**Project Title: AI Powered Sign Language Translator**

Team members:

1. Habtamu Tadese
2. Sineshaw Legese
3. Arsema Mandefro
4. Meseret Bolled
5. Abem Tigist
6. Hanna Desalegn

## Concept Note

## 1. Project Overview

## Brief Overview

## This capstone project aims to develop an AI-powered translation system that converts Amharic Sign Language (ASL) gestures into written or spoken Amharic. By utilizing machine learning and computer vision techniques, this project will enhance communication for the deaf and hard-of-hearing community in Ethiopia. [5]

## Relevance to Sustainable Development Goals (SDGs)

## This project aligns with SDG 10: Reduced Inequalities, as it aims to improve accessibility and communication for individuals with hearing impairments. Additionally, it supports SDG 4: Quality Education, ensuring that educational materials and communication are inclusive for all students.

## Problem Statement [1]

## Individuals who are deaf or hard of hearing often struggle to communicate with those who do not understand sign language, leading to social and educational barriers. This project addresses the need for an effective tool that translates ASL into written Amharic, facilitating better interactions and understanding.

## Potential Impact [2]

## The successful implementation of this translation system could significantly improve communication and access to resources for the deaf community, fostering greater inclusion in education, healthcare, and social services. It will also promote awareness of sign language and its importance in communication.

## 2. Objectives

## Data Collection and Preprocessing:

## Collect a comprehensive dataset of Amharic Sign Language gestures along with their corresponding written Amharic translations.

## Preprocess the data to ensure quality and prepare it for training the AI model.

## Model Development:

## Develop a machine learning model that accurately recognizes ASL gestures and translates them into written Amharic.

## Implement deep learning techniques, such as convolutional neural networks (CNNs), to enhance gesture recognition and translation accuracy.

## User Interface Design:

## Create an intuitive user interface that allows users to input ASL gestures (via video or images) and receive real-time translations into written Amharic.

## Ensure the interface is accessible and user-friendly for both sign language users and non-users.

## Testing and Evaluation:

## Conduct thorough testing of the translation system to evaluate its performance and usability.

## Collect feedback from the deaf community to refine the system based on their experiences and needs.

## 3. Background

Effective communication between hearing-impaired individuals and the broader community remains a significant challenge, especially in regions where tailored technological solutions are limited. In Ethiopia, Amharic Sign Language is the primary means of communication for many hearing-impaired individuals. However, due to the lack of widespread understanding of sign language among the general population, this community often faces social and informational exclusion. [  
  
While previous efforts have been made to recognize Amharic Sign Language using computer vision techniques, most are limited in scope, accuracy, or real-time usability. One such dataset includes gesture data collected from sign language teachers, annotated and preprocessed for use in object detection models like Faster R-CNN and Single Shot Multibox Detector (SSD). These models demonstrated promising results in recognizing static gestures corresponding to 10 Amharic characters.  
  
However, there remains ample room for improvement in terms of accuracy, speed, and scalability to full alphabet, word-level, or sentence-level recognition. Therefore, this study proposes building on the existing dataset and developing an improved Amharic Sign Language recognition system using alternative and advanced machine learning (ML) and deep learning (DL) techniques.

1. **Methodology**

The goal of this project is to design and evaluate improved models for Amharic Sign Language recognition, leveraging the existing annotated dataset while exploring alternative ML and DL approaches.

**Dataset Acquisition and Preparation**  
We use a pre-existing dataset composed of annotated frames extracted from videos of Amharic Sign Language gestures. The dataset includes:

* Hand gestures for 10 Amharic characters
* XML, TFRecord, and CSV formatted annotations
* Diverse backgrounds and hand positions for generalization

Initial steps include:

* Verifying and cleaning existing annotations
* Splitting data into training, validation, and testing sets
* Applying data augmentation to improve robustness (e.g., rotation, flipping, contrast)

**Model Exploration and Selection**  
To improve upon prior work, we will experiment with a range of models beyond Faster R-CNN and SSD. These include:  
  
**Classical Machine Learning Approaches:**

* Support Vector Machines (SVM) with hand-crafted features
* Random Forests / Gradient Boosting using extracted shape and motion features

**Deep Learning-Based Models [4] :**

* Convolutional Neural Networks (CNNs) for static image-based recognition
* 3D CNNs or TimeDistributed CNNs with LSTM layers for short video sequences (to capture temporal dynamics)
* YOLOv5 or YOLOv8 for real-time object detection
* Transformers (e.g., Vision Transformers or MViT) for improved contextual understanding  
  Each model will be trained and validated on the same dataset to allow for fair comparison.

**Model Training and Hyperparameter Tuning**  
Each selected model will undergo:

* Hyperparameter optimization (learning rate, batch size, number of layers, etc.)
* Evaluation using metrics such as accuracy, precision, recall, F1-score, and inference time
* Cross-validation for robustness

**Comparative Evaluation**  
All models will be benchmarked against the original SSD and Faster R-CNN results. This comparative evaluation will help identify the best-performing approach in terms of:

* Recognition accuracy
* Real-time performance
* Scalability to more characters and gesture sequences

**Future Extension Plans**  
Once an optimal approach is identified, future steps include:

* Expanding the dataset to include the full Amharic alphabet, dynamic gestures, words, and phrases
* Building a prototype application for real-time translation (e.g., webcam or mobile-based)
* Integrating speech/text synthesis for full communication support

**A diagram of a software development

AI-generated content may be incorrect.**

**Figure: Architecture Design Diagram**

**5. Architecture Design Diagram**

**Component Descriptions**

1. **User Interface (UI)**

* **Role**: The entry point for users to interact with the system.
* **Functionality**: It captures user gestures and provides feedback based on the translation results.

1. **Gesture Recognition Module**

* **Role**: Processes input from the UI to identify and interpret user gestures.
* **Functionality**: Converts physical gestures into recognized data, which is then sent to the NLP Engine for translation.

1. **Natural Language Processing (NLP) Engine**

* **Role**: Analyzes recognized gestures to generate corresponding text or speech.
* **Functionality**: Utilizes machine learning algorithms to interpret gestures and facilitate communication in natural language.

1. **Database**

* **Role**: Stores gesture mappings and translation data.
* **Functionality**: Provides the NLP Engine with necessary mappings for accurate translations and saves new mappings as they are created.

1. **Feedback System**

* **Role**: Communicates results back to the user.
* **Functionality**: Displays the translated output and any relevant feedback to enhance user experience.

1. **Data Sources**

**Data Sources for AI-Powered Amharic Sign Language Translator**

For the development of the AI-powered Amharic sign language translator, two primary datasets will be utilized. The first is the **Amharic Sign Language Recognition Dataset [7]**, which comprises video files and annotated images collected from sign language teachers. This dataset is crucial as it focuses specifically on Amharic Sign Language gestures, featuring 10 classes that represent various signs, thereby addressing a significant gap in resources for non-Western sign languages. Preprocessing steps include frame extraction from video files, resizing images for uniformity, and feature extraction to enhance gesture recognition capabilities. The second dataset, **A Generic Approach towards Amharic Sign Language Recognition [8]**, is sourced from signers at Addis Ababa University and includes a comprehensive set of 240 classes covering all Amharic alphabets and numeric characters. This dataset will undergo video processing to convert videos into sequences of frames, along with image preprocessing involving cropping and resizing. Both datasets will be instrumental in training machine learning models to improve real-time sign language recognition, thereby facilitating communication for the deaf and hard-of-hearing community.

1. **Literature Review**

The literature on AI-driven sign language translation emphasizes the effectiveness of computer vision techniques, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), in recognizing sign language gestures through visual cues such as hand movements and facial expressions. Traditional human interpreters often face challenges related to availability, especially in remote or emergency situations, and can incur significant costs, highlighting the need for automated solutions. While existing research has explored sensor-based approaches and hybrid methods that integrate visual data with wearable technology, many studies focus on established sign languages like American Sign Language (ASL). This project builds upon these findings by specifically targeting the translation of Amharic Sign Language into written Amharic, thus addressing a significant gap in the existing literature and promoting greater inclusivity for the deaf community in Ethiopia.

**Implementation Plan**

**1. Technology Stack**

* **Programming Languages**:
  + Python: The primary language for developing the AI model and integrating various components.
* **Libraries**:
  + **OpenCV**: An open-source library for real-time computer vision tasks, used for video and image preprocessing, hand detection, and frame extraction.
  + **PyTorch**: A flexible deep learning framework to build and train the model, known for dynamic computational graphs and ease of use.
  + **TensorFlow** (as an alternative): An industry-standard framework for scalable deployment, serving as a backup depending on performance needs.
* **Frameworks**:
  + **Flask**: A lightweight Python web framework to deploy the model as an API endpoint for real-time translation.
* **Model Architecture**:
  + **CNN + LSTM**: Convolutional Neural Networks (CNNs) for spatial feature extraction from video frames, combined with Long Short-Term Memory (LSTM) networks to analyze the temporal flow of gestures.
* **Development Environment**:
  + **Google Colab**: Used for training deep learning models, providing free GPU access for efficient resource use.
* **Dataset**:
  + **Amharic Sign Language Dataset**: A publicly available or custom-created dataset with labeled video frames or sequences of gestures, essential for ensuring local relevance and inclusivity.

These technologies were selected for their effectiveness, scalability, and ability to support real-time performance, which is crucial for developing an accessible sign language translation tool.

**2. Project Timeline**

Week 1 (April 22–26)

* Data Preparation
  + Finalize collection of Amharic Sign Language datasets.
  + Preprocess videos (frame extraction, resizing) using OpenCV.
  + Annotate key gestures (hand landmarks, facial expressions).
  + *Milestone: Dataset ready for training by April 26.*

Week 2 (April 29–May 3)

* Model Development
  + Build CNN+LSTM architecture in PyTorch.
  + Integrate MediaPipe for real-time hand tracking.
  + Begin initial training on Google Colab (GPU).
  + *Milestone: Functional model prototype by May 3.*

Week 3 (May 6–10)

* Training & Evaluation
  + Optimize model hyperparameters.
  + Evaluate performance using BLEU score and Word Error Rate (WER).
  + Address overfitting/data bias (e.g., augment Amharic gestures).
  + *Milestone: Model achieves acceptable validation accuracy by May 10.*

Week 4 (May 13–16)

* Deployment & Testing
  + Develop Flask API for real-time translation.
  + Create a basic UI demo (web or mobile).
  + Conduct user testing with Deaf community volunteers.
  + *Milestone: Working demo showcased by May 16.*

**Task Distribution**

|  |  |
| --- | --- |
| Task | Assigned |
| Data Preprocessing | Habtamu, Meseret |
| Model Development | Sineshaw, Arsema |
| Evaluation Metrics | Abem, Hanna |
| Deployment | All |

**3. Milestones**

1. Data Ready (April 26)
   * Complete dataset collection/preprocessing (Amharic Sign Language).
   * *Deliverable: Clean, annotated dataset in TFRecord/CSV format.*
2. Model Prototype (May 3)
   * Functional CNN+LSTM model architecture.
   * *Deliverable: Jupyter notebook with initial training script.*
3. Training Complete (May 10)
   * Model achieves acceptable accuracy on validation set.
   * *Deliverable: Performance report*
4. Deployment Demo (May 16)
   * Flask API with real-time translation demo.
   * *Deliverable: Web app prototype video/showcase.*

**4. Challenges and Mitigations**

**Quality and Availability of data**

One of the key challenges in developing an AI-powered Amharic Sign Language Translator is the **quality and availability of data**. Currently, there is a lack of publicly available, well-labeled datasets specifically for Amharic sign language. This limitation could hinder the model’s ability to generalize and perform accurately.

To mitigate this, the team plans to explore available datasets that may include relevant sign language data, while also conducting thorough research to identify and compile any existing resources. If suitable datasets are not found, the team will reach out to local organizations, educational institutions, and experts in the field to request access to or collaboration on creating a well-labeled Amharic Sign Language dataset, ensuring that the model can be trained effectively and accurately.

**Model performance**

Particularly in recognizing the nuances of hand gestures that may differ slightly between individuals or due to environmental factors like lighting and background distractions.

To address this, we will use deep learning models such as Convolutional Neural Networks (CNNs), along with preprocessing techniques like background subtraction and hand detection, to enhance input quality. The model will also be fine-tuned with iterative testing and real user feedback to improve accuracy.

**Technical constraints**

This challenge present a further challenge, especially in low-resource environments where high-performance computing is not readily available. This could impact both training and deployment. As a mitigation strategy, we aim to use lightweight and efficient models such as Mobile Net or deploy the system using TensorFlow Lite, allowing the application to run smoothly even on low-end devices. Training can also be done using cloud-based platforms with GPU access.

Lastly, the **user interface accessibility** is crucial, as it must cater to individuals who may have limited experience with technology. To overcome this, we plan to conduct usability tests with members of the target community and design an intuitive interface featuring large buttons, clear icons, and visual feedback. This ensures that the application is inclusive and easy to use for everyone.

**5. Ethical Considerations**

This project must address several ethical concerns.

**a. Data Privacy:** We ensure the data sets used are open-source and do not infringe on personal privacy. If user-generated data is added later, informed consent will be required.

**b. Bias:** Bias in training data (e.g., limited to certain ethnicities or regions) can lead to unfair predictions or poor generalization. We will use diverse datasets and continually evaluate model fairness during testing.

**c. Community Impact:** This tool aims to empower the Deaf and Hard-of-Hearing community by bridging communication gaps. We will collaborate with community members for feedback, ensuring the tool addresses real needs and doesn't misrepresent sign language.

**6.References**

1. Automated Sign Language Translation: The Role of Artificial Intelligence Now and in the Future - SciTePress, <https://www.scitepress.org/Papers/2020/101438/101438.pdf>
2. Sign Language Translation: A Survey of Approaches and Techniques, <https://www.mdpi.com/2079-9292/12/12/2678>
3. A Survey of Advancements in Real-Time Sign Language Translators: Integration with IoT Technology - MDPI, <https://www.mdpi.com/2227-7080/11/4/83>
4. Artificial Intelligence Technologies for Sign Language - PMC, <https://pmc.ncbi.nlm.nih.gov/articles/PMC8434597/>
5. AI Powered Sign Language Translator - ResearchGate, <https://www.researchgate.net/publication/375601688_AI_Powered_Sign_Language_Translator>
6. Engineers Bring Sign Language to 'Life' Using AI - Florida Atlantic University, <https://www.fau.edu/newsdesk/articles/american-sign-language.php>
7. Isayas, F. (2020). Amharic Sign Language Data Sets. Mendeley Data, V1. <https://data.mendeley.com/datasets/5d3nkyhsrf/1>
8. Isayas, F. (2022). Amharic Sign Language Recognition Using Deep Learning. \*Journal of Sensors\*, Article ID 1112169. <https://doi.org/10.1155/2022/1112169>