

Smart Meeting Assistant and Performance Tracker for Online Meetings

Aniket Kasav¹, Bhumi Nikam², Vaishnavi Patil³, Prof. V. P. Sahane⁴,
^{1,2,3}Department of Artificial Intelligence and Data Science Engineering,
SNJB's Late Sau. K. B. Jain College of Engineering,
Chandwad, Nashik, Maharashtra, India

Abstract—The widespread use of virtual meeting platforms in professional, academic, and organizational settings has reshaped modern communication and collaboration. Although tools such as Zoom, Google Meet, and Microsoft Teams offer reliable audio–video connectivity, they provide limited support for intelligent analysis of meeting content. Consequently, participants often struggle with manual note-taking, missed discussion points, unclear accountability, and subjective evaluation of participation.

This review paper examines the evolution of Smart Meeting Assistant and Performance Tracking systems that employ Artificial Intelligence (AI), Natural Language Processing (NLP), and Machine Learning (ML) to improve the effectiveness of online meetings. The study reviews existing research on speech-to-text transcription, dialogue analysis, extractive and abstractive summarization techniques, participant performance measurement, and voice-based interaction. It also identifies key research gaps and highlights the limitations of current solutions that operate as isolated or feature-specific tools.

Furthermore, the paper presents the Smart Meeting Assistant and Performance Tracker as an integrated AI-driven framework that combines real-time transcription, automated meeting summarization, speaker-level performance analysis, and voice-command control within a unified platform. By reducing cognitive load and enabling structured, data-driven insights, the proposed approach enhances meeting transparency, engagement assessment, and decision-making. This review provides a foundation for the development of next-

generation intelligent meeting systems capable of transforming virtual collaboration into an organized, efficient, and insightful process.

Index Terms— Smart Meeting Assistant, Artificial Intelligence, Natural Language Processing, Speech Recognition, Meeting Summarization, Performance Tracking, Voice Commands.

I. INTRODUCTION

Online meeting platforms such as Zoom, Google Meet, and Microsoft Teams have become essential tools for communication in today's digital environment. They support remote work, virtual learning, collaborative projects, and organizational decision-making across various domains. Although these platforms provide stable audio–video communication, they offer limited support for intelligent processing of meeting content. As a result, participants often depend on manual note-taking, post-meeting reviews, and subjective methods of evaluating participation, which increases cognitive effort and reduces overall meeting efficiency.

Despite significant advancements in artificial intelligence and cloud-based technologies, the virtual meeting ecosystem continues to face several challenges. Common issues include unstructured discussions, absence of automated documentation, lack of engagement analytics, and limited accessibility for users who are unable to attend meetings in real time. These challenges become more pronounced in lengthy or multi-participant meetings, where identifying key decisions, action items, and individual contributions is particularly difficult.

To address these concerns, intelligent meeting assistance systems have been introduced that

incorporate Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Machine Learning (ML). These systems aim to automate meeting transcription, generate concise summaries, and analyze participant behavior in real time. However, many existing solutions function as standalone tools, focusing on individual features such as transcription or summarization, rather than providing a unified platform that integrates documentation, analytics, and interaction.

Recent developments in transformer-based language models and real-time processing frameworks have enabled more accurate and context-aware analysis of meeting conversations. In addition, voice-command interfaces support hands-free interaction, improving usability and accessibility during live meetings. When combined within a single platform, these technologies have the potential to significantly enhance meeting productivity, transparency, and collaboration quality.

To overcome the limitations of current approaches, this paper reviews existing research in intelligent meeting analytics and presents the Smart Meeting Assistant and Performance Tracker as a unified, AI-powered solution. The proposed system integrates real-time transcription, automated summarization, participant performance analysis, and voice-based control to support structured, efficient, and data-driven online meetings in both academic and professional settings.

II. LITERATURE SURVEY

Recent studies emphasize the increasing role of artificial intelligence and data-driven methods in improving the productivity and effectiveness of online meetings. Initial research in this domain primarily focused on physical meeting environments, where vision-based systems were used to monitor participant activities and interaction patterns. Although these approaches showed promising results in controlled conditions, they required specialized hardware and were not suitable for scalable, web-based meeting environments.

With advancements in speech recognition technology, research attention gradually shifted toward audio-based analysis of meetings. Deep learning-based Automatic Speech Recognition (ASR) models significantly enhanced transcription accuracy,

particularly in noisy and multi-speaker scenarios. Real-time speech-to-text conversion enabled by these models has become a foundational component for advanced meeting analysis tasks, including summarization and participant performance evaluation.

Natural Language Processing has played a critical role in extracting meaningful information from meeting transcripts. Earlier extractive summarization techniques relied on statistical features such as keyword frequency and sentence ranking. In contrast, modern transformer-based models support abstractive summarization, allowing summaries to better capture context, intent, and conversational flow. Studies have shown that dialogue-aware summarization methods further improve coherence and relevance in multi-participant discussions.

Multimodal approaches that combine audio, textual, and visual inputs have also been explored to enhance meeting understanding. By analyzing facial expressions, gaze direction, and speech characteristics, these systems aim to infer participant engagement and intent more accurately. However, the high computational requirements and dependence on high-quality video data limit their practical deployment in typical online meeting scenarios.

Voice-command interfaces have gained increasing attention as they enable hands-free interaction with meeting systems, improving usability and accessibility, particularly for presenters and multitasking users. Despite these technological advancements, most existing meeting assistant solutions focus on isolated functionalities such as transcription, summarization, or recording. The lack of integration between documentation, performance analytics, and interactive control remains a significant limitation.

Overall, the reviewed literature reveals a clear research gap in the development of comprehensive, AI-driven meeting platforms that seamlessly integrate transcription, summarization, participant analytics, and voice-based interaction. The proposed Smart Meeting Assistant and Performance Tracker addresses this gap by offering a unified framework for intelligent, structured, and data-driven management of online meetings.

Table I

Comparison of AI Models Used in Smart Meeting Assistant Systems

Model	Primary Use	Strengths	Limitations
Whisper	Speech-to-Text	Robust to noise, multilingual	High computational cost
BERT	Text Understanding	Strong contextual embeddings	Not generative
T5	Text Summarization	Flexible text-to-text design	Higher latency

Based on the reviewed literature, various artificial intelligence models have been adopted for speech recognition, language understanding, and meeting summarization. Table I presents a comparative overview of commonly used models in smart meeting assistant systems.

III. EXISTING SYSTEMS AND LIMITATIONS

Existing online meeting assistant solutions mainly provide individual features such as meeting recording, speech transcription, or basic summarization. While these tools support certain aspects of virtual meetings, they do not offer a fully integrated solution for intelligent meeting analysis and performance evaluation.

The major limitations observed in current systems are as follows:

- **Fragmented functionality:** Transcription, summarization, and analytical features are often implemented as separate tools, leading to disjointed workflows.
- **Limited real-time performance tracking:** Most systems do not provide detailed analytics on participant engagement or speaking behavior during live meetings.
- **Lack of structured and actionable summaries:** Generated summaries are often unorganized or incomplete, reducing their usefulness for follow-up actions and decision-making.
- **Minimal support for voice-based interaction:** Hands-free control and accessibility features are either limited or not available in many existing solutions.

- **Poor integration with evaluation and reporting tools:** Current platforms lack seamless mechanisms to generate consolidated performance reports from meeting data.

These limitations emphasize the need for a comprehensive, AI-driven meeting assistant that integrates transcription, summarization, participant analytics, and voice-based interaction within a single, unified platform.

Table II

Comparison of Existing Meeting Assistant Systems and the Proposed Framework

Feature	Existing Systems	Proposed System
Speech-to-text	Available	Available
Meeting summarization	Partial	Comprehensive
Performance tracking	Not available	Supported
Voice commands	Limited	Supported
System integration	Fragmented	Unified
Real-time analytics	Limited	Available

IV. PROPOSED SYSTEM OVERVIEW

The Smart Meeting Assistant and Performance Tracker is developed to automate essential aspects of online meetings by combining speech processing, language understanding, and performance analytics within a single intelligent framework. Unlike conventional meeting platforms that mainly support real-time communication, the proposed system focuses on meeting intelligence, enabling systematic documentation and objective assessment of participant engagement.

During an online meeting, the system continuously captures live audio streams and processes them using Automatic Speech Recognition (ASR) to produce accurate, time-stamped transcripts. These transcripts are then analyzed through Natural Language Processing (NLP) techniques to identify key discussion points, agenda items, and decisions. By automating transcription and content analysis, the

system reduces reliance on manual note-taking and helps prevent the loss of important information.

Beyond documentation, the system assesses participant behavior by monitoring metrics such as speaking duration, frequency of interaction, and periods of silence. These quantitative indicators provide clear and objective insights into individual contribution and engagement levels, making them valuable for academic evaluation, team assessment, and performance review. In addition, the integration of voice-command functionality allows users to control meeting operations—such as starting or stopping recordings and generating summaries—without interrupting the flow of discussion, thereby enhancing overall usability and efficiency.

V. SYSTEM ARCHITECTURE

The architecture of the Smart Meeting Assistant and Performance Tracker follows a modular and layered design to ensure scalability, maintainability, and efficient real-time processing.

Fig. 1. Architecture of the Smart Meeting Assistant and Performance Tracker

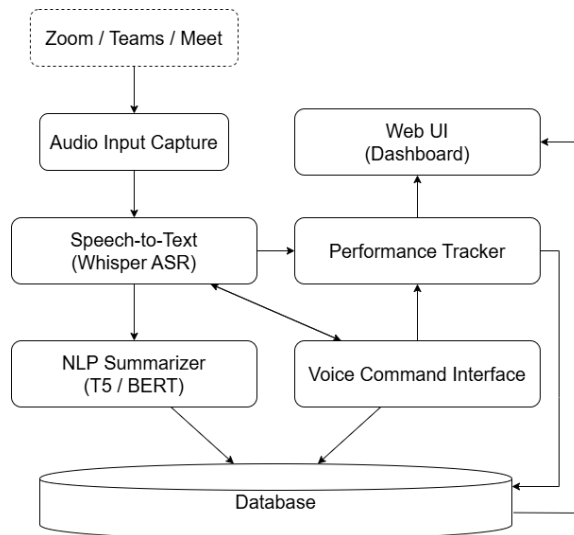


Figure 1 illustrates the layered architecture of the proposed smart meeting assistant system.

A. Frontend Layer

The frontend layer is developed using modern web technologies such as React.js or Next.js, providing an interactive and user-friendly dashboard. It displays real-time meeting transcripts, generated summaries, and participant performance analytics in a clear and

organized manner. The interface is designed to require minimal user interaction during meetings, allowing participants to remain focused on discussions while accessing essential information effortlessly.

B. Backend Layer

The backend layer manages core application logic, user authentication, and communication between different system modules. It exposes RESTful APIs and utilizes WebSocket connections to support real-time data transmission. This layer coordinates speech-to-text processing, NLP-based summarization, and performance metric computation while ensuring secure handling of user and meeting data.

C. AI Processing Layer

The AI processing layer forms the core intelligence of the system and integrates multiple machine learning components, including:

- Automatic Speech Recognition (ASR) models, such as Whisper, for accurate real-time transcription
- Transformer-based NLP models (e.g., BERT or T5) for automated meeting summarization
- Analytical modules for calculating participant engagement metrics and contribution statistics

The separation of AI components allows individual models to be upgraded or optimized without impacting the overall system architecture.

D. Data Layer

The data layer is responsible for storing user profiles, meeting metadata, transcripts, summaries, and performance records. It utilizes relational or NoSQL databases to ensure structured and efficient data storage. This design supports fast retrieval of meeting information for post-meeting analysis, reporting, and long-term record maintenance.

VI. ALGORITHMS AND TECHNOLOGIES USED

A. Speech-to-Text Transcription

Deep learning-based Automatic Speech Recognition (ASR) models are used to convert meeting audio into textual form with high accuracy, even in noisy and multi-speaker environments. The generated transcripts are time-stamped, enabling accurate alignment

between spoken content and subsequent analysis, such as summarization and performance evaluation.

B. Meeting Summarization

Meeting summaries are generated using transformer-based Natural Language Processing (NLP) models that analyze long conversational transcripts. These models identify important topics, decisions, and action items discussed during the meeting. A combination of extractive and abstractive summarization techniques is employed to enhance clarity, coherence, and readability of the generated summaries.

C. Performance Tracking

The system evaluates participant engagement by computing metrics such as total speaking duration, frequency of participation, and periods of silence. These quantitative indicators provide objective insights into individual contribution levels and overall meeting effectiveness, supporting fair assessment in academic and professional settings.

D. Voice Command Processing

Predefined voice commands enable users to control key meeting operations, including starting or stopping recordings and generating summaries. This hands-free interaction improves accessibility and usability, particularly during presentations or situations where manual interaction may interrupt the meeting flow.

Evaluation	Metrics:
The performance of smart meeting assistant systems is commonly assessed using standard evaluation metrics reported in the literature. Automatic Speech Recognition accuracy is measured using Word Error Rate (WER), which quantifies transcription errors. The quality of automated meeting summaries is evaluated using ROUGE-1 and ROUGE-L scores, which measure content overlap between generated and reference summaries. Participant performance analytics are analyzed using metrics such as speaking duration, turn frequency, and interaction count to assess engagement levels.	

VII. APPLICATION AREAS

Smart meeting assistants have a wide range of applications across various domains where effective communication and accurate documentation are essential. The key application areas of the proposed system are outlined below:

- **Education:** The system can be used to automatically transcribe online lectures and academic discussions, generate concise summaries, and assist educators in evaluating student participation and engagement.
- **Corporate Environments:** In professional settings, the system supports automated meeting minutes, performance evaluation of team members, and systematic tracking of decisions and action items.
- **Research Collaboration:** For research teams and project groups, the system enables structured documentation of discussions, ideas, and findings, ensuring transparency and continuity across multiple meetings.
- **Government and Administration:** The system can be applied to record policy meetings, committee discussions, and administrative sessions accurately, supporting accountability and reliable record-keeping.

VIII. FUTURE RESEARCH DIRECTIONS

Future enhancements of the proposed system may focus on expanding multilingual support to improve accessibility for users from diverse linguistic backgrounds. Additional research can explore the integration of emotion and sentiment analysis to better understand participant engagement and meeting dynamics. Improvements in speaker diarization techniques can further enhance accuracy in handling overlapping and multi-speaker conversations. The system may also be extended to support offline or low-connectivity environments and seamless integration with project management and productivity tools, enabling broader real-world applicability.

A. Data Privacy and Ethical Considerations

The proposed system emphasizes the responsible and ethical use of meeting data by ensuring informed user consent prior to audio recording and analysis. Meeting transcripts and performance analytics are securely stored using access-controlled databases to prevent unauthorized access. When required, sensitive information can be anonymized, and analytical insights are intended solely for constructive feedback and performance improvement. Addressing privacy, transparency, and ethical usage is essential for

building user trust and supporting the large-scale adoption of intelligent meeting assistant systems.

IX. CONCLUSION

The Smart Meeting Assistant and Performance Tracker illustrates how Artificial Intelligence can effectively enhance traditional online meetings by converting them into structured and data-driven interactions. Through the integration of real-time speech transcription, intelligent meeting summarization, participant performance analytics, and voice-based control, the system improves productivity, transparency, and accessibility. The proposed framework provides a strong foundation for the development of future intelligent collaboration platforms, supporting more efficient and insightful communication in both academic and professional environments.

ACKNOWLEDGMENT

The authors express their sincere gratitude to SNJB's Late Sau. K. B. Jain College of Engineering, Chandwad, for providing a supportive academic environment and the necessary infrastructure to carry out this research work. We are particularly grateful to our project guide, Prof. V. P. Sahane, for his continuous guidance, encouragement, and valuable technical insights throughout the preparation of this review paper.

We also acknowledge the support and constructive feedback provided by the faculty members of the Department of Artificial Intelligence and Data Science Engineering, which greatly contributed to the refinement of this work. Lastly, we extend our appreciation to our peers and all individuals who contributed directly or indirectly to the successful completion of this study.

REFERENCES

- [1] T. Hain, L. Burget, J. Dines, et al., "Transcribing Meetings with the AMIDA Systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 486–498, Mar. 2010.
- [2] E. Voss, P. Ehlen, and D. McAllester, "The CALO meeting assistant," in *Proc. Int. Conf. Intelligent User Interfaces (IUI)*, 2007, pp. 19–28.
- [3] A. Rennard, R. Bhandari, and P. Sarode, "Abstractive meeting summarization: A survey," *IEEE Access*, vol. 11, pp. 11234–11248, 2023.
- [4] A. Hannun et al., "Deep speech: End-to-end speech recognition," *arXiv preprint arXiv:1412.5567*, 2014.
- [5] S. Sarode, V. Patil, and R. Bhandari, "Attendie AI: Smart and proactive meeting solution," in *Proc. IEEE Int. Conf. Advanced Computing and Communication Systems (ICACCS)*, 2025.
- [6] C. Goo and Y. Chen, "Abstractive Dialogue Summarization with Sentence-Gated Modeling Optimized by Dialogue Acts," in *Proc. IEEE Spoken Language Technology Workshop (SLT)*, 2018, pp. 435–442.
- [7] H. Li, J. Zhu, C. Ma, J. Guo, and X. Cheng, "Multimodal Summarization for Asynchronous Collection of Text, Image, Audio and Video," in *Proc. 2017 ACM on Multimedia Conference (MM '17)*, pp. 1092–1100, 2017.
- [8] J. Renals, T. Hain, and H. Bourlard, "Multimodal approaches to meeting content extraction," *AMI/AMIDA Project Reports, Tech. Rep.*, 2007.
- [9] D. Burnett, A. van der Zee, and J. Griffiths, "WebRTC: Real-time communication framework," *IEEE Internet Computing*, vol. 25, no. 3, pp. 62–69, May–Jun. 2021.
- [10] R. Bhandari and S. Sarode, "Evaluation metrics for meeting summarization," *IEEE Access*, vol. 8, pp. 99432–99440, 2020.
- [11] X. Li, J. Feng, Y. Cheng, B. Zhu, and Y. Yan, "Topic Segmentation and Clustering for Deep Summaries," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 10, pp. 3345–3358, Oct. 2021.
- [12] N. Charif and S. J. McKenna, "Tracking the activity of participants in a meeting," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2006, pp. 1–8.
- [13] J. Allen, *Natural Language Understanding for Spoken Dialogue Systems*, MIT Press, Cambridge, MA, USA, 1995.
- [14] A. Vaswani et al., "Attention is all you need," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.
- [15] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. NAACL-HLT*, 2019, pp. 4171–4186.

[16] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1–67, 2020.

[17] OpenAI, "Whisper: Robust speech recognition via large-scale weak supervision," *arXiv preprint arXiv:2212.04356*, 2022.

[18] M. Cutler and A. Davis, "Look who's talking: Speaker diarization using audio-visual cues," in *Proc. Int. Conf. Multimodal Interfaces*, 2000, pp. 11–18.

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