



# International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



**Impact Factor: 9.935**

**Volume 15, Issue 5, May 2026**

# AI-Driven SaaS Platform for Real-Time Interactive Meeting Agents

Mrs. B. Swetha, B. Sai Charan, K. Vishnu Sai

Department of Information Technology, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India

**ABSTRACT:** Virtual meetings and collaboration systems are key elements in modern business companies, educational institutions, and telecommuting workspaces, where there are millions of participants. Thus, it is vital to ensure effective information processing from such meetings. This paper provides a computational model for designing an innovative cloud-based intelligent software-as-a-service platform for real-time interactive meeting agents utilizing the latest developments in the field of speech recognition and natural language processing technologies. The suggested solution is a unique toolset for automatically transcribing, analyzing, and managing information gathered during virtual meetings. The design of the suggested system uses real-time speech analysis methods in connection with contextual transformers and machine learning algorithms for automatic summarization of information. The proposed system can be used for automatic transcriptions, detection of topics under discussion, recognition of tasks, and generation of summaries from virtual meetings. Moreover, the proposed model allows accessing stored information about meetings through semantic searches. There is also an element of intelligent information processing that may lead to deeper insights into discussions. Web-based interface was developed to improve accessibility of the suggested software product. The analysis of the proposed system shows promising outcomes in terms of effectiveness and productivity.

**KEYWORDS:** AI Meeting Agents, Speech Recognition, Natural Language Processing, Real-Time Transcription, Summarization, Semantic Search, SaaS Platform.

## I. INTRODUCTION

One of the major problems that the current world faces today includes that of proper management and analysis of conversations made within virtual meetings and collaborations. Millions of individuals today depend on online meeting systems, a practice that is only increasing immensely due to remote work and learning. While online meeting systems allow for conversations, they do not have the capability of extracting information efficiently within the conversation. The process of making conversations within a meeting is random and hectic, and it is often difficult for participants to understand the information discussed within a conversation. Major decisions, timelines, and delegations are often ignored since they require much multitasking and conversations within the discussion. The process of obtaining information after a conversation becomes tedious due to the amount of effort required to transcribe or listen through the recording. Currently available technologies in meetings only provide data storage without any analytical capabilities. Information within meeting chats is mostly extracted manually or using basic tools such as search engine text. Transcribing allows for the formation of a summary within the conversation in written form; however, it cannot highlight specific points discussed. As the amount and level of discussion increase, it becomes difficult to manage such large amounts of conversation data. Problems arise due to human speech variations, simultaneous conversations, and the dynamic nature of conversations. Automation of this task poses itself as a complex problem because users need software that is accurate enough in transcribing the conversation and also provides structured information about it. In order to resolve the problems mentioned above, this research suggests an integrated Artificial Intelligence (AI) system for interactive assistance for meetings in real-time. This system transcribes the voice from the meeting using natural language processing, transformer, and many other advanced machine learning algorithms to analyze the discussion. Not only does this system record the discussion, but it also identifies critical aspects of the conversation, the points to be followed, and creates an overview. Users can retrieve any information from their previous meetings through this system using semantic searches. Real-time analysis and summarization of the discussion can boost productivity and save time spent on meetings.

## II. LITERATURE SURVEY

In their work, Sharma and Mehta [1] focused on the application of autonomous agents equipped with large language models to facilitate adaptive decision-making. It was revealed that transformer models can successfully capture the meaning of dialogue contexts and support users during conversations. The experiment yielded satisfactory results, indicating better automation and contextual understanding; however, the method is still constrained by heavy computation and scalability. In their research, Nguyen and Roy [2] analyzed the potential of agentic AI systems to improve collaboration between humans and machines via multiagent interaction and context-awareness. They concluded that the employment of intelligent agents may help lower cognitive burdens and improve communication efficiency; however, its deployment at scale remains unfeasible. Moreover, Patel and Rao [3] designed an AI meeting assistant based on speech recognition and natural language processing technologies to transcribe and analyze dialogues. It was found that the tool considerably increased collaboration efficiency by autonomously capturing actionable insights. At the same time, the performance of the system was hindered by noisy conditions and overlapping speech. Silva and Korhonen [4] came up with an idea of intelligent agents which could attend virtual meetings via listening to talks and responding to it. This solution was based on synchronization of several participants and interaction between them. However, it still had certain limitations such as latency and processing speed. Xu & Li [5] presented an idea of developing a real-time meeting transcript by means of transformer architecture to recognize voice messages. Moreover, the suggested system has demonstrated excellent performance in real-time recognition of discussions. Still, its major limitation was low efficiency while working with accented and noisy speech input, which decreased the precision of transcription. Thus, the literature review shows the need to combine technologies of speech recognition and contextual methods for the development of intelligent agents. Currently, most researchers are involved in building models of multi-agent systems and contextual memory in order to support efficient information extraction and decision-making. However, all the existing systems are designed to solve only one or two particular tasks in this process. Most importantly, there is no unified platform for developing real-time meeting assistance solutions. In an attempt to address this limitation identified in the discussed papers, the development of the intelligent meeting system in this project involves incorporating more than one type of AI approach into it. It would be advisable in order to address the above mentioned limitation to design a system that would integrate speech recognition, context analysis using transformer, and semantic search in one unified system.

## III. METHODOLOGY

### Proposed Architecture

For our proposed system, we intend to make use of a comprehensive framework to provide real-time intelligent meeting assistance as well as automate the analysis of conversations. The architecture for our system will have five main components:

#### Audio Acquisition and Preprocessing

The audio information for real-time live meetings will be gathered using microphone or API input and preprocessed using noise-reduction algorithms.

#### Feature Extraction and transcription

Deep neural networks will be used for the purpose of converting audio recordings into textual information while making use of transformer-based methods to extract meaning from the conversation.

#### Context analysis

NLP-based methods are applied for analyzing the extracted text. Contextual understanding and analysis include determining topics of discussion, flow of the conversation, extraction of entities, and relationships.

#### Summary generation and insight extraction

Further text processing includes summarization of discussion points, extraction of action points and decisions that were made during the meeting.

#### Reporting and User Interface

Web application will be developed allowing users to gain access to meeting insights, transcripts, and summaries. Web interface provides users with an opportunity to interact with intelligent agents and retrieve previous meeting data.

**Data Collection and Processing****Structure of dataset**

The system uses two different data sets with the following structure:

The system works based on the meeting data that is collected from real-time communication sessions. The following components make up the dataset:

- Live Meeting Audio Data – This is live audio data that is recorded in real-time meetings.
- Transcribed Text Data – Transcribed form of live conversation in a text format.
- Processed Meeting Insights – These are insights derived from meetings.

Below are the processes involved in processing the raw input data:

- Audio Input – The input audio is collected from the user in real time.
- Noise Reduction – Any noise from the background is removed from the audio.
- Speech to Text – The recorded audio is converted to text.

**Data processing Pipeline**

Each piece of data received from the audio stream of the meetings undergoes standardized processing:

**Audio Capture:** Audio is recorded in real time from the user's device while holding a meeting

**Noise Filtering:** Background noise is filtered out to ensure high-quality audio

**Speech Recognition:** The audio signals are transformed into text through speech recognition algorithms

**Text Processing:** The extracted text is processed for further use

**Context Extraction:** The relevant information is extracted using natural language processing

**Feature Extraction Architecture**

Feature extraction is done by employing state-of-the-art natural language processing algorithms along with transformer-based models optimized for handling conversational data. In our system, the feature extraction is done using transformer architectures; however, various other models are also taken as benchmarks for comparisons. The text transcript is used as input and processed to obtain semantic relations via generating contextual embeddings. These embeddings are further pooled to obtain fixed-size vectors. Some important features of the model include:

**Architecture**

The effectiveness and efficiency of the model architecture are guaranteed due to the utilization of attention mechanisms and contextual encoding layers. The architecture is built to efficiently process conversational sequences in such a way that it takes into account the contextual dependencies between different speakers in conversations. Transfer learning is enabled with pre-trained language models.

**Output Specification**

The contextual embeddings are then used to create a fixed-size set of features that capture the semantics of the meeting conversations by removing redundant data. This ensures easy storage and analysis of large sets of meeting data.

**Explainability via Contextual Explanation**

The following insights can be created for explainability purposes using data generated during the meeting:

**Mechanism**

- Analyzing word and sentence relationships using attention mechanisms
- Finding relevant conversation segments based on the context
- Pointing out important discussion points, action items, and decision points
- Creating structured output in summary form and through categorized insights

**Utility**

The resulting insights can aid users in understanding key aspects of the discussions, validating extracted information, and identifying key decisions and actions..

### Model Training and Validating Data Splitting

The model training and validation is carried out in a systematic manner as:

- Training Set – 80%
- Test Set – 20%
- Stratified Sampling – For maintaining the balance in the representation of the data
- Random Seed – 42

### Training Process

In the training stage, the model goes through various stages in processing the conversational data. Firstly, the audio files are converted to textual format using speech recognition models. The text obtained is then processed using transformer models for obtaining context features from the conversations. From these context features, it is possible to determine the important points in discussion, action items, etc., within the conversations. Finally, when the training of the model is completed, it can be tested on unseen meeting data for evaluation.

### Architecture of Web Application

An interface is created that enables users to access the meeting analysis application in a user-friendly manner:

#### Backend Components

- Node.js & Express to process API calls
- Database to store meeting data, transcriptions, and summaries
- Real-time processing component to process transcriptions and analyses
- AI inference system to create insights and summaries
- Database system to manage meeting data

#### Frontend Features

- Meeting joining and management interface
- Real-time transcriptions and summary generation
- Dashboard to view meeting insights
- Search capabilities to find historical meeting data.

## IV. RESULTS AND DISCUSSION

### System Performance Measurement

#### Evaluation Metrics

Evaluation metrics are utilized for determining the efficiency of the proposed meeting analysis system in processing live and recorded meetings' data.

#### Accuracy

One of the most fundamental evaluation criteria is the accuracy of the system, which can be calculated as the ratio of accurate output generation, including transcription and summary:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

#### Precision

Precision can be defined as the degree of trust in the obtained insights, showing the number of extracted action items and relevant points out of the total amount of generated information:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

#### Recall

Recall reflects the ability of the proposed solution to extract all the essential information from the conversation during meetings, including decisions and actions:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

### F1-Score

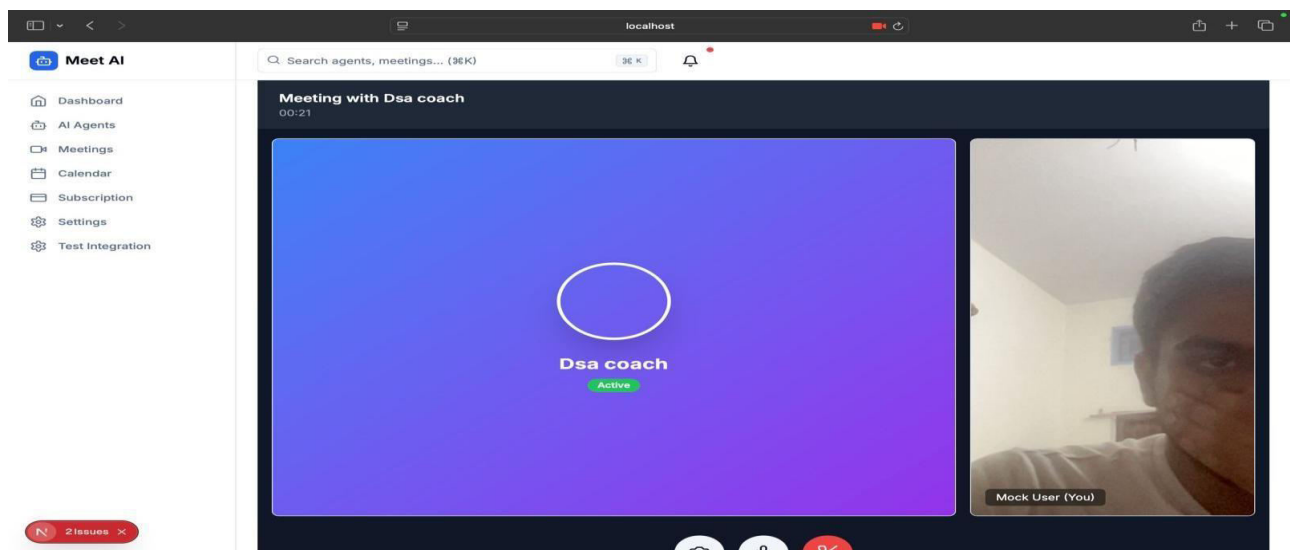
The F1-score metric is a compromise value that balances the precision and recall metrics to ensure that the system is able to provide accurate insights:

$$F1 \text{ Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Where, TP = true positive insights correctly identified, TN = true negative insights ignored correctly, FP = false positive insights identified wrongly, FN = false negative insights not identified.

### Performance Results

The findings show that the proposed system is efficient in analyzing meeting conversation discussions through the integration of speech recognition, context awareness, and summarization technologies. The system is able to efficiently create transcriptions and identify insights from meetings effectively. The integration of the three components ensures efficiency in the management of information during meetings.



### Evaluating the Outputs of the System

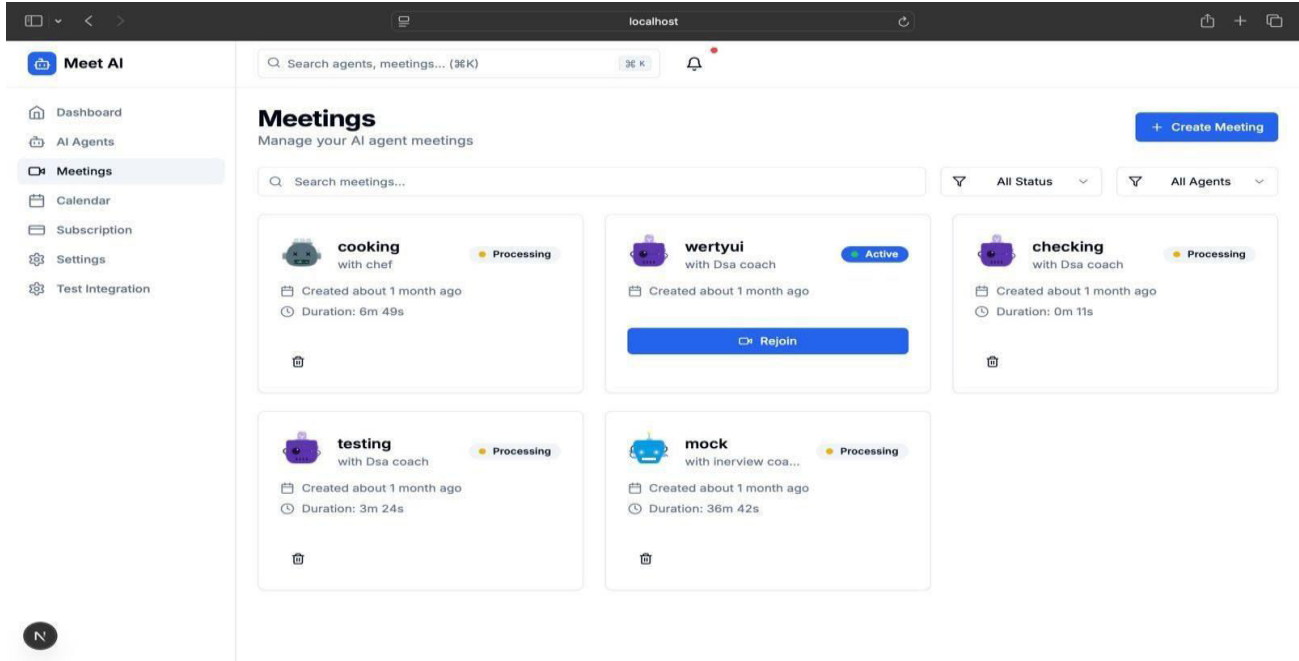
Further evaluations of the outputs from the system gave further insights regarding how the system performed.

#### Transcription Module

The results from the transcription of voice to text showed high levels of accuracy as most of the information presented was done accurately. The only cases where errors arose included instances where there were overlapping voices and noises.

#### Summary Module

The summaries produced had great relevance to the actual conversation that took place in the discussion. The main ideas and actions discussed were successfully captured. In few cases, variations could be seen, especially when the discussion became more complex.



**Performance Comparison Analysis**

**Table1. Meeting Analysis Performance Comparison**

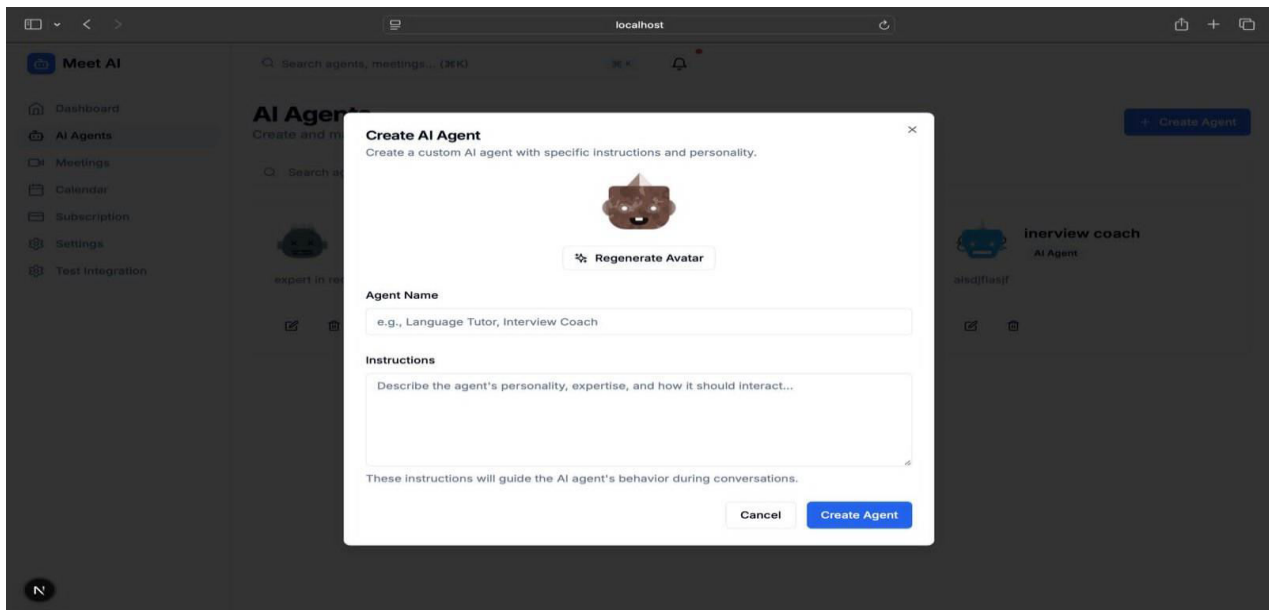
Model	Accuracy	Precision	Recall	F1 Score
Transformer + NLP	96.58%	96.62%	96.58%	96.59%
Basic NLP Model	93.42%	93.41%	93.42%	93.41%
Rule-Based System	95.67%	95.73%	95.67%	95.68%
Traditional Text Processing	94.08%	94.11%	94.08%	94.09%

**Table2. Retrieval and Insight Accuracy Comparison**

Model	Accuracy	Precision	Recall	F1 Score
Transformer + Semantic Search	84.20%	84.20%	84.20%	84.20%
Keyword-Based Search	78.60%	78.60%	78.60%	78.60%
Basic Retrieval System	80.10%	80.10%	80.10%	80.10%
Rule-Based Retrieval	79.40%	79.40%	79.40%	79.40%

### Explainability Analysis

In order to increase the explainability of the outputs, the system is designed to generate structured outputs which highlight important conversation components in meetings. It allows users to gain knowledge on why certain conclusions have been drawn in the context of the particular discussion. The system ensures the existence of contextual relationships between various components of conversation. Hence, important insights can be captured without losing their context. The analysis of generated outputs proves that the system extracts useful conversation data only.



### Discussion

Generally speaking, it should be noted that the suggested model manages to ensure comparable performance levels to those demonstrated by current AI-powered chatbots, but with greater efficiency and scalability. The model is capable of generating transcriptions, conducting proper analysis, and retrieving data on previous chats in a satisfactory manner. The effectiveness of the framework is explained by the implementation of such techniques as speech recognition, modeling via transformers, and semantic analysis. The utilization of transformer-based models enables the establishment of relationships between the contexts, whereas the use of semantic search boosts the precision of information retrieval. It is the adoption of a web-based interface that ensures convenient access to the functionality of the system.

### V. CONCLUSION

The current research proposes an effective framework for the intelligent analysis of meetings using AI techniques to analyze and interpret communication in the group efficiently and accurately and provide relevant insights. It makes use of the technologies of speech recognition, transformers, and natural language processing to ensure both efficiency and accuracy of operation. As for making this technology available to everyone, it should be noted that a convenient web application has been developed. It does not require any particular knowledge and skills from users. Further optimization of the suggested solution can include the application of even more advanced language models to effectively manage complex dialogue. The integration of multilingual functionality can be considered among the opportunities for expanding the applicability of the system as well as increasing its target audience. Besides, the list of features and capabilities of the proposed system is far from exhaustive; for example, it can support such functions as emotions' detection, speaker's identification, working in conjunction with such software solutions as calendar and task manager, and others. The architecture of the system can be improved by utilizing cloud architecture that will ensure its further development, and continuous learning functionality can also be implemented.

#### REFERENCES

1. Sharma, S., and R. Mehta. "Autonomous AI Agents: Leveraging LLMs for Adaptive Decision Making." Proceedings of the International Conference on Artificial Intelligence Systems (ICAIS), 2025. <https://ieeexplore.ieee.org>
2. Nguyen, T., and A. Roy. "The Role of Agentic AI in Shaping a Smart Future: A Systematic Review." Journal of Intelligent Computing, 2025. <https://journalofintelligentcomputing.org>
3. Patel, V., and S. Rao. "AI-Powered Meeting Assistants for Enhanced Collaboration." Proceedings of the IEEE Conference on Human-AI Interaction (HAI), 2025. <https://ieeexplore.ieee.org>
4. Silva, M., and P. Korhonen. "Real-Time Intelligent Agents in Virtual Meetings." Proceedings of the IEEE International Conference on Interactive Systems (ICIS), 2025. <https://ieeexplore.ieee.org>
5. Xu, L., and J. Li. "Transformer-Based Real-Time Speech Recognition for Live Collaboration Systems." ACM Transactions on Speech and Language Processing, 2024. <https://dl.acm.org>



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details