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

A survey on AutoNote AI: An AI-Driven System for Real-Time Meeting Transcription and PDF Creation

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 widespread adoption of digital communication technologies has fundamentally reshaped the way individuals and organizations collaborate in professional, academic, and research settings. Virtual meetings, online classrooms, and distributed team discussions have become routine components of modern workflows. Despite this shift, reliable and comprehensive meeting documentation continues to pose significant challenges. Conventional manual note-taking approaches are labor-intensive, inconsistent, and susceptible to human error, which often results in incomplete records and diminished productivity. Participants frequently struggle to simultaneously engage in discussions and record critical information, leading to the loss of important decisions, insights, and follow-up tasks.**

To overcome these challenges, this paper introduces AutoNote AI, an intelligent and fully automated system engineered to simplify the entire meeting documentation workflow. The proposed system delivers a holistic solution by combining real-time speech processing, artificial intelligence, and automated reporting into a single cohesive platform. AutoNote AI captures live audio from ongoing meetings and processes it using state-of-the-art speech recognition components, including speech-to-text conversion, background noise suppression, voice activity detection, and speaker diarization. Beyond transcription, the system employs advanced Natural Language Processing (NLP) models to perform semantic understanding and produce concise, context-sensitive summaries. Additionally, AutoNote AI automatically constructs professionally formatted PDF reports and features a built-in email delivery mechanism that securely distributes these reports to all meeting participants upon session completion.

***Index Terms*—Real-Time Transcription, Speech-to-Text, Meeting Automation, AI Summarization, Natural Language Processing, PDF Generation, Email Automation, Speaker Diarization, FastAPI, Transformer Models**

I. INTRODUCTION

The increasing prevalence of online communication platforms such as Zoom, Microsoft Teams, and Google Meet has made virtual meetings a cornerstone of modern collaboration across corporate, academic, and research domains. As organizations grow more dependent on remote interaction, the ability to generate accurate and timely meeting documentation has become critically important. Effective documentation preserves key decisions, action items, and discussion threads for future reference and organizational accountability. However, traditional manual note-taking continues to be highly inefficient, frequently producing incomplete records, factual inconsistencies, and reduced participant engagement [1].

When individuals are required to simultaneously participate in conversations and capture information manually, they face a divided attention problem that often results in missed content and diminished comprehension. Moreover, handwritten or manually typed notes lack structural uniformity, complicating information sharing and interpretation across teams. While existing digital tools have attempted to address these issues through automated transcription capabilities, most deliver only raw, unstructured transcripts that are difficult to interpret and act upon [2].

Current solutions fall short of providing a complete end-to-end workflow that encompasses structured summarization, formatted documentation, and automated distribution. The absence of integrated report generation and delivery mechanisms introduces additional manual overhead, thereby undermining the productivity benefits that automation is intended to provide. To address these shortcomings, AutoNote AI is proposed as a comprehensive intelligent meeting assistant that unifies the following capabilities:

- Real-time audio capture using advanced streaming technologies
- Speech-to-text conversion powered by AI-based transcription models

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

- 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

By consolidating these capabilities into a unified platform, AutoNote AI removes the need for human intervention in the documentation process and guarantees consistent, accurate, and well-organized records. The system enhances participant productivity while reducing cognitive burden, enabling attendees to concentrate fully on discussions rather than note-taking [3].

The remainder of this paper is structured as follows. Section II surveys five closely related works that motivate and inform the design of AutoNote AI. Section III presents a comparative analysis of these approaches. Section IV identifies the research gaps that AutoNote AI is designed to address. Section V describes the overall system architecture and design. Section **??** concludes the paper with key findings and directions for future research.

II. LITERATURE REVIEW

A. Online Meeting Summarization Systems

Schneider et al. [1] examined evaluation methodologies and operational policies for real-time meeting summarization. Their framework introduces metrics tailored to assess partial summaries generated incrementally during live meetings, with particular emphasis on achieving minimal latency. While this research advances the evaluation and efficiency of meeting summarization, it is narrowly focused on the summarization stage and does not address the production of structured outputs such as formatted reports, nor does it incorporate any automated sharing mechanism. This gap directly motivates the PDF report generation module in AutoNote AI.

B. Real-Time Transcription and Summarization Systems

Rakshitha et al. [2] introduced a system that couples real-time transcription with automated summarization and contextual analysis. This work demonstrates the potential of combining speech-to-text models with summarization algorithms to extract meaningful insights from meeting audio. However, the system does not incorporate a complete automation pipeline, particularly with respect to structured PDF generation and automated email delivery — components that are essential for practical deployment in professional environments. AutoNote AI directly addresses both of these missing elements.

C. Multimodal Meeting Datasets and Transcription Models

Chen et al. [3] presented the MISP-Meeting dataset, which incorporates multimodal signals including speech, text, and visual data to enhance transcription and summarization performance. By offering a diverse, real-world training resource, this dataset improves the robustness of machine learning models for meeting understanding. Nevertheless, the contribution is restricted to dataset development and does not extend to the design of a deployable system for real-world meeting automation. AutoNote AI draws on the multimodal insights from this work to inform its noise reduction and speaker diarization pipeline.

D. Automatic Speech Recognition (ASR) Evaluation Systems

Arriaga et al. [4] assessed end-to-end Automatic Speech Recognition (ASR) models operating under real-time conditions. Their study concentrates on improving transcription accuracy across varying audio quality levels through the application of advanced ASR architectures. While this work establishes a strong benchmark for transcription performance, it does not extend to higher-level capabilities such as summarization, document generation, or workflow automation. AutoNote AI builds upon these ASR foundations by integrating Whisper v3 and Google Speech-to-Text to deliver robust real-time transcription.

E. Transfer Learning for Natural Language Processing

Raffel et al. [5] proposed the T5 (Text-to-Text Transfer Transformer) model, establishing that a unified text-to-text learning framework can achieve leading performance across a broad range of NLP tasks, including summarization and question answering. This foundational contribution underpins the summarization capabilities of AutoNote AI, which leverages transformer-based architectures such as BERT to generate concise, context-aware meeting summaries. Insights regarding privacy considerations and continual fine-tuning strategies from this line of research also inform AutoNote AI's planned model adaptation roadmap.

III. Comparative Analysis of Existing Methods

Several important observations emerge from reviewing the surveyed works. First, no existing study presents a fully integrated, end-to-end pipeline that simultaneously encompasses real-time transcription, intelligent summarization, structured PDF generation, and automated email delivery within a single unified system. Second, the highest-performing systems rely on cloud-based infrastructure and transmit raw audio, thereby introducing data privacy concerns. Third, although individual components such as ASR and summarization have been extensively investigated, their smooth integration with document generation and distribution remains an unresolved engineering challenge. Fourth, none of the surveyed works provides domain adaptation mechanisms suited to real-world noisy, multi-speaker environments. AutoNote AI is explicitly designed to close all of these gaps.

Table I summarizes the five surveyed works across dimensions most pertinent to AutoNote AI: the task or documentation type addressed, the core methodology employed, the deployment platform, reported processing speed, accuracy metrics, and whether a complete end-to-end pipeline is provided.

IV. RESEARCH GAP

The surveyed literature reveals five persistent gaps that collectively justify the development of AutoNote AI.

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

TABLE I
COMPARATIVE SUMMARY OF RELATED WORK VS. AUTONOTE AI (PROPOSED)

Work	Task / Domain	Method	Platform	Speed	Accuracy	End-to-End Pipeline
Schneider et al. [1]	Real-time meeting summarization	Incremental summarization with evaluation metrics	Cloud server	Low latency (incremental)	Not reported	No (summarization only)
Rakshitha et al. [2]	Transcription + summarization + contextual insights	STT + summarization pipeline	Cloud / server	Near real-time	Not explicitly reported	Partial (no PDF / email)
Chen et al. [3]	Meeting transcription + summarization (multimodal)	Multimodal ML (audio, text, visual)	GPU server	Not reported	Improved via multimodal data	No (dataset only)
Arriaga et al. [4]	Automatic Speech Recognition	End-to-end ASR models	Cloud / edge	Real-time	High (clean audio)	No (ASR only)
Raffel et al. [5]	General NLP (survey / model)	T5 Transfer Learning (text-to-text)	GPU server	N/A	State-of-the-art on NLP benchmarks	No (model only)
AutoNote AI (Proposed)	**Transcription, summarization, PDF report, email delivery**	**Whisper STT + BERT NLP + FastAPI + ReportLab + SMTP**	**Android + FastAPI (cloud / edge hybrid)**	**Near real-time**	**>90% (clean audio)**	**Yes (full pipeline)**

A. Absence of a Unified End-to-End Documentation Pipeline

Each work in the survey addresses a specific, isolated component: summarization [1], transcription coupled with summarization [2], dataset construction [3], or ASR evaluation [4]. A practical meeting documentation system must simultaneously manage real-time transcription, intelligent summarization, structured report generation, and automated delivery. No existing work presents a single, jointly optimized pipeline covering all these stages within a deployable application. AutoNote AI is specifically architected to close this gap by unifying all stages into a single modular platform.

B. Lack of Structured Output Generation

Systems such as that proposed by Rakshitha et al. [2] produce raw or minimally formatted text outputs that are difficult to distribute and interpret in professional contexts. The absence of structured PDF report generation forces users to invest substantial additional effort in organizing and disseminating meeting records, thereby negating a significant portion of the productivity gains promised by automation. AutoNote AI resolves this by incorporating ReportLab and iText-based PDF generation directly into its post-meeting workflow.

C. No Automated Distribution Mechanism

None of the surveyed works integrates automated email delivery as a component of the documentation pipeline [1]–[5]. Even when accurate transcripts and summaries are produced, distributing them to stakeholders remains a fully manual task, reducing the practical value of these systems. AutoNote AI eliminates this bottleneck through SMTP-based Gmail API integration, which automatically dispatches formatted reports to all meeting participants upon session completion.

D. Insufficient Robustness in Real-World Conditions

Existing evaluations of ASR and summarization systems are predominantly conducted in controlled, low-noise settings. Real-world meetings, however, routinely involve background noise, overlapping speakers, domain-specific terminology, and fluctuating audio quality. Arriaga et al. [4] acknowledge performance degradation in adverse acoustic conditions, while Rakshitha et al. [2] do not evaluate multi-speaker diarization in sufficient depth. No surveyed work provides domain-adaptation mechanisms for these environmental variabilities. AutoNote AI addresses this through integrated noise reduction (WebRTC Noise Suppression / RNNoise), Voice Activity Detection (Silero VAD), and speaker diarization (pyannote.audio).

V. SYSTEM DESIGN AND ARCHITECTURE

A. Architectural Overview

AutoNote AI adopts a modular, layered architecture that integrates a mobile application, a backend processing server, an AI-driven text analysis layer, a database, and an email delivery service. This design prioritizes real-time performance, horizontal scalability, and a smooth end-user experience. The system is organized around four principal components:

1) An Android-based user interface for meeting interaction and audio capture
2) A FastAPI backend server for data processing and inter-component coordination
3) An AI model layer responsible for transcription, semantic analysis, and summarization
4) A database and email service layer for persistent data storage and automated report distribution

B. User Interface Layer (Android Application)

The client-side frontend is implemented as an Android mobile application built with Kotlin, serving as the primary point of interaction for end users. The application supports secure user authentication via Google OAuth 2.0 and allows users to create or join meetings with minimal friction. During a live

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

session, the application continuously captures audio streams through WebSocket (WebRTC / LiveKit) and transmits them to the backend server for processing. Real-time transcriptions and summaries are rendered dynamically within the interface, keeping participants informed of the meeting record as it develops.

C. Backend Layer (FastAPI Server)

The server-side backend is developed using FastAPI with Python 3.11, functioning as the central coordination unit of the system. It receives incoming transcription data from the mobile client, manages API request routing, and orchestrates communication between all system components. FastAPI is selected for its asynchronous request handling, high throughput, and native scalability — properties that make it well-suited to real-time, latency-sensitive applications.

D. AI Model Layer

The AI processing layer employs transformer-based Natural Language Processing architectures, including BERT and LegalBERT, to perform context-aware analysis of transcribed content. This layer is responsible for identifying key discussion points and decisions, extracting actionable items, and producing concise, structured meeting summaries. The speech-to-text module within this layer leverages Whisper v3 and Google Speech-to-Text to deliver accurate, continuous transcription even in multi-speaker environments with overlapping audio.

E. Technology Stack

Table II presents the complete technology stack employed across all layers of the AutoNote AI system.

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
Communicatio	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

VI. CONCLUSION

This paper presented AutoNote AI, an intelligent system for automated meeting documentation. Existing studies confirm that real-time transcription and summarization are both feasible and effective. However, most existing solutions lack integrated PDF generation and automated sharing capabilities. AutoNote AI addresses these limitations through a unified end-to-end pipeline. The system combines speech-to-text conversion, summarization, PDF generation, and email delivery. Additional features such as noise reduction, voice activity detection, and speaker diarization improve robustness in practical environments. The proposed approach minimizes manual effort and enhances documentation efficiency. Privacy concerns associated with cloud-based processing are recognized and motivate future on-device deployment. Future work will focus on multilingual support and integration with platforms such as Zoom, Google Meet, and Microsoft Teams. Overall, AutoNote AI represents a significant step toward smarter and more efficient meeting management.

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] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey, and I. Sutskever, “Robust Speech Recognition via Large-Scale Weak Supervision,” in *Proc. Int. Conf. Machine Learning (ICML)*, vol. 202, 2023, pp. 28492–28518.
[7] T. J. Park, N. Kanda, D. Dimitriadis, K. J. Han, S. Watanabe, and S. Narayanan, “A Review of Speaker Diarization: Recent Advances with Deep Learning,” *Computer Speech & Language*, vol. 72, p. 101317, Mar. 2022.
[8] S. Gandhi, P. von Platen, and A. Rush, “Distil-Whisper: Robust Knowledge Distillation via Large-Scale Pseudo Labelling,” *arXiv preprint arXiv:2311.00430*, 2023.
[9] A. Hard et al., “Federated Learning for Mobile Keyboard Prediction,” *arXiv preprint arXiv:1811.03604*, 2018.
[10] J. Zhang, Y. Cao, T. Chen, F. Li, and C. Lin, “An Abstractive Meeting Summarization System Based on Large Language Models,” in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, 2023, pp. 1–12.