



# Handspeak AI

Team NovaMind :

VENKATA SAI NITHIN GOLLA  
SAI HARISH PARUPUDI  
LATHA SREE SANABOYINA

## MOTIVATION :

- Scheduling and paying for interpreters in remote classes (Zoom, Teams, etc.) is a major hurdle, often excluding deaf and hard-of-hearing students. This affects their learning, participation, and sense of belonging.
- Remote learning, like any other form of education, should be accessible to all students. The barriers described directly prevent Deaf and Hard-of-Hearing students from fully participating in and benefiting from remote learning experiences.

## **Solution :**

- AI-based image recognition
- We will be using CNN for static gestures.
- NLP corrects grammar and sentence structure for natural language output.

## Key Winning Features:

- Unlike solutions requiring manual setup (e.g., Zoom's interpretation feature), the new solution could offer automatic integration with platforms like Zoom, Teams, and others, ensuring interpreters are readily available without extensive scheduling or configuration
- Provides a more efficient and potentially cost-effective way to manage interpreter services compared to manual processes.
- Deep Learning models (CNN + LSTMs/Transformers) ensure accurate sign recognition.

## Capabilities:

CNNs for Static Gesture Recognition

- Detects hand shapes and positions for letters and signs

LSTM / Transformer Models for Dynamic Signs

- Recognizes sign sequences and generates full sentences

- NLP for Language Refinement

Corrects grammar and adds contextual fluency

- Multilingual NLP Engine

Enables communication in multiple spoken languages

## Justification

- **Bridges Communication Gaps**

Empowers the deaf and hard-of-hearing to interact seamlessly

- **Accessible to All**

No need for expensive hardware or apps

- **Affordable Alternative to Human Interpreters**

Reduces long-term costs for businesses and schools

- **Cloud + Edge Architecture is Scalable**

Serves both individual users and large institutions

## Target Market

1. Deaf & mute individuals (primary users)
2. Educational institutions for hearing-impaired students
3. Hospitals and Healthcare Providers

## Value Propositions

- Real-time sign-to-speech & text translation
- Browser-based (no installation required)
- Highly accurate gesture recognition via AI
- Scalable & cost-efficient (Cloud AI)
- Freemium pricing model for broader access

## Prototype & Development Status:



### What We've Built So Far



#### Functional Flask Backend

- Handles file upload (audio/video)
- Integrates **OpenAI Whisper** for speech-to-text
- Uses **Levenshtein Distance** to match transcribed words to sign dictionary



#### Sign Language Data Dictionary

- Created using **MediaPipe + OpenCV**
- Captures and stores 3D hand gesture coordinates for each word
- Exported as a structured JSON for fast reference



#### Frontend Interface

- Built with **HTML/CSS/JS**
- Users can upload an audio file
- Returns both the **raw transcription** and **sign-translated version**



## How We Make Money:

### 1. Freemium Model

- **Free Tier** for basic access: limited number of translations per month.
- Encourages adoption by students, educators, and non-profits.
- Users can upgrade for more advanced features.

### 2. Subscription Plans

- **Individual Pro Plan:** Monthly/Yearly access to unlimited translations, video-to-text exports, and avatar support.
- **Institutional Plan:** Tailored for schools, universities, and accessibility organizations with bulk usage and API access.

### 3. API Licensing

- **Sell API Access** to EdTech platforms, telehealth services, and customer service providers.
- Enables seamless integration into third-party apps.

### 4. Enterprise Solutions

- **Custom Deployments** for large organizations (e.g., hospitals, governments).
- Includes integration, analytics dashboard, support, and training.

### 5. Advertising & Sponsorship (Optional Future Stream)

- Offer **non-intrusive ads** or branded avatar options for freemium users.
- Partnerships with NGOs or CSR programs for accessibility promotion.

## Team Members & Contributions:



### **Nithin – Team Lead & AI Developer**

- Led overall system architecture and design.
- Developed hand gesture-to-text pipeline using Google MediaPipe and OpenCV.
- Integrated audio transcription using Whisper and handled cloud deployment.



### **Latha – Backend & API Engineer**

- Built Flask-based backend and managed RESTful APIs.
- Implemented audio/video upload processing and text output handling.
- Ensured smooth data flow between frontend, backend, and AI models.



### **Harish – Frontend Developer & UI Designer**

- Designed and implemented the React-based user interface.
- Created clean, accessible layouts with upload feature, progress bar, and results display.
- Focused on user experience and responsive design for all devices.

## What We Are Asking:



### **Implementation Plan**

- Finalize MVP deployment (video + audio to text conversion working in production).
- Complete front-end integration for real-time subtitle display.
- Collect feedback from initial users (educators, interpreters, hearing-impaired users).



### **Technical Validation**

- Conduct pilot testing with sign language interpreters and inclusive education groups.
- Validate gesture accuracy, latency, and reliability across diverse backgrounds and lighting conditions.
- Improve model performance based on real-world feedback.



### **Market Entry Plan**

- Launch beta version for NGOs, schools, and accessibility forums.
- Target early adopters in inclusive education, virtual classrooms, and government accessibility mandates.
- Offer freemium model to drive adoption and gather user insights.



### **Scale-Up Strategy**

- Expand sign language support across Indian Sign Language (ISL), ASL, and regional variants.
- Add avatar-based sign rendering and multilingual text output.
- Partner with accessibility tech providers, EdTech platforms, and government initiatives (e.g., Digital India, Accessible India Campaign).

## Conclusion:

- Handspeak AI is a cloud-based intelligent system that translates sign language gestures and spoken audio into structured text output in real-time.
- By integrating OpenCV, MediaPipe, Deep Learning (CNN + LSTM), Speech-to-Text, and NLP, it ensures accurate, context-aware, and accessible communication.
- The system supports real-time translation without the need for external hardware, making it scalable and user-friendly.
- Our approach improves inclusion for the deaf, mute, and speech-impaired communities, and opens doors for broader adoption in education, healthcare, and public services.

## References:

- [1] OpenAI. (2023). Whisper Speech-to-Text Model. <https://openai.com/research/whisper>
- [2] Google MediaPipe. (2023). <https://mediapipe.dev>
- [3] Baihan, A., Alutaibi, A. I., Alshehri, M., & Sharma, S. K. (2024).  
*Sign Language Recognition Using Modified Deep Learning Network and Hybrid Optimization.*  
<https://www.nature.com/articles/s41598-024-76174-7>
- [4] Yanqiong Zhang. (2024). *Sign Language Recognition Using CNN and LSTM.*  
<https://www.researchgate.net/publication/378030108>

# THANK YOU

