MeetAI-Clone:Smart Video Conferencing With AI Notes

P.Chandra sekhar
Assistant Professor
Department of Computer Science and Engineering
Geethanjali College Of Engineering And Technology
Hyderabad, India
pchandrasekharreddy.cse@gcet.edu.in

Govindu Sai Varun
Department of Computer Science and Engineering
Geethanjali College Of Engineering And Technology
Hyderabad, India
22r11a05l2@gcet.edu.in

Mananthula Manaswini
Department of Computer Science and Engineering
Geethanjali College Of Engineering And Technology
Hyderabad, India
22r11a05m3@gcet.edu.in

Kondeti Vaishnav
Department of Computer Science and Engineering
Geethanjali College Of Engineering And Technology
Hyderabad, India
22r11a05l7@gcet.edu.in

Abstract— The recent transition towards working together in the virtual world has made it necessary to communicate, learn and work using video conferencing applications. Unfortunately, there is a lack of intelligent assistance in current systems to help capture, process and summarize the content of the discussion during video conferences, which makes such conferences less productive [1]. Although some platforms include some capabilities like basic video conference recording and transcription, they are still unable to understand the context or help the user with real-time assistance during discussion [2].

The paper introduces MeetAI, an artificial intelligence based framework for smart video conferencing aimed at increasing meeting productivity by assisting with real-time transcription and summarization of the discussion and taking notes about it. It uses a modular architecture where video communication, speech recognition, natural language processing and knowledge extraction are separated from each other. A real-time processing pipeline is used for collecting video conference data. AI-enabled capabilities such as speaker recognition, topic analysis, and contextual summarization help in creating brief and meaningful meeting minutes. Also, the proposed framework includes features for decision and action item extraction, helping the users in understanding the main takeaways from the discussions. In addition, MeetAI has the capability of performing multilingual transcription and works with cloud storage systems.

In order to ensure high usability levels of the system, an intuitive UI is provided which features live captions and summaries after meetings have been concluded. This allows easy sharing of information. Optionally, the users can integrate other external systems such as calendars and task management applications.

Evaluation of the proposed framework will be done using a prototype testing approach, with emphasis being placed on the areas of accuracy, usability, and system performance. The findings indicate that MeetAI greatly improves meeting efficiency and minimizes the manual work required for documentation purposes.

Keywords—smart video conferencing, AI-based transcription, meeting summarization, real-time collaboration, productivity improvement

I. INTRODUCTION

However, the fast-paced development of remote work and online learning, as well as the increasing demand for global collaboration, has led to the growing importance of video conferencing applications for information exchange. The organizations and individuals have to utilize them to hold meetings and lectures or engage in any other type of discussion. Nevertheless, with the rise of the number and complexity of virtual meetings, the issue of how to manage and retain the data exchanged during them has emerged as one of the major challenges [8].

At the moment, there are video conferencing apps that offer some basic functionality such as the possibility to record the video, stream live content, share screens, and transcribe the audio. Although these capabilities contribute to the process of communication between people, they cannot be viewed as sufficient means of helping users to comprehend the discussed issues and make summaries based on them. Transcriptions provided by these services are usually unstructured and do not include contextual interpretation of the phrases and sentences that can help users identify critical elements, conclusions made, and action points [1], [7].Intelligent Meeting System research areas have concentrated on various methods to perform tasks related to speech recognition, NLP, and meeting summarization. These techniques attempt to transform speech into structured information, allowing improved accessibility and management of knowledge. Nonetheless, a common approach involves creating different tools to perform one of the abovementioned functions and not provide an overall intelligent meeting system that would be able to combine several functionalities in one interface [1], [2]. Furthermore, such challenges as speaker detection, multilingualism, and context analysis become especially important in a changing meeting environment.

The recent progress in artificial intelligence as well as cloud computing can help to solve problems related to improving video conferencing. The current developments of speech-to-text technologies allow making real-time transcriptions of dialogues. Meanwhile, natural language processing allows analyzing the conversation content and extracting topics, main points, and actions to take based on the obtained information. Explainable AI makes it possible to ensure interpretability of outcomes produced by the machine learning model, which facilitates a user's understanding of the generated results [5], [6]. Although many innovations have been made in this domain, there is a dearth of systems which are able to perform all the above-discussed tasks in a coherent manner. Hence,

there is a need for the development of a framework which is capable of improving the efficiency of the meetings by automating the process of capturing valuable information from them.

The main purpose of this paper is to propose a framework called MeetAI that can facilitate video-based collaboration between individuals. In particular, it focuses on developing a platform based on artificial intelligence techniques that can enable users to extract valuable information from meetings in an efficient way. Thus, the proposed system will consist of modules that allow performing real-time transcription, context-aware summarization, speaker identification, and action item extraction tasks. This paper is structured as follows. In Section II, literature pertaining to intelligent meeting systems, transcription methods, and summarization approaches is presented. In Section III, the problem definition and the research gap that this work aims to address are stated. The MeetAI solution is presented in Section IV. In Section V, the meeting analysis approach employing artificial intelligence techniques is described. The prototype development and evaluation are described in Section VI. Section VII presents findings and future directions for improvement.

II. RELATED WORK

Intelligent Video Conferencing and Meeting Assistance Research has become quite popular in recent years owing to the fast-growing trend towards remote collaboration and communication. In most cases, research conducted in the field can be broadly classified into one of four broad categories. They include speech recognition and transcription, meeting summaries, knowledge extraction using artificial intelligence, and increased efficiency of work in the virtual environment.

Speech recognition and automatic transcription have become important components of intelligent meetings. The former involves translation of oral language into written text to facilitate further analysis of meeting contents. Owing to the advancements in deep learning technology, speech recognition and transcription software have become more accurate even under challenging conditions such as noisy environment and multiple speakers. For example, there are some interesting studies which demonstrate how current ASR systems increase accessibility and improve documentation of virtual meetings [1]. Despite their impressive capabilities, transcriptions still remain largely unstructured and devoid of context.Another important research field in making information accessible is that of meeting summarization. Summarization algorithms rely on NLP to produce short and focused summaries from long conversations. Both extractive and abstractive summarization methods have been tested for finding critical sentences, topics, and points discussed in meetings' transcripts [2], [3]. Although summarization helps overcome information overload, many of the existing algorithms lack the capability to properly understand the context, meaning of the speech, and communication process.

Knowledge extraction based on AI technologies is used in meeting systems to find critical information in meetings, such as decisions, action points, and topics of discussion. Meeting data becomes structured and transformed into meaningful knowledge through the application of NLP and machine learning models. Research studies in this area stress the role played by context analysis, identification of speakers, and semantics of discussions [4]. On top of that, efforts have been made to make AI knowledge extraction systems explainable to users [5], [6].Additionally, productivity and experiences of users in virtual collaboration environments have extensively been studied by researchers. It has been revealed that poor management of information during meetings, such as ignoring important points and decisions, could result in decreased productivity and collaboration. To solve the problems, intelligent meeting assistants are developed to automate taking notes, provide insight and assist in analyzing meetings after they occur. Researchers have found that incorporation of AI into conferencing applications could help increase user satisfaction, decrease their cognitive load and contribute to improved meeting results [7].

Notwithstanding the advancements in technology, most of currently used systems still treat transcription, summarization and knowledge extraction as separate modules. They usually fail to offer seamless processing of data in real time and context understanding as well as knowledge management. Thus, the gap between current state of affairs in development of technology and research needs to be addressed.

Introducing the MeetAI system is intended to eliminate this gap as the framework incorporates several modules into one. The MeetAI framework offers real-time transcriptions, context-aware summarizations, speaker recognition and action item extractions, which enables it to be used efficiently in virtual collaboration environments.

III. PROBLEM DEFINITION AND RESEARCH GAP

Even though there is an extensive adoption of various video conference technologies that allow people to communicate, collaborate and learn, it still remains a significant issue for them to properly capture and manage the information shared during meetings. Modern systems concentrate mainly on providing the ability to conduct real-time meetings through video streaming and audio calls. Meanwhile, the ability to process and understand the content of conversations is limited, which results in missing key information that is essential for the meeting, thus decreasing its efficiency [1], [7].

In order to enable users to analyze their meetings afterwards, existing platforms provide options to record sessions and create transcripts. The problem, however, is that the generated data does not have any context, it does not highlight any insights, and it requires extra effort from users to extract valuable pieces of information.

Speech recognition and natural language processing research has resulted in solutions for transcribing and summarizing spoken information. These solutions allow for automatic transcription of audio to text as well as the extraction of key data from conversations. Unfortunately, many of today's applications implement separate modules responsible for transcription, summarization, and knowledge extraction rather than having a coherent system with all features built-in. Moreover, issues like speaker identification, context consideration, language support, and real-time performance have been only partially solved in the current products [2], [3].

The emergence of AI technologies offers ways to improve meeting analysis with context-based summarization, topic identification, and action items. The use of machine learning algorithms allows analyzing meeting data and generating structured results that could be easily accessed and used. Finally, storing meetings in the cloud provides easy access and sharing of recordings. Unfortunately, currently, there aren't any meeting recording solutions with these advanced capabilities implemented in a single product [5], [6].

From these findings, therefore, the core challenge in current video conferencing is the lack of an effective platform where real-time transcription, context awareness, automatic summarization, and insight generation can all be accomplished at once in an intelligent manner. The current fragmented platforms are not able to provide effective ways of managing information and making decisions during the meeting.

This research will focus on addressing this challenge through the use of the MeetAI framework that encompasses real-time speech recognition, AI-based summarization, speaker identification, and action items extraction. Through this integration of different components, this technology will be able to convert raw meeting data into useful insights and structured information to help make effective decisions.

IV. PROPOSED ECOTRACE FRAMEWORK

In this section, a solution for addressing the problem is suggested through proposing the framework of MeetAI, which is an innovative framework for intelligent video conferencing that will integrate multiple modules related to real-time communication, speech processing, natural language processing, knowledge extraction, and user insights. Speech recognition and summarization were proven to be beneficial in enhancing meetings' productivity [1], [2]. Intelligent collaboration also required real-time assistance and structuring the information during meetings [3], [4]. MeetAI is a modular framework, meaning that it consists of separate but interrelated components. Modularity makes the system scalable, flexible, and easy-to-maintain. Therefore, upgrading or adding any feature will not require a complete redesign of the entire system.

A. Communication Layer

Communication layer is one of the most important elements of the MeetAI framework. It ensures the real-time communication of participants via video, voice, or screen-sharing. All the main communication functions such as streaming, transmitting, screen sharing, and managing are included in the communication layer. Through the division of communication functions into a special layer, the efficiency of the other modules is increased.

B. Speech Processing Module

The speech processing module processes oral input and converts it to text through automatic speech recognition (ASR). The module works in real-time to make it possible to convert oral communication in meetings to text. Through advanced speech modeling, this module can reduce noise interference, differentiate among speakers, and perform multilingual transcription.

C. Natural Language Processing Module

The natural language processing (NLP) module analyzes the processed text and extracts contextual and relevant information. Some of the activities undertaken by the NLP module include sentence structuring, extracting keywords, topics, semantic analysis, and many more. The main purpose of this module is to convert the unstructured transcript into structured text form.

D. Knowledge Extraction and Summarization Module

The knowledge extraction and summarization module is designed to generate useful insights from the processed data. With the use of artificial intelligence (AI), this module can generate useful insights including summaries, discussion points, decisions, and action items from the processed data. Context-aware summarization helps to retain useful information and eliminate repetition. Explainable AI will help users understand the process behind the summaries and insights generation.

E. Storage and Integration Module

The MeetAI framework also uses a cloud storage module that guarantees the security of recorded meeting data. This includes not only audio files but also meeting transcripts and generated summaries. It will facilitate efficient data access, distribution, and collaboration among multiple users. Also, it helps integrate other applications to enhance the effectiveness of workflows.

F. User Interaction Interface

User interaction module represents a means of interaction for meeting participants. When meetings occur, users are able to see live subtitles and insights about what is going on at the moment. At the same time, after a conference, the interface displays structured summaries and extracted action items that make information retention easier for participants.

G. Framework Integration

Thus, the MeetAI framework combines all above-mentioned elements into one coherent system. It allows participants to communicate with each other and process their voice into written language. Afterward, NLP module analyzes the content and extracts relevant knowledge. Data becomes available from storage and then transferred to users using the user interaction interface.

This framework helps build a flexible and extensible system that uses artificial intelligence to assist participants during virtual meetings. It processes raw meeting data and generates actionable insights to improve productivity.

V. SUSTAINABILITY EVALUATION METHODOLOGY

The proposed meeting analysis methodology is the systematic approach of analyzing video conference sessions, using MeetAI. As opposed to the traditional approaches based on recordings and transcription, the analysis methodology presented here incorporates real-time speech recognition, context-based analysis, decision making process, and AI-aided insights generation. All the above-mentioned elements are meant to improve meeting productivity in terms of scaling, precision, and understanding.

A. Definition of Analysis Units

In MeetAI, the analysis units are defined by the meeting session itself. Each session involves several participants with ongoing interactions and changing conversations. Within a single session, additional levels of segmentation can be used to analyze the information provided by speakers in separate speaker turns, sentences, or topic-specific units.

The structure of the analysis process allows collecting not only the data but also analyzing the context of conversations that occur during the session.

B. Data Collection and Change Detection

The process of analyzing starts with recording the audiovisual content provided by each participant within the course of the meeting session. The speech recognition engine transforms the speech into text while other data is gathered.

C. In case of a change in conversation – like topic changes, changes in participants in conversation, or other significant points in the conversation, then there will be a change detection mechanism that helps in identifying those changes. In case of no significant changes, for example, there is no new topic under discussion and the conversation continues with the previous topic, then there will be an incremental updating of the current context. In case of significant change, for example, new discussion topics or decision points come up, then it would trigger deeper analysis and segmentation.

D. AI Assisted Meeting Analysis

This helps in eliminating the requirement for note taking manually and ensures that all the significant outcomes from the meeting are recorded systematically. Follow-up becomes easier because of this process.

E. Hybrid Validation and User Feedback

Considering that automatic analysis by the AI does not produce absolutely flawless outputs, a hybrid validation scheme is implemented in MeetAI. It allows users to check, edit, and validate output summaries and task items after the conclusion of the meeting.

A combination of these two processes makes the analysis even more accurate and reliable. With time, the feedback received could be utilized to improve the models used by the system.

F. Methodological Properties

Several methodological properties associated with the meeting analysis methodology include:

Scalability: Process real-time with deep analysis being selectively performed to reduce computational cost.

Contextual understanding: Effectively performs segment-based analysis capturing contextual information.

Explainability: AI generated summaries and insights explain themselves.

Actionable: Actionable decisions and tasks are extracted which facilitate improved outcomes.

Integration: This methodology seamlessly integrates into other modules in the MeetAI framework including communication, storage, and user interfaces.

Using these components together, the MeetAI system converts informal conversations in meetings into organized, useful, and actionable information.

In case of significant segments in the meetings, the AI-assisted models help in conducting an analysis of the conversation. Tasks performed include extraction of topics, keywords, sentiment analysis, context-aware summarization and more.

It produces structured outputs which can be used effectively, for example, meeting summary, key points discussed, and insights from the meetings. The system will use explainable AI methods to produce insights that are transparent to the user.

D. Action Item and Decision Extraction

The process involves the extraction of actionable items from the conversation.

VI. PROTOTYPE IMPLEMENTATION AND QUALITATIVE EVALUATION

To prove the feasibility of the designed framework, a prototype of the system that showcases the integration of real-time communication, speech processing, NLU, and intelligent meeting analysis was developed. The goal of the prototyping process is to examine the validity of the design of the system and demonstrate the interaction between various modules.

The prototype system consists of three layers that include a frontend interface of the web application, the backend service layer, and the database layer. While the web app provides a user-friendly interface and basic functions such as video conferences, live captions, and meeting analysis insights, the backend takes care of all data processing tasks. Specifically, it includes modules that perform ASR, TPU, document summarization, and action items extraction.

In addition, communication and speech processing modules were incorporated into the design to mimic real-time communication and provide input data for further analysis. Specifically, live audio streams were captured from participants of the meeting and used to create transcriptions using automatic speech recognition technology. In addition, the module maintains timestamps and labels each speaker's utterances.

The natural language processing and knowledge extraction components have adopted the methodology that was mentioned earlier in Section V. As soon as meetings are taking place, the application performs the analysis of the corresponding transcripts, identifying the topics discussed, extracting main points from them, and creating summaries accordingly. Decision-based processing is used in the implementation; the software creates incremental summaries during continuous discussions in the same context and initiates more in-depth analysis in case of drastic topic changes and decision moments.

Furthermore, the developed prototype includes the intelligent meeting assistant component, which allows generating the real-time captions as well as the summary report after the meeting. Users will be able to retrieve structured results in the form of the summary report, highlighting the key points of discussions and extracted action items. Moreover, it is possible to integrate the tool with other applications such as calendars and task managers to manage action items as tasks.

Since the aim of the implementation is to test the feasibility of the system architecture, not its scalability, it is reasonable to evaluate the prototype according to the following qualitative criteria:

Modularity: Capability of the framework to isolate communication, processing, and analysis into separate components.

Usability: Ability of the user interface to convey insights from both real-time and post-meeting analysis.

Analysis Accuracy: Quality and relevancy of generated summaries, main points, and action items.

Efficiency: Capability of the decision-making processor in limiting redundancy and repetition of computations.

Extensibility: Capability to provide support for further extensions such as multilingual support, advanced analysis, etc.

It appears that qualitative assessment shows that the MeetAI framework is capable of combining real-time transcription, intelligent analysis, and user interaction into a single unit that is efficient, modular, and flexible. In particular, it is possible to use modularity to make individual components independently extensible while preserving inter-module communication. Moreover, intelligent assistant user interface shows how insights can be effectively communicated to users.

It should be noted that while the created prototype proves the concept, it requires additional modifications and enhancements in order to be used on a wider scale. Some directions for future improvements may include increased model accuracy, handling many meetings simultaneously, improved multilingual capabilities,

and user studies focused on increased productivity and efficiency of the system.

VII. DISCUSSION

Qualitative evaluation of the MeetAI prototype proves that implementation of real-time communication, speech processing, and AI-driven analysis is indeed feasible. The modular organization of the system makes it possible to use the same technology in different combinations to satisfy the evolving requirements of intelligent collaboration and virtual meeting tools.

First and foremost, the modularity of the proposed framework should be considered among its advantages. The separation of communication, speech processing, NLU, knowledge extraction, storage, and user interaction into separate modules is extremely efficient and ensures that the system is able to evolve with time due to the constantly emerging innovations. Many experts agree that modular design enhances both maintainability and extensibility of real-time communication platforms and intelligent applications [3], [4].

Finally, one must not forget about the context-awareness of the proposed meeting analysis algorithm. Traditional video conferencing software considers meetings merely as recordings and transcripts without recognizing their inherent dynamism and contextual nature. Meeting analysis by topic, speaker, and insight segmentation is much more effective in this regard.

The decision-based processing technique further improves system efficiency by minimizing unnecessary reprocessing. Rather than continually analyzing all the meeting information, the system detects any notable change like the shift in topic or any decision point and analyzes those specific changes thoroughly. Thus, the process becomes computationally efficient without compromising the quality of output, making it suitable for real-time implementations.

Moreover, the framework focuses on the interpretability of the AI-powered analysis results. By adopting interpretable AI techniques, MeetAI guarantees that the outputs generated are comprehensible and easily understandable to the user. This aspect is crucial in the professional and organizational settings, where users depend on AI-assisted analyses for their decision-making processes [5], [6].

Finally, the provision of both real-time and post-meeting assistance makes the framework user-friendly. The live captioning feature helps participants during the meeting, whereas the summary generation and action item extraction assist in the post-meeting phase.

However, despite these advantages, there are some important limitations associated with the present study. First of all, in order to prove the possibility of the proposed architecture implementation, an emphasis was put on showing its architectural features. Thus, it is still unclear whether the implemented prototype could demonstrate good performance in practice or under large-scale use conditions. Secondly, the results presented by the system strongly depend on the quality of speech recognition and natural language processing algorithms.

A significant drawback of the developed solution relates to its dependence on the automatic conversational data analysis, where subtle nuances of human interaction could get lost, as well as users' accents and emotions. In spite of the fact that the system has feedback tools for enhancing the quality of analysis, reaching the necessary level of accuracy remains difficult.

For future researches, it is necessary to pay attention to large-scale studies and improving the speech recognition and natural language processing models. In addition, the issue of the multilinguality should be considered. Finally, user-centered studies will provide useful information about participants' interaction with AI-generated meeting summaries and action items.

VIII. CONCLUSION AND FUTURE WORK

This paper introduced MeetAI, which is a modular framework that can be used to improve video conferencing solutions by leveraging the use of real-time communication, meeting analysis using artificial intelligence (AI), and smart information management. Through a combination of speech recognition, natural language processing, and knowledge extraction, the proposed framework can overcome some of the inherent limitations of existing meeting analysis solutions by offering valuable insight into meetings.

Existing video conferencing applications concentrate on facilitating communication and simple recording functions without adding value in terms of information management. The proposed framework uses context-aware meeting analysis that captures the conversational flow during meetings, recognizes key discussion areas, and creates structured and meaningful meeting summaries. This process makes it easy for meeting attendees to understand the information being provided since it is well-structured and organized.

In addition to meeting analysis, the proposed framework utilizes a decision-driven processing process that can make meeting analysis processes more efficient since computing operations would only be executed when a change in context occurs, such as when there is a shift in topics discussed or when decisions are made. With decision-driven processing, computation costs are minimized without reducing the quality of the output obtained.

Furthermore, the incorporation of real-time transcription, live captioning, and post-meeting summary functionality adds continuous assistance throughout the entire meeting process. Insights will therefore be available not only during meetings but also afterwards, making meetings more productive, accountable, and collaborative. Moreover, the system's modularity also ensures that additional functionality like multilingual capabilities, analytics, and productivity tool integrations can be added easily in the future.

To demonstrate the viability of the presented architecture, a proof-of-concept implementation was developed to evaluate MeetAI's feasibility. Qualitative evaluation shows that this system is indeed successful in integrating the three key components into one complete system.

Despite the high level of feasibility of the suggested model, there are still some directions that can be explored further in the future research. Among the issues to consider include large-scale assessment of its performance in a real-life setting, increasing accuracy of the speech recognition algorithms for a range of languages and accents, and creation of more sophisticated AI models that would take into account context of usage.

Thus, in conclusion, MeetAI suggests a scalable and versatile framework for intelligent video-conferencing systems, incorporating the ideas of combining communication, analysis based on artificial intelligence algorithms, and user-centered interface design.

REFERENCES

[1] [1] D. Jurafsky and J. H. Martin, *Speech and Language Processing*, 3rd ed., Pearson, 2023.

[2] [2] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 6645–6649.

[3] [3] G. Penn and X. Zhu, "A critical reassessment of evaluation baselines for speech summarization," in *Proc. Annual Meeting of the Association for Computational Linguistics (ACL)*, 2018.

[4] [4] Q. McNemar et al., "Automatic meeting summarization: A survey," *ACM Computing Surveys*, vol. 52, no. 5, 2019.

[5] [5] D. V. Minh et al., "Explainable artificial intelligence: A comprehensive review," *Artificial Intelligence Review*, 2022.

[6] [6] S. Ghaffarian et al., "Explainable artificial intelligence in decision support systems," *International Journal of Information Management*, 2023.

[7] [7] C. Gutwin, S. Greenberg, and M. Roseman, "Workspace awareness in real-time distributed groupware," *Human–Computer Interaction*, vol. 11, no. 3, pp. 411–446, 1996.

[8] [8] Microsoft, "The future of work: Good habits, productivity, and collaboration in virtual meetings," Microsoft Work Trend Index Report, 2022.