

16<sup>TH</sup> ITU ACADEMIC CONFERENCE

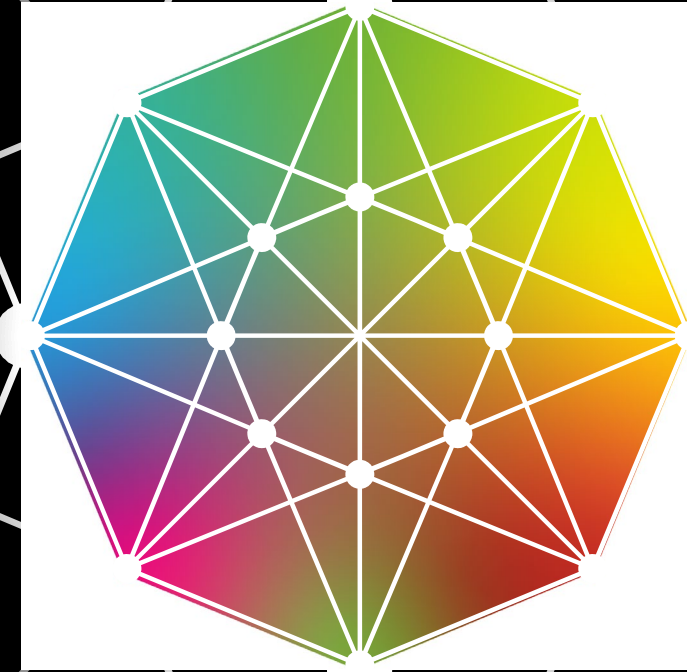
# ITUKALEIDOSCOPE

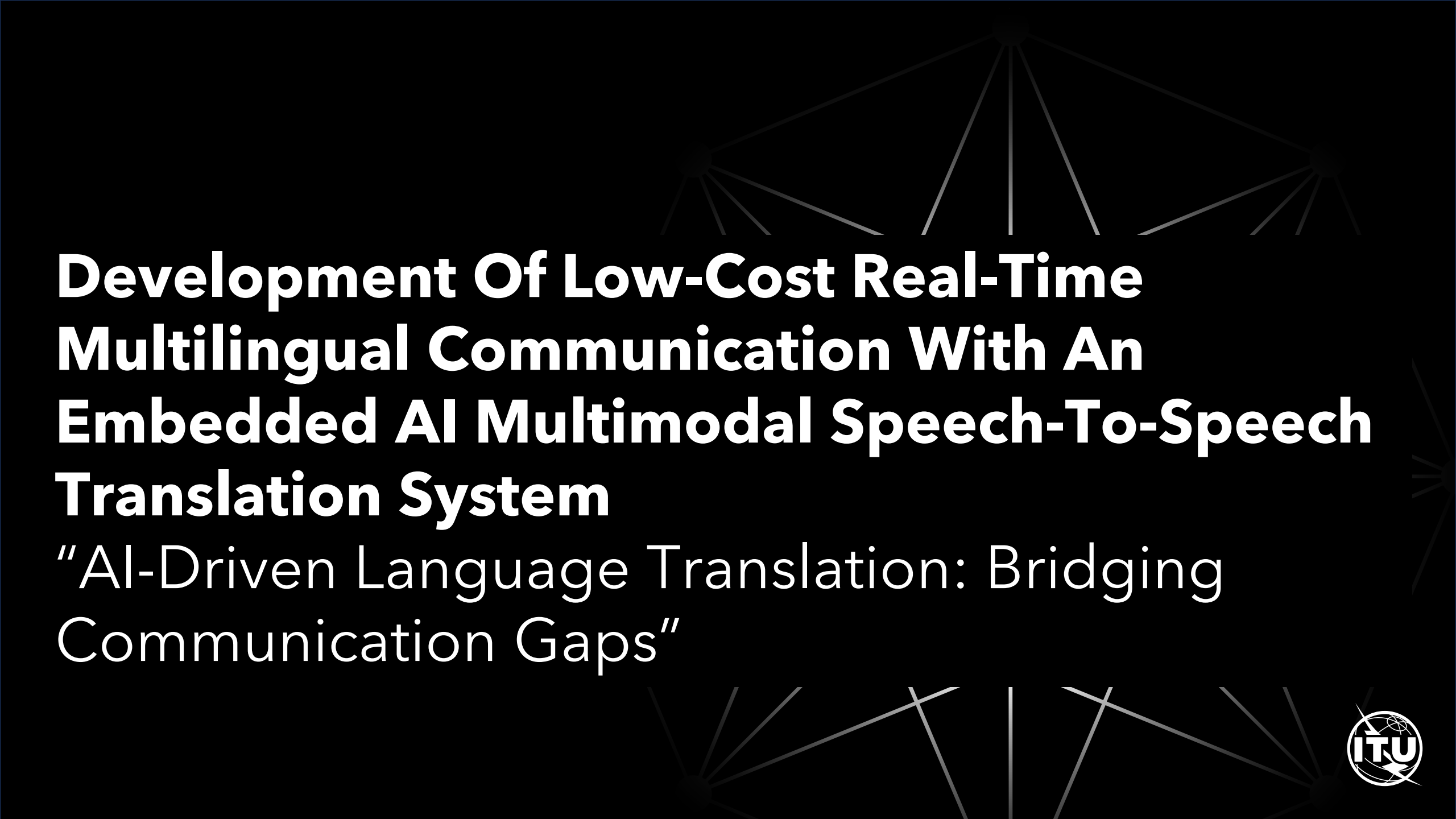
GENEVA2026

*AI and frontier  
technologies for good*

7-9 July 2026  
Geneva, Switzerland

Hosted by  **AI for Good**  
Global Summit





# **Development Of Low-Cost Real-Time Multilingual Communication With An Embedded AI Multimodal Speech-To-Speech Translation System**

"AI-Driven Language Translation: Bridging  
Communication Gaps"



## **Professor James AGAJO**

Federal University of  
Technology Minna Nigeria

Session #S4.2

Translation has long been essential for promoting **global communication, knowledge sharing, and cross-cultural understanding.**

**Translation methods have evolved from manual interpretation and human translators to AI-powered machine translation.**

## Problem Context

Language barriers continue to hinder effective communication across critical sectors, including:

- **Healthcare** – Misinterpretation may result in incorrect diagnosis or treatment.
- **Business** – Communication delays reduce productivity and collaboration.
- **Education** – Limited access to multilingual learning resources.
- **Tourism & Diplomacy** – Difficulty in seamless cross-cultural interaction.

## Challenges with Existing Translation Systems

- Human interpreters are expensive and not always available.
- Text-based translators often introduce delays and translation errors.
- Offline translation systems have:
  - Limited processing capability.
  - Reduced contextual understanding.
  - Lower translation accuracy.
- Many AI translation systems require a **stable internet connection**, making them unsuitable for real-time communication in remote or resource-constrained

## Research Gap

- There is a need for an **efficient, Low cost, context-aware, and low-latency multilingual translation system** capable of supporting accurate real-time communication with minimal dependence on internet connectivity.

### Takeaway:

- *Developing intelligent real-time translation systems is essential for breaking language barriers and enabling seamless global communication.*

## **Online Mode:**

Cloud-based AI translation using Gemini 2.5 Flash.

**Offline Mode:** Retrieval of previously translated phrases from an **SD card cache** and execution of an **embedded lightweight translation**

Ensures **continuous communication in low-connectivity or no-network environments.**

## **Multimodal Speech-to-Speech Communication**

- Supports a complete **Speech → Text → AI Translation → Text-to-Speech → Speech** workflow.

# Key Advantage

Edge Computing

Faster response with reduced cloud dependency.

Embedded AI

Intelligent language detection and translation.

Low Cost & Portable

ESP32-based handheld solution

Low Latency

Easily supports additional languages and AI services.

Scalable

Easily supports additional languages and AI services.

- Performs local caching, API management, and system control on the ESP32
- reducing latency, bandwidth consumption, and cloud service costs.
- Reducing reliance on high-end computing hardware.

- Embedded web server enables secure device configuration and translation cache management.

- **Multimodal Speech-to-Speech Communication**
- Supports a complete **Speech → Text → AI Translation → Text-to-Speech → Speech** workflow.

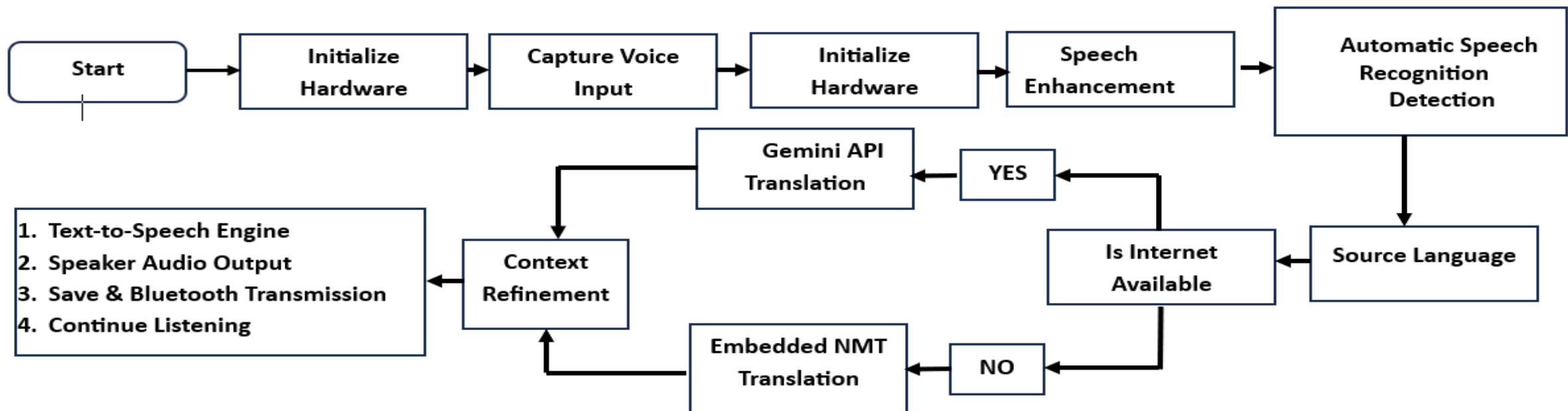
# Process involve

This research developed a :

- low-cost embedded AI multimodal speech-to-speech translation system that integrates automatic speech recognition, neural machine translation,

- Contextual language Understanding, and text-to-speech synthesis into a unified edge computing platform.

Unlike existing cloud-based translation systems, the proposed solution provides low-latency, privacy-preserving, and energy-efficient multilingual communication with minimal internet dependency



# Real-Time Translation Process

## Integrated system flow

1. Gather speech data
2. Improve sound quality
3. Execute streaming STT
4. Make use of predictive translation
5. Modify the model according to the user
6. Produce speech that has been translated
7. When in group mode, synchronize devices

## Key performance metrics

- Completely offline operation
- Robust-Noise accuracy
- Personalized translation quality
  - Multi-user synchronization;
  - Latency  $\leq 50$  ms

## Embedded AI Speech-to-Speech Translation Algorithm

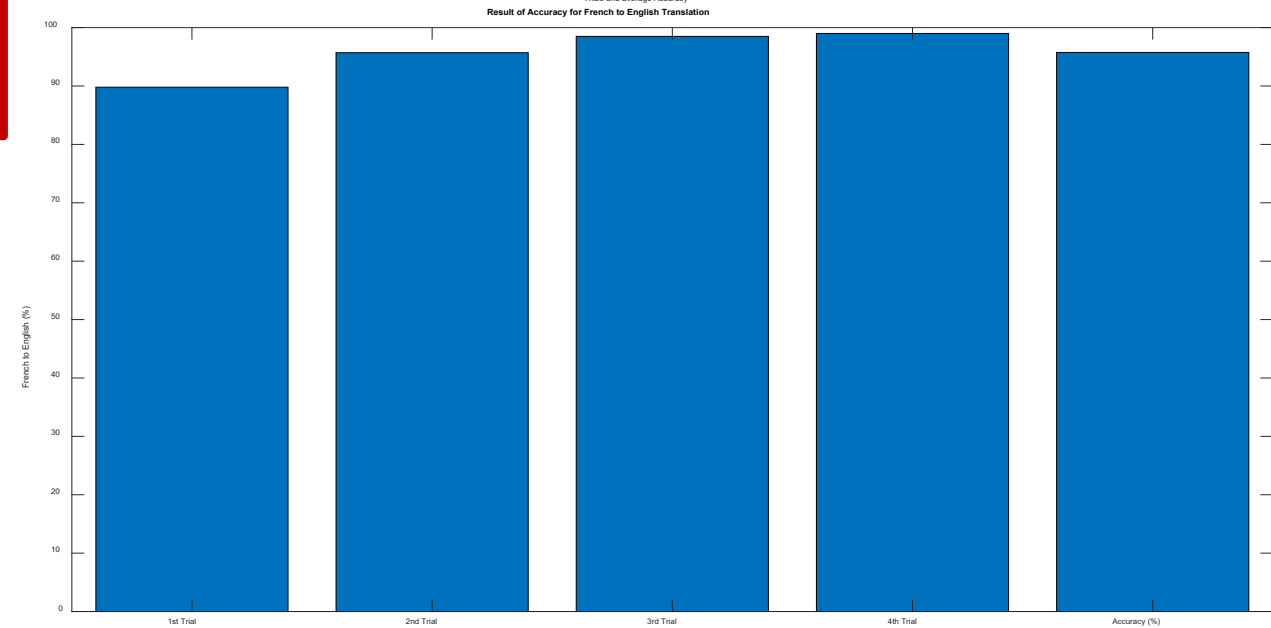
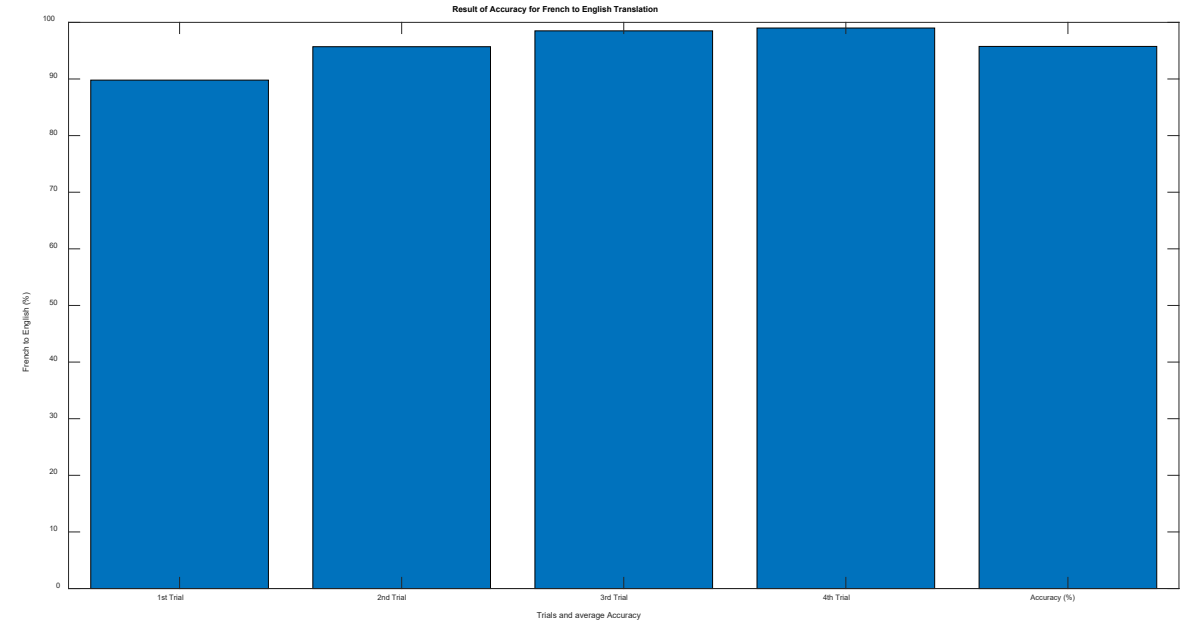
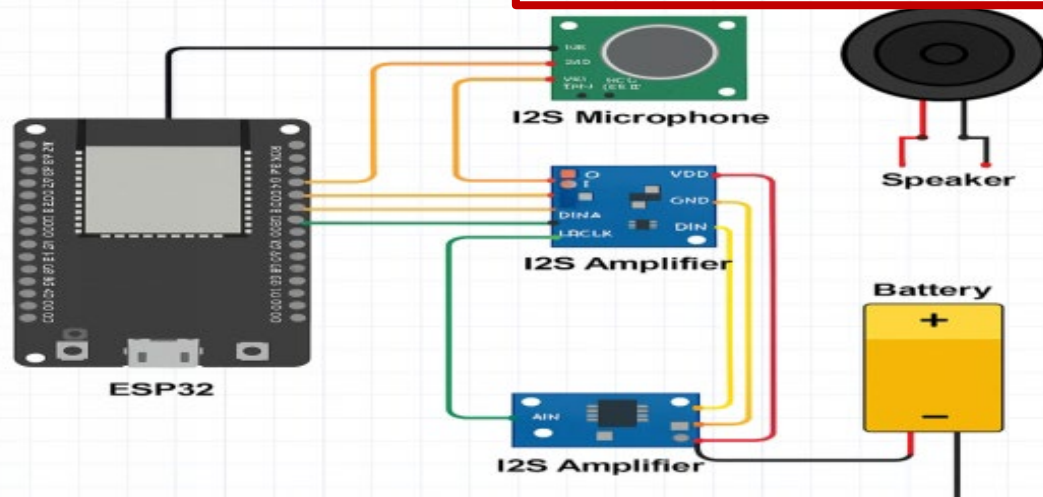
- Step 1: System Initialization
- Step 2: Capture Speech
- Step 3: Speech Enhancement
- Step 4: Automatic Speech Recognition (ASR)
- Step 5: Language Identification
- Step 6: Translation Mode Selection
- Step 7A: Online Translation
- Step 7B: Offline Translation
- Step 8: Context Refinement
- Step 9: Output Selection
- Step 10: Speech Synthesis
- Step 11: Audio Playback

# Result and Discussion

Robust performance across multiple metrics: including:

- accuracy
- speech clarity
- latency
- power usage
- accuracy in noisy environments.

- Accuracy results indicate strong translation performance, with an average of 97.7% for English to French
- and 95.75% French to English across four trials, showing consistent improvement from 95.60% to 99.00% and 89.80% to 99.00% respectively



# Conclusion

Developed a low-cost, portable embedded AI translation system that integrates Edge Computing (ESP32) with cloud-based AI (Google Gemini 2.5 Flash) to enable real-time multilingual speech-to-speech communication.

Implemented a hybrid online–offline architecture, incorporating SD card translation caching and offline speech capabilities, ensuring continuous multilingual communication even in environments with limited or no Internet connectivity.

Enhanced translation efficiency and user experience through intelligent language detection, contextual AI translation, Bluetooth connectivity, and natural speech synthesis using Google Cloud Text-to-Speech and the MAX98357A audio module.

Provides a scalable and affordable solution for multilingual communication in education, healthcare, tourism, emergency response, and rural communities, with the potential for future integration of on-device AI models supporting fully offline real-time translation.

# Thank you!

