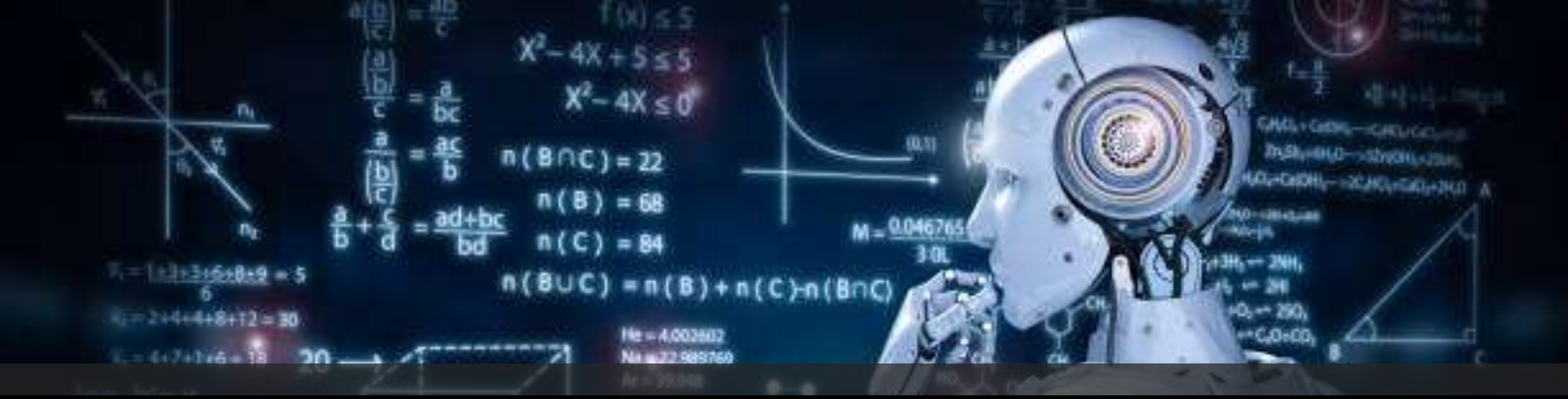


# Lip To Speech…

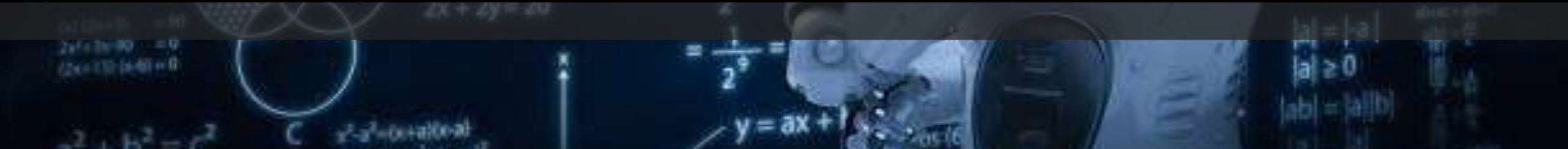
## Deep Learning

---


Afeef 2AB19CS037

Nayil 2AB19CS008

Guide: Prof. Rizwan Sheikh

# Content

- Company Profile
- Introduction
- Problem Statement
- Objective
- Flow Chart/ Architecture
- Challenges
- Maximizing Online Tools and Resources for a Successful Internship Experience
- How Lip To Speech Synthesis Works
- Software and Hardware requirements
- References

# COMPANY PROFILE

Varcons Technologies Pvt Ltd, was incorporated with a goal, “To provide high quality and optimal Technological Solutions to business requirements to their clients”.

Varcons Technologies Pvt Ltd is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Varcons Technologies Pvt Ltd specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

# Products of Varcons Technologies

**Android Apps**

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java and/or Kotlin.

**Web Application**

It is a client–server computer program in which the client (including the user interface and client- side logic) runs in a web browser.

**Web design**

It is encompasses many different skills and disciplines in the production and maintenance of websites.

# Services provided by Varcons Technologies

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On Job Training
- Software Train

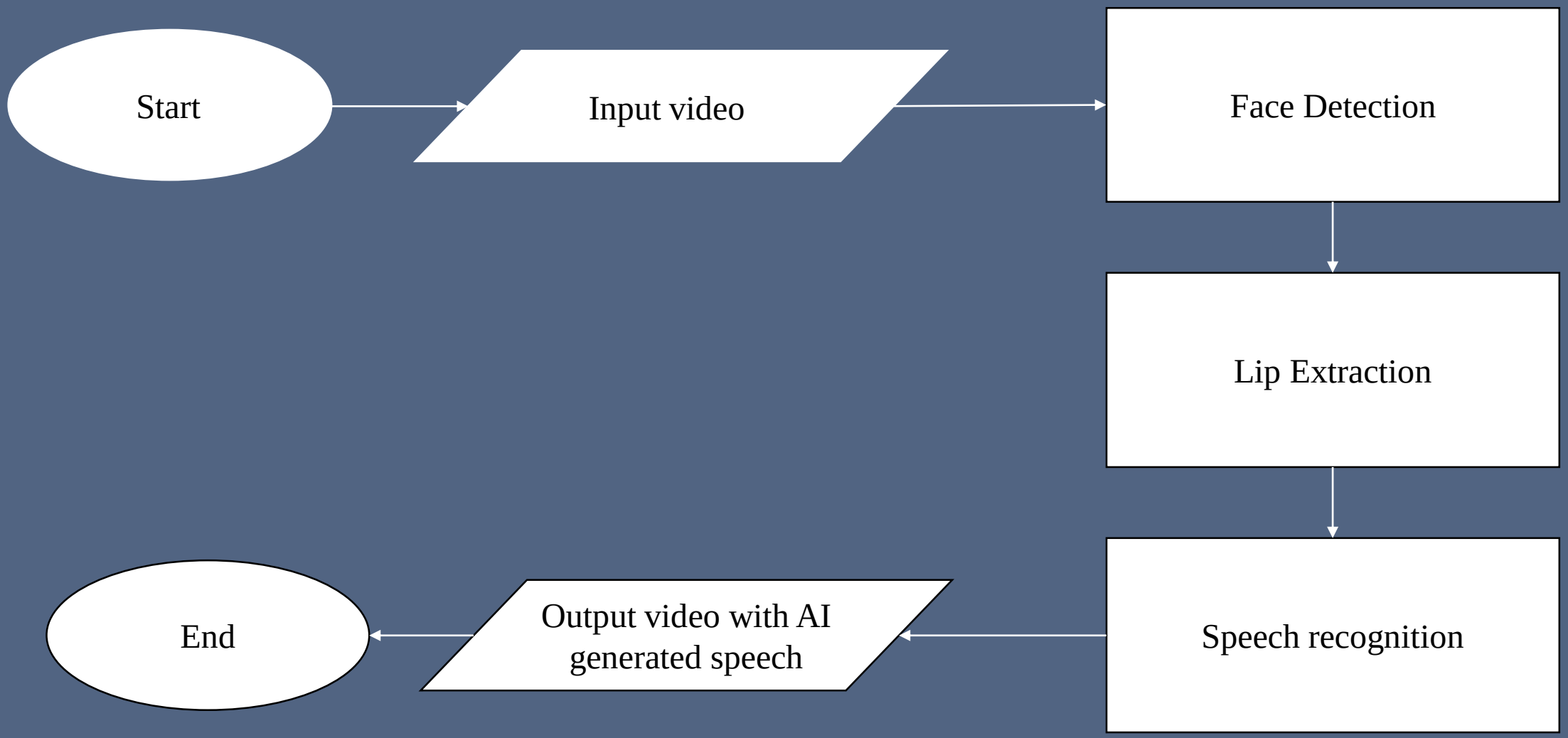
# Introduction

- We as humans pay high attention to lip movements to “visually hear” the speech in highly noisy environments. Facial actions, specifically the lip movements, thus reveal a useful amount of speech information.
- Our project motivation is to use the same thing as humans to create the speech based on the lip movements of the speaker in the video.

# Problem Statement

Learning the speech pattern of the speaker based on lip movement using CNN

# Lip To Speech Flowchart



# Challenges

- Lack of recognition: Despite doing good work, interns may not always receive the recognition they deserve, which can be discouraging.
- It’s our first job kind of experience. Reaching late for classes and late submissions of assignments was tolerated by our college professors. But in a professional workspace it isn’t acceptable and hence managing time can be really difficult for us.
- Balancing work, academics, and personal life can be challenging, especially for interns who are still figuring out how to manage their time effectively in a professional workspace.
- We accepted an internship thinking that there won’t be much pressure and the compensation would be sufficient for it. But after we actually start, we realize the pay is far lesser than the work deserves.
- Communication is one of those internship challenges that can actually hinder our internship time.

# Making the Most of Online Resources: Our Journey to Completing a Challenging Task

- W3Schools is a popular online resource for learning web development, and we used it to learn various concepts and techniques related to web development.
- YouTube is another great resource that we used to watch video tutorials and learn from experts in the field.
- Online forums like Stack Overflow were also helpful in troubleshooting issues and finding solutions to problems we encountered during our work.
- We also used online tools like code editors, color palette generators, and image compression tools to streamline our workflow and improve our productivity.
- By using these resources, we were able to learn new skills, solve problems more efficiently, and ultimately complete our task successfully.

# Requirements

## Software

Frameworks:
1. Django
2. Flutter
3. Tensorflow
4. dlib

Programming Languages:
1. Python 3.10.x
2. Dart

Tools:
Visual Studio Code,
Jupyter Notebook,

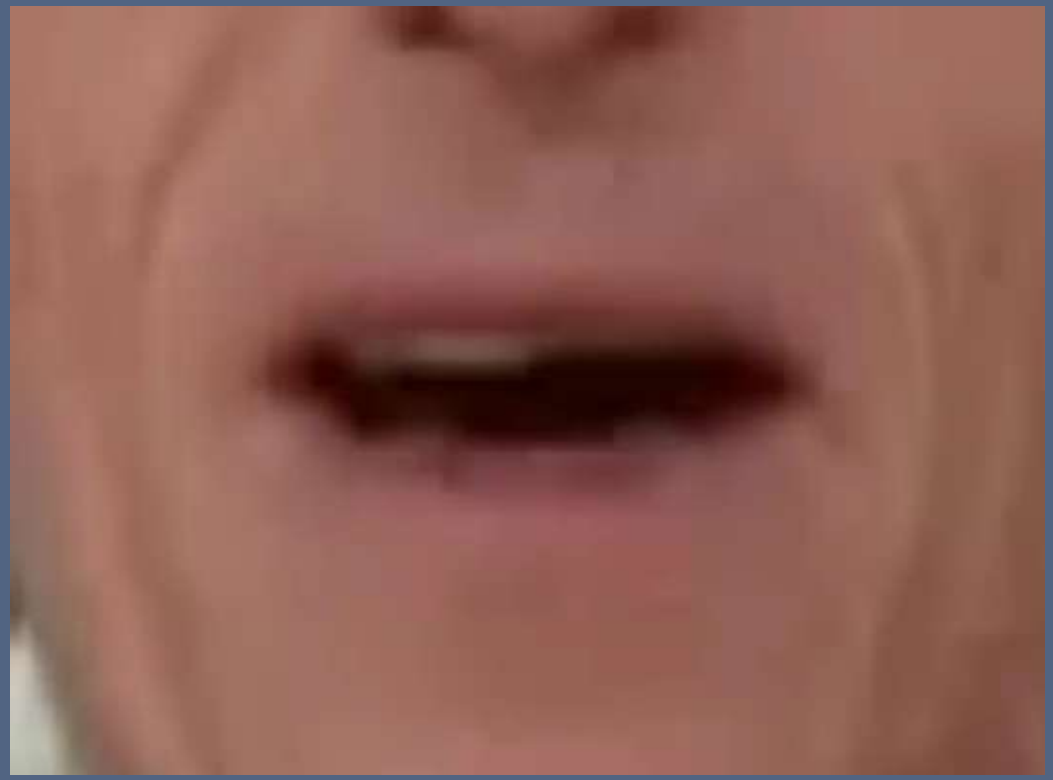
Operating System:
Windows Native - Windows 7 or higher (64-bit)

## Hardware

1. Intel Core i3 @3.7GHz or higher .
2. 8GB of RAM
3. NVIDIA® GPU card with CUDA® architectures 3.5, 5.0, 6.0, 7.0, 7.5, 8.0 and higher.

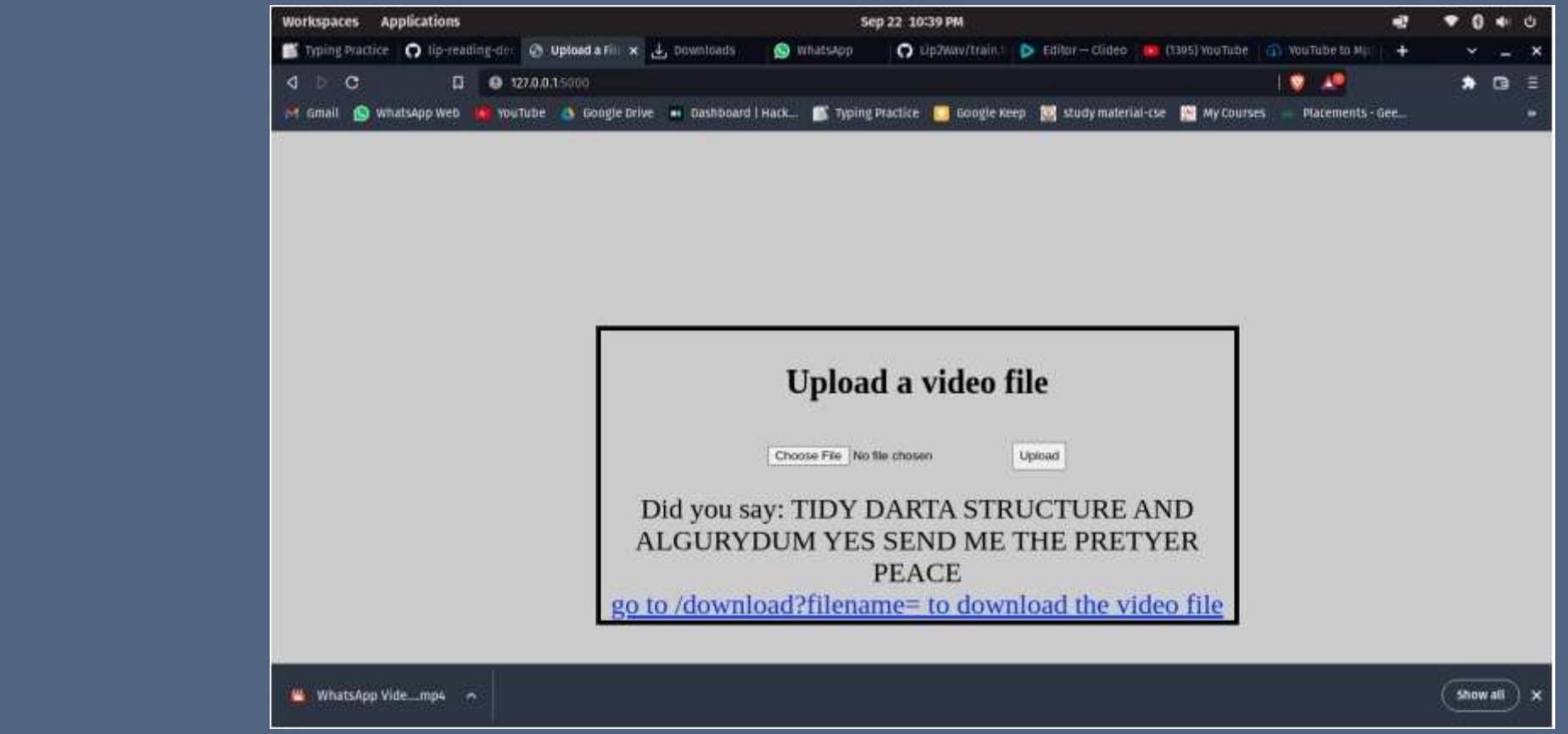
| Video Input | Processed Input | Speech Output |
| --- | --- | --- |
|  |  |  |

## Ground Truth

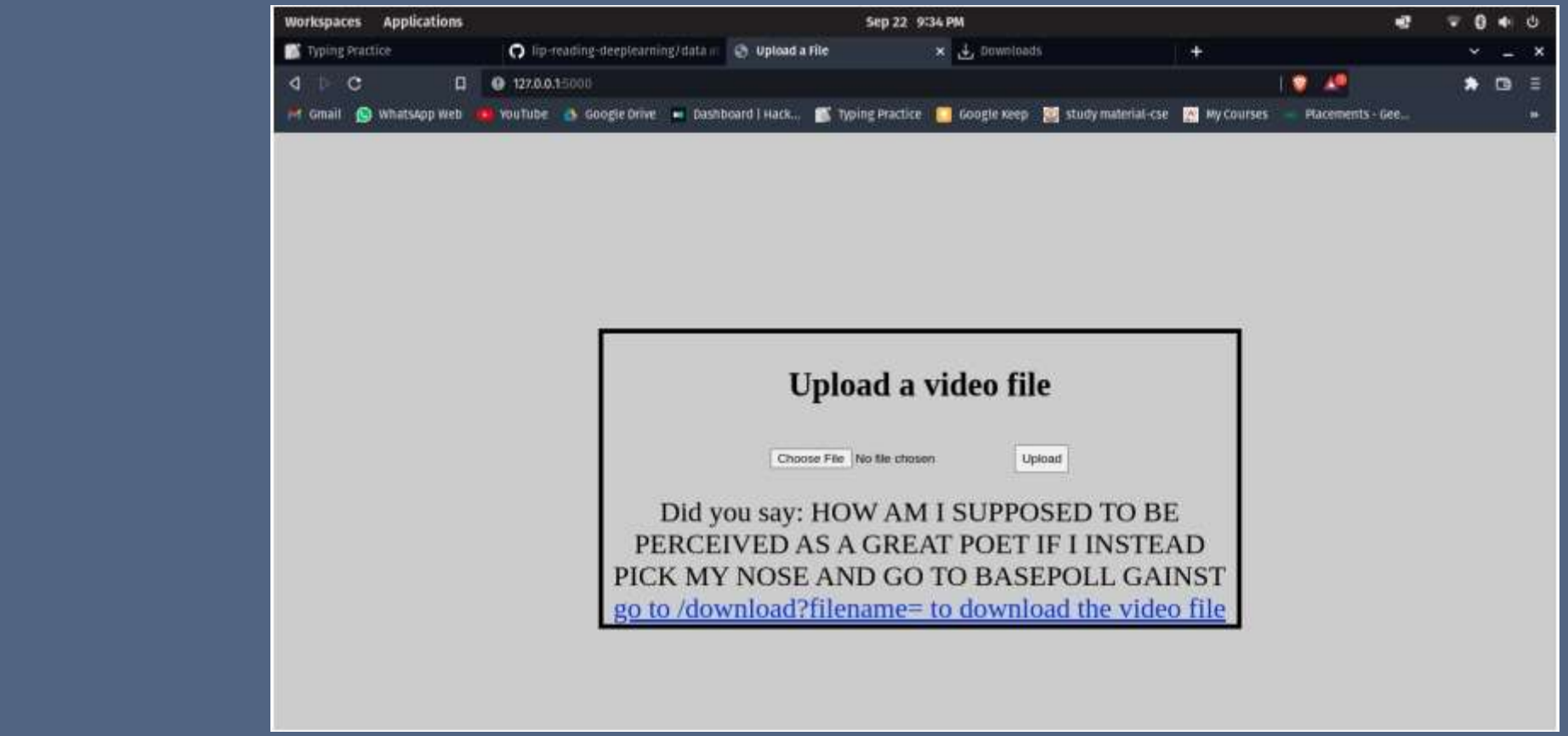


## Predicted Value

# Front End


Workspaces
Applications
Sep 22 10:39 PM
Upload a File
Downloads
WhatsApp
Editor — Clideo
(1395) YouTube
127.0.0.1:5000
Gmail
WhatsApp Web
YouTube
Google Drive
Dashboard | Hack...
Typing Practice
Google Keep
study material-cse
My Courses
Placements - Gee...
Upload a video file
Choose File
No file chosen
Upload
Did you say: TIDY DARTA STRUCTURE AND ALGURYDUM YES SEND ME THE PRETYER PEACE
go to /download?filename= to download the video file
WhatsApp Vide...mp4
Show all

# User Interface



# Skills Learnt During Internship

- Supervised learning: The computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs.
- Unsupervised learning: No labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can be a goal in itself (discovering hidden patterns in data) or a means towards an end (feature learning).
- Reinforcement learning: A computer program interacts with a dynamic environment in which it must perform a certain goal (such as driving a vehicle or playing a game against an opponent). As it navigates its problem space, the program is provided feedback that's analogous to rewards, which it tries to maximize.
- Data Modeling and Evaluation: Data modeling involves understanding the underlying structure of the data and then finding patterns that are not obvious to the naked eye. The data needs to be evaluated using an algorithm that is suitable for the data. For example, the type of machine learning algorithms to use such as regression, classification, clustering, dimension reduction, etc. depends on the data

# There are several Soft Skills as well. Few of them are: -

- Teamwork
- Problem Solving Skills
- Work Ethics
- Adaptability Skills
- Communication skills
- Responsibility
- Time Management

# Conclusion

- In conclusion, LipTo Speech is a groundbreaking technology that has the potential to greatly improve the quality of life for individuals with speech impairments. The technology is highly effective, portable, and easy to use, making it accessible to a wide range of individuals.
- While there are some challenges and limitations associated with the technology, ongoing research and development efforts are likely to address these issues and further enhance the usability and effectiveness of LipTo Speech in the future.

# References


1. Triantafyllos Afouras, Joon Son Chung, and Andrew Zisserman.The conversation: Deep audio-visual speech enhancement. arXiv preprint arXiv:1804.04121, 2018.
2. Ariel Ephrat and Shmuel Peleg. Vid2speech: speech reconstruction from silent video. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5095–5099. IEEE, 2017.
3. A. Torfi, S. M. Iranmanesh, N. Nasrabadi and J. Dawson, "3D Convolutional Neural Networks for Cross Audio-Visual Matching Recognition," in IEEE Access, vol. 5, pp. 22081-22091, 2017, doi: 10.1109/ACCESS.2017.2761539.
4. K. R. Prajwal, R. Mukhopadhyay, V. P. Namboodiri and C. V. Jawahar, "Learning Individual Speaking Styles for Accurate Lip to Speech Synthesis," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 13793-13802, doi: 10.1109/CVPR42600.2020.01381