DEPI ROUND 3 PROPOSAL

OCTOBER 12, 2025

Prepared by

Prepared for
eyouth X DEPI – round 3
admission@debi.gov.eg

Sarah Ahmed Mabrouk El-Shinnawy sarah.shinnawy@gmail.com 01026292901 Mariam Ahmed Abd El-Fattah El-Bhnasawy 01201904535 mariam.bhnasawy@gmail.com	Nouran Shaban El-shora 01093310564 nouranelshora9@gmail.com Alaa Kamal Farooq Esmail (team leader) 01113844729 Alaakamal7957358@gmail.com

HAND GESTURE RECOGNITION SYSTEM

1. Project description

The proposed project develops a real-time bilingual sign language recognition system capable of detecting and translating American Sign Language (ASL) and Arabic Sign Language (ArSL) gestures into text. Leveraging computer vision techniques such as Convolutional Neural Networks (CNNs) and hand landmark tracking, the system captures hand gestures via a camera and outputs readable text on-screen.

Designed for **practical communication scenarios**—including healthcare, education, and public services—this solution enhances accessibility for individuals with hearing or speech impairments. The project establishes a **scalable foundation** for future integration with **conversational AI** and **text-to-speech technologies**, enabling intelligent, two-way communication. By addressing both English and Arabic sign languages, the system meets a **unique bilingual accessibility need**, positioning it as an innovative tool in the field of assistive technologies.

Team Member	*Role & Responsibilities*
Marim	Lead on Arabic Sign Language (ArSL) dataset. Responsible for collecting images, preprocessing the Arabic dataset, and assisting in model training.
Noran	Model training specialist for real-time prediction. Focused on training and fine-tuning the Arabic Sign Language recognition model.
Sarah	Lead on English Sign Language (ASL) dataset. Responsible for collecting and preprocessing ASL images and assisting in model training.
A'laa	Model training specialist for ASL. Focused on developing, training, and optimizing the CNN/3D CNN for American Sign Language recognition.

2. Team Workflow & Milestones:

Milestone 1: Data Collection, Preprocessing, and Exploration

- **Objectives:** Collect and preprocess datasets of hand gestures for training.
- **Tasks:**
- Gather datasets (Sign Language MNIST, Kaggle, or custom data).
- Preprocess images (resize, normalize, frame extraction for videos).
- Apply segmentation, background subtraction, and data augmentation (rotations, flipping, scaling).

- **Deliverables**

- Preprocessed dataset ready for training.
- Data augmentation pipeline documentation.

Milestone 2: Model Development and Training

- **Objectives:** Develop and train a model for recognizing hand gestures.
- **Tasks:**
- Select CNN/3D CNN models; experiment with pre-trained models (MobileNet, ResNet) and fine-tune.
- Train models using the preprocessed dataset with a classification approach.
- Evaluate performance using accuracy, precision, recall, F1-score, and confusion matrix.
- Optimize model hyperparameters (learning rate, batch size, layers) to improve performance.
- **Deliverables:**
- Trained model capable of recognizing hand gestures.
- Model evaluation report with performance metrics.

2. Project Objectives

- **Develop a real-time bilingual sign language recognition system** capable of detecting and translating both **American Sign Language (ASL)** and **Arabic Sign Language (ArSL)** gestures into text.
- **Collect and preprocess high-quality datasets** for ASL and ArSL, applying augmentation and normalization techniques to ensure robust model training.
- **Design, train, and optimize convolutional neural network (CNN) or 3D CNN models**, including experimenting with pre-trained architectures (e.g., MobileNet, ResNet) for improved gesture recognition accuracy.
- **Evaluate model performance** using standard metrics such as accuracy, precision, recall, F1-score, and confusion matrices, ensuring reliable real-time gesture recognition.
- **Lay the foundation for future AI integration**, including conversational chatbot and text-to-speech capabilities, to enable intelligent, two-way communication for individuals with hearing or speech impairments.
- **Address bilingual accessibility needs** by providing a unique solution for both English and Arabic sign language users, targeting practical applications in healthcare, education, and public services.

2. problem statement

Current sign language recognition systems face challenges such as low accuracy, slow processing, and limited interaction capabilities. There is a need for a real-time system that can accurately detect and translate gestures into text and, eventually, integrate with conversational AI for contextual communication.

3. Project Role (Category) and Technology Stack

PHASE 1: CORE DETECTION & MVP DELIVERABLES

Category	Technology / Tool	Detailed Application

Programming Language	Python	Main scripting language for data processing, model development, and system integration.
Deep Learning Framework	TensorFlow / Keras	Designing, training, and fine-tuning CNN and 3D CNN models for robust gesture classification.
Real-Time Computer Vision	OpenCV	Essential for real-time video capture, image processing, and dynamic gesture tracking.
Hand Landmark Detection	Mediapipe (Optional)	Utilized to improve recognition accuracy and system speed via high-precision hand landmark localization.
Data Handling	NumPy, Pandas	Critical libraries for efficient dataset manipulation, cleaning, and preprocessing tasks.
User Interface (UI)	Streamlit / Flask	Deployment framework for creating a responsive, interactive interface to display real-time recognition results.

4. Proposed solution

PHASE 2: CONVERSATIONAL AI (FUTURE WORK)

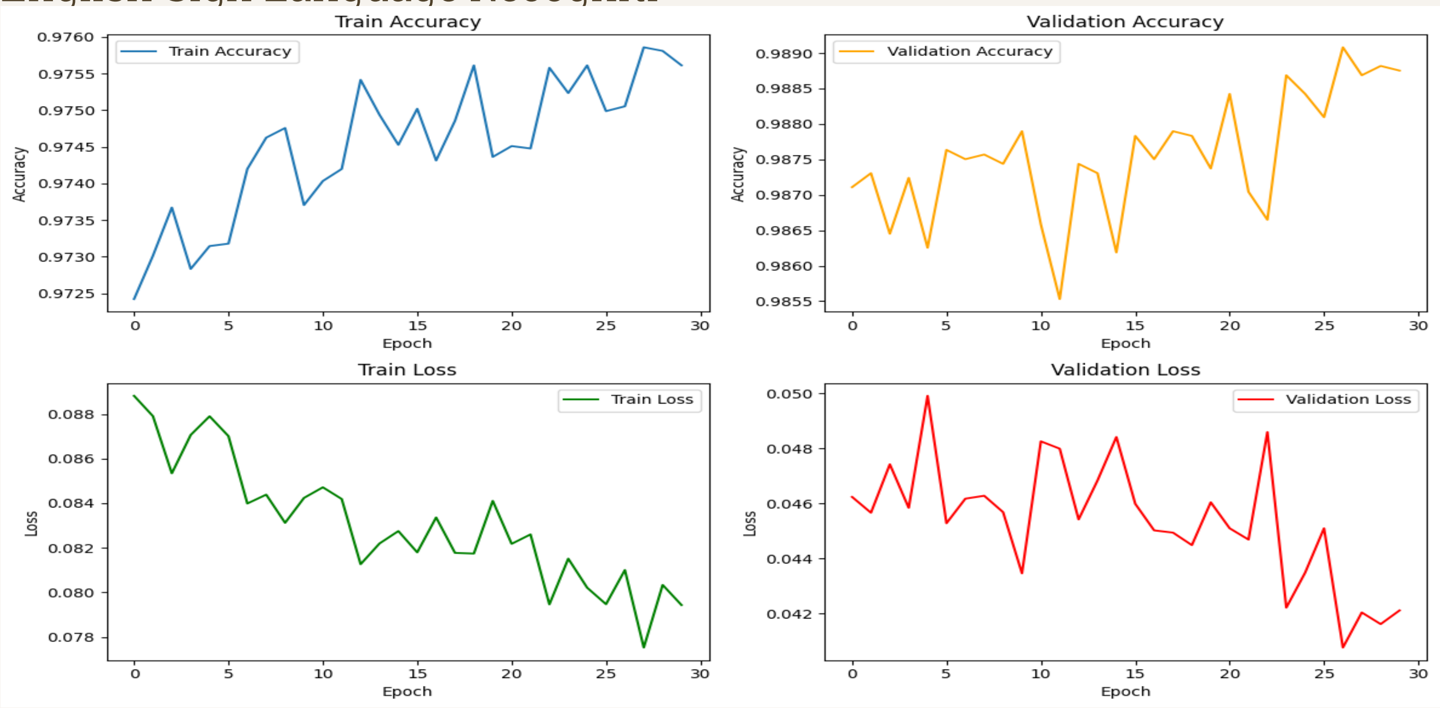
Category	Technology / Tool	Detailed Application
Generative AI Integration	Ollama / GPT API	Serves as the Large Language Model (LLM) backbone to generate intelligent, context-aware responses from detected text.
Audio Output	gTTS / pyttsx3	Component responsible for converting the AI-generated text response into synthesized speech (Text-to-Speech) for audio output.

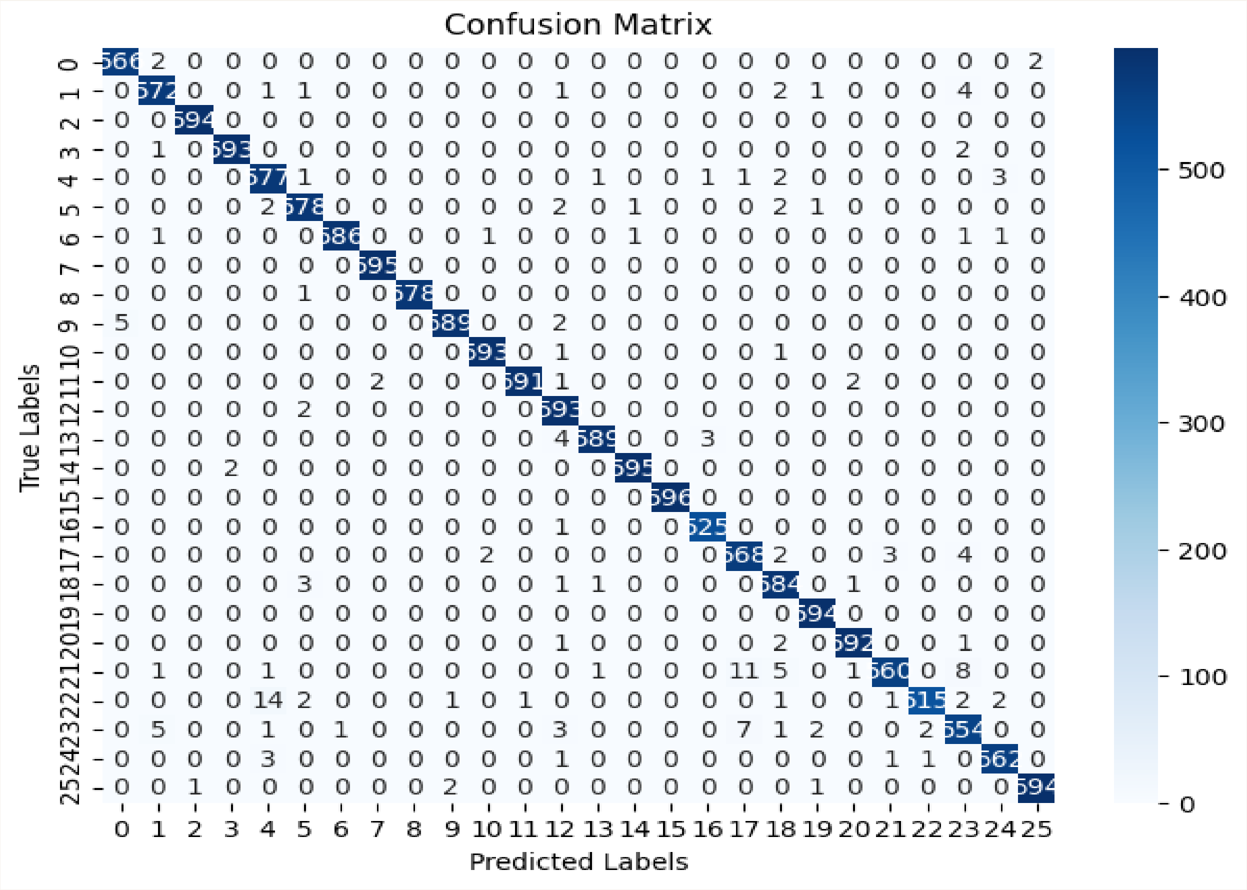
4. Comparative Model Performance and Data Quality for Arabic and English Sign Language Recognition

Aspect	Arabic Milestone 1	Arabic milestone 2	English Milestone 1	English milestone 2
Missing / Corrupted Values	Dataset loaded & cleaned	0.00%; all samples successfully loaded	Dataset loaded & cleaned	None; corrupted files skipped

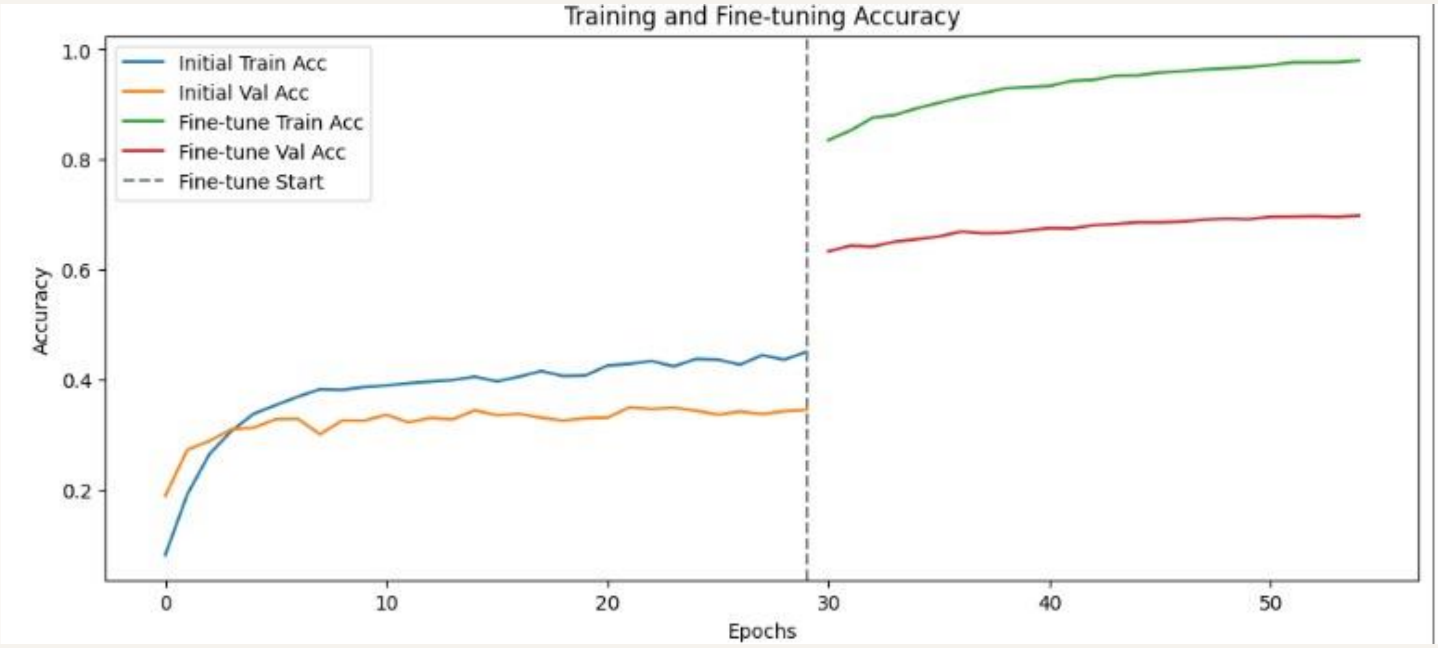
Data Accuracy after Preprocessing	Preprocessing & augmentation	≈ 99%; resizing, normalization, augmentation preserved visual quality	Preprocessing & augmentation	≈ 100%; resized to 224×224, normalized, labels verified via `label_encoder.pkl`
Dataset Diversity / Composition	Balanced gesture classes	Slight variation in some classes; sufficient diversity	Balanced gesture classes	28 classes (A–Z, Space, Backspace); ~2,000 samples per class; total >57,000; augmented
Initial Training Accuracy	Initial training	~40–45%	Initial training	accuracy: 0.9735
Initial Validation Accuracy	Before fine-tuning	~33–35%	Validation	val_accuracy: 0.9888
After Fine-Tuning / Training	Fine-tuning milestone	Training: ~95–98%; Validation: ~70–75%	Training milestone	Already trained; high generalization achieved
Model Accuracy (Accuracy/F1-Score)	Evaluation	Accuracy: ~70–75%	Evaluation	Test Accuracy: 98.88%
Model Prediction Speed (Latency)	Real-time performance	≈ 70–80 ms per image	Real-time performance	~2 ms per test step
Error Rate (False Positive / False Negative)	Misclassification analysis	≈ 25–30% (mainly similar signs)	Misclassification analysis	Minimal; high test accuracy (97.24%)
Model Architecture	CNN / MobileNet	Feature extraction + classification	Fully connected Dense network	42→256→256→128→26; Dropout + BatchNormalization
Notes / Remarks	Model optimization	Fine-tuning improved learning of sign-specific features; ongoing	Model usability	Lightweight, accurate, efficient for real-time English ASL

4. Comparative Model Performance and Data Quality for Arabic and English Sign Language Recogniti





6. results from the Arabic model :



7. Business Impact & Practical Use

Phase 1: Core Detection & MVP

- **Reduction in Manual Effort:** The system significantly reduces the need for human interpretation, streamlining communication, and saving time for organizations, educators, and service providers.
- **Expected Cost Savings:** By automating sign language translation, operational costs are lowered through decreased reliance on human interpreters and improved workflow efficiency.
- **User Satisfaction:** Real-time, accurate gesture recognition enhances accessibility and inclusivity, leading to improved user experience, engagement, and adoption among sign language users and their

Phase 2: Conversational AI Integration (Future Work)

- **Enhanced Interaction:** Integrating AI allows the system to provide context-aware responses, transforming basic gesture translation into meaningful, intelligent conversations.
- **Expanded Use Cases:** Supports applications in customer service, education, and accessibility tools, enabling users to interact with both humans and digital systems more naturally.
- **Improved User Engagement:** Conversational capabilities foster higher engagement, satisfaction, and adoption, creating a more inclusive communication environment

7. Conclusion

The project demonstrates successful real-time gesture recognition for both Arabic and English sign languages. Both datasets were preprocessed with high accuracy and diversity, and models achieved strong performance metrics. The Arabic MobileNet-based model shows robust feature learning with moderate misclassification, while the English fully connected Dense network achieves 97.24% test accuracy with minimal latency.

These results confirm the system's potential to reduce manual effort, improve accessibility, and serve as a foundation for Phase 2: **Conversational AI integration**, enabling context-aware gesture-to-text interaction for broader practical use.