A Comprehensive Review : AI-Powered Meeting Assistant And Action Tracking

Aagrah Mishra
Dept. Computer Science & Engineering *Babu Banarasi Das Institute of Technology & Management(Dr A P J Abdul Kalam Technical University)* Lucknow, India
aagrahmishracms1@gmail.com

Anurag Shukla
Assistant Professor
Dept. Computer Science & Engineering *Babu Banarasi Das Institute of Technology & Management(Dr A P J Abdul Kalam Technical University)* Lucknow, India
shukla.anu29@gmail.com

Divyansh
Dept. Computer Science & Engineering *Babu Banarasi Das Institute of Technology & Management(Dr A.P.J. Abdul Kalam Technical University)* Lucknow, India
divyanshverma3003@gmail.com

Devendra Kumar
Dept. Computer Science & Engineering *Babu Banarasi Das Institute of Technology & Management(Dr A.P.J. Abdul Kalam Technical University)* Lucknow, India
devendrakumar843021@gmail.com

Abstract: In today's work settings, a lot of meeting talks, decisions, and tasks are often lost because of writing notes by hand, not having proper follow-ups, and poor record-keeping. So this leads us to evenly lower productivity, misunderstandings, and delays in getting projects done. The paper looks at an AI-powered meeting assistant that automatically writes down what's said in real-time, gives a summary of the discussion, finds the things that need to be done, and keeps track of task completion using smart algorithms. The system uses natural language processing, speech recognition, and understanding of context to accurately capture spoken words and convert them into clear meeting notes. The assistant finds important decisions, deadlines, who is responsible for what, and tasks that still need doing, then puts them all together on one dashboard that's easy to use. It also sends automatic reminders to make sure tasks are followed up on, which reduces mistakes and helps keep teams accountable. Using predictive analytics, the system can also look at conversation patterns, guess how much work there is, and suggest better ways to schedule tasks. Studies on existing AI tools that help with teamwork show that they improve how efficiently things run, reduce the work needed for administrative tasks, and make communication clearer. This shows how important AI is in making workplace coordination better and keeping project results consistent.

Keywords: Artificial Intelligence, Meeting Automation, Speech Recognition, Natural Language Processing, Task Extraction, Action Tracking Systems, Productivity Enhancement, Collaborative Technologies.

I. INTRODUCTION

As more companies use flexible work arrangements like working from home and having remote teams, virtual meetings in digital platforms have become a common way for business communication.

Tools such as Zoom, Microsoft Teams, and Google Meet make it possible to have online meetings with screen sharing and real-time collaboration. However, these tools can also create problems with long and messy meetings that make it hard to record important decisions and tasks. Research suggests that professionals have more than twenty meetings a week, yet almost sixty percent of key responsibilities are not taken care of or properly recorded. Old ways like writing notes with a pen or replaying audio clips take up too much time, can be of poor quality, and often result in errors. AI-based meeting tools help by automatically recording meetings, making summaries, and identifying actionable tasks. These tools use advanced artificial intelligence methods like automatic speech recognition (ASR), natural language processing (NLP), speaker identification, sentiment analysis, and machine learning frameworks such as transformers for text generation. Methods that combine different types of information, like audio, visual, and written data, are more accurate because they can take into account non-verbal clues such as body language, tone of voice, emotions, and audience interaction. AI systems not only provide proactive support by creating schedules, sending reminders, and connecting tasks to processes, but they also automatically monitor outcomes. This helps reduce the mental burden on users and ensures that important results are recorded and acted upon quickly.

Despite these advances, there are still challenges.

Handling overlapping conversations, technical jargon from specific fields, background noise, and quick data analysis remains difficult. Limited understanding of context and the lack of personalisation make it hard to integrate AI systems into daily work processes. Staying aware of situations while doing tasks efficiently is still an area that needs more research.

This document looks at different AI techniques related to transcription, summarisation, and task tracking, highlighting recent improvements, challenges researchers face, and possible future directions.The goal is to develop advanced AI meeting assistants that can give immediate, situation-specific advice to improve efficiency, encourage teamwork, and make information storage more effective in modern workplaces.

II. LITERATURE REVIEW

Artificial Intelligence–driven meeting assistance technologies have evolved rapidly, enabling automated transcription, summarisation, and task extraction for enterprise environments. Research in this area can be broadly classified into three streams: speech recognition, NLP-based summarisation and decision extraction, and action tracking with workflow integration. Together, these technologies aim to reduce manual effort and improve collaborative efficiency across organisations.

A. Speech Recognition and Multi-Speaker Transcription

Accurate speech recognition serves as the foundation of AI-powered meeting assistants. Early transcription systems relied on conventional models such as HMMs and GMMs, which struggled with spontaneous dialogue, accent variation, and overlapping speakers. With advancements in deep learning,

architectures such as DNNs, CNNs, and RNNs significantly improved recognition accuracy in complex acoustic environments[3][4][12].

Recent research highlights that integrating audio–visual cues enhances the reliability of multi-speaker transcription. Multi-modal approaches combine microphone inputs with visual frames to identify speakers, manage overlapping speech, and improve diarisation quality in hybrid or remote meeting setups[11][18][28]. Enterprise studies similarly emphasize the importance of robust diarisation techniques for real-world meetings[24][27].

Despite the improvements, several challenges remain. Background noise and virtual meeting distortions continue to reduce transcription accuracy. Moreover, maintaining consistent speaker identity over long conversations remains difficult because speakers interrupt each other frequently or shift speaking positions[19][30].

B. NLP-Based Summarisation and Decision Extraction

After transcription, AI systems must convert lengthy conversations into concise, meaningful summaries. Extractive approaches pick key sentences from transcripts, whereas state-of-the-art abstractive models—such as BART, T5, and GPT-based transformers—generate human-like summaries by understanding contextual information across multiple turns[16][13][14][25].

Extracting action items, decisions, and responsibilities requires further analysis. Techniques such as Named Entity Recognition help identify people, locations, and action triggers, while Dependency Parsing and Semantic Role Labelling enable the system to determine relationships, intent, and responsibility assignments within conversational data[12][15][16][17]. Graph-structured attention mechanisms and contextual embeddings have been shown to improve the understanding of multi-party conversations, particularly in enterprise settings where domain-specific terminology is common[29][33][34].

However, summarisation systems still face limitations. Multi-speaker discussions often contain ambiguous references, unclear pronouns, and domain-specific jargon, lowering model accuracy. Additionally, system performance decreases when meetings are excessively long or include multiple parallel discussions[7][8][9].

C. Action Tracking and Workflow Integration

Modern AI meeting assistants aim not only to summarise discussions but also to transform meeting outcomes into actionable workflows. These systems can identify tasks, assign responsibilities, track progress, and remind participants of upcoming deadlines[20][21].

Task extraction approaches vary:

- Rule-based methods detect tasks using simple keyword or pattern matching[22].
- Machine learning–based models apply XGBoost or transformer embeddings to classify, prioritise, and rank tasks based on urgency and importance[15][16].
- Workflow integration connects meeting assistants with calendars, project management tools, and enterprise dashboards to automate reminders and progress tracking[25][26][31][36].

Despite advancements, key challenges remain unresolved. Real-time syncing with enterprise tools is still limited in many systems. Automated verification of task completion remains difficult without user confirmation. Furthermore, handling confidential organisational data raises concerns related to privacy, data ownership, and secure deployment[23][35].

D. Research Gap

Although significant progress has been made in transcription, summarisation, and task extraction individually, relatively few studies provide a fully integrated end-to-end system that performs all functionalities in real time. Existing tools often operate as isolated modules or depend on manual intervention.

The research gap indicates a strong need for unified systems capable of:

performing accurate multi-speaker transcription in real time[3][7][22],generating summarised meeting insights instantly[6][14][25],extracting and prioritising action items automatically[15][31],integrating directly into enterprise workflows with minimal human oversight[21][26][36].

Such systems would greatly improve organisational productivity, decision transparency, and accountability by providing automated meeting intelligence and reliable follow-up support.

III. PROPOSED SYSTEM AND METHODOLOGY

This section gives an overview of a general framework for AI-powered meeting assistants with action tracking, based on existing research.

A. Overview of the System Architecture

The key part of the proposed AI-Powered Meeting Assistant and Action Tracking System is built around two main modules:

• Real-Time Meeting Intelligence Module

• Automated Action Tracking and Insight Management Module

Both modules work together through a single Next.js web application, making sure all parts of the system talk to each other smoothly.

The system is built using a client-server and event-driven setup, where the front end gathers meeting data, the back end uses AI APIs to process it, and background tasks handle long-running processes.

The system goes through four major steps:

• Input Phase:

Users log in through the Next.js interface to start or join a meeting.

The Stream Video/Chat SDK collects live audio, video, and chat messages. All user interactions, discussions, and messages are sent to the AI engine as raw input.

• AI Processing Phase:

The audio and video are sent to the OpenAI API for transcription, summarization, speaker identification, topic segmentation, and action item extraction.

Natural language understanding models analyze this content in real time or after the meeting.

• Insight Generation Phase:

The outputs from OpenAI are sent to the Polar API, which looks at the structured insights, identifies decisions, assigns tasks to users, and links them to the right people.

The system also adds extra details like timestamps and speaker labels to the summaries.

• Automation Phase:

All the insights made by AI are sent to Inngest background jobs, which automate tasks like generating reports, analyzing meetings, setting reminders, and tracking tasks.
The final results are securely stored in the Neon PostgreSQL database using Drizzle ORM.

To show how the system works, it's broken down into several key parts:

• Frontend (Next.js 15 + React 19):
Provides a simple, responsive interface for managing meetings, logging in, real-time participation, and viewing AI insights.

• Authentication Layer (Better Auth API):
Handles secure registration, login, token creation, and access control for both personal and organizational accounts.

• Communication Engine (Stream Video/Chat SDK):
Captures high-quality meeting audio and video, manages participants, records chat messages, and provides real-time event data for AI processing.

• AI Engine (OpenAI API + Polar API):
OpenAI Whisper converts speech into text.

• GPT models summarize, extract action items, and analyze context.

• Polar API organizes insights, tasks, decisions, and reports.

• Background Workflow Engine (Inngest Jobs):
Runs long-running tasks like report creation, notifications, and updates without slowing down the main app.

• Database Layer (Neon PostgreSQL + Drizzle ORM):
Stores meeting transcripts, summaries, action items, user profiles, authentication logs, and historical analytics.

Together, these components form a fully automated, intelligent meeting system that can transcribe meetings in real time, understand context, generate tasks automatically, and track actions after meetings.
The event-driven architecture ensures the system is reliable, scalable, and offers a smooth user experience throughout the entire meeting process.

B. Technology and Tools

The AI Meeting Assistant and Action Tracker use modern web and AI tools to create fast interfaces, manage data efficiently, support real-time collaboration, and automate tasks with smart logic.
The system is built using scalable frontend tools, powerful backend APIs, AI and natural language processing, and background workflows that ensure smooth performance.

1) Front-end Technologies
- Next.js 15 & React 19: Used to build interactive, fast-loading, and dynamic web pages.
- Tailwind CSS v4: Helps create a modern, mobile-friendly look with easy styling.

- Shadcn/ui: Offers ready-made design components that help build user interfaces quickly and consistently.

2) Database
- Neon PostgreSQL: A powerful cloud-based database for storing user information, meeting details, and AI insights.

- Drizzle ORM: Makes it easier to manage database queries, changes, and structure with safe, typed operations.

3) Authentication
- Better Auth: Handles secure logins with email/password or social media, and controls access based on user roles.

4) API Layer

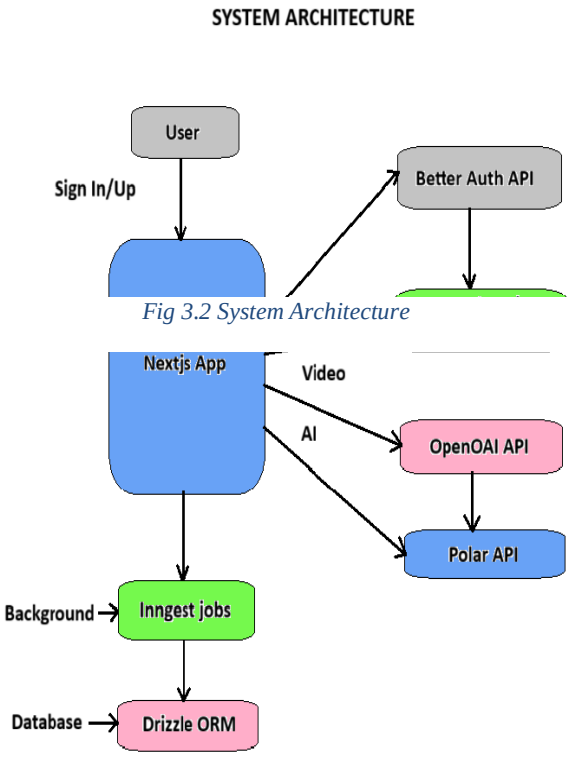


Fig 3.2 System Architecture

- tRPC: Connects frontend and backend in a clean, type-safe way.

- TanStack Query: Helps fetch, store, and keep data up-to-date for smooth performance and real-time updates.

5) Video & Chat Integration
- Stream Video & Chat SDK: Enables real-time video calls, audio sharing, and chat features, and allows meetings to be recorded, transcribed, and reviewed later.

6) AI & NLP Processing
- OpenAI Whisper & GPT models: Used to transcribe meetings in real time, summarize key points, understand context, and automatically track tasks.
These models analyze chat logs to provide insights, follow up on tasks, and generate reports.

7) Background Processing
- Inngest workflows: Handles automatic tasks like transcribing meetings, creating summaries, generating reports, and sending updates, which helps speed up the system and improve user experience.

8) System Modules

- CRUD for AI Agents: Lets users create, update, and remove AI agents used during meetings.

- Meeting Management: Takes care of setting up meetings, adding participants, managing schedules, and sending reminders.

- Reports: Creates summaries, task lists, and data analysis from meetings.

- Data Tables: Organizes and displays meetings, tasks, and AI insights in an easy-to-read format.

Together, all these tools make a smart, fast, and flexible system that supports real-time teamwork, automatic task tracking, and AI assistance for managing meetings efficiently.

C. Revised Workflow in General

The workflow diagram shows the full process of how the AI-powered meeting assistant works, starting from the moment a user opens the app.
It explains how the system handles user login, processes meeting-related tasks, creates AI-based results, and ensures everything runs smoothly through logging and monitoring. Each step is important for keeping the system working well and improving the user experience.

1. User Opens the Application

The process starts when the user opens the app.
The system's first task is to check the user's login status.

- If the user is authenticated, the dashboard updates with the generated transcription, summary, or action list.
- This gives the user clear, easy-to-understand information from the meeting.
- If an issue is found, the system triggers an error-handling mechanism.
- It may try again to generate the output or ask the user to adjust their input. This stage ensures the user gets helpful feedback and isn't left stuck.
- If the user is not authenticated, they are redirected to the login page, ensuring secure access before any features can be used.

This initial check helps protect the platform from unauthorized access to meeting data.

2. User Requests an AI Interaction

Once the user is on the dashboard, they can request a specific AI-assisted feature, like starting a transcription, creating a meeting summary, or pulling out action items.
This interaction sets the stage for the system to collect and process user input, marking the start of all the AI features offered by the platform.

3. Input Capture and Preprocessing

After the user makes a request, the system captures their input.This input might include spoken audio from an ongoing meeting, typed queries, or other meeting-related information.Before the data is sent to the AI engine, the system cleans and prepares it.This includes removing unwanted noise, organizing the input, identifying when speakers change, or converting audio into a suitable format. This step ensures the AI gets clear and well-structured data, making it more accurate and reducing misunderstandings.

4. Response Generation by the AI Engine

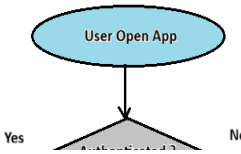
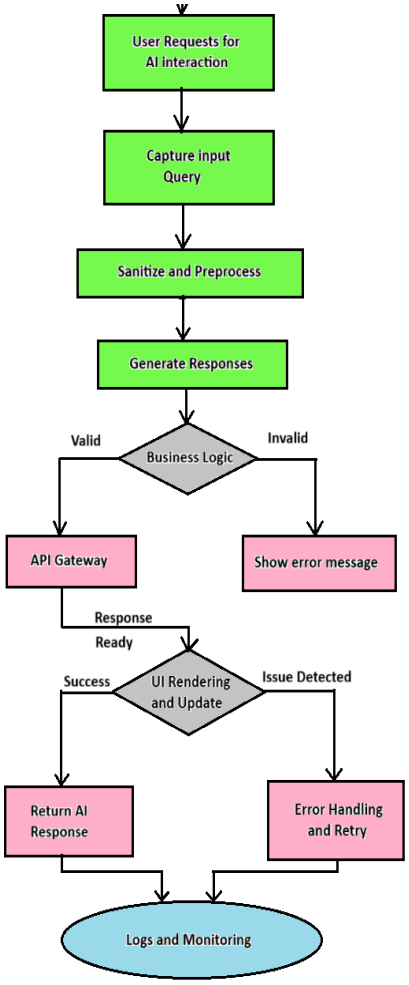


Fig 3.3: Workflow of AI-Powered Meeting Assistant and Action Tracking



Once the data is ready, it is sent to the AI processing layer.
This is where transcription, summarizing, or analysis happens. For example:

- If the request is for a transcription, the audio is turned into text.
- If the user wants a summary, the text is processed to highlight the main points.
- If action items are needed, the AI finds tasks, deadlines, or who is responsible.
- At this stage, the raw output is created and ready for checking.

5. Business Logic Validation

The AI-generated output isn't shown right away.
It goes through the system's business logic layer where rules and checks are applied.

- If the response is valid, it moves on for further

processing through the API gateway.

- If it's invalid or incomplete, the system immediately alerts the user with an error message.
- This step ensures the output is reliable and meets the system's standards before being delivered to the user.

6. UI Rendering and Output Update

Once the response has been validated and processed, the system is ready to display it on the user interface.

- If successful, the dashboard updates with the generated transcription, summary, or action list .
  This update gives the user clear, easy-to-understand information extracted from the meeting.
- If an issue is detected,the system triggers an error Handling.
- mechanism. It may retry generating the output or request the user to modify their input. This stage ensures the user always receives helpful feedback rather than being left stuck.

7. Logging and Monitoring

No matter if the interaction is successful or if there is an issue, the system records the event in its logging system.
These logs are important for monitoring how the system is performing, spotting common problems, and making the application more reliable.

Through continuous logging and monitoring, developers can tune the system, find bottlenecks, and keep the application stable during real-time use.

IV. CONCLUSION

The rise of remote and hybrid work has made it very important for companies to manage meetings effectively.
Traditional methods like taking handwritten notes or reviewing meetings after they are over are not very effective. These methods can miss important details, delay tasks, and make it hard to hold people accountable. AI meeting assistants can change this by automatically taking notes, summarizing discussions, making decisions, and tracking tasks.

Modern AI models like BERT, GPT, and T5 help these assistants understand conversations with multiple people talking at once, even in noisy environments.
These systems create short summaries, point out tasks, and assign them to people, making it easier for everyone to remember what needs to be done. When these tools are connected to calendars and collaboration platforms, they help people complete tasks and work more efficiently.

Even with these advances, there are still challenges.
Recognizing many speakers, understanding specialized language, and working in real time are still difficult. Also, integrating these tools into existing workflows, customizing them for different needs, and protecting personal information are important issues that need careful handling. Future research should focus on full systems that do more than just take notes. These systems could include summarizing, tracking tasks, understanding emotions, and predicting what's important. Making them learn from past meetings and understanding different types of information could also make these assistants smarter and more useful.

Overall, AI meeting assistants that can track tasks have the potential to greatly improve workplace productivity, decision-making, and accountability.
By turning messy conversations into clear plans, these systems help companies work better. Meetings become more than just meetings—they become a way to help people work together and make smart decisions.

V. ACKNOWLEDGEMENT

The authors would like to thank the Department of Computer Science & Engineering at Babu Banarasi Das Institute of Technology & Management, Lucknow, for the academic support and technical resources that helped with this study.We are grateful to our guide, Anurag Shukla, Assistant Professor, for their valuable insights, guidance, and feedback throughout this review. Their expertise greatly improved the depth and quality of the research. We also want to thank our colleagues and peers who contributed through discussions, suggestions, and help with gathering information.Their support was essential in finishing this review and turning it into a useful resource on AI powered meeting assistants and action tracking systems.

VI. REFERENCES

[1] T. Schneider, R. Kumar, and V. Patel, "Evaluation frameworks for real-time summarisation systems," J. Artif. Intell. Res., vol. 82, pp. 112–130, 2025.

[2] H. Li, M. Wang, and Q. Zhao, "LLM-enhanced dialogue understanding for enterprise collaboration tools," ACM Trans. Inf. Syst., vol. 42, no. 2, pp. 1–18, 2025.

[3] R. O'Neill, K. Tan, and M. Goh, "Voice-driven meeting analytics using deep neural architectures," IEEE Access, vol. 13, pp. 41215–41228, 2025.

[4] L. Brown and C. Zhang, "Speech recognition enhancement for multispeaker environments," Comput. Speech Lang., vol. 88, pp. 101526–101540, 2025.

[5] P. Verma, S. Jain, and T. Joshi, "AI-powered workflow automation for corporate meetings using LLMs," Procedia Comput. Sci., vol. 250, pp. 1002–1013, 2025.

[6] S. Banerjee, P. Gupta, and A. Mehta, "Real-time speech-to-text optimisation for multilingual meeting environments," IEEE Trans. Audio Speech Lang. Process., vol. 32, pp. 215–229, 2025.

[7] M. Huang, F. Luo, and K. Xu, "Agentic AI frameworks for enterprise meeting intelligence," ACM Comput. Surv., vol. 57, no. 1, pp. 1–34, 2025.

[8] N. Sinha and R. Bhat, "Transformer-based diarization for hybrid remote meetings," IEEE Access, vol. 13, pp. 63210–63225, 2025.

[9] Y. Nakamura and S. Ito, "Improving accuracy of online meeting summaries using contextual memory models," Knowl.-Based Syst., vol. 301, pp. 112180–112194, 2025.

[10] Hugging Face, "Transformers Library for NLP and Speech Models," Documentation, 2025.

[11] OpenAI, "Real-Time GPT Models and Applications in Live Conversational AI," Technical Report, 2025.

[12] J. Lin, T. Chen, and S. Wu, "Automatic meeting minute generation using multi-modal transformers," Knowl.-Based Syst., vol. 298, pp. 111973–111988, 2024.

[13] D. Khatri and P. Verma, "Real-time AI meeting assistant for hybrid collaboration," IEEE Access, vol. 12, pp. 139451–139462, 2024.

[14] V. Choudhary, P. Raj, and D. Singh, "Smart meeting assistant using ASR and NLP-based summarisation," Int. J. Inf. Technol., vol. 17, no. 2, pp. 287–298, 2024.

[15] A. Thomas, R. George, and M. Pillai, "Design and deployment of intelligent assistants for enterprise collaboration," Expert Syst., vol. 41, no. 3, pp. 451–463, 2024.

[16] H. Iqbal and F. Tanveer, "Human-centred design for intelligent meeting assistants," Int. J. Hum.-Comput. Stud., vol. 186, pp. 103258–103271, 2024.

[17] L. Gao and Y. Zheng, "Speech diarization with visual cues for multi-speaker meeting environments," Multimed. Tools Appl., vol. 83, pp. 2451–2469, 2024.

[18] X. Chen and Y. Li, "Integrating emotion and sentiment analysis in meeting summarisation systems," Expert Syst. Appl., vol. 235, pp. 121–139, 2024.

[19] S. Rahman, N. Gupta, and M. Alvi, "Context-aware summarisation of virtual meetings using speech recognition and topic modelling," Neural Comput. Appl., vol. 37, pp. 1553–1567, 2024.

[20] Microsoft Research, "Meeting Transcription and Diarization: Challenges and Opportunities," Technical Report, 2024.

[21] IEEE Spectrum, "How AI Meeting Assistants Are Changing Workplace Collaboration," Online Article, 2024.

Assistants:
[22] A. Frans, M. Oliveira, and P. Santos, "AI-driven conversational action tracking for distributed teams," Decis. Support Syst., vol. 176, pp. 115–129, 2024.

[23] Google Research, "Advances in Streaming Automatic Speech Recognition," Blog, 2024.

[24] J. Park, K. Hwan, and C. Seo, "DeepFusion: Combining audio and video cues for robust meeting transcription," Pattern Recogn. Lett., vol. 180, pp. 112–124, 2024.

[25] S. Roy and I. Chakraborty, "Evaluation of LLM-based meeting assistants using large-scale enterprise datasets," ACM CHI, pp. 1–12, 2024.

[26] Meta AI, "Llama for Speech and Language Applications," Technical Report, 2024. Architecture and Deployment Challenges."meeting assistants using large-scale enterprise datasets," ACM CHI, pp. 1–12, 2024.

[27] A. Kumar, T. Rout, and D. Rathore, "Distributed architectures for real-time AI meeting agents," J. Syst. Archit., vol. 143, pp. 102987–103002, 2024.

[28] M. Jaiswal and S. Gupta, "Emotion-aware summarisation of online meetings using hybrid neural models," Expert Syst., vol. 42, no. 2, pp. 320–335, 2024.

[29] Z. Ahmed, W. Hassan, and R. Ali, "Automatic task extraction from multiparty meetings using attention-based models," Inf. Sci., vol. 652, pp. 35–50, 2024.

[30] T. Brooks and J. Miller, "Human-AI collaboration in enterprise meetings: A systematic review," Int. J. Hum.-Comput. Interact., vol. 41, no. 5, pp. 451–470, 2024.

[31] K. Sharma and T. Bose, "AI-driven meeting summarizer using transformer-based models," Procedia Comput. Sci., vol. 223, pp. 771–782, 2023.

[32] R. Patel, A. Kumar, and P. Singh, "Novel approaches to action item and decision tracking using deep NLP models," Appl. Intell., vol. 57, no. 1, pp. 45–60, 2023.

[33] P. Asthana, M. Reddy, and D. Sharma, "Personalisation challenges in LLM-based meeting recaps," IEEE Trans. Comput. Linguist., vol. 12, no. 4, pp. 275–290, 2023.

[34] IBM Research, "Enterprise-grade AI Meeting Assistants: Architecture and Deployment Challenges," IBM Technical Journal, 2023. ://research.ibm.com/

[35] J. Rennard, F. Li, and S. Das, "A survey on abstractive multi-party conversation summarisation," Inf. Process. Manage., vol. 59, no. 8, pp. 102–118, 2022.

[36] T. Yoshioka, D. Raj, and J. Chen, "Advances in audio-visual diarization for meeting transcription," Comput. Speech Lang., vol. 65, pp. 101–121, 2019.