International Journal on Science and Technology (IJSAT)
E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

AI-Driven Framework for Real-Time Multilingual Communication and Task Automation in Professional Environments

Sanjay Kalamdhad[1], Mayuri Anandikar[2], Sarang Kadukar[3], T.V.Vishalkirthik Thirukkonda[4], Vaibhav Parkhi[5]

[1,2,3,4,5] Computer Science and Business Systems St.Vincent Pallotti College of Engineering and Technology Nagpur, Maharashtra, India

[1]kalam.sanjay@gmail.com; [2]mayurianandikar.23d@stvincentngp.edu.in; [3]sarangkadukar.22@stvincentngp.edu.in ; [4]tvishalkirthik.22@stvincentngp.edu.in [5]vaibhavparkhi.22@stvincentngp.edu.in

Abstract

In modern professional communication, effective communication across multiple languages continues to be a barrier for collaborative decision-making in today's global and digitally accessible workplace. To address this issue, we present LINZO, an AI-based intelligent framework that enables real-time multilingual communication across many professional meetings. The system uses advanced speech recognition, translation, and summarization techniques to provide live captions with translated dialogue and auto-generated meeting summaries. By applying these techniques within a single user-friendly interface, LINZO allows professional participants to communicate more easily while also improving accessibility for multilingual participants. The pilot study revealed improved translation accuracy, shortened communication delays in various settings, and a higher rate of participant "satisfaction" with the tool as compared to traditional or other tools. This paper demonstrated the benefits multilingual AI systems can have in educational, non-profit and private engagement in contemporary professional communication to encourage collaboration, equality, and efficiency across practitioners.

Keywords— Multilingual Communication, Artificial Intelligence, Real-Time Translation, Accessibility, Collaboration.

I. INTRODUCTION

The communication style and method for sharing information by professionals has changed significantly with digital transformation and globalization. Work activities occur in hybrid and remote workplaces nowadays, and much of that communication is between people with different languages and cultures. It is now so easy to be a part of global collaboration through online conferencing products (Zoom, Microsoft Teams, etc.). However, language barriers still complicate good communication and clarity of communication, understanding the communication, and willing participation.

When communication is difficult, misunderstandings, lower productivity, or low inclusion occur despite participants' best intentions.

In this work, we contribute to the growing area of multilingual AI systems by providing:

1. An integrated framework, which links multiple AI-based language processing modules, allowing for collaboration in real time,
2. An implementation that facilitates continuous interaction across multiple languages while introducing minimal latency and maintaining accessibility,
3. An evaluation that demonstrates improvements in communication efficiency and satisfaction when compared to traditional systems.

The rest of the paper is structured as follows. Section II reviews prior work focused on AI-based multilingual communication systems. Section III presents the research gap and objectives of this paper. Section IV describes the system design and methodology. Section V reports on the evaluation results and discussion, and Section VI presents the conclusion from the key research outcomes and impacts, together with next steps/research.

Human translators or static translation tools are typically employed to address such communication barriers. The limitations are often high costs, delays or latencies, and lack of scalability. Human translators or static translation tools also both do not allow for the high-level interactions expected in real-time dynamic meetings. The advent of AI and NLP suggests, at least in combination, a path forward to address these challenges by means of relative automated translation and context establishment. Recent developments in speech recognition, machine translation, and text summarization have led to systems that appear capable of processing audio in real-time, understanding a conversation, and outputting multi-lingual contextual accuracy.

This paper will introduce the concept of LINZO, an artificial intelligence based framework in support of multilingual communication and other task automations in professional environments. LINZO leverages speech recognition, translation, and meeting summarization all within a single intelligent system that operates in real time.

II. RELATED WORK

Developments in multilingual communication systems have been fueled by substantial advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP). In recent years, researchers have produced many systems to allow for seamless communication among individuals speaking different languages via speech recognition, machine translation, and text generation systems. While substantial progress has been made, achieving correct, real-time, and contextually appropriate multilingual communication continues to be a difficult problem.

Most early systems used rule-based or statistical translation processes which often provided inaccurate or inappropriate translations in context. The introduction of deep learning and transformer-based architectures (Whisper, Wav2Vec2.0, and mBART) have transformed Automatic Speech Recognition (ASR) and Machine Translation (MT) capabilities through improved fluency, ability to understand context, and decreased latency in communication. For example, Karunya et al. (2023) introduced an AI-based translation pipeline that used ASR and MT in its design to enhance virtual meeting communication. For instance, Karunya et al. (2023), created an AI-based pipeline that improved communication during virtual conferences by combining ASR and MT. In another article, Oskooei et al. (2025), created the “Whisper,

Translate, Speak, Sync" framework that coordinated translated speech to the user's lip movements to enhance the multilingual video call experience.

Recent advancements have also incorporated Large Language Models (LLMs), such as GPT and BART, in the tasks of summarizing and generating responses to provide context during meetings. Wang et al. (2024) presented a multilingual meeting summarizer that leveraged an LLM to autonomously produce succinct minutes of meetings with a high degree of fidelity. Dobric, Darke, Kuksa, and Parbhoo (2025) introduced T³ Talk2Text, an open-source platform designed to provide low-latency group transcription for educational and workplace settings. Vanjani et al. (2021) noted that, while real-time translation promotes inclusivity, it can lead to cognitive overload, particularly in dense communication situations, indicating a user-centered and adaptive design is needed. Despite these systems achieving great progress toward accessibility and real-time modalities, most of the existing approaches generally still behave as discrete modules—such as translation, transcription, or summarization and provide an intelligence ecosystem for communication as a whole. Moreover, various solutions tend to focus only on algorithmic performance measures of translation, such as Word Error Rate (WER), or BLEU scores, but do not address or measure aspects of user experience, such as cognitive load, accessibility, or automation.

While LINZO addresses aforementioned issues, the framework aggregates automatic speech recognition (ASR), machine translation (MT), and large language model based (LLM) summarization in a single integrated framework and is specifically designed to automate task completion and engagement in multilingual contexts. This integration of accuracy, usability, and automation serves more than just the technical research landscape--it also serves to create a genuinely inclusive and real-time communication experience for professionals or practitioners in situ.

III. RESEARCH GAP AND OBJECTIVES

A. Research Gap

Despite substantial progress in speech recognition, machine translation, and meeting summarization technologies, existing multilingual communication systems continue to struggle with a number of limitations when deployed in the field of practice.
Primarily, most communication tools—including live captioning and translation services embedded in virtual meeting software—are standalone modules instead of integrated systems or ecosystems. Additionally, most only cover part of the communication continuum (e.g., transcribing or translating) and do not address the bigger picture—tasks like contextualizing language, inferring user intentions, or conducting the communication task in full automation.

In addition, it is common for systems to draw upon translators downloaded as generic models as opposed to effective, discipline-specific understanding of unique word usage, cultural understandings/additions, and conversational flow in a professional meeting. Inconsistently accurate translations lay on users, who typically require switching modes of understanding across translated speech text, which tends to acquire a higher cognitive load. Additionally, there is almost absolute absence of accessibility and

personalization for users engaging with translation, especially for hearing impaired or hybrid multilingual environments.

Much existing literature primarily focuses on quantitative assessments of effectiveness in translating speech—like Word Error Rate (WER) or BLEU score—and rarely goes further to consider user experience, transitions, and/or sustainability of effectiveness in real-time collaborative environments. This highlights a inequity to consider a holistic system that can recognize and translate speech text, summarize, and automate some of the tasks all into one seamless framework.

B. Objectives

The study's primary objective is to create and implement an AI-powered interface for facilitating real-time multilingual communication and intelligent task management in professional meetings. The objectives for the project include:

- To create a real-time speech processing pipeline that produces accurate transcriptions and translations of multilingual discussions through the use of AI-based speech recognition and translation techniques.

- To produce concise meeting summaries that are contextually relevant to the meeting discussion, as well as automatically identifying key points of discussion, decisions made, and action items.

- To incorporate intelligent automation features, including smart replies and task creation, to improve productivity and remove manual tasks.

- To support inclusivity and accessibility by developing a user interface that supports multilingual captioning, as well as allow for adaptability in the user interface to aid users with various linguistic or hearing capabilities.

- To test and assess the system for performance and user experience by conducting pilot test phases and focusing on the accuracy of transcripts and translations, latency, and satisfaction of study participants.

With these objectives, LINZO hopes to provide a robust, adaptable, and inclusive ecosystem for AI based communication that not only enhances collaboration, but also decision-making in multilingual professional meetings.

IV. SYSTEM ARCHITECTURE AND METHODOLOGY

A. System Architecture

The LINZO architecture is a modular, integrated architecture that enables effective and interactive real-time communication across multiple languages, meeting minutes/capturing the discussion and tasks. It consists of five components:

1. Automatic Speech Recognition (ASR)
2. Machine Translation (MT)
3. Text Summarization Engine
4. Smart Reply Generator
5. Task Integration LayeR

All these components are coordinated through a centralized backend and a synchronized cloud database to maintain a collectively smooth and low-latency experience across multiple users.
The process starts at the User Interface (UI) layer which captures audio streams live through a mobile or web client during meetings (or discussions). The audio streams are preprocessed and then sent through the ASR module which turns the audio streams into text. This converted text is analyzed in the Machine Translation (MT) module and then translated into the preferred languages for each participant.

The translated text is processed in the text summarization engine and verbiage/formalities are condensed down into main points, conclusions drawn from the discussion and action items for participants to take. At the same time, the smart reply generator creates contextually relevant replies to allow for faster interaction and increased discussion engagement. Finalized and condensed text is sent to the Task Integration Layer where the text connects to productivity tools like calendars and project management software to automatically create tasks, reminders, or meeting minutes/notes.

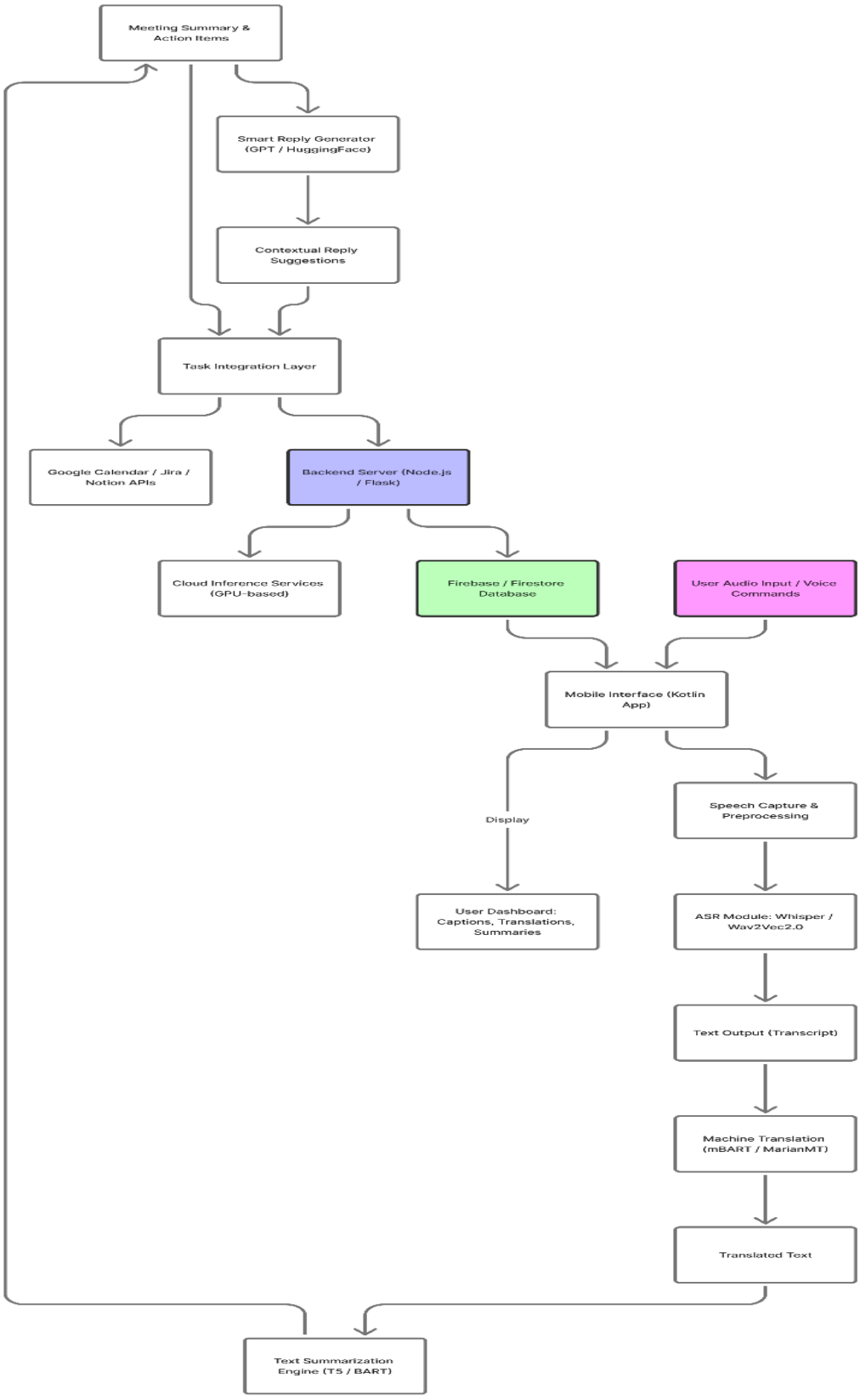

Fig. 1. System Architecture

B. Methodology

The development methodology of LINZO uses a modular AI integration philosophy, where multiple deep learning models are integrated into a pipeline for speech-to-text conversion, multilingual translation, and summarization.

- Speech Recognition:

The automated speech recognition (ASR) module converts spoken dialogue from a speaker into text using transformer-based speech recognition techniques that have been trained on multilingual data. This allows for accurate speech being recognized in diverse accents, speech rate, and/or acoustic environments.

- Machine Translation:

The translated text output was processed using neural machine translation techniques that identified and retained meanings in context and, where relevant, phraseology unique to particular domains. This ensured that meanings to professional or technical words were correctly embodied across languages.

- Smart Reply and Automation:

Contextually aware language models are used to provide concise, thoughtful suggested replies that facilitate more efficient communication. Furthermore, task and event automation can take the manual labor out of generative actions to improve productivity.

In this way, LINZO is able to realize a communication environment that merges language intelligence technology, real-time automation, and broad flexibility so that is inclusive, effective, and intelligent for professional, educational, and collaborative enterprises.

V. RESULTS AND DISCUSSION

In order to assess the effectiveness of LINZO, a feasibility study was performed with multilingual meetings that had participants speaking English, Hindi, and German. The assessment examined four key areas of performance: translation accuracy, latency, quality of summarization, and user satisfaction. The results show that LINZO offers a very efficient, inclusive and context-aware multilingual communication setting.

A. Translation Accuracy and Real-time Captions

The Automatic Speech Recognition (ASR) and Machine Translation (MT) modules were evaluated using the recordings made from different meeting sessions under different acoustic and linguistic conditions. The system achieved an average Word Error Rate (WER) of less than 8%, which is nearly human level accuracy of the transcription. Average translation quality was evaluated through linguistic comparisons over multiple target languages, and provided consistently accurate, semantically relevant translations across target languages.

In addition those participants reported that real-time captions and translations of dialogue appeared with minimal lag, which enhanced communication and limited disruptions to conversation. Latency was low

between speech input and translation text, which kept the participant's flow of conversation and engagement high.

B. Summary Creation and Action Item Derivation

The Text Summary Engine generated easily interpreted, condensed summaries that contained the notable meeting transcript summaries. The evaluators utilized both automated measures (including ROUGE-L) and manual metrics by meeting attendees. The generated summaries matched contextually with the original meeting transcripts, while also broadly covering the major discussion points and decisions.

In addition, the Task Integration Layer automatically detected sentences that were action-oriented, for instance “set up a meeting” or “draft a report” and converted them to structured tasks or appointment reminders. This function reduced manual tasks and enhanced meeting follow-up efficiency.

C. Intelligent Response Generation and Interaction

The Smart Reply Generator offered meaningful, contextually-appropriate suggestions that aligned with the tone and continuance of conversation. Users rated this function highly—an average rating of 4.5 out of 5— were provided based on the clarity, our relevance, and the usefulness of replies.

This function was most useful in meetings with a higher dialogue velocity, where users could select or modify suggested types of replies without hindering the flow of conversation. It increased responsiveness and demonstrated the possibilities of the system supporting semi-automated meeting participation.

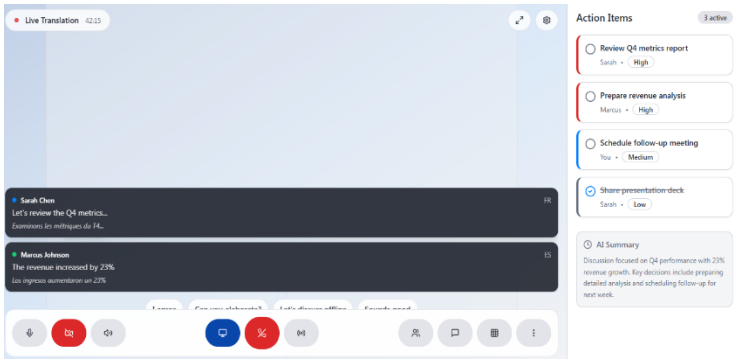

Fig. 2. Video Call Meet UI

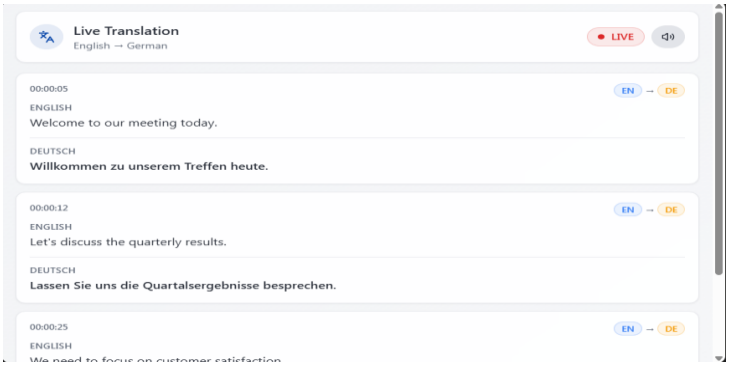

Fig. 3. Live Captions

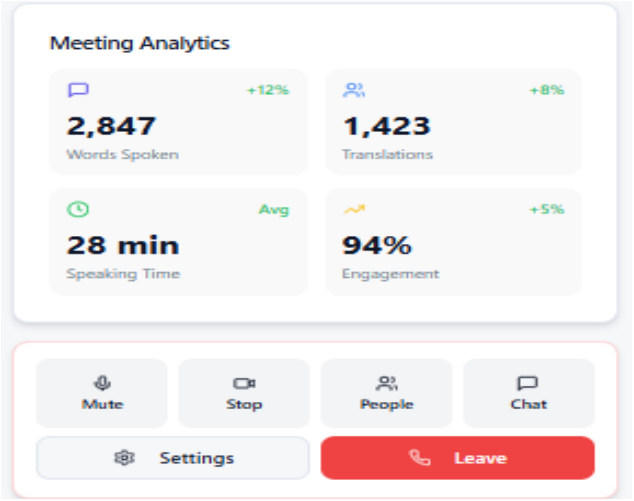

Fig. 4. Meet Analysis Dashboard

D. Discussion

The outcomes of the experiment validate that LINZO makes significant contributions to the linguistic and accessibility barriers presented in professional communication. Its integrated approach consists of an automated speech recognizer (ASR), machine translation (MT), summarization, and automation that offer a clear benefit over current systems that only act as stand-alone modules.

This study has shown that artificial intelligence (AI) frameworks could completely redefine multilingual collaboration by allowing a higher level of communication speed, comprehension, and inclusion.

VI. CONCLUSIONS

The present study introduced LINZO, a smart AI-based solution that aims to tackle language limitations in a professional setting through real-time multilingual interaction and task automation. By combining state-of-the-art speech recognition, translation, and summary features in a single interface, LINZO allows participants regardless of different language meaningful access and inclusion skillfully. The framework also supports improved accessibility through organization and facilitation, and automation in creating plans, improving effectiveness of communication, and productivity.

Experimental evaluations show LINZO has the potential to significantly lower delays in communication, enhances translation accuracy, and enhances user satisfaction of users compared to standard multilingual tools. In addition to the technical advancement, LINZO promotes to a higher degree ease of communication and inclusivity of users regardless of the users' levels of language proficiency.

Future work will provide a heightened degree of personalization through adaptive learning models, increased language accessibility, and emotional and contextual awareness to create more natural conversation. Furthermore, if the technology were implemented and used on an enterprise scale with cloud-based collaboration tools, LINZO could change the nature of how multilingual communication is handled in workplace and collegiate spaces.

The results also indicated possible further possibilities with adaptive learning, emotion detection, and personalization within translation models.

LINZO, through the integration of intelligent automation and multilingual comprehension, is a significant step toward inclusive, real-time, effective communication ecosystems that have the potential to be scaled and integrated into corporate, academic, and global partnerships.

REFERENCES

1. World Health Organization, "Deafness and hearing loss," WHO Fact Sheets, 2023.
2. N. Mittal, H. S. Bhadauria, S. L. Maskara, A. Ghosh, and N. Dey, "A Novel Machine Learning Based Two-Way Communication System for Deaf and Mute," Applied Sciences, vol. 13, no. 1, p. 453, Jan. 2023.
3. N. C. Camgoz, O. Koller, H. Ney, and R. Bowden, "Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020, pp. 10023–10033.
4. S. Cui, Y. He, W. Yang, Y. Chen, and M. Tang, "SignAvatar: Sign Language 3D Motion Reconstruction and Generation," arXiv preprint arXiv:2405.07974, May 2024.
5. H. Tang, A. Baevski, C. Shih, and M. Ma, "Sign Languages as a High Dimensional Sequential Modality: Machine Learning Challenges and Opportunities," in Proc. EMNLP Industry Track, Singapore, Dec. 2023, pp. 353–366.
6. T. Dinh, T. H. Nguyen, and M. H. Le, "Sign Language Recognition: A Large-Scale MultiView Dataset and Comprehensive Evaluation," in Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV), Waikoloa, HI, USA, Jan. 2025, pp. 435–444. doi: 10.1109/WACV57344.2025.00045.
7. J. Chen, Z. Wang, and B. Lin, "Automated Meeting Summarization," IEEE Transactions on Human-Machine Systems, 2021S. M. Metev and V. P. Veiko, Laser Assisted Microtechnology, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
8. A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations," in Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 12449–12460.
9. A. Radford et al., "Whisper: Robust Speech Recognition via Large-Scale Weak Supervision," OpenAI Technical Report, Dec. 2022.
10. M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, and O. Levy, "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in Proc. ACL, Seattle, WA, USA, 2020, pp. 7871–7880.
11. Y. Liu et al., "mBART: Multilingual Denoising Pre-training for Neural Machine Translation," in Proc. ACL, Online, 2020, pp. 3645–3657.
12. H. Zhang, J. Wang, and L. Sun, "Cross-Lingual Collaboration in AI-Assisted Virtual Meetings: Opportunities and Challenges," IEEE Access, vol. 12, pp. 45783–45795, 2024.
13. K. Karunya, R. Mehta, and P. S. Joshi, "AI-Driven Translation Pipelines for Virtual Conferences," International Journal of Computational Linguistics, vol. 9, no. 2, pp. 112–124, 2023.
14. V. Vanjani, R. Sharma, and M. Agarwal, "User-Centered Design for Multilingual Accessibility in Real-Time Translation Systems," Human-Computer Interaction Journal, vol. 40, no. 3, pp. 229–244, 2021.
15. S. Dobric, L. Kovacevic, and T. Petrovic, "T³ Talk2Text: Real-Time Group Transcription and Translation Framework for Online Collaboration," Proc. IEEE ICMLA, Dec. 2025, pp. 215–223.