Infosys Springboard Virtual Internship 6.0

AI-Powered Real-Time Speech Translation For Multilingual Content

QalamAI

BATCH - 03

Agenda

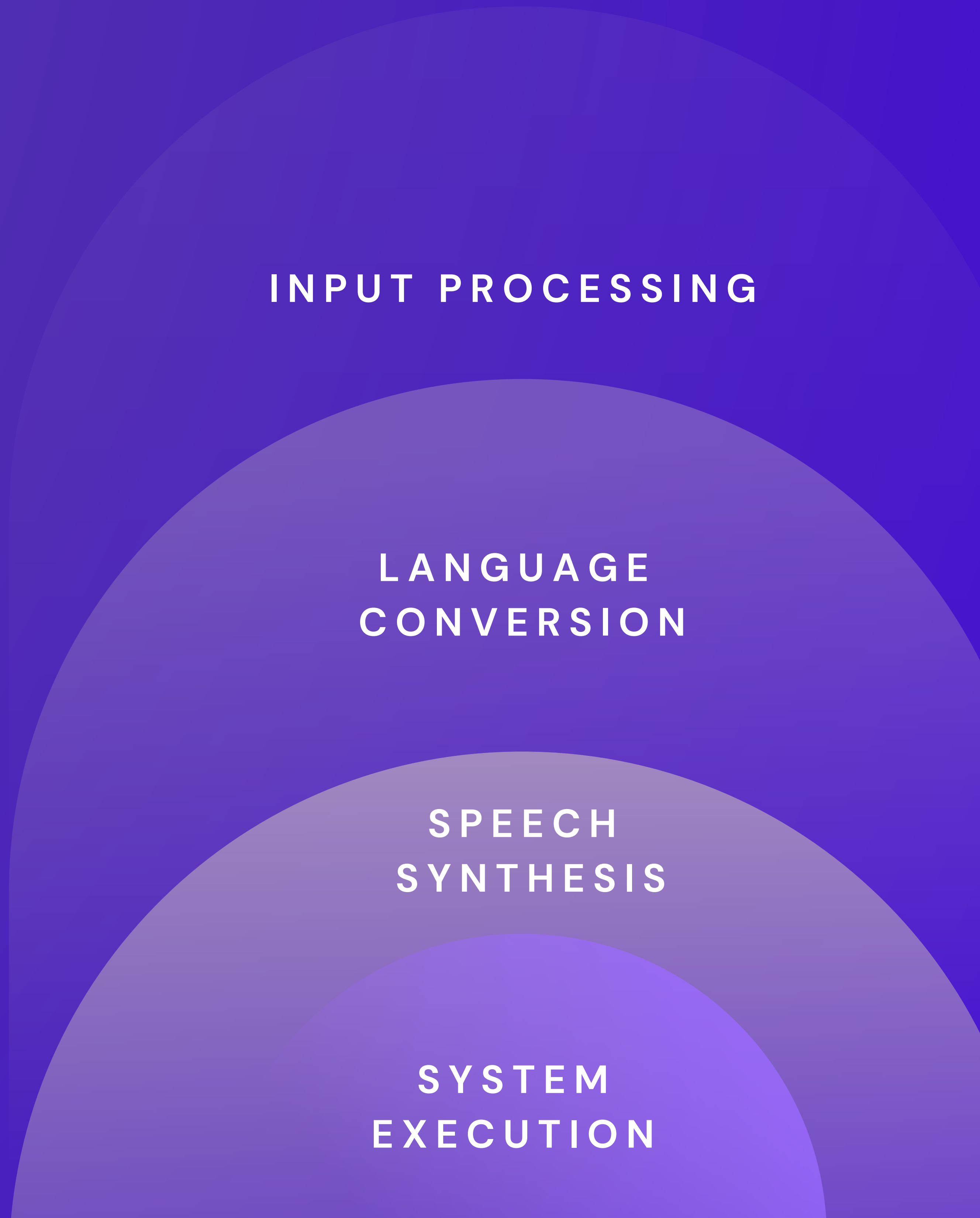
Content

- Introduction
- About The Dataset
- Implementation Overview
- Demo Workflow
- Challenges & Solutions
- Detailed Pipeline
- Future Improvements
- Learning & Skills Gained
- Impact & Real-World Applications
- Conclusion

Introduction

AI-Powered Real-Time Speech Translation For Multilingual Content is an AI-powered tool that converts spoken language into translated text or speech in real time.

- Integrates Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Speech Synthesis modules.
- Aims to bridge communication barriers by providing multilingual, accessible speech translation.
- Developed as part of an 8-week Infosys Springboard internship focusing on end-to-end AI project development.



INPUT PROCESSING

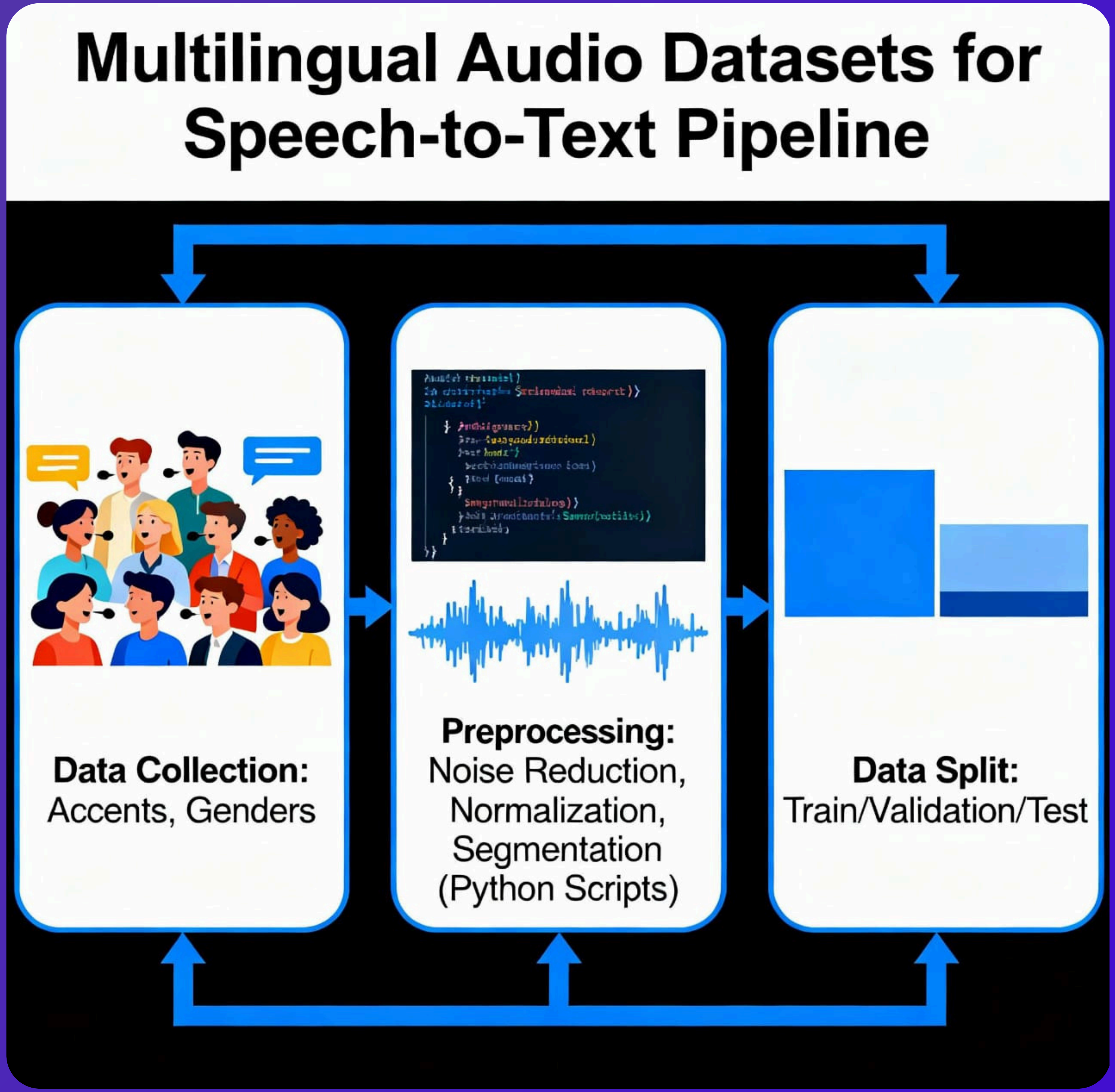
LANGUAGE CONVERSION

SPEECH SYNTHESIS

SYSTEM EXECUTION

About The Dataset

- Utilized multilingual audio datasets for speech-to-text training and evaluation (e.g., Common Voice, LibriSpeech, multilingual text corpora).
- Dataset contains diverse accents, genders, and regional variations to ensure high translation accuracy.
- Cleaned and preprocessed using Python scripts — noise reduction, normalization, and segmentation.
- Data split for training, validation, and testing phases.



Multilingual Audio Datasets for Speech-to-Text Pipeline

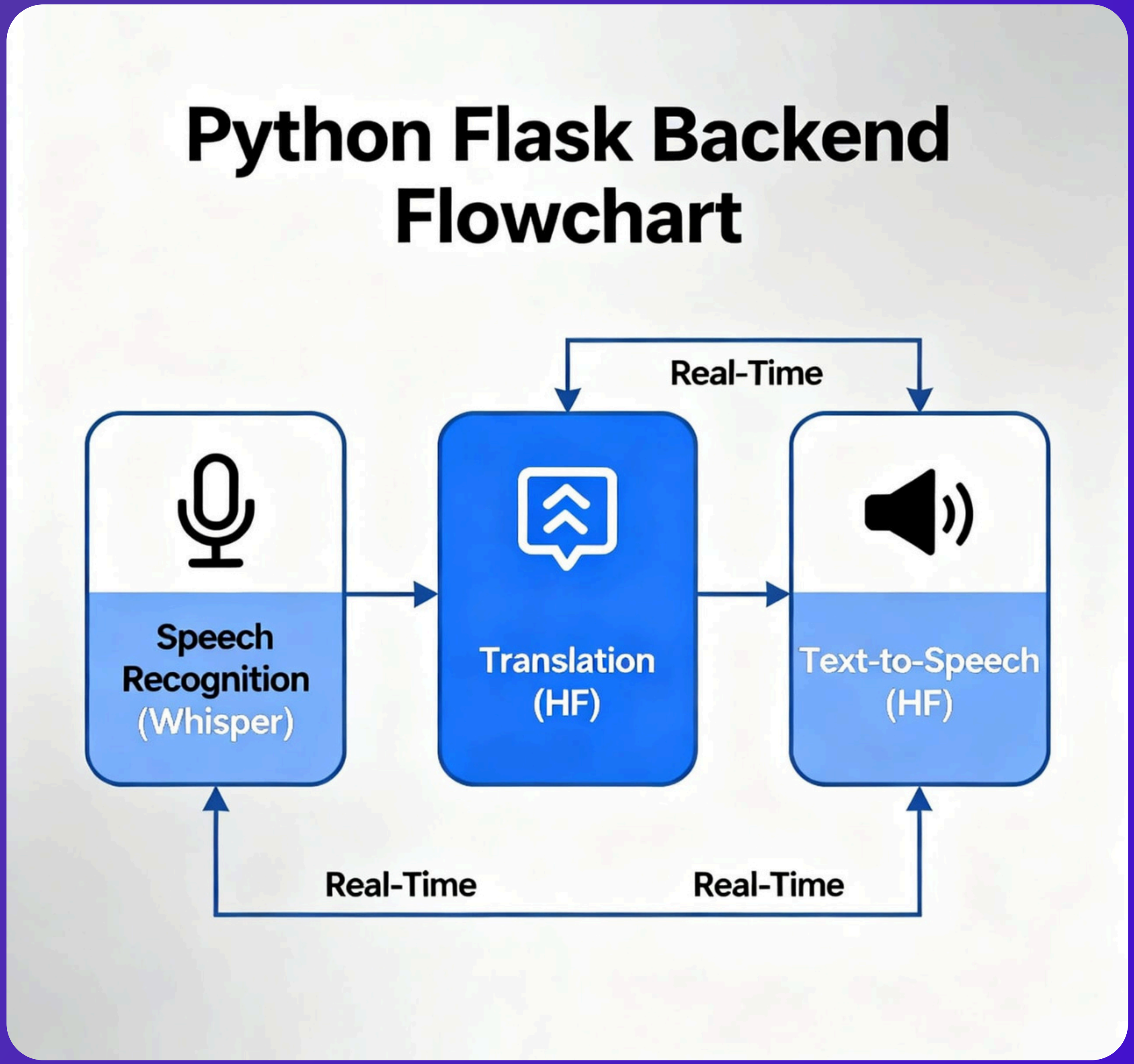
Data Collection: Accents, Genders

Preprocessing: Noise Reduction, Normalization, Segmentation (Python Scripts)

Data Split: Train/Validation/Test

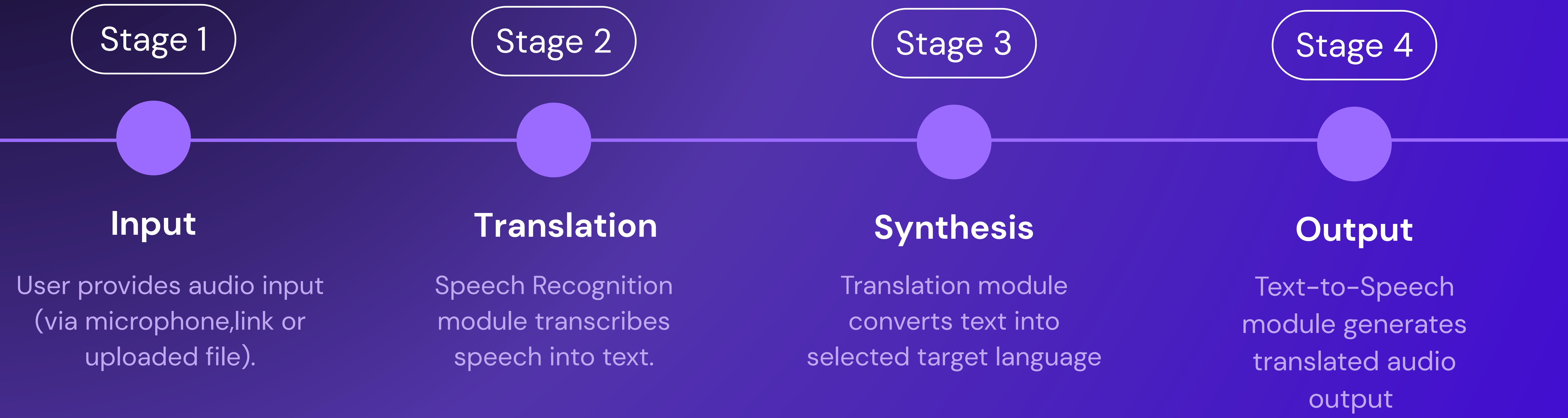
Implementation Overview

- Implemented using Python, Flask.
- Leveraged OpenAI Whisper / Hugging Face Transformers for transcription and translation.
- Backend pipeline integrates speech recognition → translation → text-to-speech output.
- Real-time processing with modular design for scalability and easy integration.



Demo Workflow

Stages Of AI-Powered Real-Time Speech Translation For Multilingual Content



Stage 1

Input

User provides audio input (via microphone,link or uploaded file).

Stage 2

Translation

Speech Recognition module transcribes speech into text.

Stage 3

Synthesis

Translation module converts text into selected target language

Stage 4

Output

Text-to-Speech module generates translated audio output

Challenges & Problem

Data Noise

- Background noise affected transcription accuracy

Model Latency

- Slow translation for long audio clips

Integration Issues

- Flask and React connection delays

Resource Constraints

- Limited computing power

Solutions To The Problems

Data Noise

- Used noise filtering and pre-processing techniques

Model Latency

- Optimized pipeline and used lightweight model variants

Integration Issues

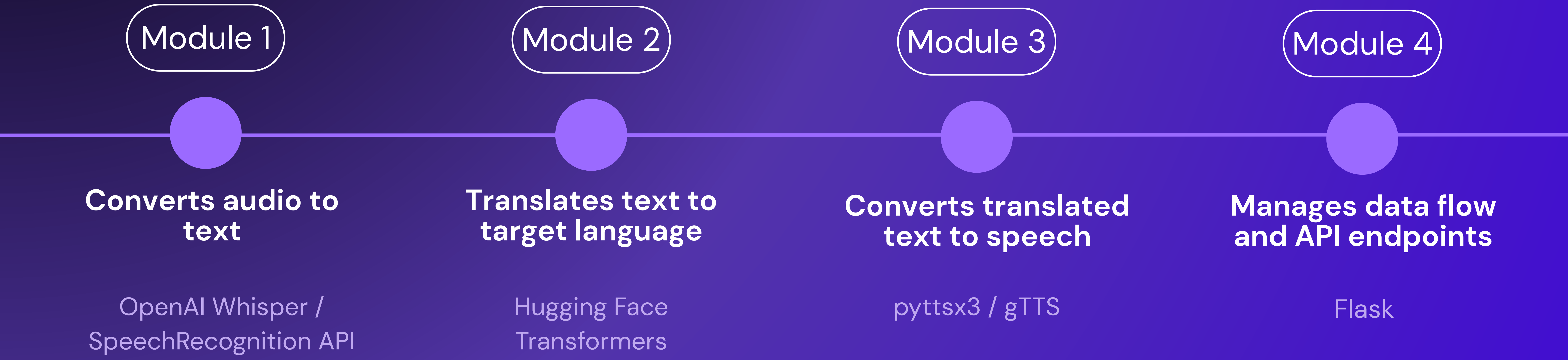
- Implemented async API handling

Resource Constraints

- Used batch processing and model quantizationr

Detailed Pipeline

Stages Of AI-Powered Real-Time Speech Translation For Multilingual Content



Module 1

Converts audio to text

OpenAI Whisper / SpeechRecognition API

Module 2

Translates text to target language

Hugging Face Transformers

Module 3

Converts translated text to speech

pyttsx3 / gTTS

Module 4

Manages data flow and API endpoints

Flask

Future Improvements

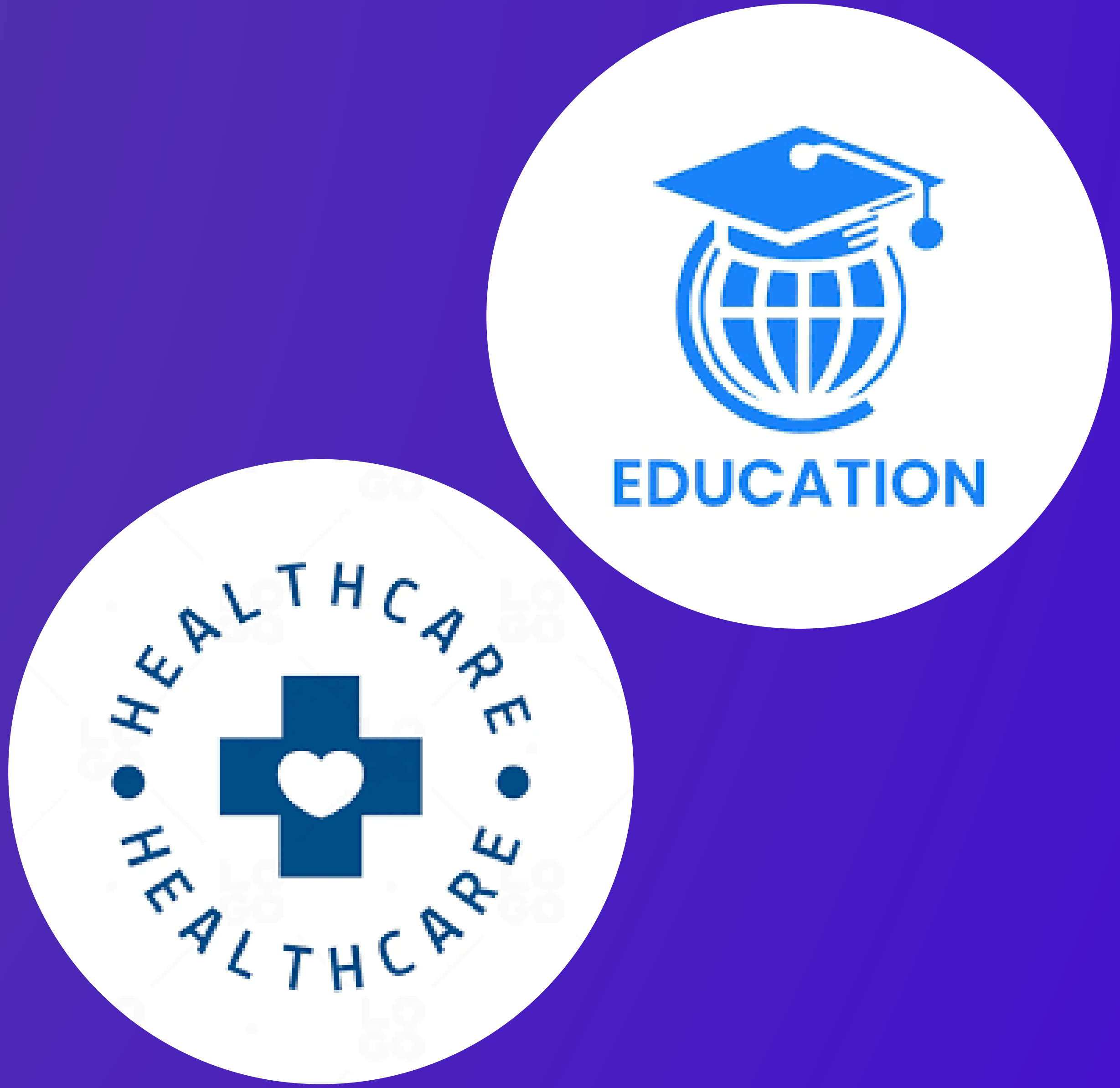
- Integrate real-time streaming translation (live speech to live translation).
- Add support for more regional Indian languages.
- Implement offline mode using compressed models.
- Enhance UI/UX for accessibility and multilingual support.
- Deploy project as a cloud-hosted web application (AWS / Render / Vercel).

Learnings & Skills Gained

- Hands-on experience with AI/ML model integration and API development.
- Learned real-world team collaboration, task division, and documentation.
- Improved technical skills in Python, Flask, and NLP frameworks.
- Enhanced understanding of model optimization and deployment techniques

Impact & Real-World Application

- Useful for education, travel, healthcare, and accessibility tools.
- Can assist differently-abled users through speech-driven interfaces.
- Reduces language barriers in global communication and online learning.



Conclusion

The project successfully demonstrated a modular AI speech translation system.
Provided real-time multilingual communication capability with simple interface.
Strengthened our understanding of AI workflows, teamwork, and innovation.
Sets the foundation for future enhancements and larger-scale applications.

Infosys Springboard Virtual Internship 6.0

THANKS FOR WATCHING

BATCH - 03