

DEKUT HACKATHON-

2024



Accessible AI

Team Nooglers



ABOUT US

We're a small but mighty team of three, bringing our unique skills together to tackle exciting projects. We love working together, learning from each other, and turning great ideas into reality

Baranidharan S
Sai Nivedh V
Roshan T



OUR PROJECT



The "Accessible AI" is an innovative web application designed to bridge the communication gap between hearing individuals and the deaf or hard-of-hearing community.

Our application uses advanced speech recognition, natural language processing, and video processing technologies to convert spoken or written language into Indian Sign Language (ISL).

The application supports multiple input options, including YouTube URLs, local video or audio files, and direct text input. The backend processes involve extracting audio, transcribing speech-to-text using Google's Speech-to-Text API, parsing and processing the text, and converting the processed text into ISL.



OUR VISION

- Our application aims to convert digital content into Indian Sign Language (ISL) to address communication barriers faced by the deaf and hard-of-hearing community. Leveraging modern technologies like natural language processing (NLP), machine learning, and cloud-based services, the system provides seamless conversion of spoken and written content into ISL videos.
- It processes YouTube videos, local audio/video files, and direct text input. Using Google's Speech-to-Text API, the system transcribes spoken language into text, which is then accurately translated into ISL through advanced NLP techniques, enhancing communication in a diverse society.



OUR DATASET

- The effectiveness and accuracy of the Speech-to-ISL conversion system heavily rely on the quality and comprehensiveness of the datasets used.
- The “NLP_videos.csv” dataset which helps in the text-to-ISL conversion process. It contains a detailed list of words along with their corresponding video URLs, start times, and end times, which are used to generate the ISL video clips.
- The “alphabet” dataset is also an component of dataset in which it helps in text-to-ISL conversion process, particularly for spelling out words or handling characters that do not have direct ISL video representations. This dataset contains GIF files representing each letter of the alphabet in ISL



METHODOLOGY

SPEECH-TO-TEXT CONVERSION

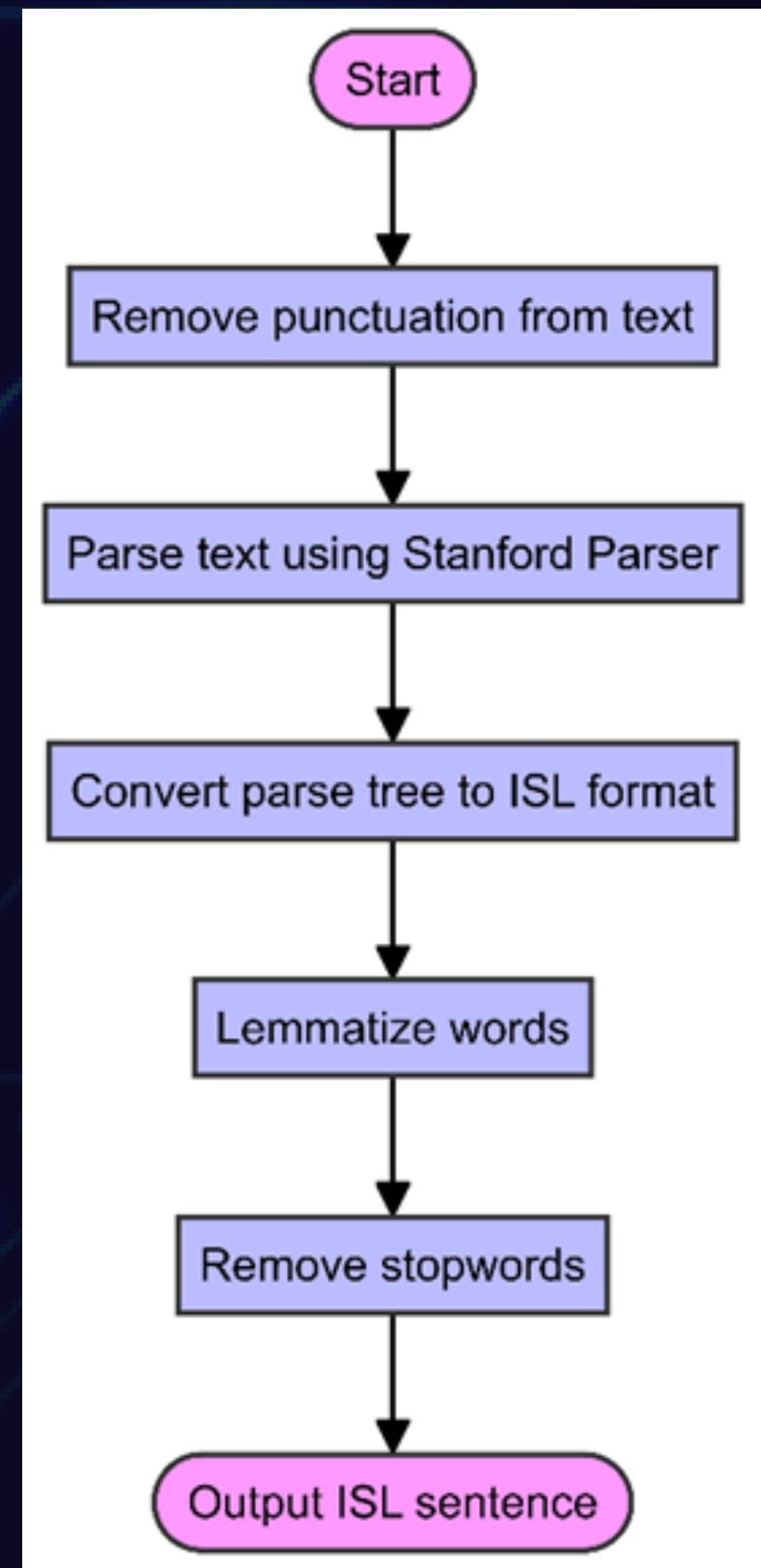
The speech-to-text conversion component is integral to the project, transforming audio input into transcribed text that can be processed further for conversion into Indian Sign Language (ISL). This process involves several key steps, including audio extraction, uploading to Google Cloud Storage, and using Google's Speech-to-Text API to generate transcriptions.

The first step in the speech-to-text conversion process is audio extraction. Depending on the input type, which could be a YouTube URL, a local video file, or an audio file, the system extracts the audio component. For YouTube videos, the yt-dlp library is used to download the audio track. If the input is a local video file, FFmpeg is utilized to extract the audio. The extracted audio is saved in WAV format, which is necessary for the subsequent processing steps.



NATURAL LANGUAGE PROCESSING (NLP)

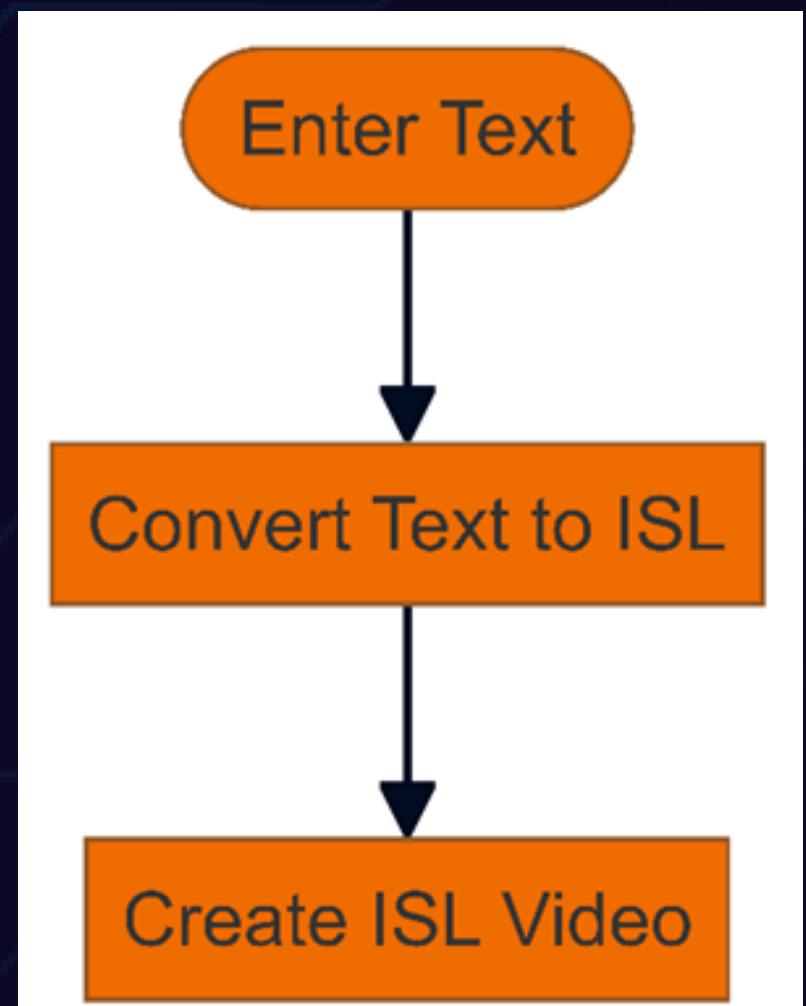
The first step in our NLP pipeline is text parsing using the Stanford Parser, which generates parse trees to represent grammatical structures. These trees break down the text into parts like nouns, verbs, and adjectives, organized hierarchically. Next, we manipulate these parse trees to align with Indian Sign Language (ISL) grammar, which often follows a topic-comment structure, by reordering elements accordingly. After tree manipulation, we perform lemmatization with the WordNet Lemmatizer to reduce words to their base forms, ensuring consistency. This process converts inflected forms to their simplest forms, such as "running" to "run" and "geese" to "goose," creating a more interpretable dataset for ISL conversion.





TEXT-TO-ISL CONVERSION

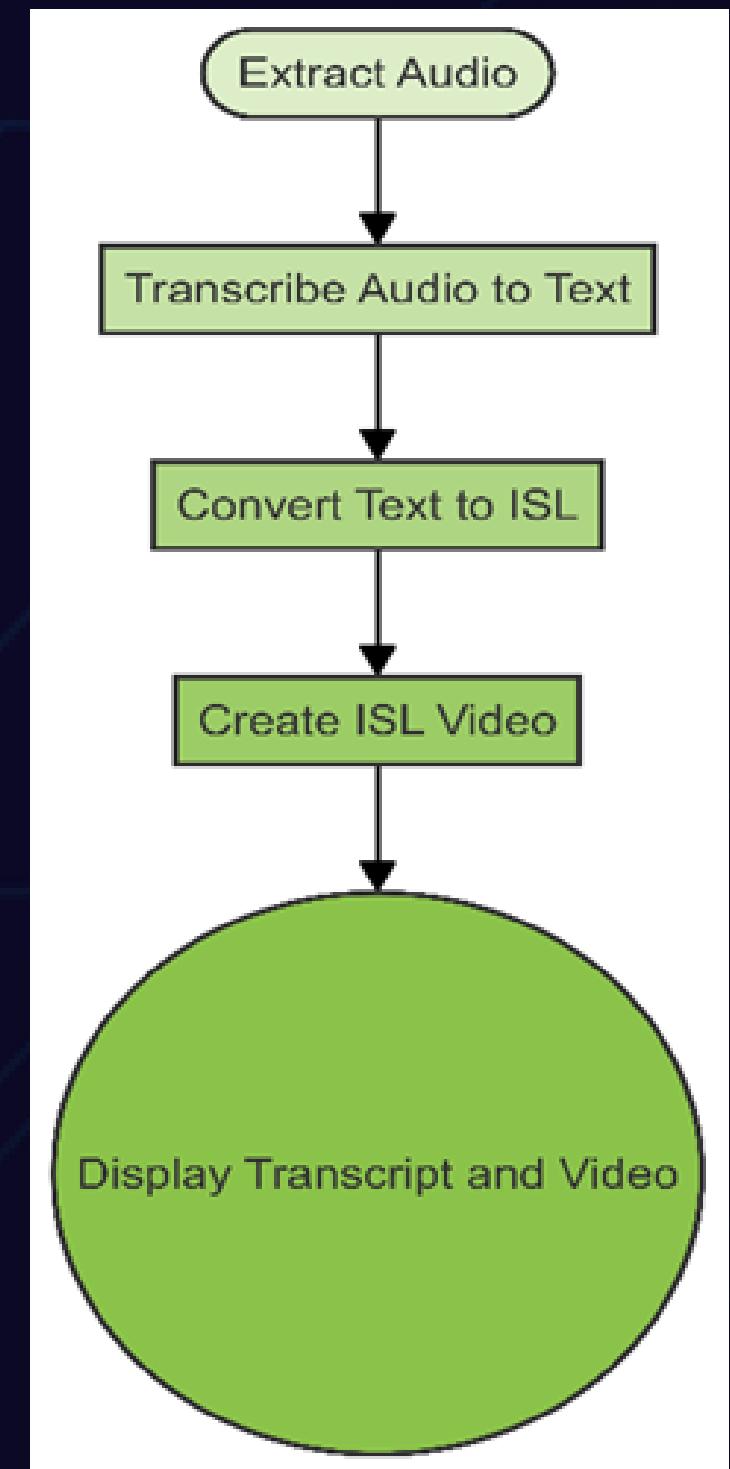
The Text-to-ISL conversion component transforms transcribed text into Indian Sign Language (ISL) video content through several steps. First, text preprocessing cleans and tokenizes the text, removing extraneous elements. Next, the system adapts the cleaned text to match ISL grammar, often shifting from English's subject-verb-object (SVO) structure to ISL's subject-object-verb (SOV) structure. Finally, the system generates video sequences by mapping each word to its corresponding ISL sign and synthesizing the video content, including non-manual signals and timing adjustments, into a coherent video.





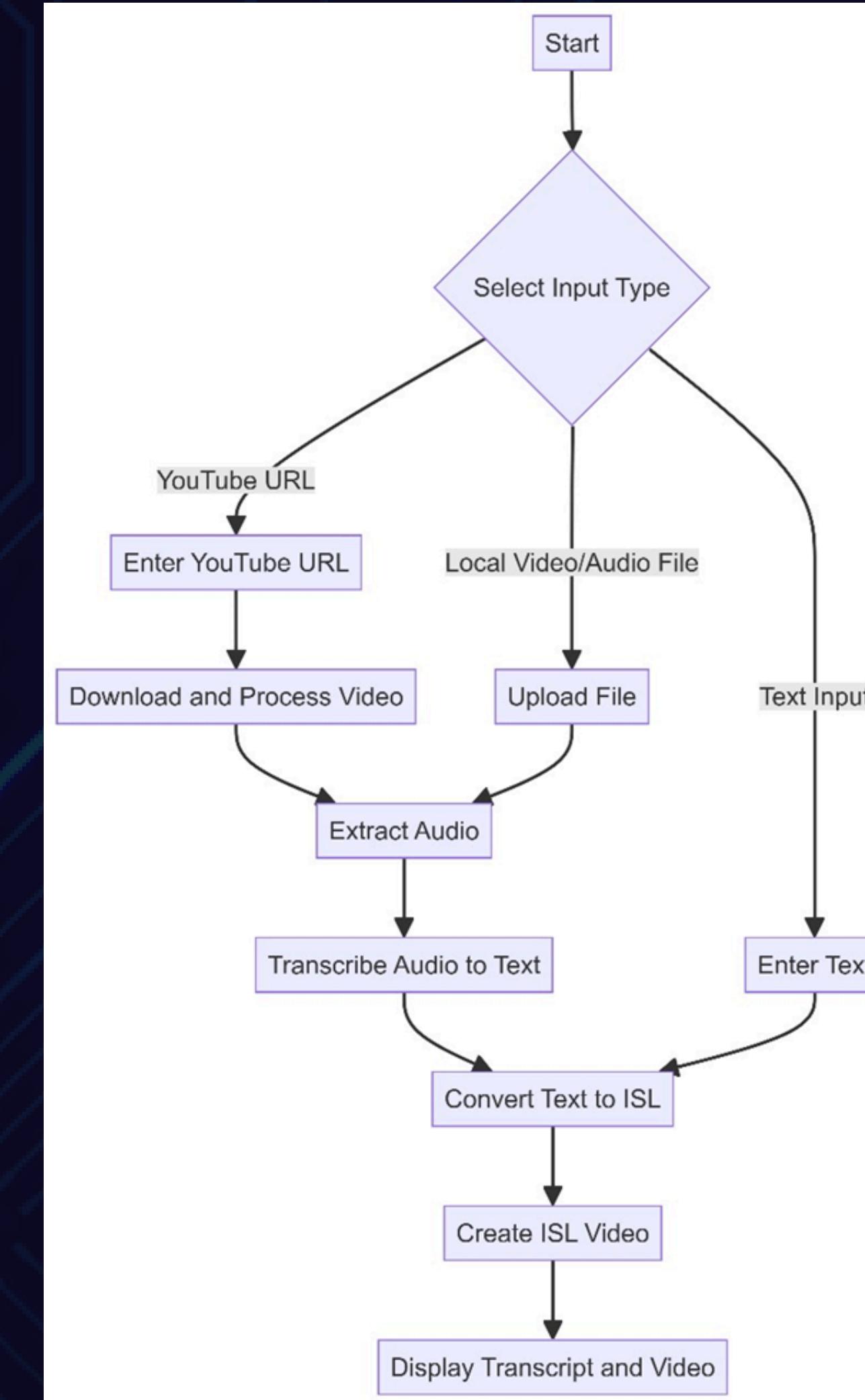
VIDEO PROCESSING

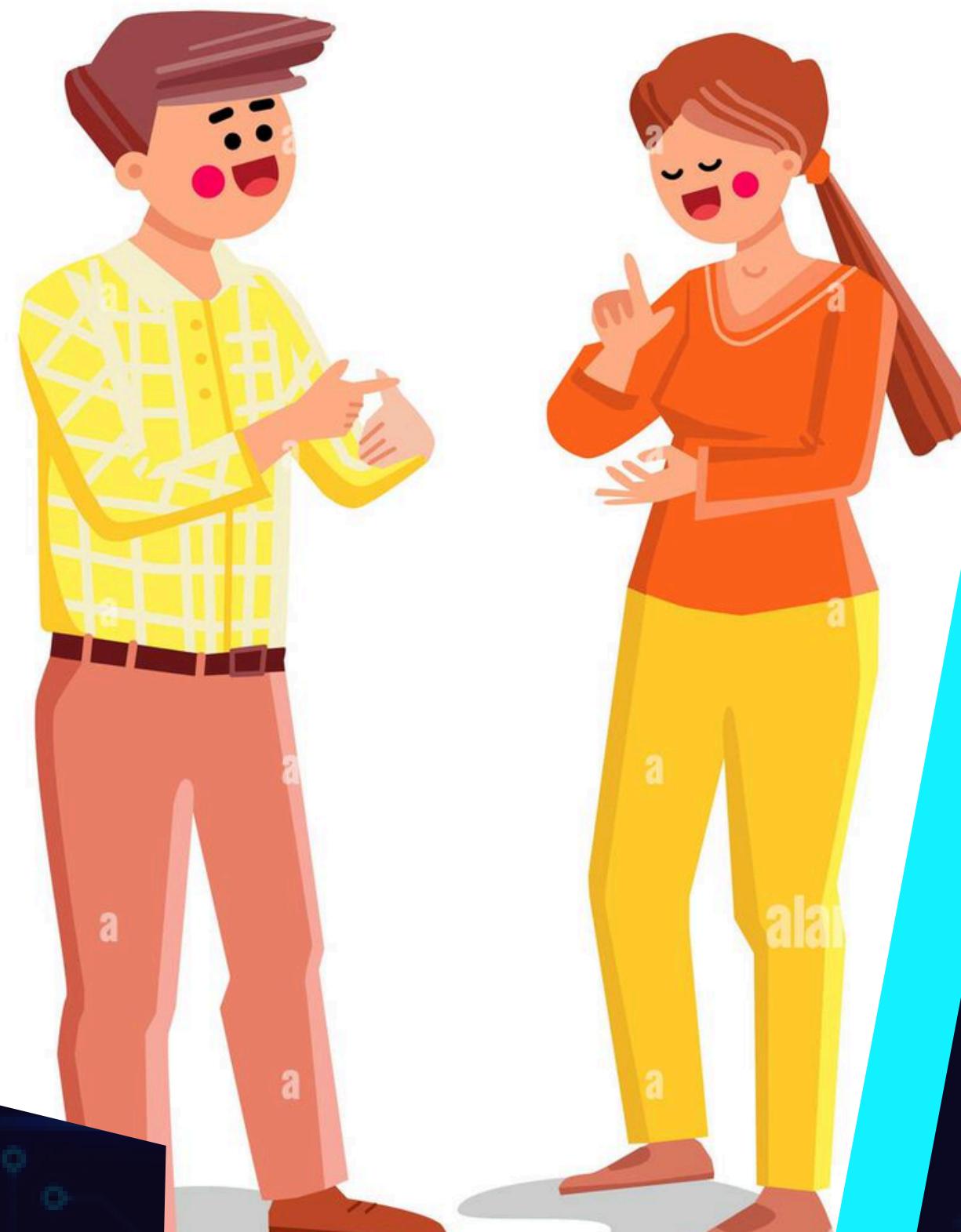
The Video Processing component synthesizes the final output video to represent input speech in Indian Sign Language (ISL). This involves several stages: video generation, non-manual signal integration, timing adjustments, and final compilation. Initially, the system translates adapted text into ISL signs using a pre-recorded library of ISL video clips. Each word or phrase in the text is mapped to its corresponding sign, retrieving the appropriate clip from the library. For words or phrases without direct signs, fingerspelling or contextual interpretation is used. These clips are then sequentially arranged to form a cohesive video.





Workflow





CONCLUSION

In conclusion, the Speech-to-Indian Sign Language (ISL) conversion system represents a significant advancement in bridging the communication gap between the hearing and the deaf or hard-of-hearing communities. This Application covers technologies in speech recognition, natural language processing, and video synthesis to deliver an end-to-end solution capable of translating spoken English into fluent ISL.



TEAM NOOLERS

THANK YOU

For more detailed reference and Demo video check out our Github repo :

[Accessible AI : Nooglers](#)