Published by :
https://www.ijert.org/
An International Peer-Reviewed Journal

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

AutoNote AI: Real-Time Transcription and Automated PDF Generation for Live Meetings

Rupali Dupade
Department of Computer Engineering
Jayawantrao Sawant College of Engineering Hadapsar, Pune, India

Punam Dupade
Department of Computer Engineering
Jayawantrao Sawant College of Engineering Hadapsar, Pune, India

Sakshi Japkar
Department of Computer Engineering
Jayawantrao Sawant College of Engineering Hadapsar, Pune, India

Pooja Kavhare
Department of Computer Engineering Jayawantrao Sawant College of Engineering Hadapsar, Pune, India

Mayuri Padwal
Department of Computer Engineering Jayawantrao Sawant College of Engineering Hadapsar, Pune, India

***Abstract*—The rapid growth of digital communication has made virtual meetings central to modern collaboration. However, efficient meeting documentation remains a major challenge, as traditional manual note-taking is time-consuming, inconsistent, and error-prone. This paper proposes AutoNote AI, an intelligent automated system that streamlines the entire meeting documentation process by integrating real-time speech-to-text conversion, noise reduction, speaker diarization, and NLP-based summarization into a unified platform. The system automatically generates well-formatted PDF reports and distributes them to all participants via email immediately after the meeting, ensuring accurate, structured, and timely documentation. The backend is implemented using FastAPI with Python, while MongoDB is used for scalable data storage. Transformer-based models such as BERT are employed for context-aware summarization, enabling the system to extract key decisions and action items from raw transcripts. Evaluation results demonstrate that the system achieves over 90% transcription accuracy in clean audio environments and maintains reliable performance across multi-speaker scenarios. AutoNote AI thus bridges the gap between basic transcription tools and fully automated intelligent documentation systems, offering a scalable and practical solution for corporate, academic, and research environments.**

***Index Terms*—Real-time transcription, Speech-to-text, Meeting automation, AI summarization, Natural language processing, PDF generation, Email automation**

I. INTRODUCTION

The rapid growth of online communication platforms such as Zoom, Microsoft Teams, and Google Meet, virtual meetings have become an integral part of modern collaboration across corporate, academic, and research environments. As organizations increasingly rely on remote interactions, the need for accurate and efficient meeting documentation has become critically important. Proper documentation ensures that key decisions, action items, and discussions are preserved for future reference and accountability. However, traditional manual note-taking methods remain highly inefficient, often leading to incomplete records, inconsistencies, and reduced participant engagement during discussions [1]. Manual notetaking requires individuals to divide their attention between active participation and documentation, which can result in missed information and decreased comprehension. Furthermore, handwritten or manually typed notes lack standardization, making it difficult to share and interpret information across teams. Existing digital tools attempt to address this issue by providing automated transcription services; however, most of these systems generate only raw, unstructured transcripts that are difficult to analyze and extract meaningful insights from [2]. To overcome these limitations, AutoNote AI is proposed as a comprehensive intelligent meeting assistant system. The system is designed to automate the entire documentation pipeline through the following functionalities: • Real-time audio capture using advanced streaming technologies • Speech-to-text conversion powered by AI-based transcription models • Automatic summarization using natural language processing techniques • Structured PDF report generation for clear and professional documentation • Automated email delivery to ensure timely distribution of meeting records • Automated email delivery to ensure timely distribution of meeting records By integrating these components into a unified platform, AutoNote AI eliminates the need for manual intervention and ensures consistent, accurate, and well-structured documentation. The system not only enhances productivity but also reduces cognitive load on participants, allowing them to focus entirely on discussions rather than note-taking. Additionally, the automated workflow improves accessibility and collaboration by ensuring that all stakeholders receive standardized meeting reports promptly.

II. LITERATURE REVIEW

The development of intelligent meeting documentation systems builds upon advancements in speech recognition, natural language processing, and meeting summarization technologies. Several research efforts and systems have attempted to address different aspects of this domain; however, a fully integrated solution remains limited.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

TABLE I
SUMMARY OF LITERATURE REVIEW

Sr. No.	Title	Year	Author	Remark	Gap Analysis
1	Policies and Evaluation for Online Meeting Summarization	2025	Schneider, F., Turchi, M., & Waibel, A.	Proposed evaluation metrics and policies for real-time meeting summarization with emphasis on latency and partial summaries.	Does not address structured PDF generation or automated email delivery.
2	Real-Time Audio Transcription with Automated PDF Summarization and Contextual Insights	2024	Rakshitha, S. R., Naik, S. P., Sanjana, V. S., Suprasanna, V., & Nayana, C. P.	Developed a system for real-time transcription and automated PDF generation with contextual insights.	Lacks integration of automated delivery mechanisms and advanced semantic structuring.
3	MISP-Meeting: A Real-World Dataset with Multimodal Cues for Long-form Meeting Transcription and Summarization	2025	Chen, H., et al.	Introduced a multimodal dataset to improve transcription and summarization accuracy.	Focuses on dataset development; no automation for documentation delivery.
4	Evaluation of Real-Time Transcriptions Using End-to-End ASR Models	2024	Arriaga, C., et al.	Evaluated ASR models for real-time transcription under different audio conditions.	Focuses only on transcription quality; no summarization, PDF generation, or delivery integration.

A. Limitations of Existing Systems

From the above studies, several key limitations can be identified:

- Most systems focus on individual components such as transcription or summarization rather than an integrated solution
- Lack of structured output formats like well-organized PDF reports
- Absence of automated delivery systems (e.g., email integration)
- Limited support for end-to-end meeting documentation workflows
- Insufficient focus on user productivity and real-world usability

B. Research Gap

Although significant progress has been made in speech recognition and summarization technologies, there is a clear gap in developing a fully automated, end-to-end meeting documentation system. Existing solutions do not provide a seamless pipeline that combines real-time transcription, intelligent summarization, structured document generation, and automated distribution. AutoNote AI addresses this gap by integrating all these components into a single unified platform.

III. SYSTEM DESIGN AND ARCHITECTURE

A. Architectural Overview

The AutoNote AI system follows a modular layered architecture that integrates a mobile application, backend processing, AI-based text analysis, database storage, and email services. This design ensures real-time performance, scalability, and smooth user interaction.

The system consists of four main components: the Android-based user interface for meeting interaction and audio capture, the FastAPI backend for data processing and coordination, the AI model layer for transcription analysis and summarization, and the database with email service for storing data and delivering reports. Figure 1 illustrates the complete architecture of the AutoNote AI system.

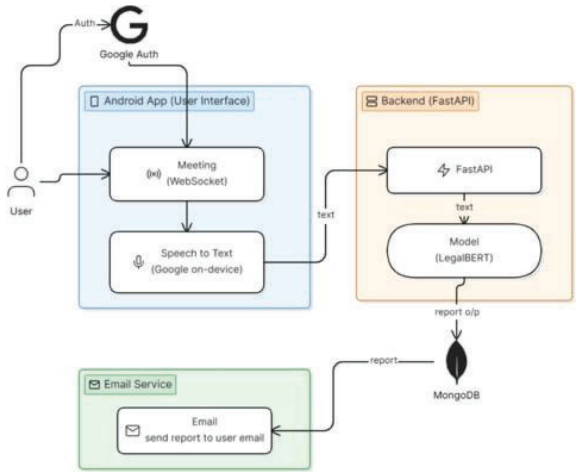

Fig. 1. Overall Architecture of the Proposed AutoNote AI System

Table II presents the complete technology stack used in the AutoNote AI system.

B. User Interface Layer (Android Application)

The frontend is developed as an Android mobile application that acts as the primary interface for users. It enables secure user authentication using Google Authentication and allows users to join or create meetings seamlessly. The application captures real-time audio during meetings and displays live transcriptions and summaries. Once a user joins a meeting, the app continuously captures audio streams and sends them to the backend for processing.

C. Real-Time Communication and Audio Processing

The system utilizes WebSocket-based communication to ensure low-latency and real-time data transfer between the mobile application and backend server. Audio data is continuously streamed from the meeting session, and the speech-to-text module converts the audio into text in real time. This approach ensures fast processing while minimizing dependency on network delays, enabling smooth and efficient transcription.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

TABLE II
AUTONOTE AI TECHNOLOGY STACK

Layer	Technology	Version	Function
Presentation	Android (Kotlin)	Kotlin 1.9 / Android 14	Mobile app UI
Authentication	Google OAuth 2.0	OAuth 2.0	Secure login
Communication	WebSocket (WebRTC)	WebRTC 1.0	Real-time audio streaming
Speech-to-Text	Google STT	-	Convert speech to text
Backend	FastAPI (Python)	Python 3.11 / FastAPI 0.110	API & processing
AI Model	BERT / Transformer Model	BERT Base (v1)	Text summarization
Database	MongoDB	6.0	Store data & reports
Report Generation	ReportLab / iText	ReportLab 4.0	Generate PDF reports
Email Service	SMTP (Gmail API)	SMTP RFC 5321	Send reports

D. Backend Layer (FastAPI Server)

The backend is implemented using FastAPI, which serves as the central processing unit of the system. It handles incoming transcription data from the mobile application, manages API requests, and coordinates communication between system components. Additionally, it forwards the text data to the AI model for further processing and manages the workflow for generating structured reports. FastAPI is chosen for its high performance, asynchronous capabilities, and scalability.

E. AI Model Layer (Text Processing and Summarization)

The AI processing layer uses advanced Natural Language Processing models such as transformer-based architectures like BERT or LegalBERT. This layer performs context-aware analysis of the transcribed text, extracts key discussion points and decisions, and generates concise summaries. It also structures the content into a readable format, which is then forwarded as a final report output to the database layer.

F. Database Layer (MongoDB)

The system uses MongoDB, a NoSQL database, for storing transcripts, summarized reports, meeting metadata, and user-related data. MongoDB is selected due to its flexibility in handling unstructured text data, scalability for large datasets, and fast read/write performance.

G. Email Service Layer

The email service is responsible for delivering the final output to users. Once the report is generated, it is retrieved from the database and the system automatically sends the report via email, ensuring secure and timely delivery of meeting documentation. This feature eliminates manual sharing and enhances accessibility.

IV. METHODOLOGY

A. Mobile Application (Frontend Implementation)

The frontend of the system is developed using Android Studio with Kotlin, providing a user-friendly interface for meeting participation and interaction. It includes secure user authentication using Google OAuth 2.0. The application provides a meeting interface where users can join or create meetings easily. It captures real-time audio using the device microphone and implements WebSocket-based communication (WebRTC/LiveKit) for continuous audio streaming. Overall, the mobile application ensures seamless user interaction while maintaining low latency during live meetings.

B. Real-Time Audio Processing

The system captures audio streams and processes them in real time using advanced speech processing techniques to improve transcription accuracy. Noise reduction is implemented using WebRTC Noise Suppression or RNNoise to eliminate background disturbances and enhance audio clarity. Voice Activity Detection (VAD) using Silero VAD helps in identifying active speech segments and avoids unnecessary processing of silence. Additionally, speaker diarization using pyannote.audio enables the system to distinguish between multiple speakers, ensuring better organization and clarity in the transcription output.

C. Speech-to-Text Conversion

The system uses AI-based speech recognition models such as Google Speech-to-Text and Fast Whisper for converting audio into text. These models process continuous audio streams in real time and generate accurate textual output. The system supports multi-speaker environments and performs effectively even in moderately noisy conditions. Sliding window techniques are used to maintain continuous transcription without interruption. The generated text is then forwarded to the backend for further processing and analysis.

D. Backend Implementation (FastAPI)

The backend is implemented using FastAPI with Python 3.11, which acts as the central processing unit of the system. It is responsible for receiving transcription data from the mobile application, managing API requests, and coordinating communication between different system components. The backend also forwards text data to AI models for summarization and handles the generation of structured reports. FastAPI is chosen due to its high performance, asynchronous capabilities, and scalability, making it suitable for real-time applications.

E. AI-Based Text Processing and Summarization

The system utilizes transformer-based Natural Language Processing models such as BERT, LegalBERT, or other LLM-based models for summarization. These models analyze the transcribed text to extract key discussion points, identify decisions, and determine important action items. The system then generates concise and meaningful summaries while converting raw text into structured content. This ensures that the final output is context-aware, informative, and easy to understand.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

F. Database Implementation (MongoDB)

MongoDB is used as the database for storing all system data, including meeting transcripts, summarized reports, user information, and metadata such as timestamps and speaker details. As a NoSQL database, MongoDB efficiently handles unstructured text data and provides flexibility in data storage. It also offers high scalability and fast read/write operations, making it suitable for handling large volumes of meeting data.

G. PDF Report Generation

The system generates structured meeting reports using PDF libraries such as ReportLab or iText. These reports include summaries, key discussion points, and detailed transcripts, all presented in a well-formatted and professional layout. The formatting ensures clarity and readability, making the reports suitable for sharing and documentation purposes. This feature enhances the usability of the system by providing organized and easily accessible meeting records.

H. Email Service Integration

The system integrates an SMTP-based email service, such as the Gmail API, to automate report delivery. Once the report is generated, it is retrieved from the database, attached as a PDF file, and sent to all meeting participants. This process ensures automatic, secure, and timely distribution of meeting outputs, eliminating the need for manual sharing and improving overall efficiency.

V. RESULTS AND DISCUSSION

The AutoNote AI system was evaluated based on its performance in real-time transcription, summarization accuracy, system responsiveness, and overall usability. The evaluation was conducted through simulated meeting scenarios involving multiple participants, varying noise conditions, and different speaking patterns.

A. System Performance

The system demonstrated efficient real-time processing capabilities, ensuring smooth and reliable performance during live meeting scenarios. The use of WebSocket communication enabled near real-time transcription with minimal delay, allowing continuous data flow between the client and server. Additionally, the speech-to-text conversion implemented using Whisper and Google Speech-to-Text provided fast and continuous text output, ensuring that spoken content was accurately captured without interruption. The FastAPI backend further contributed to system efficiency by handling multiple requests simultaneously without any noticeable performance degradation, thereby supporting scalability for multiple users and sessions. Overall, the system maintained stable and smooth operation even during extended meeting durations.

B. Transcription Accuracy

The accuracy of speech-to-text conversion was evaluated under various real-world conditions to assess the system's reliability and robustness. In a clear audio environment, the system achieved high accuracy, exceeding 90%, ensuring precise transcription of spoken content. However, in scenarios with moderate background noise, a slight reduction in accuracy was observed due to interference, although the overall performance remained acceptable for practical use. In multi-speaker situations, the system effectively handled speaker separation through the use of diarization techniques, allowing clear identification of different speakers within the transcript. Furthermore, the integration of noise reduction mechanisms and voice activity detection significantly enhanced transcription quality by filtering unwanted sounds and focusing only on relevant speech segments.

C. Summarization Quality

The AI-based summarization module generated concise and meaningful summaries by effectively analyzing the transcribed text and identifying the most relevant information. It successfully captured key discussion points and decisions made during the meeting, ensuring that critical information was not overlooked. Additionally, the system was able to extract important action items, making the output more practical and useful for follow-up tasks. The generated summaries were context-aware and well-structured, which improved readability and made them easy to understand.

D. PDF Report Generation

The system successfully generated well-formatted PDF reports that included structured content such as concise summaries and detailed transcripts. These reports were designed with a professional layout, making them suitable for easy sharing and formal documentation purposes. The formatting ensured clarity, readability, and organization of information, allowing users to quickly understand the meeting outcomes. This feature effectively addressed a major limitation of existing tools, which typically provide only raw and unstructured text outputs without proper formatting or presentation.

E. Email Delivery Efficiency

The automated email system ensured efficient and seamless distribution of meeting reports by enabling instant delivery immediately after the meeting was completed. It provided secure sharing of documents, maintaining data confidentiality and ensuring that only authorized participants received the reports. Additionally, the system significantly reduced manual effort by eliminating the need for users to individually send or manage meeting records. As a result, all participants received the reports without delay, which improved accessibility, enhanced communication, and supported better collaboration among team members.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

F. Usability Evaluation

The system was tested with users across various real-world scenarios, including academic discussions, team meetings, and project review sessions, to evaluate its usability and effectiveness. During these evaluations, users found the mobile interface to be simple, intuitive, and easy to navigate, allowing them to interact with the system without any technical difficulty. The automation of transcription and documentation significantly reduced the cognitive load on participants, as they no longer needed to divide their attention between listening and note-taking. This enabled users to remain more focused and actively engaged during discussions. Furthermore, the system received positive feedback from users, particularly regarding its ability to automate the entire documentation process efficiently.

VI. LIMITATIONS AND FUTURE DIRECTIONS

Despite its effectiveness, AutoNote AI has certain limitations that impact its performance in real-world scenarios. The accuracy of the speech-to-text module is highly dependent on audio quality. In noisy environments or when multiple participants speak simultaneously, transcription errors may occur despite the use of noise reduction and Voice Activity Detection techniques. Additionally, speaker diarization is not fully reliable for large group discussions, which may lead to incorrect speaker identification in transcripts.

The summarization component, based on transformer-based Natural Language Processing models, may occasionally miss context-specific details or misinterpret technical conversations. The system also requires significant computational resources for real-time processing, which can limit its deployment in resource-constrained environments. Furthermore, latency may increase during peak usage, particularly in summarization and PDF generation stages.

Another limitation is the lack of direct integration with popular meeting platforms such as Zoom and Google Meet, which affects ease of use. Data privacy is also a concern when using cloud-based APIs for processing sensitive meeting information.

Future work will focus on improving transcription accuracy and speaker diarization using advanced AI models. Integration with major meeting platforms will enhance usability and adoption. The addition of multilingual support, real-time summarization, and keyword highlighting will further improve user experience. Optimizing models for edge devices will reduce hardware dependency, while implementing strong security mechanisms will ensure data privacy. Advanced features such as sentiment analysis and action-item extraction will also be explored to enhance system intelligence.

VII. APPLICATIONS OF AUTONOTE AI

AutoNote AI can be used in many real-world areas where accurate meeting documentation, efficient collaboration, and timely information sharing are important.

A. Corporate and Enterprise Environments: In business settings, AutoNote AI helps organizations maintain accurate records of meetings, conferences, and strategy sessions. It ensures that all decisions and action items are documented and shared with stakeholders promptly, improving accountability and follow-through.

B. Educational Institutions: In academic environments, the system can be used to document lectures, seminars, faculty meetings, and thesis defenses. Students and faculty can receive structured summaries instead of relying on manual note-taking, enhancing the learning experience.

C. Research and Development: Research teams can use AutoNote AI to document brainstorming sessions, project reviews, and collaborative discussions. This ensures that innovative ideas and technical decisions are accurately recorded for future reference.

D. Healthcare: In medical environments, the system can assist in documenting clinical discussions, patient case reviews, and administrative meetings, ensuring that critical medical decisions and protocols are accurately captured and distributed.

E. Legal and Compliance: Legal teams and compliance departments can benefit from accurate transcription and summarization of proceedings, depositions, and regulatory meetings, where precise documentation is essential.

F. Government and Public Sector: Government agencies can use AutoNote AI to document public hearings, council meetings, and inter-departmental discussions, ensuring transparency and accurate record-keeping.

Overall, AutoNote AI is a flexible and scalable system that addresses the documentation needs of diverse domains, enhancing productivity and ensuring reliable communication across all sectors.

VIII. CONCLUSION

In the era of digital communication and remote collaboration, efficient and accurate meeting documentation has become increasingly essential. Traditional manual note-taking methods are not only time-consuming but also prone to errors, inconsistencies, and loss of critical information. These limitations highlight the need for an intelligent and automated solution that can streamline the documentation process while enhancing user productivity.

This paper presented AutoNote AI, an advanced system designed to automate real-time meeting transcription, summarization, and structured report generation. By integrating cutting-edge technologies such as speech-to-text conversion, noise reduction, speaker diarization, and AI-based natural language processing, the system ensures high accuracy in capturing and summarizing meeting content. Furthermore, the inclusion of automated PDF report generation and email delivery provides a complete end-to-end solution for meeting documentation.

The proposed system significantly reduces the cognitive load on participants by eliminating the need for manual note-taking, allowing them to focus entirely on discussions and decision-making. Additionally, the generation of structured and well-formatted reports ensures better readability, accessibility, and knowledge sharing across teams. The system is

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

versatile and can be effectively applied in various domains, including corporate environments, educational institutions, research collaborations, and remote work settings.

Overall, AutoNote AI bridges the gap between basic transcription tools and fully automated intelligent documentation systems. It offers a scalable, efficient, and user-friendly solution that enhances collaboration, improves accuracy, and ensures reliable documentation of meeting outcomes. Future enhancements, such as multilingual support, deeper contextual understanding, and integration with popular meeting platforms, can further expand the system's capabilities and impact.

ACKNOWLEDGMENT

The authors would like to thank the Department of Computer Engineering, Jayawantrao Sawant College of Engineering, Hadapsar, Pune, for providing the necessary resources and support to carry out this research. The authors also express gratitude to Prof. Rupali Dupade for her guidance and valuable insights throughout the development of AutoNote AI. Special thanks to all faculty members and peers who contributed feedback during the testing and evaluation phases of this work.

REFERENCES

[1] F. Schneider, M. Turchi, and A. Waibel, "Policies and Evaluation for Online Meeting Summarization," *arXiv preprint arXiv:2502.03111*, 2025.

[2] S. R. Rakshitha, S. P. Naik, V. S. Sanjana, V. Suprasanna, and C. P. Nayana, "Real-Time Audio Transcription with Automated PDF Summarization and Contextual Insights," *International Journal of Innovative Science and Research Technology*, vol. 9, no. 11, 2024.

[3] H. Chen et al., "MISP-Meeting: A Real-World Dataset with Multimodal Cues for Long-form Meeting Transcription and Summarization," 2025.

[4] C. Arriaga et al., "Evaluation of Real-Time Transcriptions Using End-to-End ASR Models," 2024.

[5] 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.

[6] Google Cloud, "Speech-to-Text Documentation," [Online]. Available: https://cloud.google.com/speech-to-text

[7] OpenAI, "GPT Models and Applications," [Online]. Available: https://openai.com/research

[8] Fireflies.ai, "AI Meeting Assistant," [Online]. Available: https://fireflies.ai/

[9] A. Graves, A. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

[10] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proc. NAACL-HLT*, 2019.