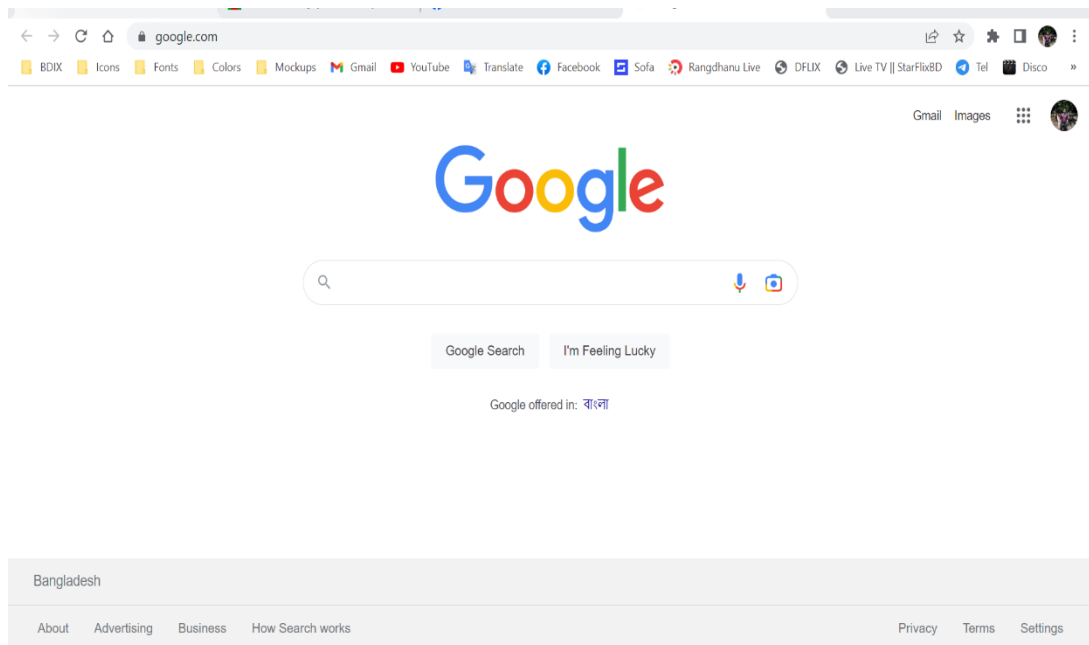


Figure 4.2.11: Output of Open Google Search Engine using Saathi-(VA)

Also here people will ask to Saathi 'গুগোল চালু করো ' or 'গুগোল ওপেন করো' in Bengali and Saathi do open Google Search Engine.

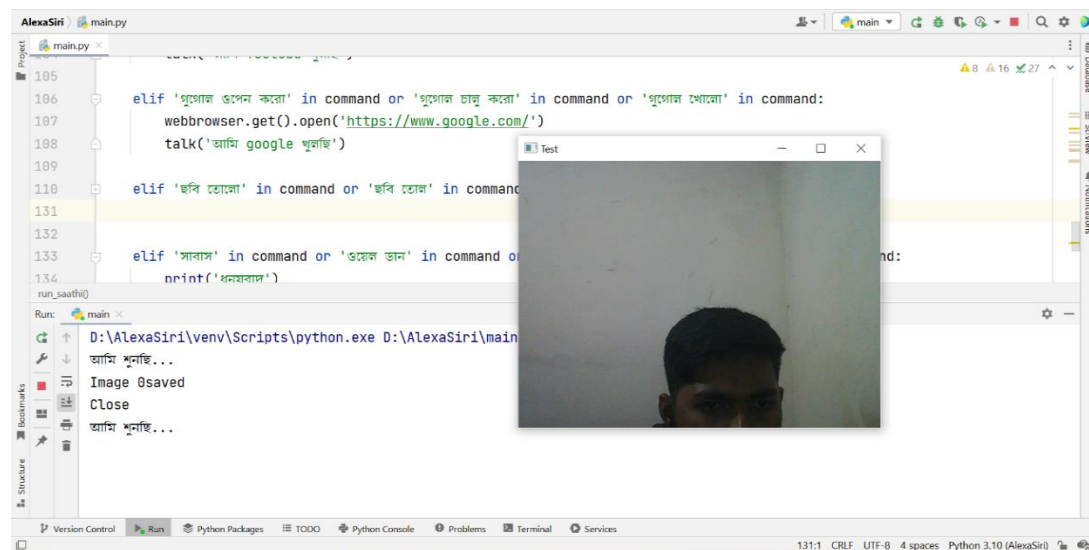


Figure 4.2.12: Output of Captured Photo from Camera using Saathi-(VA).

In here, People tell Saathi through command 'ক্যামেরা চালু করো' or 'ক্যামেরা ওপেন করো' and Saathi open the camera by webcam, get capture by keyboard key 'space' and press ESC to close after captured.

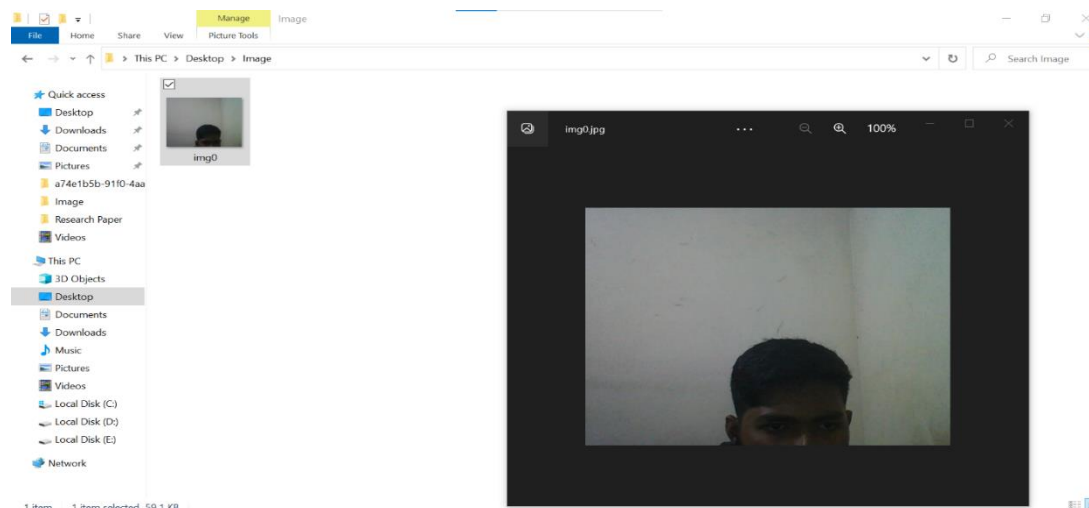
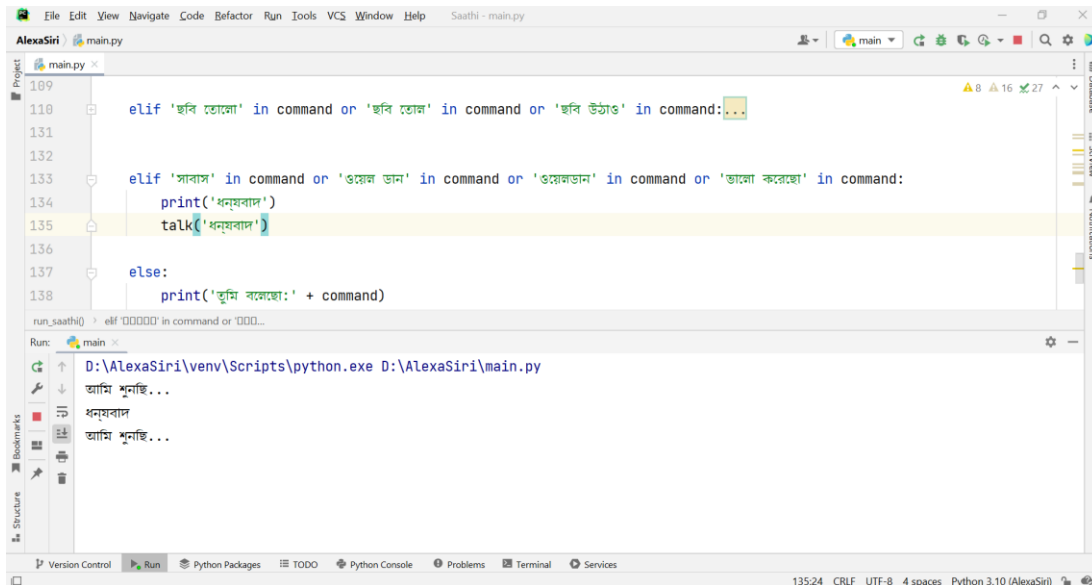


Figure 4.2.13: Output of saved photo in Folder.



Later, the captured photo will save to attach folder name 'Image' and the address gave here 'C:/Users/av617/Desktop/Image/img'.



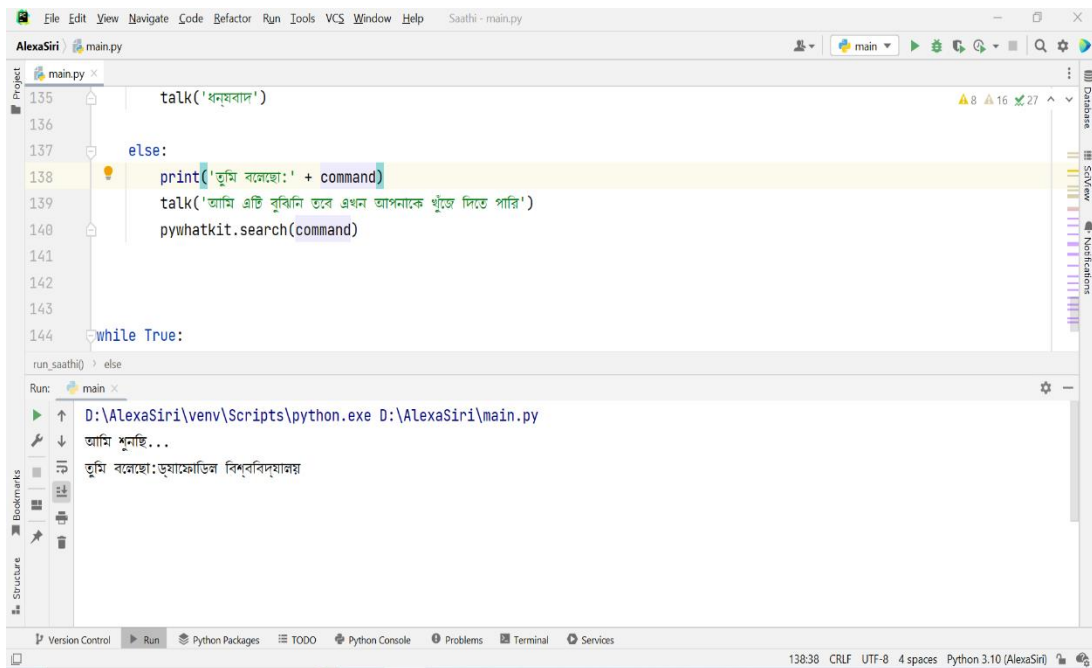
```
109
110 elif 'ছবি তোলা' in command or 'ছবি তোল' in command or 'ছবি উঠাও' in command:...
131
132
133 elif 'সাবাস' in command or 'ওয়েল ডান' in command or 'ওয়েলডান' in command or 'ভালো করেছে' in command:
134     print('ধন্যবাদ')
135     talk('ধন্যবাদ')
136
137 else:
138     print('তুমি বলছো: ' + command)

run_saathi() > elif '00000' in command or '000...
```

```
Run: main x
D:\AlexaSiri\venv\Scripts\python.exe D:\AlexaSiri\main.py
আমি শুনছি...
ধন্যবাদ
আমি শুনছি...
```

Figure 4.2.14: Output of Praising Saathi using Saathi-(VA)

Can appreciate Saathi through command in Bengali 'সাবাস' or ' ভালো করেছে' and Saathi will reply 'ধন্যবাদ'.



```
135     talk('ধন্যবাদ')
136
137 else:
138     print('তুমি বলছো: ' + command)
139     talk('আমি এটি বুঝিনি তবে এখন আপনাকে খুঁজে দিতে পারি')
140     pywhatkit.search(command)
141
142
143
144 while True:
```

```
run_saathi() > else

Run: main x
D:\AlexaSiri\venv\Scripts\python.exe D:\AlexaSiri\main.py
আমি শুনছি...
তুমি বলছো: ডুয়াকোডিন বিশববিদ্যালয়
```

Figure 4.2.15: Output of search command which is not included using Saathi-(VA)

When the user commands the companion but if the companion does not understand the command, it will say " আমি এটি বুঝিনি তবে এখন আপনাকে খুঁজে দিতে পারি " and immediately search for that command in Google.

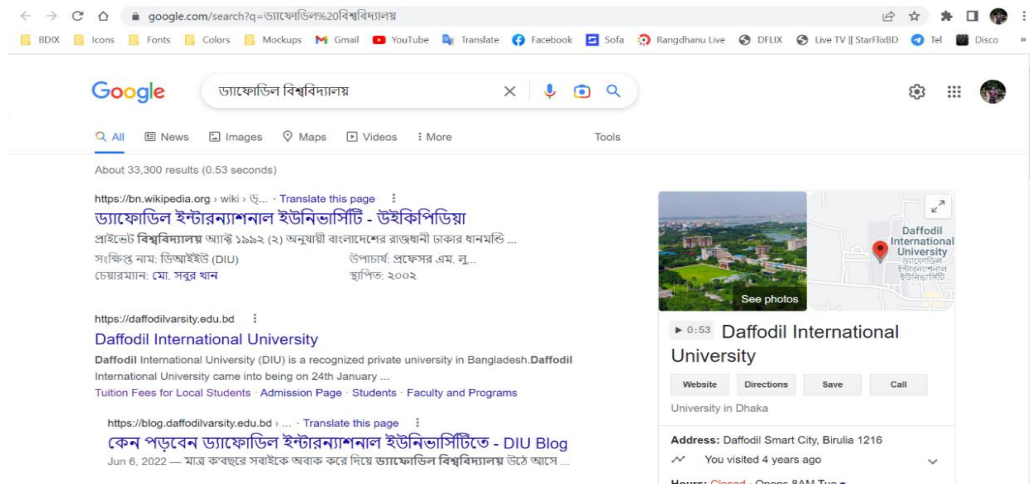


Figure 4.2.16: Output of search command by Saathi-(VA)

Here you can see the output In Google Search that a user command 'ড্যাফোডিল বিশ্ববিদ্যালয়' as Saathi bring here.

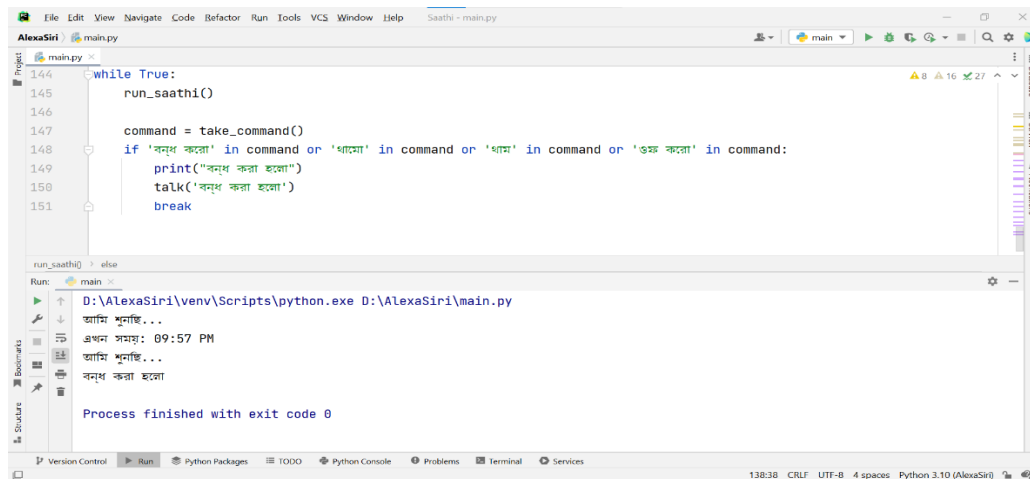


Figure 4.2.17: Output command of Saathi-(VA) for turn off.

At last Saathi will turn off with reply in Bengali ' বন্ধ করা হলো ' when the user completed his/her task and tell Saathi to ' বন্ধ করো ' or 'থামো' or 'ওফ করো' which are shown in Figure 16 for close.

### 4.3 Discussion

Voice-activated assistants have a promising future. There's no denying that voice assistants are currently a part of our everyday lives and will only grow more ubiquitous in the future. With more and more gadgets including speech recognition and more people discovering how useful these technologies can be. With the advent of 5G and developments in machine learning, it's possible that voice assistants will one day become indispensable. Together with Machine Learning methods, it can power a helpful personal assistant that can multitask across programs to improve people's daily lives. But there are obstacles to overcome before we get there, such as substantial investment, advancements in technology, and consumer trust that this item in their lives does not pose a threat to their privacy. The idea was to make the device as simple to operate as feasible by using only vocal commands. The physically challenged, as well as any user interested in voice recognition, will benefit greatly from this program. In the classroom, learning about these systems can help students learn about AI, NLP, ML, HCI, and UI/UX design. The proposed method can interpret voice instructions when disconnected from the internet, allowing customers to save money on monthly data plans. This factor alone makes the software considerably quicker than competitors like Apple's Siri, Google's Assistant, etc. In addition, the solution may do a wide range of operations, such displaying the time and date, playing music or movies, making phone calls, checking the weather and temperature, searching the web, and so on. As a prototype for several cutting-edge programs, this document is useful in its own right. Consequently, based on the literature review and analysis of the current system, we have concluded that the suggested system will not only ease interaction with other programs and modules, but will also assist us to keep it structured. There is still much ground to cover in the realm of automation, but device capabilities can help us develop a new generation of voice-controlled devices and usher in continual fresh revolution in this field. Use this paper as a template for a wide variety of cutting-edge projects. We use a total of twelve different variations of regularly used instructions to train the proposed system. Any user, regardless of age or gender, who speaks Bengali can use their most natural commands to operate the system. Utilizing recent innovations in mobile phone networks and speech recognition software, the new service is now

available. We only made our solution available on smart phones because their prevalence in rural regions far outweighs that of laptops. They did, however, demonstrate the significance of pleasure motives in influencing whether or not consumers will recommend a product to others. Finally, they highlight the significance of entertainment as a link between functional drives (such ease of use) and real product adoption. It would be worthwhile to explore how this tool might be used for both functional and decorative purposes in homes in the future. The user can issue commands and receive responses in a rural language. These devices were developed to understand and carry out certain user commands. The increased availability of smartphones in rural areas has allowed people who do not know standard Bengali to communicate fluently in their preferred regional dialect. It's also made to aid those with visual impairments. The flexibility to outsource only the necessary work is what makes hiring a virtual assistant so enticing. The equipment you purchase may need you to use specific AI service providers. The findings show that these approaches have promise for application in voice recognition scenarios. We tweaked it so that it now just reacts to the user's voice and not any ambient noise. The modular design of the project makes it easy to understand and flexible for use in different contexts. More features can be added to the application without sacrificing its current capabilities. All required Python packages are up and running, and development and deployment have taken place in the IDE VS Code (IDE). Together with the Python 3.x code, we gathered data on ambient noise levels. Only application-based operations are supported by the system. This will lead to even greater market fragmentation.

## **CHAPTER 5**

### **IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY**

#### **5.1 Impact on Society**

Voice assistant has a lot of social impact. Using this many people can easily able to do many things. If we think about how these Voice Assistants are changing our lives, there are pretty positive results. It is not just about accessing everything hands-free, but people are enjoying these talks. They are engaging with their voice assistant as if they were human. People can command anything they want to do. And hearing those command the voice assistant can do those stuff. Voice assistant actually a blessing for blind people. They can easily do their stuff by a simple command when they need. Their life is now so easily and comfortable by using voice assistant.

Voice assistant Not only just changing our live, it also reduces our time quickly to manage other works.

There is a good chance that each of us is familiar with voice assistant technology, and may have even utilized it in some way or another. The world's most influential technology companies have each released devices that support voice technology. Apple's Siri, the Google Assistant, Amazon's Alexa, and Microsoft's Cortana are just some of the digital assistants currently available on the market.

Over the course of the past few years, an increasing number of businesses have been investigating the potential of personal voice assistant technology for app development and business in general. Taking into account all of the benefits, this has resulted in a rise in both the frequency of use and the level of awareness among customers. In recent research, the global auditing organization PwC found that out of 1000 customers (aged 18–64), 90% of the subjects had awareness about speech technology, and the majority (72%) had also used a voice assistant. The research was conducted on consumers in different countries around the world.

Research conducted by Juniper indicates that there are approximately 3.25 billion voice assistants in use today, and the company predicts that figure will rise to 8 billion

personal voice assistants in use by the year 2023. Voice assistants are becoming more useful in business environments as a result of recent advancements in artificial intelligence assistive technologies.

More people who own businesses need to give some thought to the numerous benefits that can come from using a voice-activated personal assistant, which are some of the things that I will go over in the following paragraphs.

Chatbots that are integrated with voice assistant technology are an excellent tool for companies in the e-commerce business to use to excite their customers by giving them the convenience of being able to shop online using any device. In addition, businesses receive data that is acquired from customer information depending on the customer's preferences, device, cart or purchase history, access location, search history, and other information. The information gathered can be put to use in the development of targeted marketing strategies and enhanced search engine optimization (SEO) for the website of your company.

Support around the clock is something that customers want. It's not uncommon for them to seek assistance at unusual hours, and when that support isn't readily available, it may be a very frustrating experience for them. Voice assistants are helpful in preventing awkward situations like these from occurring. The use of vacation or sick time is not necessary for a digital talking assistant, which eliminates any disruptions to customer care and interactions.

If you run a hotel that makes use of voice assistant technology, for example, when a guest is uncomfortable with the temperature in their room during the night, rather than having to call the front desk or fiddle with instruments, the guest can use the in-room smart speaker instead of having to do either of those things. Because of this, you no longer have a need for night support workers because you are protected by personal assistant technology, and your 24-hour-a-day, seven-day-a-week customer care is ensured.

One further significant advantage that personal voice assistants bring to businesses is that they help to simplify the processes that are required to use digital assistants in a company. This is a really useful feature. These talking assistants continue to function normally even in the face of developing technologies and advanced learning. They are constantly accessing reports, conducting data analysis, and making sure that crucial systems have the most recent updates.

Freeing up human time and resources can be accomplished by delegating routine activities to a voice-activated personal assistant that can be automated. In addition to this, it is able to accomplish these tedious duties in an effective manner with no errors, which frequently results in an increase in customer satisfaction. Voice assistants are delegated to handle routine activities, freeing up humans to focus more of their efforts and time on responsibilities that require human interaction to get the desired results in terms of business solutions and services.

In addition, the implementation of voice assistants across your workforce will not only have an effect on the experience that your customers receive, but it will also increase the level of general productivity inside your company. According to a survey published by Gartner, artificial intelligence assistant technology will not only generate \$2.9 trillion in corporate value by the year 2021, but it will also recoup 6.2 billion employee working hours.

You may make your clients' lives easier by facilitating particular actions with the help of Alexa Skills and Google Actions. Your company will benefit from the streamlining of day-to-day activities that are continually being watched when you implement artificial intelligence assistant technology. Using certain voice commands, you can initiate a variety of actions, such as remembering crucial dates and deadlines, scheduling appointments, booking appointments, and so on.

## 5.2 Impact on Environment

The nearly eight billion people that inhabit our planet means that a lot more has to happen for significant environmental action, but every one of us can still make a difference by making little adjustments.

Over the past decade, many have looked to technology in the hopes that it can aid in cleaning up the world. Technologists are going all out to make the world habitable for future generations, with initiatives ranging from ocean cleanups to a machine that eliminates CO2 emissions from the atmosphere.

Some environment-friendly technology helps in its own modest manner on the internet, fine-tuning the minor things that consumers can genuinely handle rather than relying on massive contraptions and gear straight out of science fiction. Sharing leftovers, monitoring water consumption, and planting a forest's worth of trees each month are all achievable without leaving your house. Now that they've here, voice assistants are ready to help save the world, too.

Voice is proving to be an invaluable tool in the field of education. Education is without a doubt one of Voice's strong suits, and it can be used for a variety of purposes. These range from assisting children who struggle with studying to providing emergency medical personnel with additional information that could save lives. It is only natural that people have started to use this power for the greater benefit. Alexa provides skills like as Environmental News and Climate Change News, which provide daily updates on the most recent scientific findings and current events. Google Assistant has a popular action called Climate Change Trivia. Google Assistant also has a skill called Climate Change Trivia. It does not appear that there are any flash briefings on the topic as of right now, which means that you have a fantastic opportunity to develop your very own flash briefing!

Voice-enabled Alexa skills are proliferating in response to a global movement encouraging individuals to lessen their environmental impact.



One useful skill is learning how to live in a more environmentally friendly manner on a daily basis. Knowing what can and can't be recycled is one of the most challenging aspects of living sustainably, but with the help of the Recycle Track skill, you'll be well on your way. Tell Alexa what you want to recycle, and she'll tell you where to take it. A child can take on the role of Captain Earth in an exciting interactive adventure, where the goal is to protect Earth from evil forces. This narrative is a great approach to introduce children to the concepts of reuse and recycling while also entertaining them.

Both the Google Assistant and Alexa are able to take control of your lights and other electronic devices, and many apps come equipped with optional timers that turn off the gadgets after a predetermined amount of time (or even as soon as you leave the room).

In particular, Google suggests that you connect your environmentally conscious assistant with a smart thermostat, LED lighting, and sprinklers in order to assist in reducing unnecessary usage. They also recommend setting up leak detectors such as LeakSmart so that Google Assistant may quickly turn off your water supply in the case that it is being used in an inefficient manner.

### **5.3 Ethical Aspects**

As the use of voice assistants grows to encompass a bigger audience number of users, users in a variety of age groups, and users in a variety of geographical locations, the ethics of speech technology have become essential problems for organizations as well as voice artificial intelligence providers. As a direct result of the proliferation of voice interfaces, it is likely that we will see an increase in the examination of how and when data is gathered, how it is utilized, and how businesses are working to ensure that their voice assistants satisfy rising ethical standards. Additionally, it is likely that we will see an increase in the examination of how businesses are working to guarantee that their voice assistants satisfy rising ethical standards. If you and the other members of your organization haven't begun discussing this matter just yet, it could be a good idea to start doing so as soon as possible in order to avoid being taken aback by something unexpected.

What exactly is going on with the manner in which you are speaking? AI space isn't unique. It is natural for individuals to start asking about the unintended consequences of adopting a newly developed technology whenever that technology has achieved extensive adoption and widespread appeal. This is because widespread adoption and widespread appeal are indicators of widespread appeal. They might start to ask whether that technology is being used in the public's best interest or whether certain components of it, such as data collection and tracking, are having unanticipated and undesirable consequences on people.

In most cases, when they are made aware of the concerns, respected businesses swiftly respond to questions and take measures to keep one step ahead of any complaints that may come their way. Before companies can take any action, they need to be informed of the potential hazards and given a strategy to reduce those risks. Only then will they be allowed to go forward. When it comes to the success of any endeavor, transparency is frequently the single most critical component. The greater the amount of communication that takes place, the more users will trust the business.

When developing a voice assistant, there are a variety of broad ethical factors that need to be thought through and accounted for. There are still certain general ethical issues to take into account despite the fact that these difficulties are continuously evolving and that each organization should assess what is most beneficial for the clients they serve. Whether the company already has a voice assistant that it is looking to improve or is just beginning the process of planning, addressing ethical considerations early on will result in a more robust voice artificial intelligence design. This is true whether the company already has a voice assistant or is just beginning the planning process.

When developing a voice assistant, the following are four ethical considerations you should keep in mind:

- ✓ Concerning privacy and the gathering of data
- ✓ Suggestive language
- ✓ Use by children
- ✓ Cultural biases

## 5.4 Sustainability Plan

In our research we are developing A voice assistant which will make people life easy. As we know that People's days in the workplace can be a nightmare of number crunching, folder shuffling, and phone calls that never end. In the workplace, most redundant tasks have been removed thanks to technological advancements and new ideas, but a few remain. Artificial intelligence (AI) was the impetus for this whole project, which relied heavily on automation. Workflows that were once rife with redundancy have been gradually simplified by AI across the board, from customer service to daily administration. Now, a new partnership between AI and Voice Technology promises to skyrocket productivity.

Even though speech recognition and a great many other varieties of voice technology have been available for some time, the number of people using these technologies has seen an exponential rise in the number of people using these technologies in the recent years. This is the case even though the number of people using these technologies has been available for quite some time. The most recent few years have been particularly noteworthy with regard to this expansion. It is feasible that making use of a piece of technology that can be operated in any way, shape, or form simply by verbal contact will result in a nearly limitless number of positive outcomes. This is because verbal contact can be used to operate the technology in any way, shape, or form. This is due to the fact that the technology can be operated in any way, shape, or form through the use of vocal interaction. When other people are also in a position to profit from our effort, we will then know that our activity has attained a degree of sustainability. Until then, we won't be able to determine whether or not it has. We won't know for sure if it has or not until that time comes around, so please be patient. Until that amount of time has gone, we won't be able to say with complete assurance that it can be maintained for the long term.

## CHAPTER 6

### CONCLUSION, RECOMMENDATION AND FUTURE WORKS

#### 6.1 Summary of the Study

Voice recognition is being integrated into a growing number of software programs and electronic devices, which has resulted in an increase in the number of people who are becoming aware of the advantages that may be gained from using such tools. It is evident that the technology will soon be available everywhere, and with the coming of 5G and developments in machine learning, it is possible that voice assistants may eventually become tools that humans cannot effectively function without. It is possible to combine it with methods of machine learning to create a smart assistant that is capable of acting on a variety of applications and will make living a more comfortable experience for humans. This combination will be possible because it is possible to combine it with methods of machine learning. On the other hand, before we can get to that stage, we will have to first triumph over a number of challenges. These hurdles include making a significant financial investment, advancing technological capacity, and assuring customers that the product they use on a regular basis does not in any way endanger their privacy. The only input that was necessary was a person's voice, and the idea was that doing so would make using the device as easy and uncomplicated as is humanly feasible. This program may also be useful for users who are unable to move their bodies appropriately, and it will be useful for users who are interested in voice recognition. If students are made more aware of systems such as these, it will help them get a deeper comprehension of topics such as artificial intelligence, neural networks, natural language processing, machine learning, and human computer interaction. They may also learn how to improve the user experience when it comes to the development of applications from this. Customers are able to save money on data bundles thanks to the newly designed technology, which can execute voice instructions even when users are not connected to the internet. This opens up new financial opportunities for businesses. This element also helps to make it substantially speedier when compared to other programs like Apple's Siri, the Google assistant, and so on and so forth. In addition, the solution is able to effectively carry out a number of duties, such as

presenting the current date and time, playing music or movies, making phone calls, locating the current weather and temperature, searching the internet for information, and so on and so forth. This single sheet of paper has the ability to act as a template for a wide variety of additional applications that are far more complex. The findings of the literature review and the analysis of the existing system led us to the conclusion that the proposed system would not only make the process of interacting with other modules and programs easier, but it would also make it easier for us to keep the system's orderliness intact. As a result of this, we have come to the realization that the suggested system would be an improvement over the existing one. Even though there are still a lot of uncharted territories in the field of automation, the capabilities of devices can help us develop a new generation of voice-controlled devices and bring about a continuous new revolution in the industry of automation. Even though there are still a lot of unexplored territories in the field of automation. This body of work has the potential to act as a template for a wide variety of other applications that are more advanced. We make use of twelve typical commands and variants of those commands in order to train the system that has been proposed. Anyone who uses the system and is fluent in Bengali, regardless of their age or gender, will be able to interact with it by following the directions that are most frequently used in the language. This forward-thinking solution makes use of recent developments in voice recognition software as well as mobile phone networks to provide superior quality results. Because of the reduced percentage of people who own computers in rural areas, we came to the conclusion that the only viable platform for our system would be mobile, and more specifically, smart phones. Despite this, the findings highlighted how essential the pursuit of pleasure is in determining whether or not individuals will share a product with others. In their conclusion, they emphasize the necessity of having fun as a bridge between strictly utilitarian drivers (such as convenience and utility) and genuine engagement with the product. They say this bridge is necessary in order to facilitate genuine engagement.

## **6.2 Conclusion**

This article presents the design and implementation of a Bengali virtual assistant that goes by the name of "Saathi." This is where the specifics of the system for the virtual assistant are detailed. The Bengali command will be understood by the proposed system, and either laptops or desktop computers will be able to carry out the answer. The system is able to take commonly used commands in Bengali from persons of varying ages and genders, and it has an accuracy rate of over 98% in its responses to those orders. The most significant benefit is that it is able to learn on its own and does not rely on any external API. However, the most significant obstacle is the absence of a suitable label dataset written in Bengali. In order to train the suggested system, just twelve often used commands and their variants are used. To make the system more resilient, we need more complicated commands, each with their own unique variation. In addition, the application that we have built is only compatible with desktop computers; thus, it is necessary to create a mobile app for the Bengali virtual assistant. This is because smart phones are used more frequently in rural areas than desktop computers.

## **6.3 Implication for Further study**

The future of voice assistants seems bright from a technological and consumer standpoint. There is no way to deny the assumption that voice assistants are already an amazing example of human invention and will continue to develop into one in the foreseeable future. In point of fact, they are already beginning to make their way into our regular lives in some fashion or another. In the future we will collect more data from different people with different accent. For this reason, our project “সাথী” will help more people and engaged with more people. We will also build an application for both android and IOS to use it more easily.

## REFERENCES

- [1] Chowdhury, Saadman Shahid, et al. "Domain specific intelligent personal assistant with bilingual voice command processing." *TENCON 2018-2018 IEEE Region 10 Conference*. IEEE, 2018.
- [2] Burbach, Laura, et al. "" Hey, Siri", " Ok, Google", " Alexa". Acceptance-Relevant Factors of Virtual Voice-Assistants." *2019 IEEE International Professional Communication Conference (ProComm)*. IEEE, 2019.
- [3] M.N. Sabab, M.A.R. Chowdhury, S.M.I. Nirjhor, J. Uddin, "Bangla speech recognition using 1D-CNN and LSTM with different dimension reduction techniques", *International Conference for Emerging Technologies in Computing* pp. 158-169, Springer, Cham August 2020.
- [4] O. Sen, M. Fuad, M.D.Islam, J. Rabbi, M.D. Hasan, M.Baz, M. Masud, M. Awal, A.A. Fime, M. Fuad, and T. Hasan, M.D. Iftee, " Bangla Natural Language Processing: A Comprehensive Review of Classical, Machine Learning, and Deep Learning Based Methods" arXiv preprint arXiv:2105.14875 2021.
- [5] Poushneh, Atieh. "Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors." *Journal of Retailing and Consumer Services* 58 (2021): 102283.
- [6] Nasirian, Farzaneh, Mohsen Ahmadian, and One-Ki Daniel Lee. "AI-based voice assistant systems: Evaluating from the interaction and trust perspectives." (2017).
- [7] Beirl, Diana, Y. Rogers, and Nicola Yuill. "Using voice assistant skills in family life." *Computer-Supported Collaborative Learning Conference, CSCL*. Vol. 1. International Society of the Learning Sciences, Inc., 2019.
- [8] Tuzovic, Sven, and Stefanie Paluch. "Alexa-What's on my shopping list? Investigating consumer perceptions of voice-controlled devices." *SERVSIG Conference Proceedings 2018 Paris: Opportunities for services in a challenging world*. SERVSIG, 2018.
- [9] Tractica (2020). Tractica. <https://tractica.omdia.com/newsroom/press-releases/voice-and-speech-recognition-software-market-to-reach-6-9-billion-by-2025/>.
- [10] Yan, Chen, et al. "A Survey on Voice Assistant Security: Attacks and Countermeasures." *ACM Computing Surveys* 55.4 (2022): 1-36.
- [11] Haas, Gabriel, et al. "Keep it Short: A Comparison of Voice Assistants' Response Behavior." *CHI Conference on Human Factors in Computing Systems*. 2022.

# Alex Sarker

## ORIGINALITY REPORT

11%

SIMILARITY INDEX

9%

INTERNET SOURCES

4%

PUBLICATIONS

6%

STUDENT PAPERS

## PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	<a href="https://dspace.daffodilvarsity.edu.bd:8080">dspace.daffodilvarsity.edu.bd:8080</a> Internet Source	1%
3	<a href="https://chatbotsjournal.com">chatbotsjournal.com</a> Internet Source	1%
4	<a href="https://www.voicesummit.ai">www.voicesummit.ai</a> Internet Source	1%
5	S Subhash, Prajwal N Srivatsa, S Siddesh, A Ullas, B Santhosh. "Artificial Intelligence-based Voice Assistant", 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4), 2020 Publication	1%
6	Submitted to Ahsanullah University of Science and Technology Student Paper	<1%
7	<a href="https://medium.com">medium.com</a> Internet Source	<1%