**SHRI G.S. INSTITUTE OF TECHNOLOGY AND SCIENCE INDORE (M.P.)**



## DECLARATION

I,

Brijeshwari Harode declare that this project report titled, “**You Tube Video**

**player assistant**” and the work presented in it are my own. We confirm that:

* This work was wholly while in candidature for a master’s degree at this University.
* No part of this project has previously been submitted for a degree or any other qualification at this University or any other institution.
* Where we have consulted the published work of others, this is always clearly at tributed.
* Where we have quoted from the work of others, the source is always given. With the exception of such quotations, this project is entirely our own work.
* We have acknowledged all main sources of help.

Signed: Date:

**SHRI G.S. INSTITUTE OF TECHNOLOGY AND SCIENCE INDORE (M.P.)**



## RECOMMENDATION

The project report entitled “**You tube video assistant**” submitted by **Neeraj Malviya**, students of MCA final year in the session 2023-24, towards the fulfillment of the degree of **Master of Computer Applications** of Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal is a satisfactory account of their work and is recommended for the award of degree.

**Mr. Upendra Singh Dr. K. K. Sharma**

**Project Guide Head of Department**

Department of Information Department of Information Technology Technology

**SHRI G.S. INSTITUTE OF TECHNOLOGY AND SCIENCE INDORE (M.P.)**



## CERTIFICATE

The project report entitled “Linear regression Model” submitted by **Neeraj Malviya** students of MCA final year in the session 2023-24, towards partial fulfillment of the degree of **Master of Computer Applications** of Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, is a satisfactory account of their work and is approved for the award of the degree.

**Internal Examiner External Examiner**

Date: Date:

**SHRI G.S. INSTITUTE OF TECHNOLOGY AND SCIENCE**

**INDORE (M.P.)**



## ACKNOWLEDGEMENT

Every work accomplished is a pleasure and a sense of satisfaction, however a number of people always motivate, criticize and appreciate a work with their objective ideas and opinions, hence we are heartily pleased to acknowledge all those people who have helped us in the successful completion of this project. With great pleasure, we express our heartfelt gratitude to our esteemed guide, **Mr. Upendra Singh**, Professor Department of Computer Technology & Application, S.G.S.I.T.S. Indore. Their persistent encouragement, perpetual motivation, everlasting patience, and valuable technical inputs in discussions have enabled the successful completion of this project. Their invaluable help, advice, and constant encouragement helped us a lot and provided the impetus to the progress of the project. We extend our profound indebtedness to the Head of the department, **Dr. K.K. Sharma**, the words lose their worth for his invaluable guidance, continuous encouragement, and cooperation in every respect.

We sincerely wish to express our gratitude to all the members of staff of M.C.A. who have extended their cooperation at all times and have contributed in their own way to developing the project. Successful completion of a project is not an individual effort. It is an outcome of the cumulative effort of a number of persons, each having his own importance to the objective. We are thankful to our parents for being a constant source of encouragement in all our endeavors. Indeed, it is their support that helps us through the ups and downs of life. The support and suggestions of our friends are worth appreciation and thankfulness. A blend of gratitude, pleasure, great satisfaction and indebtedness is what we feel to convey to all those who have directly or indirectly contributed to the successful completion of our project work.

**SHRI G.S. INSTITUTE OF TECHNOLOGY AND SCIENCE**

**INDORE (M.P.)**



###### ABSTRACT

The project aims to develop a You tube video player Assistant for windows based system.

Assistant draws its inspiration from virtual assistants like Cortana for windows and Siri for iOS.It has been designed to provide a user friendly interface for carrying out a variety of tasks by employing certain well-defined commands.Users can interact with the YouTube Video assistant through voice commands .Youtube video player assistant takes the voice input through our microphone and it converts our voice into computer understandable language and gives the required solutions and answers which are asked by user has questioned.This connect with youtube to provide result that user has questioned .This project works on voice input and gives output through playing video on screen .The main agenda of our youtube video assistant is that it makes people smart and give instant and computed results.

**TABLE OF CONTENTS**

**Chapter No. TITLE Page No.**

**ABSTRACT** v

**INTRODUCTION** 1

1.2. DESIGN 3

1.3.You Tube video ASSISTANT 3

1.3.1 WHAT IS VOICE assistant 3

1.3.2 WHY DO WE NEED IT 3

1.3.3 WHERE TO USE IT

**LITERATURE SURVEY 10**

2.1. RELATED WORK 5

**3 METHODOLOGY 12**

3.1. EXISTING SYSTEM

3.2. PROPOSED SYSTEM 8

3.3. OBJECTIVE OF THE PROJECT 10

3.4. SOFTWARE AND HARDWARE REQUIREMENT 12

3.4.1. SOFTWARE REQUIREMENTS 14

3.4.2. HARDWARE REQUIREMENTS 15

3.4.3. LIBRARIES

3.5. PROGRAMMING LANGUAGES 16

3.5.1. PYTHON 17

3.6. SYSTEM ARCHITECTURE 18

3.7. ALGORITHMS USED 19

3.7.1. SPEECH RECOGNITION MODULE

3.7.2. SPEECH TO TEXT & TEXT TO SPEECH CONVERSION

3.7.3. PROCESS & EXECUTES THE

REQUIRED COMMAND

3.8. SYSTEM DESIGN 21

3.8.1 USE CASE DIAGRAM 22

3.8.2 COMPONENT DIAGRAM 23

3.8.3 SEQUENCE DIAGRAM 24

**4 RESULTS AND DISCUSSION 27**

4.1. WORKING 27

**5 CONCLUSION 28**

5.1. CONCLUSION 25

5.2. FUTURE WORK 26

**APPENDICES** 27

A. SOURCE CODE 31

B. SCREENSHOTS 33

**REFERENCES**

##### 

**Chapter 1**

**Introduction**

* + The very first voice activated product was released in 1922 as Radio Rex. This toy was very simple, wherein a toy dog would stay inside a dog house until the user exclaimed its name, “Rex” at which point it would jump out of the house. This was all done by an electromagnet tuned to the frequency similar to the vowel found in the word Rex, and predated modern computers by over 20 years
  + In the 21st century, human interaction is being replaced by automation very quickly. One of
  + the main reasons for this change is performance. There’s a drastic change in technology rather
  + than advancement. In today’s world, we train our machines to do their tasks by themselves or to
  + think like humans using technologies like Machine Learning, Neural Networks, etc. Now in the
  + current era, we can talk to our machines with the help of virtual assistants.
  + Virtual assistants are software programs that help you ease your day to day tasks, such as
  + showing weather reports, giving daily news, searching the internet etc. They can take
  + commands by voice. Voice-based intelligent assistants need an invoking word or wake word to
  + activate the listener, followed by the command. We have so many virtual assistants, such as
  + Apple’s Siri, Amazon’s Alexa and Microsoft’s Cortana and Amazon's Alexa and this has been
  + an inspiration for us to do this as a project. This system is designed to be used efficiently on
  + Youtube Voice assistants are programs on digital devices that listen and respond to verbal

###### 

**1.2 DESIGN**

a) The You tube voice assistant takes an input word which is called as "signal word" to

be activated. so, it takes in the signal word and starts operating for the

user commands.

b) Converting the speech into text will be processed by the assistant.

c) The converted text is now processed to get the required results.

d) The text given by the user should contain one or two keywords that

determine what query is to be executed. If the keyword doesn’t match any

of the queries in the code then the assistant asks the user to speak again.

e) Finally, the output to the user's query will be given by converting speech

to text

**1.3.Youtube video VOICE ASSISTANT**

Our assistant “NOVA” extends to helps us when working on a system in which it is

installed. We can access by calling the wake word “alexa".

***1.3.1 WHAT IS VOICE ASSISTANT***

A voice assistant, also known as an intelligent personal assistant or a connected speaker,

is a new type of device that is based on natural language speech recognition and is

offered by popular companies like Apple, Amazon, and Google. We got inspired by that

and created one our self.

***1.3.2 WHY DO WE NEED IT***

Usually, typing out and searching or doing day-to-day tasks becomes hectic. But our life

doesn’t need to be like that. One can ask for help to voice assistants. They let the users to

perform a task using a speech command, as well as retrieve information via voice

synthesis

Following are the reasons to have a voice assistant.

Minimal Effort

It’s easier to say a few words than type them on a small smartphone screen.

Eyes Free

One can be as blind as a bat, but a voice assistant will always help you. Our ears are

enough. One can also ask the bot about something while cooking at the same time.

Fast response

Imagine how much time you have to spend to find some information on a website? Or

how many clicks do you need to make before you find the thing you need in a mobile

application? Voice assistants don’t generate such difficulties. One can ask a question

and you have the answer.

**CHAPTER 2**

**LITERATURE SURVEY**

**2.1 RELATED WORK**

This field of virtual assistants having speech recognition has seen some major advancements

or innovations. This is mainly because of its demand in devices like smartwatches or fitness

bands, speakers, Bluetooth earphones, mobile phones, laptop or desktop, television, etc.

Almost all the digital devices which are coming nowadays are coming with voice assistants

which help to control the device with speech recognition only. A new set of techniques is being

developed constantly to improve the performance of voice automated search.

As the amount of data is increasing exponentially now known as Big Data the best way to

improve the results of virtual assistants is to incorporate our assistants with machine learning

and train our devices according to their uses. Other major techniques that are equally important

are Artificial Intelligence, Internet of Things, Big Data access and management, etc. With the

use of voice assistants, we can automate the task easily, just give the input to the machine in

the speech form and all the tasks will be done by it from converting your speech into text form to

taking out keywords from that text and execute the query to give results to the user.

Machine Learning is just a subset of Artificial Intelligence. This has been one of the most helpful

advancements in technology. Before AI we were the ones who were upgrading technology to do

a task but now the machine is itself able to counter new tasks and solve it without need to

involve the humans to evolve it.

This has been helpful in day-to-day lifestyle. From mobile phones to personal desktops to

mechanical industries these assistants are in very much demand for automating tasks and

increasing efficiency

**P.Krishnaraj** (2021) et al. presented a project on Portable Voice Recognition with GUI

Automation, This system uses Google’s online speech recognition system for converting

speech input to text along with Python. Therefore, this project has a GUI and is also has a

portable framework. Accuracy of this text to speech (TTS) engine is comparatively less and

also lacks IoT.

**CHAPTER 3**

**METHODOLOGY**

**3.1 EXISTING SYSTEM**

**3.2 PROPOSED SYSTEM**

We are proposing a system in an efficient way of implementing a Personal

voice assistant, Speech Recognition library has many in-built functions, that will let

the assistant understand the command given by user and the response will be

sent back to user in voice, with Text to Speech functions. When assistant captures

the voice command given by user, the under lying algorithms will convert the voice

into text. And according to the keywords present in the text (command given by

user), respective action will be performed by the assistant.

**3.3 OBJECTIVE OF PROJECT**

Main objective of building personal assistant software for Youtube video search(a virtual assistant) is using

semantic data sources available on the web, user generated content and providing knowledge

from knowledge databases. The main purpose of an intelligent virtual assistant is to answer

questions that users may have. This may be done in a business environment, for example, on

the business website, with a chat interface. On the mobile platform, the intelligent virtual

assistant is available as a call-button operated service where a voice asks the user “What can I

do for you?” and then responds to verbal input. Virtual assistants can tremendously save you

time. We spend hours in online research and then making the report in our terms of

understanding.

**functionality**

The proposed system will have the following functionality:

(a) The system will keep listening for commands and the time for listening is

variable which can be changed according to user requirements.

(b) If the system is not able to gather information from the user input it will

keep asking again to repeat till the desired number of times.

(c) Features supported in the current version include playing music,any video on youtube.

**3.4 SOFTWARE AND HARDWARE REQUIREMENTS**

***3.4.1 Software Requirements:***

1. Python 3.5 & Above
2. Windows 7 And Above
3. Visual studio or google colab

***3.4.2 Hardware Requirements:***

1. Processor: Intel Core i5
2. RAM: 4GB
3. OS:Windows / Mac
4. Microphone

***3.4.3 Libraries:***

**Pyttsx3-** It is a text to speech conversion library in python which is used to

convert the text given in the parenthesis to speech. It is compatible with

python 2 and 3. An application invokes the pyttsx3.init() factory function to get

a reference to a pyttsx3. it is a very easy to use tool which converts the

entered text into speech. The pyttsx3 module supports two voices first is

female and the second is male which is provided by “sapi5” for windows.

Command to install: - pip install pyttsx3

It supports three TTS engines: -

sapi5- To run on windows

nsss - NSSpeechSynthesizer on Mac OS X

espeak – eSpeak on every other platform

**Speech\_recognition-** It allows computers to understand human language.

Speech recognition is a machine's ability to listen to spoken words and

identify them. We can then use speech recognition in Python to convert the

spoken words into text, make a query or give a reply. Python supports many

speech recognition engines and APIs, including Google Speech Engine,

Google Cloud Speech API.

Command to install :- pip install SpeechRecognition

**Datetime**- This module is used to get the date and time for the user. This is a built-in

module so there is no need to install this module externally. Python Datetime module

supplies classes to work with date and time. Date and datetime are an object in

Python, so when we manipulate them, we are actually manipulating objects and not

string or timestamps.

**pyttsx3** Text to Speech (TTS) library for Python 2 and 3. Works without internet connection or delay. Supports multiple TTS engines, including Sapi5, nsss, and espeak.

[**PyWhatKit**](https://pypi.org/project/pywhatkit/) is a Python library with various helpful features. It's easy-to-use and does not require you to do any additional setup. Currently, it is one of the most popular library for WhatsApp and YouTube automation. New updates are released frequently with new features and bug fixes.

**3.5. PROGRAMMING LANGUAGES**

***3.5.1 PYTHON***

Python is an OOPs (Object Oriented Programming) based, high level,

interpreted programming language. It is a robust, highly useful language focused

on rapid application development (RAD). Python helps in easy writing and

execution of codes. Python can implement the same logic with as much as 1/5th

code as compared to other OOPs languages. Python provides a huge list of

benefits to all. The usage of Python is such that it cannot be limited to only one

activity. Its growing popularity has allowed it to enter into some of the most

popular and complex processes like Artificial Intelligence (AI), Machine Learning

(ML), natural language processing, data science etc. Python has a lot of libraries

for every need of this project. For this project, libraries used are speech

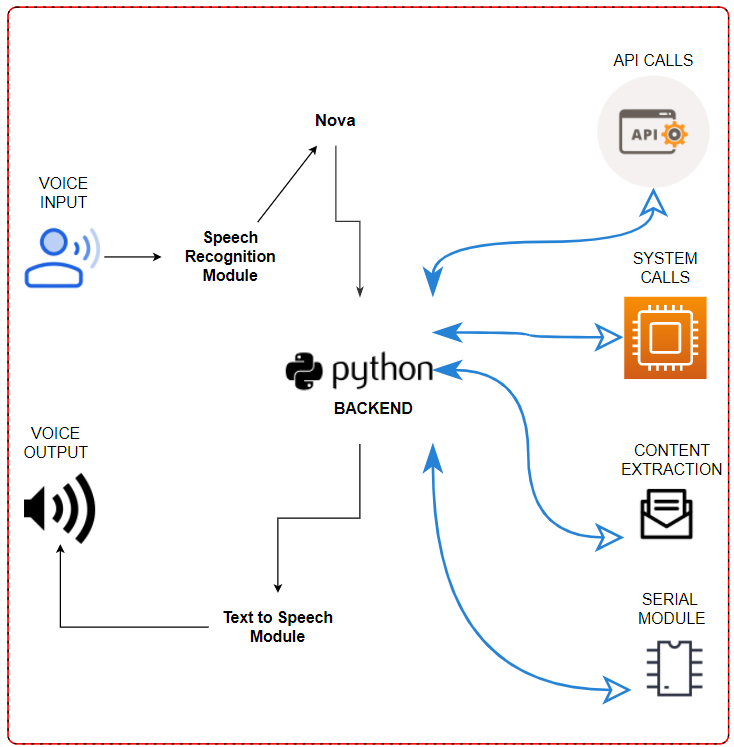
recognition to recognize voice, Pyttsx for text to speech, selenium for web

automation etc.

It’s owing to the subsequent strengths that Python has

**\**

**3.5.2 SYSTEM ARCHITECTURE**

****

**3.7 ALGORITHMS USED**

***3.7.1 SPEECH RECOGNITION MODULE***

1. The class which we are using is called Recognizer.
2. It converts the audio files into text and module is used to give the output in speech.

**Energy threshold** function represents the energy level threshold for sounds. Values

below this threshold are considered silence, and values above this threshold are

considered speech.17

Recognizer instance.adjust\_for\_ambient\_noise(source, duration = 1), adjusts the

energy threshold dynamically using audio from source (an AudioSource instance) to

account for ambient noise.

***3.7.2 SPEECH TO TEXT & TEXT TO SPEECH CONVERSION***

1. Pyttsx3 is a text-to-speech conversion library in Python. And can change the Voice,
2. Rate and Volume by specific commands.
3. Python provides an API called Speech Recognition to allow us to convert audio into
4. text for further processing converting large or long audio files into text using the
5. Speech Recognition API in python.
6. We have Included sapi5 and espeak TTS Engines which can process the same.

***3.7.3 PROCESS & EXECUTES THE REQUIRED COMMAND***

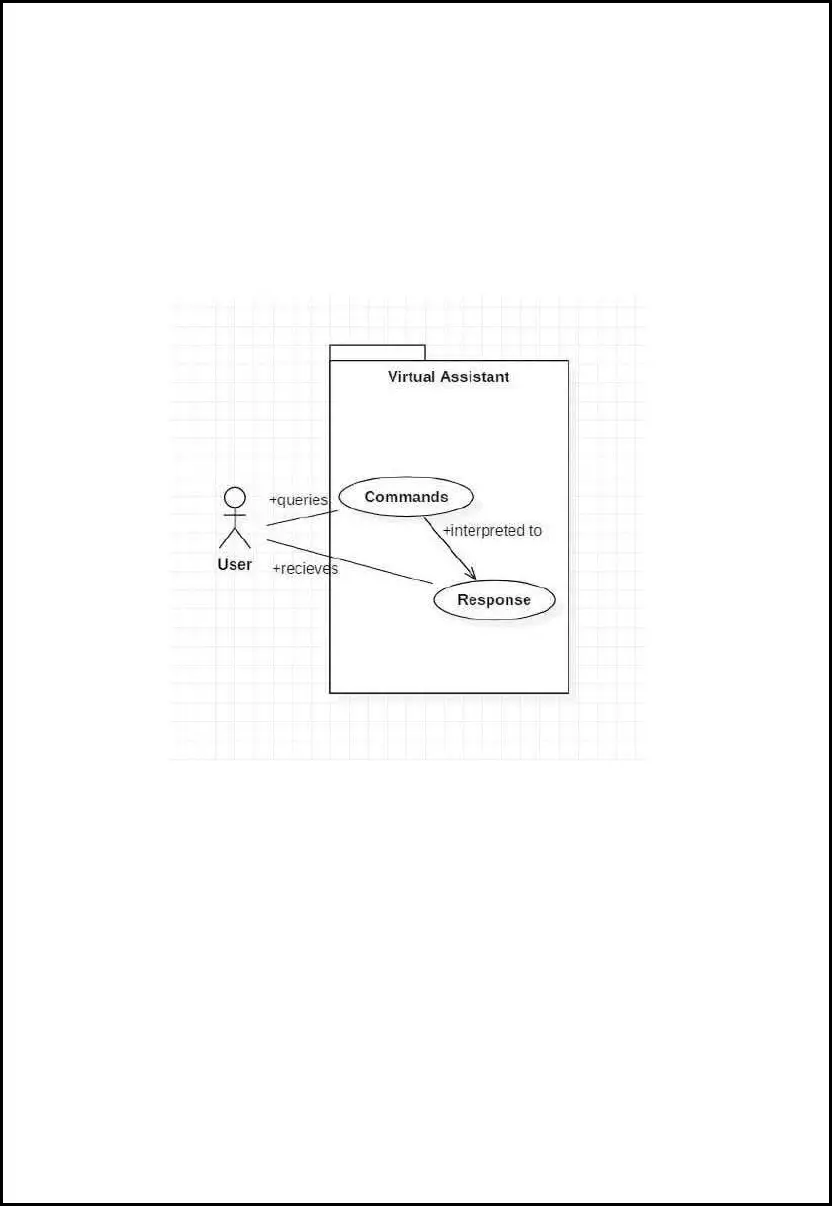
The said command is converted into text via speech recognition module and further

stored in a temp.

Then, Analyze the user’s text via temp and decide what the user needs based on

input provided and runs the if conditions

Then, Commands are executed



**Use case diagram**

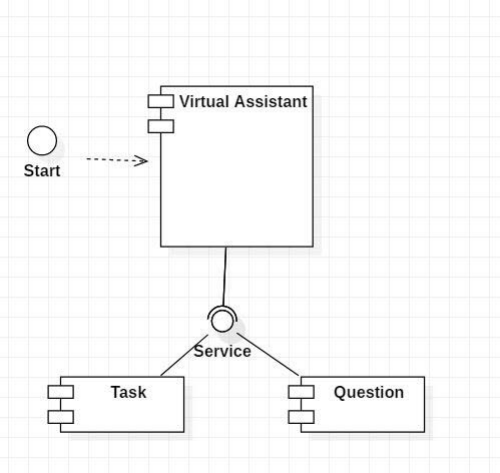
In this project there is only one user. The user queries command to the system. System

then interprets it and fetches answer. The response is sent back to the user**.**

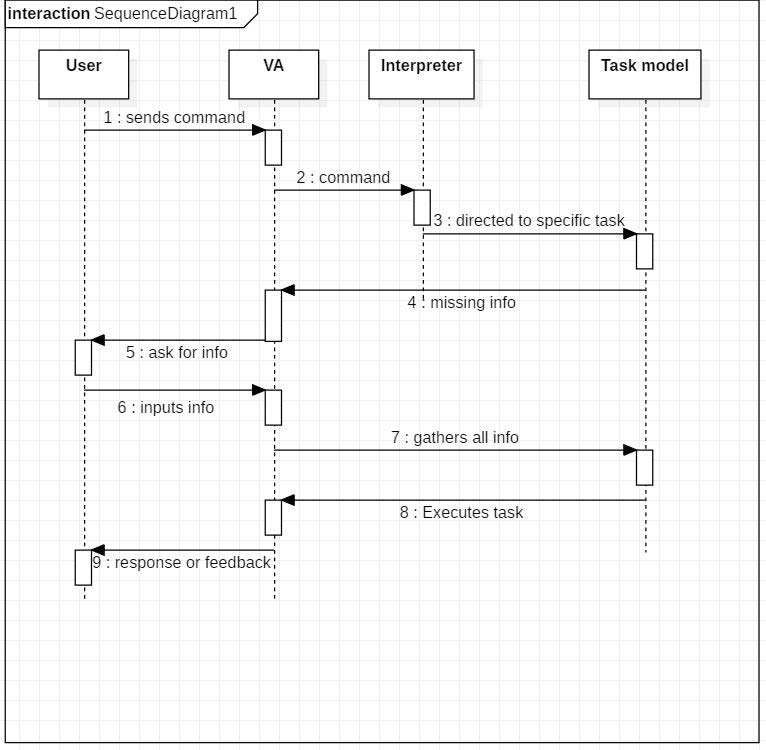
**3.8.2 COMPONENT DIAGRAM:**

**You tube video player assistant**



****

**SEQUENCE DIAGRAM:**

****

**Fig 3.4. Sequence Diagram**

**CHAPTER 4**

**RESULTS AND DISCUSSION**

The project work of the voice assistant has been clearly explained in this report, how

useful it is and how we can rely on a you tube video voice assistant for performing any/every task which the

user needs to complete and how the assistant is developing everyday which we can hope

that it'll be one of the biggest technology in the current technological world. Development of

the software is almost completed form our side and it's working fine as expected which was

discussed for some extra development. So, maybe some advancement might come in the

near future where the assistant which we developed will be even more useful than it is now

**4.1. WORKING**

It starts with a signal word. Users say the names of their voice assistants for the same reason.

They might say, “Hey Siri!” or simply, “Alexa!” Whatever the signal word is, it wakes up the

device. It signals to the voice assistant that it should begin paying attention. After the voice

assistant hears its signal word, it starts to listen. The device waits for a pause to know you’ve

finished your request. The voice assistant then sends our request over to its source code. Once

in the source code, our request is compared to other requests. It’s split into separate commands

that our voice assistant can understand. The source code then sends these commands back to

the voice assistant. Once it receives the commands, the voice assistant knows what to do next.

If it understands, the voice assistant will carry out the task we asked for. For example, “Hey

devices receive, the better and faster they get at fulfilling our requests. The user gives the voice

input through microphone and the assistant is triggered by the wake up word and performs the

STT (Speech to Text) and converts it into a text and understands the Voice input and further

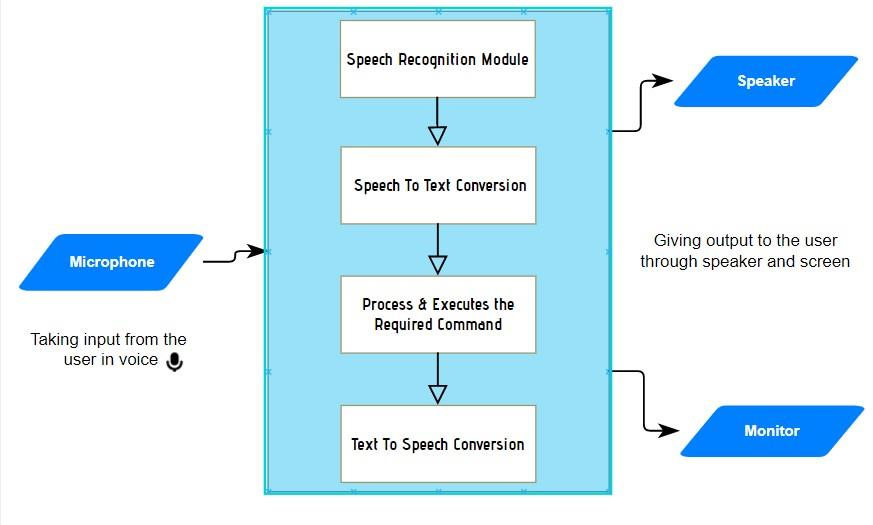
performs the task said by the user repeatedly and delivers it via TTS (Text to Speech) module

**CONTENT EXTRATION**

This can Perform content extraction from YouTube, using the pywhatkit

web driver module from youtube which provides all the implementations for the youtube like

searching for a specific video to play, to get a specific information from youtube



**You tube video player assistant**

1) Must provide the user any information which they ask for: -

The user might need any information which will be available on the youtube but searching

for that information and reading that takes a lot of time but with the help of a voice assistant, we

can complete that task of getting the information sooner than searching and reading it. So, this

is a small proof that a voice assistant help

**CHAPTER 5**

**CONCLUSION**

**5.1. CONCLUSION**

As stated before, "you tube voice assistant is one of the biggest problem solver" and you can see that in

the proposals with the examples that it is in fact one of the biggest problem solver of the current

world. We can see that voice assistant is one of the major evolving artificial intelligence in the

current world once again on seeing the proposal examples because at the past, the best feature

which a voice assistant had was telling the date and searching the web and giving the results

but now look at the functions that it can do so with this, we can say that it is a evolving software

in the current world. The main idea is to develop the assistant even more advanced than it is

now and make it the best ai in the world which will save an ample of time for its users. I would

like to conclude with the statement that we will try our best and give one of the best voice

assistants which we are able to.

**5.2. FUTURE SCOPE**

We are entering the era of implementing voice-activated technologies to remain relevant and

competitive. Voice-activation technology is vital not only for businesses to stay relevant with their target

customers, but also for internal operations. Technology may be utilized to automate human operations,

saving time for everyone. Routine operations, such as sending basic emails or scheduling appointments,

can be completed more quickly, with less effort, and without the use of a computer, just by employing a

simple voice command. People can multitask as a result, enhancing their productivity. Furthermore,

relieving employees from hours of tedious administrative tasks allows them to devote more time to

strategy meetings, brainstorming sessions, and other jobs that need creativity and human interaction.

1) Sending Emails with a voice assistant:

Emails, as we all know, are very crucial for communication because they can be used for any

professional contact, and the finest service for sending and receiving emails is, as we all know, GMAIL.

Gmail is a Google-created free email service. Gmail can be accessed over the web or using third-party

apps that use the POP or IMAP protocols to synchronize email content.

To integrate Gmail with Voice Assistant we have to utilize Gmail API. The Gmail API allows you to

access and control threads, messages, and labels in your Gmail mailbox

**APPENDICES**

**Source Code**

**import speech\_recognition as sr**

**import pyttsx3**

**import pywhatkit**

**import datetime**

**listener=sr.Recognizer()**

**engine=pyttsx3.init()**

**voices=engine.getProperty('voices')**

**engine.setProperty('voice',voices[1].id)**

**def talk(text):**

**engine.say(text)**

**engine.runAndWait()**

**def take\_command():**

**try:**

**with sr.Microphone() as source:**

**print('listening...')**

**voice=listener.listen(source)**

**command=listener.recognize\_google(voice)**

**command=command.lower()**

**if'alexa'in command:**

**command=command.replace('alexa','')**

**print(command)**

**except:**

**pass**

**return command**

**def run\_alexa():**

**command=take\_command()**

**print(command)**

**if'play' in command:**

**song=command.replace('play','')**

**talk('playing' + song)**

**pywhatkit.playonyt(song)**

**elif'time' in command:**

**time = datetime.datetime.now().strftime('%I:%M %p')**

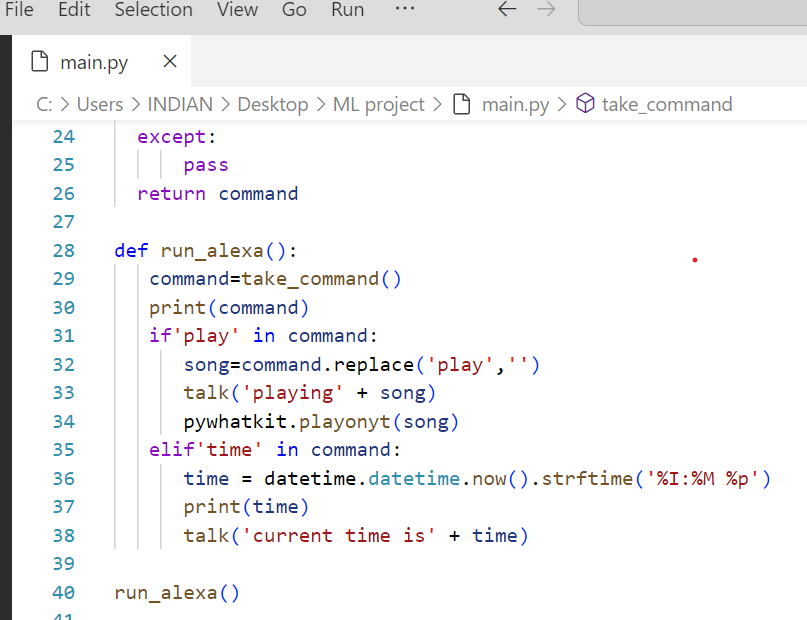
**print(time)**

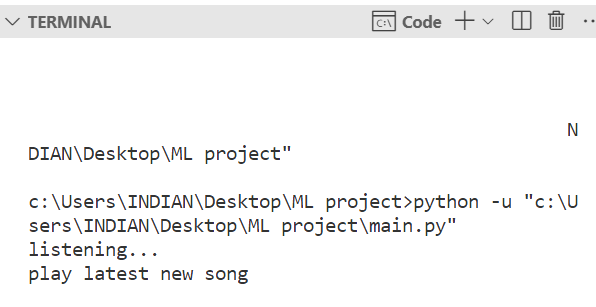
**talk('current time is' + time)**

**run\_alexa()**

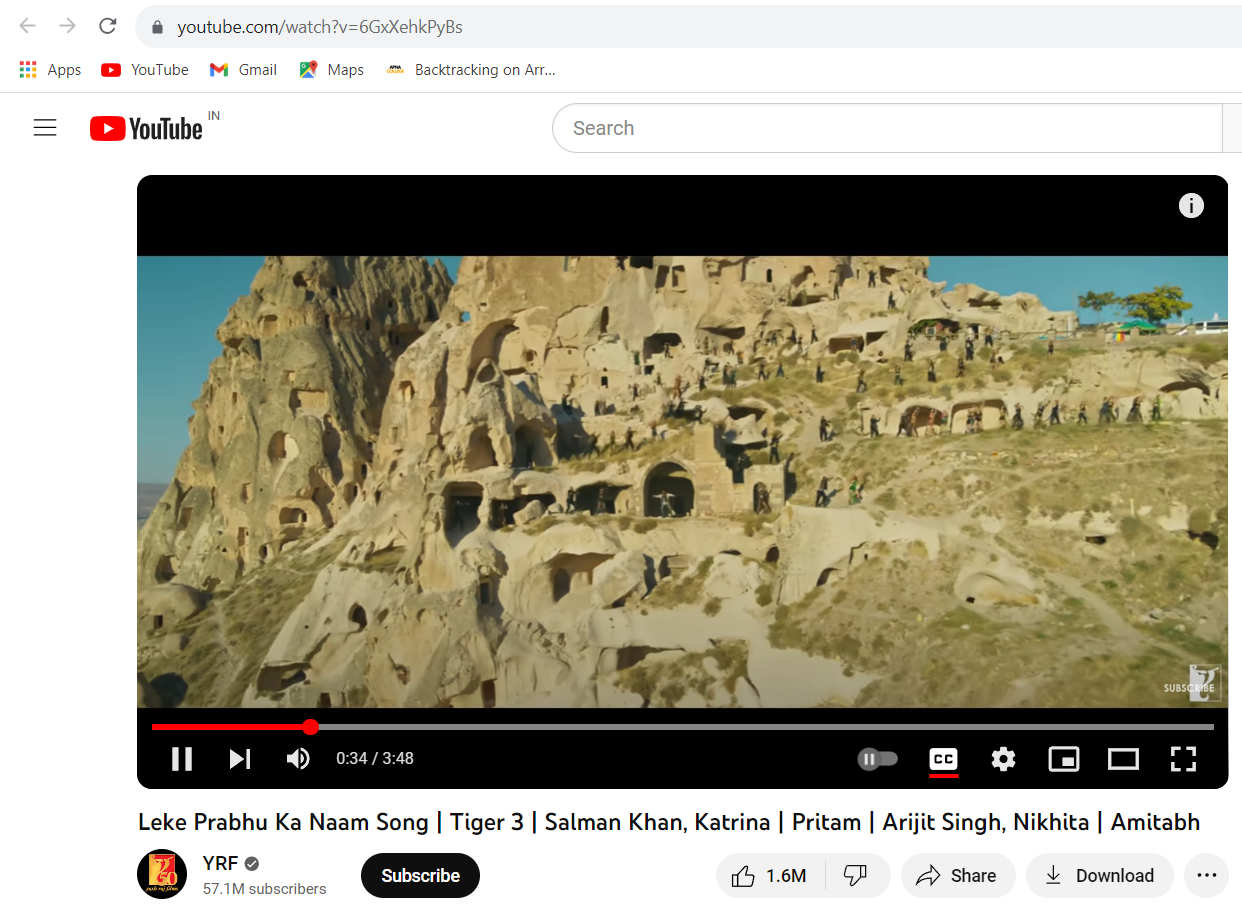
**Screenshots**

****

****

****

**Play video on youtube screenshot**

****

**REFERENCES**

1 Geeks for geeks

<https://www.geeksforgeeks.org/python-programming-language/>

2.pywhatkit 5.4

<https://pypi.org/project/pywhatkit/>

3.speech recognition

<https://pypi.org/project/SpeechRecognition/>

4.pyttsx 3

<https://pypi.org/project/pyttsx3/>