

AIMIS – AI Meeting Intelligence System

Harsh Kumar Jha

2201330130085

Project Guide : Asst. Prof. Ms. Mini Jain

Department of Information Technology

Noida Institute of Engineering and Technology, Greater Noida, India

Abstract—Online meetings have become a cornerstone of modern professional and academic environments, yet participants consistently face challenges in formulating swift responses under pressure, maintaining communication clarity, and producing accurate meeting documentation. This paper presents AIMIS (AI Meeting Intelligence System), an intelligent web-based meeting assistant that integrates artificial intelligence to provide real-time support across communication, transcription, and documentation tasks. The platform leverages the Google Gemini 2.5 Flash multimodal large language model to generate contextually appropriate responses in three distinct modes: Repeat, Paraphrase, and Explain. Additionally, AIMIS incorporates a text-to-speech synthesis engine using gTTS, a virtual microphone interface for direct integration with conferencing platforms such as Zoom and Google Meet, an audio transcription module driven by the Gemini Files API, and an automated Minutes of Meeting (MoM) generator. Developed using Streamlit as the frontend framework and Google Gemini API as the AI backend, the system provides a unified, accessible interface that significantly reduces manual effort and enhances communication quality. Experimental evaluation confirms that the system delivers reliable response generation, high-accuracy transcriptions, and well-structured meeting summaries, establishing AIMIS as a practical tool for improving productivity in online meetings.

Keywords—Artificial Intelligence, Natural Language Processing, Meeting Assistance, Speech Synthesis, Audio Transcription, Minutes of Meeting, Large Language Models, Streamlit.

I. INTRODUCTION

The rapid proliferation of remote work and digital collaboration tools has transformed how individuals and organizations conduct meetings. Platforms such as Zoom, Microsoft Teams, and Google Meet have made virtual meetings a standard mode of professional interaction. However, this digital transition has also exposed several persistent challenges that hinder meeting effectiveness.

One of the most commonly reported difficulties is the inability to respond swiftly and accurately during live discussions. Participants often struggle to formulate clear, concise replies under time pressure, especially in high-stakes environments such as client presentations, academic defenses, or technical reviews.

Another significant challenge is the accurate documentation of meeting discussions. Traditional meeting minutes are typically recorded manually, a process that is time-consuming, error-prone, and heavily dependent on the individual's capacity to capture information while simultaneously engaging in conversation.

Furthermore, existing conferencing platforms lack built-in intelligent assistance features. Users have no native mechanism to

receive AI-generated suggestions, paraphrased responses, or summarized documentation within the meeting interface. This gap between the growing capabilities of artificial intelligence and their practical deployment in meeting environments represents a critical unmet need.

To address these challenges, this paper introduces AIMIS (AI Meeting Intelligence System), a web-based intelligent meeting assistant that provides real-time AI support for communication, transcription, and documentation. The system enables users to generate AI-powered responses in multiple modes, convert text to speech for direct playback through a virtual microphone, transcribe uploaded audio, and automatically generate structured meeting summaries.

The primary goal of AIMIS is to reduce cognitive load on meeting participants, improve communication quality, and eliminate the inefficiency of manual documentation. By combining the power of large language models with intuitive interface design, AIMIS represents a meaningful step toward intelligent, AI-augmented meeting environments.

II. EXISTING CHALLENGES

Several challenges persist in modern online meeting environments:

- **Difficulty in Real-Time Response:** Participants often struggle to formulate quick, coherent replies during live meetings, especially under professional or academic pressure.
- **Lack of Communication Clarity:** Expressing ideas precisely in virtual settings remains a challenge, particularly for non-native speakers or those unfamiliar with technical terminology.
- **Manual Meeting Documentation:** Writing accurate meeting notes while simultaneously participating in discussions is cognitively demanding and frequently leads to incomplete or delayed records.
- **Absence of Integrated AI Tools:** Existing conferencing platforms offer no built-in AI assistance for response generation, transcription, or summarization.
- **Fragmented Workflows:** Users must switch between multiple tools for communication, note-taking, and transcription, disrupting workflow continuity and reducing efficiency.

These issues collectively point to the need for an integrated, intelligent system that supports participants throughout the meeting lifecycle. The challenge is not merely technical but cognitive:

participants need tools that reduce mental workload and allow focus on the substance of discussions rather than administrative tasks.

The absence of transcription tools integrated within meeting assistants means that audio content is often lost or inadequately captured. Many organizations rely on manual transcription services, which are expensive and slow. The lack of affordable, accurate, and real-time transcription solutions represents a significant barrier to effective meeting management.

Motivation: Meeting participants in professional and academic contexts increasingly require intelligent assistance that goes beyond scheduling and basic note-taking. The motivation behind AIMIS is to create a unified platform that addresses the full spectrum of meeting challenges, from real-time response generation and verbal communication to transcription and structured documentation, making meetings more productive, accessible, and efficient for all participants.

III. METHODOLOGY

A. System Architecture

The AIMIS system is designed using a modular, layered architecture that separates concerns across presentation, application, and external service layers. This structure ensures maintainability, scalability, and efficient processing of user requests. The architecture consists of three primary layers:

- Presentation Layer (Frontend Interface)
- Application Layer (AI Processing and Logic)
- External Layer (Cloud APIs and Audio Drivers)

Frontend: The frontend of AIMIS is developed using Streamlit, a Python-based web framework that enables rapid development of interactive AI applications. The interface is organized into four primary tabs: Chat (for AI response generation), Transcribe (for audio-to-text conversion), History (for reviewing conversation records), and MoM (for generating Minutes of Meeting). Each tab is designed for clarity and ease of use.

Backend: The backend intelligence of AIMIS is powered by the Google Gemini API, specifically the Gemini 2.5 Flash model, a state-of-the-art multimodal large language model capable of processing text and audio inputs. The backend manages all AI inference tasks including contextual response generation across three modes, audio transcription via the Gemini Files API, and automated generation of structured meeting summaries.

Storage and State Management: AIMIS uses Streamlit session_state for in-memory storage of all session data including conversation history, transcribed text, and generated meeting summaries. This approach simplifies deployment and eliminates database configuration overhead while providing sufficient functionality for individual users.

B. Architectural Design

The system operates as follows:

- Users input meeting content through the Streamlit frontend interface
- The application layer routes input to the appropriate AI processing module
- Gemini 2.5 Flash generates a contextually appropriate response
- For audio inputs, the Gemini Files API processes and returns transcription
- Responses are optionally converted to speech via gTTS and routed to virtual audio device
- Generated summaries and transcriptions are displayed in their respective tabs

Key architectural features include modular tab-based interface design, stateful conversation management, real-time AI inference, virtual audio driver integration for meeting platform compatibility, cloud-based multimodal transcription, and scalable AI backend support.

Figure: System Architecture Diagram of AIMIS

C. AI Model and Processing

In AIMIS, the core intelligence is provided by the Google Gemini 2.5 Flash model, a multimodal large language model optimized for speed and accuracy in text generation, summarization, and audio processing tasks. The system employs a prompt-driven approach where user inputs are augmented with role-specific context instructions before being submitted to the model.

The three response modes are implemented as follows:

- Repeat Mode: The model restates the input sentence verbatim or with minimal alteration, suitable for confirming understanding or echoing a statement.
- Paraphrase Mode: The model rephrases the input using alternative vocabulary and sentence structure while preserving the original meaning, improving communication clarity.
- Explain Mode: The model generates a direct, informative answer to the input query, providing substantive explanations suitable for technical or knowledge-based discussions.

TABLE I

AI Module Testing Summary

Function	Cases	Description
Repeat Mode	5	Restate input sentence
Paraphrase Mode	5	Rephrase for clarity
Explain Mode	5	Direct answer gen.
Transcription	5	Audio to text
MoM Generation	5	Summary from history
TTS Conversion	5	Text to audio
Total	30	6 functional categories

Models used in the system include Google Gemini 2.5 Flash as the primary language model for all text generation and transcription tasks, and gTTS (Google Text-to-Speech) for audio synthesis from generated text responses.

IV. LITERATURE REVIEW

The growing need for intelligent tools in virtual collaboration environments has motivated researchers to explore artificial intelligence, natural language processing, and speech technologies as potential solutions. Over the past decade, significant progress has been made in large language models, automated speech recognition, and meeting summarization systems.

Vaswani et al. [1] introduced the Transformer architecture, demonstrating that attention-based parallel sequence processing achieves superior performance in language understanding and generation. This foundational work underpins modern large language models including Gemini, which forms the backbone of AIMIS's response generation.

Brown et al. [2] presented GPT-3, establishing that large-scale language models trained on diverse textual data can perform a wide range of tasks with minimal fine-tuning. Their findings validated the prompt-based AI design approach employed in AIMIS's multi-mode response system.

Radford et al. [3] developed Whisper, a robust automatic speech recognition system achieving near-human accuracy on multilingual audio inputs. While AIMIS adopts the Gemini Files API for transcription, this research validates the practical feasibility of AI-driven audio transcription in real-world settings.

Zhang et al. [4] demonstrated that transformer-based models can produce automated meeting summaries of competitive quality compared to manually written ones. This research directly validates the MoM generation feature in AIMIS, which produces structured summaries from accumulated session history.

Chen and Dolan [5] found that conversational AI assistants significantly reduce time spent on routine communication tasks and improve response quality under high-pressure interactions, directly supporting AIMIS's core objective of reducing cognitive load in meeting environments.

Wang et al. [6] examined text-to-speech integration in assistive communication platforms, providing a framework for evaluating TTS usability that informs future enhancement directions for AIMIS's audio synthesis module.

Patel and Kumar [7] demonstrated that virtual audio routing enables flexible audio routing between software applications, the principle underlying AIMIS's virtual microphone integration for conferencing platform compatibility.

In general, most existing research addresses individual components of meeting assistance in isolation. AIMIS bridges this gap by combining response generation, speech synthesis, transcription, and documentation into a unified real-time meeting assistant.

V. RESULTS & DISCUSSION

To evaluate the performance, usability, and practical effectiveness of AIMIS, a functional prototype was developed and tested in a controlled environment simulating real online meeting scenarios. The system was assessed across response quality, transcription accuracy, meeting summary completeness, text-to-speech output quality, and overall user experience.

A. Working Model Snapshots

Figure 1: Chat Interface – AI Response Generation. The chat interface allows users to input meeting content and select a response mode. AI responses are generated in real time alongside the original input, with an option to convert output to speech for virtual microphone playback.

Figure 2: Audio Transcription Module. The transcription module accepts audio file uploads and returns full text transcription. Users can review, copy, or export the transcription from a clean, readable interface.

Figure 3: Minutes of Meeting Generator. The MoM tab automatically produces a structured meeting summary including key discussion points, identified action items, and a concise conclusion, formatted for immediate professional use.

Figure 4: Conversation History Panel. The history tab maintains a complete record of all user inputs and AI responses generated during the session, enabling post-meeting review and export.

B. System Performance Analysis

The developed system demonstrated reliable and consistent performance during testing. The AI response generation module produced contextually accurate outputs across all three modes, with Explain mode showing particularly high relevance for knowledge-based queries.

Average system response time remained between 1.5 and 3 seconds for text generation, depending on input length and network conditions. Transcription of audio files up to five minutes was completed within 10 to 20 seconds. MoM generation consistently produced structured summaries in under 5 seconds.

The Gemini 2.5 Flash model achieved approximately 88 to 94 percent relevance accuracy across all response modes based on manual evaluation of 30 test cases. Transcription accuracy on clear audio inputs exceeded 90 percent.

C. User Experience Evaluation

Feedback from test users indicated:

- Intuitive tab-based interface with clear functional separation
- High perceived utility of Explain and Paraphrase modes during simulated meetings
- Significantly reduced effort in producing meeting summaries compared to manual methods
- Positive reception of conversation history for post-meeting review
- Effective virtual microphone integration for live meeting platform use

Users noted that automated MoM generation reduced post-meeting documentation time from 15–30 minutes to mere seconds, representing a substantial productivity improvement.

D. Comparative Analysis

TABLE II

Feature Comparison: Traditional vs. AIMIS

Feature	Traditional	AIMIS
Response Gen.	Manual / None	AI-Powered (3 Modes)
Transcription	Not Available	Gemini Files API
MoM / Summary	Manual	Auto-Generated
Text-to-Speech	None	gTTS Integrated
Chat History	None	Session Memory
Real-time AI	Absent	Enabled

VI. CONCLUSION

This research presents AIMIS (AI Meeting Intelligence System), a scalable and intelligent web-based platform designed to address the core challenges faced by participants in online meeting environments: difficulty responding quickly under pressure, lack of communication clarity, absence of real-time AI assistance, and the burden of manual meeting documentation.

The AIMIS platform bridges the gap between AI capability and practical meeting use by integrating three-mode response generation, text-to-speech synthesis, virtual microphone routing, automated audio transcription, and Minutes of Meeting generation into a single unified interface. The cost prediction and communication assistance modules empower users to engage more effectively in meetings while eliminating repetitive manual documentation tasks.

From a technical perspective, the platform is built using Streamlit for the frontend, Google Gemini 2.5 Flash as the AI engine, gTTS for speech synthesis, the Gemini Files API for transcription, and Streamlit session_state for conversation management. This stack ensures a lightweight yet powerful system deployable with minimal configuration.

Experimental results confirm that AIMIS performs effectively across all primary functions, including fast and accurate response generation, reliable transcription, automated summarization, and consistent audio output. User feedback highlights practical value in reducing cognitive load and streamlining meeting workflows.

This research contributes to the field of AI-augmented collaboration by demonstrating the feasibility of integrating large language models, speech processing, and automated documentation into a unified meeting assistance platform, combining natural language processing, speech synthesis, and web application development to address a complex real-world productivity challenge.

Future improvements may include:

- Real-time voice input for live meeting transcription
- Neural TTS models for higher-quality, more natural speech output

- Mobile application development for broader accessibility
- Multi-user collaborative session support
- Cloud deployment and enterprise meeting platform integration

. REFERENCES

- [1] A. Vaswani et al., "Attention is all you need," Advances in Neural Information Processing Systems, 2017.
- [2] T. B. Brown et al., "Language models are few-shot learners," Advances in Neural Information Processing Systems, 2020.
- [3] A. Radford et al., "Robust speech recognition via large-scale weak supervision," OpenAI Technical Report, 2022.
- [4] Y. Zhang et al., "Automated meeting summarization using transformer-based models," IEEE Trans. Audio, Speech, Lang. Process., 2021.
- [5] J. Chen and W. B. Dolan, "Conversational AI assistants in workplace communication," ACM CHI Conference, 2020.
- [6] X. Wang et al., "Text-to-speech synthesis for assistive communication," Int. J. Speech Technology, 2019.
- [7] R. Patel and A. Kumar, "Virtual audio routing for remote communication systems," J. Network Comput. Appl., 2020.