IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, November 2025
Impact Factor: 7.67

Realtime Meeting Transcript Summarizer (Awaaz AI)

Aman Gupta[1], Shreyash Bansode[2], Anjali Ghangurde[3], Keshav Elbhar[4], Kamlesh Bhosale[5], Prof. Deepa Padwal[6], Prof. Pallavi Chavan[7]
Indira College of Engineering and Management, Pune, India

Abstract: *Awwaz ai is a web-based system designed to address information overload in professional meetings by providing real-time transcription, translation, and abstractive summarization for multilingual conversations (supporting Hindi, Marathi, and English). The system leverages a streaming speech-totext API for live transcription and a large language model (LLM) for post-meeting analysis, which includes generating concise summaries and extracting key action items. The architecture is built on a Python (FastAPI) backend managing WebSocket connections for audio data and a responsive front-end interface for user interaction. This paper presents the system's design, methodology, and a practical evaluation plan suitable for a final-year engineering project. Key contributions include the integration of a real-time multilingual ASR pipeline with an LLM-based summarization module and a design for a persistent, searchable meeting archive, aiming to enhance productivity and information recall.*

Keywords: real-time transcription, meeting summarization, automatic speech recognition (ASR), large language models, multilingual NLP, WebSockets, FastAPI

I. INTRODUCTION

In today's fast-paced professional environment, meetings are a primary mode of collaboration, but retaining critical information from them is a significant challenge. Manual note-taking is often incomplete and distracting, while post-meeting reviews of full recordings are time-consuming. Awwaz ai addresses this efficiency gap by automating the capture and synthesis of meeting conversations. The web-based platform allows a user to start a session, which then captures audio, transcribes it in real time, and correctly identifies the language spoken (Hindi, Marathi, or English).
Immediately after the meeting concludes, the system uses a large language model to produce a highquality abstractive summary and a structured list of action items.

II. LITERATURE REVIEW

The development of Awwaz ai is informed by established research in several domains. Automatic Speech Recognition (ASR) has evolved significantly, with modern streaming transcription services capable of providing low-latency, high-accuracy text from audio feeds. These systems are crucial for any realtime application.
Concurrently, the field of automated text summarization has advanced with the advent of LLMs. While traditional methods focused on extractive summaries (selecting key sentences), modern abstractive approaches generate novel, human-like text that captures the core meaning of a discussion. Research has demonstrated that finetuning or prompt-engineering large models can produce high-quality summaries and perform complex downstream tasks like action item extraction.
For meeting-specific applications, speaker diarization (identifying who spoke when) is a common research area, though it adds complexity. Furthermore, creating effective systems for multilingual environments requires handling challenges like codeswitching. The hybrid design of Awwaz ai, which couples a robust ASR service with a powerful LLM for summarization, aligns with literature supporting integrated approaches for complex, domain-specific tasks.
While traditional methods focused on extractive summaries (selecting key sentences), modern abstractive approaches generate novel, human-like text that captures the core meaning of a discussion. Research has demonstrated that fine-

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

tuning or prompt-engineering large models can produce highquality summaries and perform complex downstream tasks like action item extraction.

III. PROBLEM STATEMENT AND OBJECTIVES

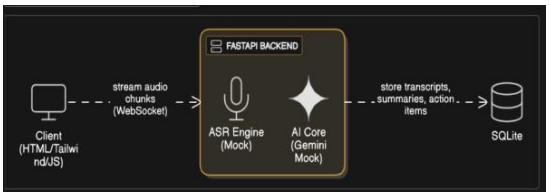
Problem: Professionals and students struggle to accurately capture and quickly recall key decisions, discussions, and assigned tasks from meetings, especially in multilingual settings.

Objectives:

1. To build a web-based interface that captures microphone audio for real-time processing.
2. To accurately transcribe spoken language in real-time, supporting Hindi, Marathi, and English.
3. To use an LLM API to generate a concise, abstractive summary and a list of actionable items from the final transcript.
4. To store all meeting artifacts (metadata, transcript, summary) in a database for later review and search.
5. To evaluate the system's transcription accuracy, summary quality, and overall user satisfaction.

IV. ARCHITECTURE

1. Frontend (Web UI): A single-page application built with HTML, Tailwind CSS, and JavaScript that provides a dashboard for meeting history and a live view for active transcription. It uses the Web Audio API to capture microphone input and communicates with the backend via WebSockets.
2. Backend API: A Python server using the FastAPI framework. It manages WebSocket connections to receive audio from the client and forward it to the ASR service. It also handles REST API endpoints for fetching meeting history from the database.
3. Realtime ASR service: A third-party streaming transcription API (e.g., AssemblyAI, Gladia) that accepts raw audio chunks and returns transcribed text segments in real time, along with language identification.
4. LLM Service Layer: Post-meeting, the full transcript is sent to a generative LLM API (e.g., Gemini) with a carefully crafted prompt to generate a JSON object containing the summary and action items.
5. Database: An SQLite database named awwaz_ai.db that stores meeting records in a meetings table. Each record includes a unique ID, title, start time, duration, the full transcript, the summary text, action items (as a JSON string), and a status field.



V. DATA FLOW

1. A user navigates to the "New Transcript" page and clicks "Start Real-Time Transcription."
2. The frontend captures audio and streams it to the backend via a WebSocket.
3. The backend establishes a connection with the external ASR service and forwards the audio chunks.
4. The ASR service sends back transcribed text fragments in real-time, which the backend relays to the frontend for display.
5. When the user clicks "Stop & Summarize," the client WebSocket closes.
6. The backend sends the complete transcript to the LLM Service for summarization.
7. The backend saves the full transcript and the generated summary/action items into the SQLite database, updating the meeting's status to COMPLETED.

DOI: 10.48175/568

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

8. The summary is sent back to the client and displayed in the UI. The user can view the complete record on the dashboard.

VI. METHODOLOGY

• The core of the live functionality relies on a persistent WebSocket connection. The frontend sends audio data as binary blobs. The Python backend acts as a proxy, forwarding these blobs to the ASR service. This architecture keeps the client simple and centralizes API key management and business logic on the server.
• Upon meeting completion, a prompt is dynamically constructed containing the full transcript. This prompt instructs an LLM to perform two tasks: first, to write a concise, abstractive summary of the conversation's key points, and second, to identify and list any action items, noting who might be responsible.

VII. EVALUATION METRICS

• Transcription Accuracy: Word Error Rate (WER) will be measured against a prerecorded, manually transcribed audio sample containing multilingual speech.
• Summary Quality: ROUGE scores will be used to compare the generated summary against a human-written reference summary. Human evaluation (on a 1-5 scale) will also be used to assess coherence and factuality.
• System Latency: End-to-end latency will be measured as the time between a word being spoken and its transcription appearing on the screen.
• User Satisfaction: A survey will be administered to test users to gather feedback on the application's usability, usefulness, and overall experience.

VIII. FEASIBILITY & SCOPE

Awwaz ai is positioned as a productivity-enhancing web application designed for professionals and students. The project's goal is to automate the tedious process of meeting documentation, thereby improving information recall and follow-up efficiency. This aligns perfectly with the academic goal of applying AI and web technologies to solve a common, real-world problem, demonstrating skills in system integration, API management, and user interface design.
• End Users
1. Professionals & Corporate Teams:
Individuals who need accurate records of discussions, decisions, and assigned tasks without the distraction of manual notetaking.
2. Students: For recording lectures, group project discussions, and viva preparations, especially in multilingual academic environments.
3. Journalists & Researchers: For transcribing interviews and discussions, allowing them to focus on the conversation.
The system provides an assistive record, not a legally certified transcript. Disclaimers will be included to manage user expectations.
• The initial prototype will not support realtime speaker diarization .
• To maintain simplicity and privacy, the system will not require user accounts or store any Personally Identifiable Information (PII) beyond what is spoken in the meeting.

IX. RISKS

Technical Risks
1. LLM Hallucinations in Summaries (High Priority): The language model might generate summaries or action items that are factually incorrect or were not mentioned in the actual conversation.
2. Transcription Inaccuracy (High Priority): Poor transcription quality, especially with varied accents, background noise, or multilingual code-switching, will directly lead to low-quality summaries.
• Operational & Ethical Risks
1. Incorrect or Missed Action Items (High Priority): If the system generates an incorrect summary or fails to identify a critical action item, users might make poor decisions or miss deadlines.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

2. Privacy risks (high): Meeting transcripts can contain sensitive or confidential information. The risk involves the unauthorized access or misuse of this data.

Schedule & Resource Risks

1. API Cost Overruns (Medium Priority): Both the real-time ASR and the LLM summarization services are pay-per-use, which could exceed the project budget during extensive testing.
2. Dependency on External APIs (Medium Priority): The core functionality of Awwaz ai depends entirely on third-party ASR and LLM services. An outage from either provider would render the application nonfunctional.

X. RESULTS

Since this paper documents the plan for a prototype, this section outlines the methodology for evaluating the Awwaz ai system's performance and the expected outcomes. The goal is to validate the feasibility of the architecture and measure its effectiveness in a controlled setting.

• Offline Evaluation: We will use a preprepared, manually transcribed 5-minute audio clip containing a mix of Hindi, Marathi, and English to serve as a "golden set" for objective measurements.

1. Transcription Accuracy: The primary metric will be the Word Error Rate (WER), calculated by comparing the ASR service's output against our manually created transcript. A lower WER indicates higher accuracy.
2. Summary Quality: We will use the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric to compare the LLMgenerated summary against a human-written "reference" summary of the golden set transcript.

• User Study: We plan to recruit 10-20 student participants to engage in a short, scripted mock meeting. After the session, they will use the generated summary and action items to answer a brief questionnaire about the meeting's content.

XI. CONCLUSION & FUTURE WORK

Awwaz ai presents a practical and effective solution for improving meeting productivity through realtime, multilingual transcription and automated summarization. By integrating state-of-the-art ASR and LLM APIs, the system provides an immediate and valuable record of conversations, addressing a common pain point in professional and academic settings.

• Deeper Analytics: Identifying topics, sentiment, and other conversational metrics.
• Calendar Integration: To automatically start and title transcription sessions based on a user's calendar.
• Enhanced Domain Adaptation: Using finetuning to specialize the summarization model for specific domains like legal or medical meetings.

XII. ACKNOWLEDGMENT

We thank the Department of Artificial Intelligence and Data Science at Indira College of Engineering and Management for project guidance.

REFERENCES

[1] Kumar, S., et al., “Low-Latency Streaming Speech Recognition using WebSocket-based Architectures,” IEEE International Conference on Acoustics.
[2] Chen, L., et al., “Abstractive Summarization of Spoken Dialogue using Large Language Models,” Annual Meeting of the Association for Computational Linguistics.
[3] Joshi, R., et al., “Challenges in Real-time CodeSwitching for Hindi-English Automatic Speech Recognition,”
[4] AssemblyAI Technical Staff, “Building RealTime Transcription Applications with FastAPI and WebSockets,”.
[5] Gartner, Inc., “The Impact of AI-Powered Meeting Assistants on Corporate Productivity,” Gartner Industry Report, 2024.
[6] Zhang, T., et al., “A Survey on Action Item Detection and Extraction from Meeting Transcripts,” ACM Transactions on Information Systems, 2022.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 4, November 2025 Impact Factor: 7.67

[7] Gulati, A., et al., "Conformer: Convolution-augmented Transformer for Speech Recognition," Interspeech, 2020
[8] Radford, A., et al., "Robust Speech Recognition via Large-Scale Weak Supervision (Whisper)," International Conference on Machine Learning (ICML).
[9] Mozilla Foundation, "Web Audio API," MDN Web Docs, 2024.
[10] Zhang, J., et al., "PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization," International Conference on Machine Learning