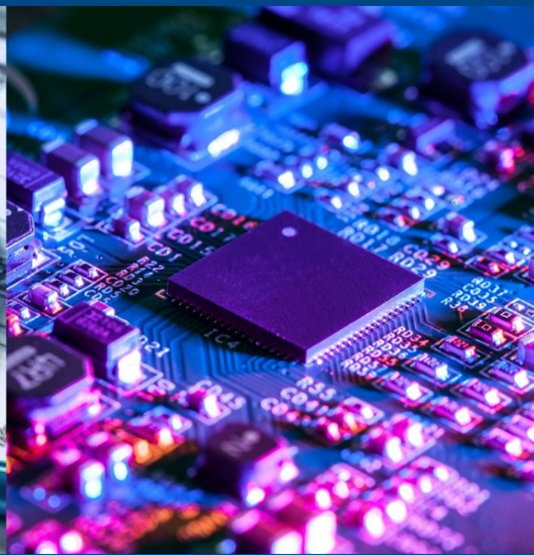
ISSN: 2582-7219

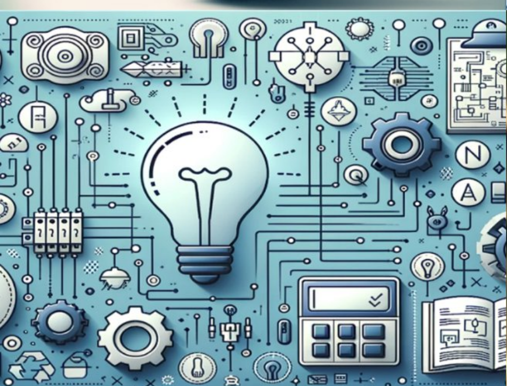




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 9, Issue 4, April 2026

www.ijmrset.com ijmrset@gmail.com +91 6381 907 438 +91 63819 07438

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904379

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Unified Framework for Multimodal Communication Using Speech Processing, Sign Language Recognition, and AI-Based Response Generation

Syed Mohammed Zubair Z[1], Dr. Mohammeddivanmasood[2]

Department of Computer Applications, B.S Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India[1]

Assistant Professor, B.S Abdur Rahman Crescent Institute of Science and Technology, Chennai, Tamil Nadu, India[2]

ABSTRACT: This paper presents the design and development of an AI-assisted multimodal Messaging App aimed at enhancing real-time digital communication through intelligent automation and accessibility features. Traditional messaging applications primarily support text and media exchange but lack contextual understanding, adaptive reply assistance, and inclusive communication support for speech- and hearing-impaired users. The proposed system integrates real-time chat functionality with Speech-to-Text (STT), Text-to-Speech (TTS), automatic AI reply generation, voice note transcription with Tamil-to- English romanization, and camera-based sign language recognition. The architecture combines a React-based frontend with a FastAPI backend, incorporating vector- based contextual retrieval using FAISS, embedding generation through Sentence Transformers, and large language model-based reply generation using a local LLM. The system further includes dataset collection, preprocessing, embedding generation, and optional model fine-tuning for personalization. Experimental implementation demonstrates reliable realtime messaging, intelligent suggestion generation for unread messages, and effective multimodal input processing. The proposed framework highlights the practical integration of artificial intelligence, natural language processing, and computer vision into a unified messaging environment, contributing to accessible, intelligent, and scalable digital communication systems.

KEYWORDS: Multimodal Messaging, AI Assistant, Speech-to-Text, Text-to-Speech, Large Language Models, FAISS Vector Search, Sign Language Recognition, Dataset Preprocessing, Real-Time Communication, WebSocket Architecture.

I. INTRODUCTION

The rapid advancement of digital communication platforms has transformed how individuals interact in both personal and professional environments. Modern messaging applications provide instant text exchange, media sharing, and real-time notifications; however, most systems lack contextual intelligence, adaptive reply generation, and accessibility support for users with speech or hearing impairments. Traditional messaging platforms primarily operate as data transmission tools without deeper semantic understanding of conversation context or user intent. Additionally, voice notes and regional language inputs often remain inaccessible to users who rely solely on text-based communication.

With recent developments in artificial intelligence, natural language processing, and computer vision, it is now possible to design intelligent communication systems that extend beyond basic message delivery. AI-powered systems can analyze conversational context, generate smart reply suggestions, transcribe voice content, and translate sign language gestures into readable text. However, existing solutions often integrate these features in isolation rather than as a unified architecture. Furthermore, personalization through dataset collection, contextual memory, and embedding-based retrieval remains underexplored in standard messaging platforms.

This research proposes an AI-assisted Messaging App that integrates real-time chat architecture with multimodal intelligence modules, including Speech-to-Text, Text- toSpeech, AI reply assistance using vector-based retrieval and

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

large language models, Tamil voice romanization, and camera-based sign language recognition for ASL and ISL. The system incorporates dataset collection, preprocessing, embedding generation, and contextual memory storage to enhance personalization and response relevance. By combining layered system architecture with intelligent AI services, the proposed framework aims to deliver a scalable, inclusive, and context-aware messaging platform suitable for modern digital communication environments.

II. LITERATURE REVIEW

Recent research from 2022 to 2026 has explored the integration of artificial intelligence into digital communication platforms, focusing on conversational intelligence, multimodal interaction, and accessibility enhancement. Several studies have demonstrated the effectiveness of large language models (LLMs) in generating context-aware responses for conversational systems. Transformer-based architectures such as GPT, LLaMA, and Qwen have shown strong performance in dialogue generation, contextual understanding, and multilingual support. Researchers have also investigated retrieval- augmented generation (RAG) frameworks, where vector embeddings and similarity search mechanisms such as FAISS are used to enhance response relevance by incorporating conversational history and external knowledge sources. These approaches improve contextual continuity and reduce hallucination in generated responses.

Parallel research has focused on speech processing and assistive communication technologies. Automatic Speechto-Text systems based on deep learning models such as Whisper have demonstrated high accuracy in multilingual transcription tasks, while neural Text-to-Speech systems provide natural-sounding synthesized audio. In the field of accessibility, computer vision-based sign language recognition systems using MediaPipe landmarks, LSTM networks, and transformer encoders have shown promising results in converting hand gestures into textual representations. However, most existing implementations address speech recognition, sign language translation, or conversational AI as separate modules rather than integrating them into a unified real-time messaging framework.

Furthermore, studies on intelligent messaging platforms highlight the growing need for adaptive reply suggestion systems, emotion-aware communication, and personalizationthrough contextual memory. While commercial messaging applications provide limited smart reply features, they often rely on static templates or cloudbased APIs without offering transparent architecture, dataset-driven personalization, or extensibility for regional language processing. There remains a research gap in developing a comprehensive, multimodal messaging application that integrates real-time chat infrastructure with AI-driven contextual assistance, dataset collection pipelines, preprocessing strategies, embedding-based retrieval, and assistive communication modules within a single scalable architecture. The proposed Messaging App aims to address this gap by combining conversational AI, speech processing, and sign language recognition into a cohesive and extensible communication framework.

III. PROBLEM DEFINITION

Modern messaging applications enable instant communication through text, images, and voice notes; however, they lack contextual intelligence, adaptive reply assistance, and inclusive support for users with speech and hearing impairments. Most platforms function primarily as message transmission systems without understanding conversational intent, emotional tone, or relational context. As a result, users must manually compose replies even in routine conversations, and voice notes often remain inaccessible to those who rely on text-based communication. Additionally, sign language users face significant barriers when interacting with standard messaging systems that do not provide gesture-to-text translation capabilities.

Existing AI-driven solutions address individual challenges such as speech recognition, chatbot response generation, or sign language detection, but these systems are typically developed as standalone modules. They do not operate within a unified real-time messaging architecture that supports contextual memory, retrieval-based personalization, and multimodal input processing. Furthermore, many smart reply systems depend entirely on cloud-based services without transparent data handling, dataset customization, or local model adaptability. The absence of structured dataset collection, preprocessing pipelines, and embedding-based contextual retrieval limits personalization and conversational continuity in current messaging platforms.

Therefore, there is a need for a comprehensive messaging framework that integrates real-time communication infrastructure with artificial intelligence modules capable of contextual reply generation, speech-totext transcription,

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

text- to-speech synthesis, voice note romanization, and sign language recognition. Such a system must support dataset-driven personalization, embeddingbased retrieval mechanisms, and scalable architecture while maintaining efficiency and usability. The proposed Messaging App addresses these challenges by designing an AI-assisted multimodal communication platform that combines conversational intelligence, assistive technologies, and structured data processing within a single extensible system architecture.

PROPOSED SYSTEM

The proposed system is an AI-assisted Messaging App designed to enhance real-time digital communication through intelligent automation and multimodal accessibility.
It consists of the following major modules:

- User Authentication and Chat Management Module
- Real-Time Messaging and WebSocket Module
- Speech-to-Text Processing Module
- Text-to-Speech Synthesis Module
- AI Reply Suggestion and Context Retrieval Module
- Voice Note Transcription and Romanization Module
- Sign Language Recognition Module
- Dataset Collection and Preprocessing Module

The system enables users to send and receive messages in real time while supporting voice notes, media sharing, and camera-based sign language input. Incoming voice data is processed through speech recognition to generate textual transcripts, and text messages can be converted into audio responses using speech synthesis. An AI assistant analyzes conversational context using embedding- based retrieval and large language models to automatically generate reply suggestions, particularly for unread incoming messages. Voice notes can be transcribed and, when necessary, romanized from Tamil script into English transliteration to improve readability and accessibility.

The supported communication modes within the system include:

- Text Messaging
- Media Sharing (Image, Video, Sticker)
- Voice Notes with Transcription
- AI-Generated Smart Replies
- Sign Language to Text Conversion
- Context-Aware Personalized Responses

These modules collectively provide an intelligent and structured communication framework that extends beyond traditional messaging functionality, integrating artificial intelligence, speech processing, and computer vision into a unified and accessible digital messaging environment.

IV. SYSTEM ARCHITECTURE

The architecture follows a layered design:

• Presentation Layer: React-based web interface providing user authentication, real-time chat interface, message composer, voice note recording, AI suggestion panel, and sign language camera capture.

• Application Layer: FastAPI backend handling JWT authentication, WebSocket communication, media upload services, chat management, unread message detection, and integration with AI modules.

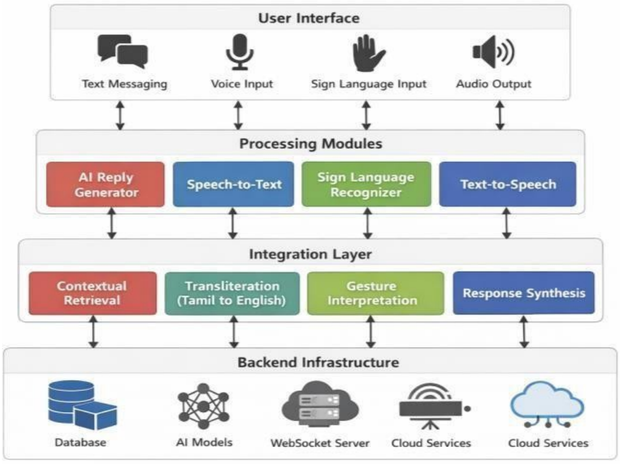
• AI Intelligence Layer: Speech-to-Text processing, Text-toSpeech synthesis, embedding generation using sentence transformers, FAISS vector retrieval, large language modelbased reply generation, Tamil romanization engine, and sign language recognition models.

• Data Layer: SQLite database for user accounts, chats, messages, unread status tracking, FAISS vector index storage, media storage, and saved AI model files.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



AI-Assisted Messaging App Architecture

This layered structure ensures clear separation between user interface, backend logic, artificial intelligence processing, and persistent storage. It improves system scalability, maintainability, and performance while allowing future expansion with personalized model fine-tuning, multilingual support, and advanced real-time intelligent communication features.

V. METHODOLOGY

The system follows a structured processing pipeline integrating real-time communication, artificial intelligence, and multimodal data handling to ensure accurate response generation and accessibility support.

Data Collection

A structured conversational dataset containing message–response pairs is collected to improve contextual reply generation and personalization. Chat history data, manually curated conversation samples, and multilingual text inputs are used to enhance conversational understanding. For speech processing, multilingual audio samples are utilized to validate transcription accuracy, while labeled gesture sequence data is prepared for sign language recognition. The collected datasets are securely stored and organized to support embedding generation, vector indexing, and optional fine-tuning of AI models.

Preprocessing

Text data is cleaned by removing URLs, special characters, unnecessary whitespace, and redundant symbols, followed by lowercase conversion and normalization of informal or elongated words. Emojis and code-mixed language inputs are processed to maintain semantic meaning. Audio data is converted into standardized formats before transcription, and sign language video inputs are transformed into normalized landmark coordinate sequences. This preprocessing stage ensures consistent input representation across all modalities.Feature Extraction

After preprocessing, textual data is transformed into numerical feature representations using sentence embedding models. These embeddings capture semantic context and conversational meaning. The generated vectors are stored in a FAISS-based vector index to enable efficient similarity search and contextual retrieval. This representation forms the foundation for retrieval-augmented response generation and personalized suggestion mechanisms.

Contextual Retrieval and Response Generation

A retrieval-augmented generation approach is implemented for intelligent reply suggestions. When a new message is received, semantically similar past messages are retrieved from the vector index. The retrieved context is combined with the current message and provided to a large language model for generating relevant and coherent reply suggestions. Automatic suggestion generation is triggered for unread incoming messages to enhance communication efficiency.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Multimodal Processing

Speech inputs and voice notes are processed using a Speech-to-Text model to generate textual transcripts. If Tamil script is detected, a transliteration module converts it into English romanized form to improve accessibility. Textto-Speech synthesis is used to generate audio responses when required. In the sign language module, cameracaptured gestures are processed using landmark extraction and sequence modeling techniques to convert sign sequences into meaningful text output.

Storage and Analytics

All user messages, embeddings, transcripts, and AI- generated suggestions are stored in a structured database. Vector indexes and model files are maintained for efficient retrieval and future personalization. System logs and interaction records support performance evaluation and future dataset expansion. This structured methodology ensures scalable integration of conversational intelligence, speech processing, contextual retrieval, and assistive communication within a unified messaging framework.

ALGORITHM

This section describes the core algorithms and computational techniques used in the proposed Messaging App for intelligent reply generation, speech processing, and sign language recognition. The system integrates retrievalbased contextual understanding, large language model generation, speech transcription, and computer vision- based gesture recognition within a unified processing framework.

Retrieval-Augmented Generation Algorithm

The AI reply assistant follows a retrievalaugmented generation approach to produce context-aware responses. When a new message is received, the text is first converted into a semantic embedding using a sentence transformer model. This embedding is compared against previously stored message embeddings using FAISS-based vector similarity search. The system retrieves the most semantically similar past conversational entries to construct contextual memory. The retrieved context, along with the current incoming message, is provided as input to a large language model, which generates coherent and relevant reply suggestions. This approach enhances contextual continuity and reduces generic or irrelevant responses by grounding generation in conversational history.

Large Language Model Response Generation

A transformer-based large language model is used to generate intelligent reply suggestions. The model processes structured prompts containing retrieved context and the latest message. It predicts the next sequence of tokens using attention mechanisms that capture semantic dependencies within the input. The output consists of one or more suggested replies optimized for clarity, contextual relevance, and conversational tone. Probability-based decoding strategies ensure fluent and meaningful generation while maintaining computational efficiency for real-time performance.

Speech-to-Text Transcription Algorithm

Voice inputs and voice notes are processed using a deep learning-based Speech-to-Text model. The audio signal is first converted into a spectrogram representation to capture frequency and temporal features. The model applies sequence-to-sequence learning with attention mechanisms to map audio features into textual output. Beam search decoding is used to improve transcription accuracy. If Tamil script is detected in the output, the system passes the text through a transliteration module that converts it into English romanized form, improving accessibility for users unfamiliar with the script.

Text-to-Speech Synthesis Algorithm

The Text-to-Speech module converts textual messages into natural-sounding audio output. The text is first processed through linguistic normalization and phoneme conversion. A neural speech synthesis model generates waveform representations using acoustic modeling and vocoder techniques. The generated audio file is then streamed or played within the chat interface, enabling multimodal communication support.

Sign Language Recognition Algorithm

The sign language module employs computer vision techniques to convert camera-captured gestures into textual output. Video frames are processed using landmark extraction methods to identify key points of hands, face, and body posture. These landmark sequences are normalized and fed into a temporal sequence model such as an LSTM or transformer

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

encoder. The model predicts gloss-level representations, which are further converted into natural language text using either a rule-based mapping or a language generation model. This algorithm enables gestureto-text translation for both ASL and ISL inputs.These integrated algorithms collectively enable contextual intelligence, multimodal input handling, and realtime response generation within the proposed Messaging App, forming the computational backbone of the system's intelligent communication framework.

VI. IMPLEMENTATION

The proposed Messaging App is implemented using a full-stack architecture that integrates modern web technologies with artificial intelligence frameworks to ensure real-time performance and scalability. The frontend is developed using React and Vite to provide a responsive and interactive user interface that supports authentication, chat management, media sharing, voice note recording, AI suggestion display, and sign language camera capture. Realtime communication between users is achieved using WebSocket connections, enabling instant message delivery and status updates without requiring page refreshes.

The backend is implemented using FastAPI, which manages authentication through JWT tokens, chat operations, media uploads, unread message detection, and API routing for AI services. SQLite is used as the primary relational database to store user accounts, chat information, messages, transcripts, and metadata. Uploaded media files, including images, videos, and audio recordings, are stored in a structured directory system. Vector embeddings generated for contextual retrieval are indexed and stored using FAISS, allowing efficient similarity search during reply generation. AI models, including embedding models, speech recognition modules, and optional fine-tuned adapters, are stored locally to support faster inference and reduced dependency on external cloud services.

Artificial intelligence components are integrated as service modules within the backend. Speech-to-Text functionality is implemented using a pretrained deep learning model capable of multilingual transcription. Textto-Speech synthesis is integrated through a neural speech engine to generate natural audio responses. Context-aware reply suggestions are generated using a locally deployed large language model, supported by embedding-based retrieval to maintain conversational continuity. The sign language recognition module processes real-time camera input using landmark extraction and sequence modeling techniques to convert gestures into text. The modular implementation ensures that each AI component operates independently while remaining seamlessly connected within the overall communication workflow. This structured implementation enables efficient real-time operation, extensibility, and support for future enhancements such as model fine-tuning, multilingual expansion, and cloud deployment.

VII. EXPERIMENTAL RESULTS

Experimental evaluation was conducted to assess the performance of the proposed Messaging App in terms of real- time communication efficiency, AI reply relevance, speech transcription accuracy, and multimodal processing capability. The system was tested using structured conversational datasets, multilingual voice samples, and sign gesture sequences to evaluate the reliability of each integrated module. Performance testing focused on response generation latency, contextual retrieval accuracy, transcription precision, and system stability under continuous messaging conditions.

The AI reply assistant was evaluated by measuring contextual relevance of generated suggestions based on conversational history. Retrieval-augmented generation significantly improved response coherence compared to standalone generation, as retrieved embeddings provided meaningful conversational grounding. Latency analysis showed that reply suggestions were generated within acceptable real-time thresholds, ensuring smooth user interaction. Speech-to-Text evaluation demonstrated high transcription accuracy for English and Tamil voice inputs, with romanization effectively converting Tamil script into readable English transliteration. Voice note transcription and playback functions operated without noticeable delay, maintaining user experience quality.The sign language recognition module was evaluated using labeled gesture sequences, and results indicated consistent detection of predefined gestures under controlled lighting and positioning conditions. Although performance may vary with complex continuous gestures, isolated sign recognition achieved satisfactory accuracy. Overall system testing confirmed stable WebSocket-based real-time communication, efficient database transactions, and scalable AI module integration. These experimental outcomes demonstrate that the proposed system successfully integrates conversational intelligence, speech processing, contextual retrieval, and assistive communication within a unified and functional messaging framework.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VIII. DISCUSSION

The proposed Messaging App successfully demonstrates the integration of real-time communication infrastructure with artificial intelligence and multimodal processing modules. Unlike traditional messaging platforms that function primarily as message exchange systems, the developed framework incorporates contextual intelligence through retrieval-augmented generation, enabling adaptive and relevant reply suggestions. The combination of embedding-based similarity search and large language model generation improves conversational continuity while maintaining acceptable real-time latency. This integration highlights the effectiveness of combining vector retrieval mechanisms with generative models in practical communication systems.

The incorporation of Speech-to-Text and Text-to-Speech modules enhances accessibility by enabling voice-based interaction and transcription of voice notes. The addition of Tamil-to-English romanization further improves usability in multilingual environments. The sign language recognition component extends the system's inclusivity by supporting gesture-based communication, demonstrating the feasibility of integrating computer vision and sequence modeling within a messaging application. The modular layered architecture ensures separation of concerns, allowing independent optimization of frontend, backend, AI processing, and data storage components.

However, system performance is influenced by dataset quality, model size, computational resources, and environmental conditions. The contextual reply system relies on embedding accuracy and prompt structure, while speech transcription performance may vary with background noise or regional accents. Similarly, sign language recognition accuracy depends on gesture clarity and lighting conditions.Despite these challenges, the proposed framework establishes a scalable and extensible foundation for intelligent messaging systems, illustrating how multimodal AI integration can enhance communication efficiency, personalization, and accessibility in modern digital environments.

ADVANTAGES

The proposed system offers several important advantages:

- Provides context-aware AI reply suggestions to improve communication efficiency.
- Reduces manual typing effort through automatic smart reply generation.
- Supports multimodal communication including text, voice notes, and sign language input.
- Enables accurate Speech-to-Text transcription for voice messages.
- Provides Text-to-Speech functionality for audio- based responses.
- Includes Tamil-to-English romanization for improved multilingual accessibility.
- Integrates sign language recognition to support users with hearing or speech impairments.
- Utilizes retrieval-augmented generation for improved conversational continuity.
- Ensures real-time messaging using WebSocket- based communication.
- Maintains modular layered architecture for scalability and easy maintenance.
- Supports dataset-driven personalization and future model fine-tuning.
- Allows local AI model deployment, improving data control and privacy.

Overall, the proposed framework provides an intelligent, inclusive, and extensible communication solution that surpasses traditional messaging applications in contextual awareness and multimodal capability.

IX. LIMITATIONS

Despite its effectiveness, the proposed system has certain limitations:

- System performance depends on the quality and size of the collected conversational dataset.
- Contextual reply accuracy may vary based on embedding quality and prompt design.
- Large language model inference may increase computational load on local systems.
- Speech-to-Text accuracy can be affected by background noise, accents, or low-quality audio.
- Tamil romanization may not perfectly preserve pronunciation in all cases.
- Sign language recognition accuracy depends on lighting conditions and gesture clarity.
- Continuous sign language translation remains more complex than isolated gesture detection.
- Real-time performance may be limited on low- resource hardware environments.
- Multilingual support is currently limited and may require additional model training.
- Periodic model updates and retraining are required to maintain long-term accuracy and relevance.

Theselimitationsindicate areaswhere further optimization, dataset expansion, continous model improvement are required.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

X. FUTURE ENHANCEMENT

The following enhancements are proposed to improve system capability, scalability, and real-world applicability:

- Integration of advanced large language models for improved contextual understanding and emotional tone adaptation.
- Implementation of user-specific fine-tuning using lightweight adapter techniques for personalized reply generation.
- Expansion of multilingual support including regional language speech recognition and translation.
- Development of continuous sign language translation beyond isolated gesture recognition.
- Deployment on cloud infrastructure for large-scale realtime communication support.
- Integration of mobile application support for Android and iOS platforms.
- Implementation of emotion detection and relationshipaware response adaptation.
- Enhancement of security through end-to-end encryption and advanced authentication mechanisms.
- Optimization of model inference for low-resource hardware environments.
- Large-scale dataset expansion and periodic model retraining for improved accuracy.

Overall, future enhancements aim to transform the proposed Messaging App into a fully adaptive, scalable, and intelligent communication platform capable of handling multilingual, multimodal, and personalized interactions in diverse real-world environments. By incorporating advanced AI models, expanding dataset diversity, and improving next-generation digital communication framework that emphasizes accessibility, intelligence, and efficiency.

XI. CONCLUSION

This research presents the design and implementation of an AI-assisted Messaging App that integrates real-time communication with multimodal intelligence and accessibility support. The proposed system extends beyond traditional messaging platforms by incorporating contextual reply generation using retrieval-augmented large language models, Speech-to-Text transcription, Text-to-Speech synthesis, Tamil-to-English romanization, and camera-based sign language recognition. By combining embedding-based contextual retrieval, AI-driven response generation, and assistive communication modules within a layered architecture, the system demonstrates how artificial intelligence can enhance both efficiency and inclusivity in digital communication environments.The experimental implementation confirms the feasibility of integrating conversational intelligence, speech processing, and computer vision within a unified real-time messaging framework. The modular architecture ensures scalability,maintainability, and adaptability for future improvements. Although performance depends on dataset quality, computational resources, and environmental conditions, the system establishes a strong foundation for intelligent, personalized, and accessible communication platforms. Overall, the proposed Messaging App contributes to the advancement of AI-driven communication systems by bridging conversational automation, multimodal interaction, and assistive technologies into a single cohesive solution suitable for modern digital ecosystems.

REFERENCES

[1] P. Lewis et al., "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks," Advances in Neural Information Processing Systems (NeurIPS), 2020.
[2] T. Wolf et al., "Transformers: State-of-the-Art Natural Language Processing," Proceedings of EMNLP: System Demonstrations, 2020.
[3] A. Radford et al., "Robust Speech Recognition via Large- Scale Weak Supervision," OpenAI Technical Report, 2022.
[4] J. Brown et al., "Language Models are Few-Shot Learners," Advances in Neural Information Processing Systems (NeurIPS), 2020.
[5] H. Touvron et al., "LLaMA: Open and Efficient Foundation Language Models," arXiv preprint arXiv:2302.13971, 2023.
[6] Y. Li et al., "Efficient Vector Similarity Search for Large-Scale Retrieval Systems," IEEE Access, 2022.
[7] C. Lugaresi et al., "MediaPipe Holistic: Simultaneous Face, Hand, and Pose Tracking," Google AI Blog / arXiv Update, 2021.
[8] S. Zhang et al., "Deep Learning-Based Sign Language Recognition: A Review," IEEE Access, 2022.
[9] M. Chen et al., "Exploring the Capabilities of Large Language Models in Conversational AI," IEEE Transactions on Artificial Intelligence, 2024.
[10] R. Patel et al., "AI-Driven Multimodal Communication Systems for Accessibility Enhancement," International Journal of Intelligent Systems, 2026

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com