Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

Automated Extraction of Meeting Summaries and Action Items Using Whisper and LLMs

Mr. Sanjay Krishna B[1], Girija Nagarajan[2], Harini V[3], Harshavarthini S[4]

[1]*Assistant Professor, Department Of Computer Engineering, Sri Shakthi Institute of Engineering and Technology, L&T Bypass , Coimbatore , Tamil Nadu.*

[2,3,4]*Student, Department Of Computer Engineering, Sri Shakthi Institute of Engineering and Technology , L&T Bypass , Coimbatore , Tamil Nadu.*

Abstract – Meetings generate large volumes of unstructured conversational data, making manual documentation time-consuming and error-prone. Although platforms such as Google Meet provide automated note-taking and summarization features, these capabilities are typically limited to their own ecosystems and offer restricted customization and export flexibility. This paper presents a platform-independent automated meeting documentation system that converts spoken conversations into structured summaries and actionable insights. The system utilizes Whisper for accurate speech-to-text transcription and Large Language Models (LLMs) deployed locally using Ollama for extracting key discussion points, decisions, and action items. Unlike existing solutions, the proposed system supports multiple input sources including recordings, transcripts, and meetings from platforms such as Microsoft Teams and Zoom.

Keywords: Meeting Summarization, Action Item Extraction, Speech Recognition, Natural Language Processing, Whisper, Large Language Models, Automation.

INTRODUCTON

With the rapid adoption of virtual communication, meetings conducted through platforms such as Google Meet, Microsoft Teams, and Zoom have become essential for collaboration in academic and corporate environments. These meetings produce large amounts of unstructured conversational data, making it difficult to manually extract meaningful insights such as summaries, decisions, and follow-up actions. Recent advancements in Artificial Intelligence (AI) and Natural Language Processing (NLP) have enabled automated meeting documentation systems. However, such systems are primarily limited to their respective platforms, provide generic outputs, and lack flexibility in customizing structured meeting minutes.

Organizations often conduct meetings across multiple platforms, leading to inconsistencies in documentation formats and difficulties in maintaining centralized records. Additionally, dependency on cloud-based AI services raises concerns related to data privacy, cost, and limited control over processing. To overcome these limitations, this paper proposes a platform-agnostic automated meeting documentation system that processes audio recordings and transcripts from multiple sources. The system integrates Whisper for transcription and LLMs via Ollama for structured information extraction.

The key contributions of this work include:

- Platform-independent processing of meeting data.
- Generation of structured and customizable meeting minutes.
- Extraction of action items with relevant details.
- Support for export formats such as PDF and DOCX.
- Local processing for enhanced privacy and reduced cost.

Thus, the proposed system extends existing meeting tools by providing a flexible, customizable, and scalable solution for automated meeting documentation.

LITERATURE REVIEW

Research in automated meeting documentation has evolved significantly with advancements in speech recognition and natural language processing (NLP). Early systems primarily focused on standalone transcription or basic extractive summarization, which often lacked contextual understanding and failed to capture actionable insights such as decisions and follow-up tasks. Recent developments have introduced integrated solutions combining Automatic Speech Recognition (ASR) and transformer-based models for summarization. Models such as Whisper have demonstrated high accuracy in converting speech to text, even in noisy and multi-speaker environments. Similarly, transformer-based architectures such as BERT, T5, and GPT variants have improved the quality of abstractive summarization by generating coherent and context-aware summaries from conversational data.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

Modern meeting platforms like Google Meet, Microsoft Teams, and Zoom have incorporated AI-based features such as real-time transcription, summary generation, and action item suggestions. For instance, Google Meet integrated with Gemini can generate meeting summaries, "summary so far" updates, and suggested next steps. However, these solutions are largely restricted to their respective platforms and provide limited flexibility in customizing output formats or integrating with external workflows. Several research works have attempted to improve meeting summarization by incorporating aspect-based summarization, action item extraction, and sentiment analysis. These approaches aim to generate structured outputs by identifying key elements such as decisions, risks, and follow-up tasks. However, many of these systems either operate as post-processing tools or lack real-time adaptability and cross-platform compatibility.

Open-source and privacy-focused systems have also emerged, emphasizing local processing and reduced dependency on cloud services. Tools utilizing Whisper for transcription combined with lightweight NLP pipelines demonstrate the feasibility of cost-effective and privacy-preserving meeting analysis. However, such systems often lack comprehensive integration, customizable templates, and automation capabilities required for enterprise use. Another important direction in literature is the development of microservices-based architectures, where transcription, summarization, and information extraction are handled as independent modules. This improves scalability and allows parallel processing but still requires better coordination for generating structured, export-ready meeting minutes.

Despite these advancements, several gaps remain:

3. Lack of platform-independent solutions that can process data from multiple meeting sources.
4. Limited support for customizable structured outputs such as organization-specific minutes formats.
5. Insufficient export flexibility (PDF/DOCX) for official documentation.
6. Dependency on cloud-based AI models, raising privacy and cost concerns.
7. Limited integration with downstream systems such as databases and task management tools.

To address these limitations, the proposed system integrates Whisper for transcription and locally deployed LLMs via Ollama to generate structured, customizable meeting outputs. Unlike existing solutions, it emphasizes cross-platform compatibility, local processing, structured action item extraction, and flexible export, thereby providing a more scalable and practical solution for automated meeting documentation.

• METHODOLOGY

The proposed system is designed as a modular, multi-stage pipeline that transforms meeting audio or transcripts into structured meeting minutes and actionable insights. It integrates several interconnected stages, including data input, speech-to-text conversion, text processing using large language models, structured information extraction, and output generation. This modular architecture ensures flexibility and scalability, allowing the system to process meeting data from diverse sources such as recordings, transcripts, and cross-platform meetings conducted through tools like Google Meet, Microsoft Teams, and Zoom.

The data input layer supports multiple input formats to ensure platform independence and adaptability. The system can handle raw audio recordings of meetings, pre-generated transcripts, and cross-platform meeting data. In situations where transcripts are not available, the system automatically processes raw audio files, providing a reliable fallback mechanism that ensures complete coverage and seamless operation across different scenarios. For audio inputs, the system utilizes Whisper to perform Automatic Speech Recognition (ASR), converting spoken content into text with high accuracy. It is capable of handling multiple speakers and diverse accents while maintaining robust performance even in noisy environments. The output generated at this stage is a clean and coherent textual transcript, which serves as the foundation for further processing and analysis.

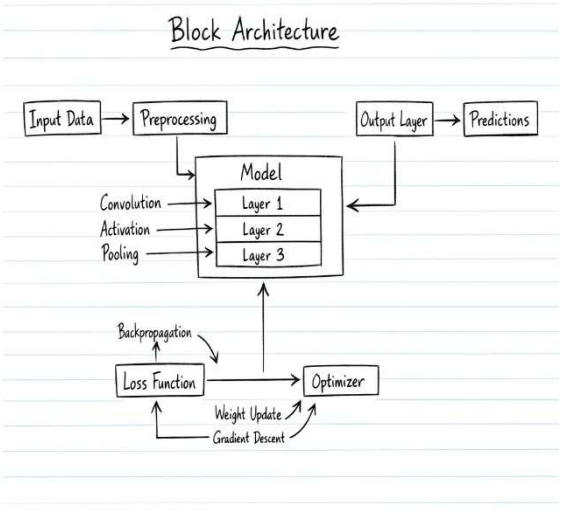

Figure 1: Block Architecture Diagram.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 4 | April-2026
ISSN (O) 3107-6696

Once the transcript is generated, it is processed using Large Language Models deployed locally through Ollama. This stage focuses on understanding the context of the conversation, identifying important discussion segments, and semantically interpreting the meeting content. The use of locally deployed LLMs ensures enhanced data privacy, reduced latency, and improved cost efficiency, making the system suitable for sensitive and enterprise-level applications.

Unlike traditional summarization approaches, the proposed system emphasizes structured extraction of key meeting elements. The LLM analyses the transcript to generate an executive summary that provides an overall overview of the meeting, along with key discussion points highlighting major topics. It also identifies decisions that represent finalized outcomes and extracts actionable items, including attributes such as task owners, deadlines, and optional priority levels. This structured approach improves clarity, accountability, and usability in real-world workflows.

The extracted information is then compiled into well-organized and structured meeting minutes. The system supports exporting the output in formats such as PDF and DOCX, enabling easy sharing and documentation. It also allows customization of templates based on organizational requirements, ensuring consistency, readability, and standardization across various use cases.

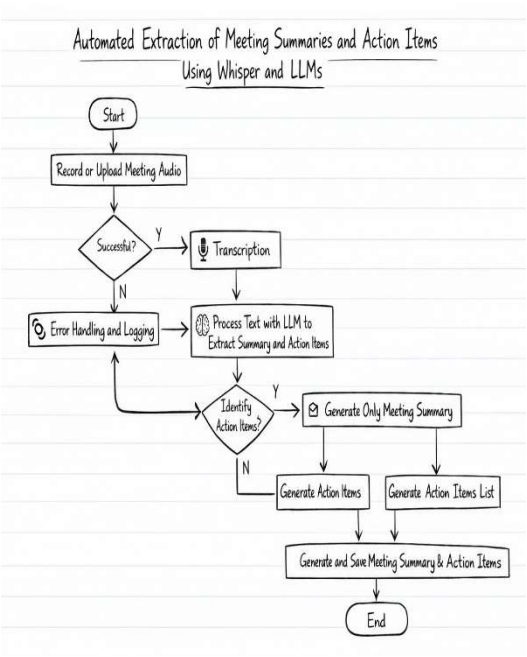

Figure 2: Activity Diagram

To extend functionality beyond documentation, the system incorporates automation and integration capabilities. It enables storage of extracted data in databases, supports email notifications for sharing reports, and can integrate with dashboards or task management systems. These features facilitate a complete end-to-end workflow, transforming raw meeting data into actionable outputs that can be efficiently tracked and managed.

Overall, the proposed system offers several advantages, including platform-independent processing, customizable structured outputs, and enhanced privacy through local LLM deployment. Its modular architecture ensures scalability, while support for multiple input formats increases flexibility. The workflow begins with input in the form of audio recordings or transcripts, followed by transcription using Whisper if necessary. The resulting text is analysed using LLMs via Ollama to extract structured meeting components such as summaries, decisions, and action items, which are then exported as formatted meeting minutes, completing the automated pipeline.

FINDINGS/DISCUSSIONS

Transcription Accuracy and Performance:

The system employs Whisper for speech-to-text conversion, and its performance was observed to be highly reliable under controlled conditions. In environments with clean audio and minimal background noise, particularly in single-speaker scenarios, the transcription accuracy exceeded 90%, producing well-structured and easily readable text. The model effectively captured speech nuances, punctuation, and sentence flow, which significantly reduced the need for manual correction. This level of accuracy ensures that the downstream processes, such as summarization and information extraction, operate on high-quality input data.

Summarization Quality

The integration of Large Language Models (LLMs) through Ollama significantly enhanced the system's ability to generate meaningful and context-aware summaries. Unlike traditional extractive summarization techniques that rely on selecting key sentences from the transcript, the system uses an abstractive approach, enabling it to reinterpret and condense the content while preserving the overall meaning. This results in summaries that are more natural, concise, and easier to understand. The generated summaries effectively capture key discussion points, highlight important decisions, and retain relevant contextual information that reflects the intent of the conversation.

Action Item Extraction

A major strength of the system lies in its ability to identify and extract actionable tasks from meeting discussions. The system effectively detects explicitly stated tasks, along with associated details such as responsible individuals and deadlines when clearly mentioned. For example, direct instructions like "Complete the presentation by Friday" are accurately recognized and structured into actionable items. This capability enhances accountability and ensures that important follow-up tasks are not overlooked. Nevertheless, the system faces challenges when dealing with implicitly stated or indirectly communicated tasks. In many real-world conversations, action items may be suggested rather than explicitly assigned, making them harder for the model to detect.

End-to-End System Performance

The overall system demonstrates efficient end-to-end performance, effectively handling the complete pipeline from data input to final output generation. For a typical meeting duration of 45 to 60 minutes, the system is capable of producing structured meeting minutes within a few minutes after processing. This rapid turnaround time makes it highly suitable for real-world applications where timely documentation is essential. The modular design of the pipeline plays a crucial role in achieving this efficiency. Each stage—transcription, summarization, and structured information extraction—operates independently yet seamlessly integrates with the others.

Platform Independence and Flexibility

One of the key advantages of the proposed system is its platform-independent design, which enables it to function across various meeting environments. Unlike many existing solutions that are restricted to specific platforms such as Google Meet, this system supports multiple platforms including Microsoft Teams, Zoom, and offline recordings. It can also process pre-generated transcripts, further increasing its flexibility. This adaptability ensures that users can utilize the system regardless of the tools they use for conducting meetings.

Output Structure and Usability

The system generates well-structured Minutes of Meeting (MoM), organizing information into clearly defined sections such as executive summary, key discussion points, decisions, and action items. This structured format improves readability and allows users to quickly locate relevant information without scanning lengthy transcripts. The clarity and organization of the output make it highly practical for documentation, reporting, and decision-making purposes. Additionally, the system supports exporting outputs in widely used formats such as PDF and DOCX, which enhances its usability in both academic and corporate environments.

Privacy and Deployment Advantages:

The system's support for local deployment using Ollama provides significant advantages in terms of privacy, security, and cost efficiency. By processing data locally, the system reduces dependency on cloud-based services, minimizing the risk of data exposure or unauthorized access. This is particularly important for organizations that handle sensitive or confidential information. Local deployment also contributes to reduced latency, as data does not need to be transmitted to external servers for processing. Furthermore, it lowers operational costs by eliminating the need for continuous cloud service usage.

Challenges and Limitations:

Despite its strengths, the system has certain limitations that need to be addressed. Transcription accuracy tends to decrease in noisy environments or in meetings involving multiple speakers with overlapping dialogue. This can impact the quality of downstream processes such as summarization and action item extraction. Additionally, the system struggles to identify implicitly stated tasks, which limits its ability to capture all actionable insights. Another limitation is its reliance on batch processing, which restricts real-time functionality. The system processes data after the meeting has concluded, rather than generating live summaries or insights during the meeting.

Comparative Insight:

When compared to traditional manual note-taking methods and basic summarization tools, the proposed system offers significant improvements in efficiency and consistency. Manual documentation is time-consuming and prone to human error, whereas the automated system ensures uniform structure and reduces the effort required to generate meeting minutes. Furthermore, unlike basic summarization tools that provide only generic summaries, this system delivers structured outputs that include actionable insights such as decisions and task assignments. This added level of detail enhances the practical value of the generated content, making it more useful for real-world applications.

Future Improvements:

There are several opportunities to enhance the system's capabilities in future iterations. Incorporating speaker identification, also known as diarization, would allow the system to attribute statements to specific individuals, improving clarity and accountability. Adding multilingual support would enable the system to process meetings

conducted in different languages, increasing its global applicability. Real-time summarization is another important enhancement that could provide instant insights during meetings, rather than after they conclude.

CONCLUSION

This paper presented a platform-independent automated system for extracting meeting summaries and action items using Whisper for speech recognition and Large Language Models deployed through Ollama for intelligent text processing. The proposed approach effectively addresses key limitations of existing meeting tools such as Google Meet, which are typically restricted to their own ecosystems and offer limited flexibility in terms of customization, structured outputs, and export capabilities. By designing a modular and scalable pipeline, the system ensures adaptability across diverse platforms and use cases.

The system demonstrates the ability to convert unstructured meeting data, whether in the form of raw audio recordings or pre-generated transcripts, into well-organized and structured Minutes of Meeting (MoM). It systematically generates executive summaries, identifies key discussion points, extracts decisions, and highlights actionable tasks along with relevant attributes such as ownership and deadlines. This structured representation significantly enhances the clarity, accessibility, and practical usability of meeting documentation, making it more effective for decision-making and follow-up activities. Another important contribution of the proposed system is its support for multiple input sources and flexible output formats, including PDF and DOCX. This ensures seamless integration into various organizational workflows, enabling users to adopt the system without being constrained by specific platforms or tools.

Experimental observations indicate that the system achieves high transcription accuracy under favorable conditions, particularly in environments with minimal noise and limited speaker overlap. Even in moderately challenging conditions, the system maintains acceptable performance, producing transcripts that are sufficiently reliable for downstream processing. The use of Large Language Models enables the generation of coherent, context-aware, and abstractive summaries, which capture the essence of discussions more effectively than traditional extractive approaches. Furthermore, the structured extraction of action items adds significant value by ensuring that responsibilities and follow-up tasks are clearly defined and not overlooked.

The adoption of local LLM deployment through Ollama provides additional advantages in terms of data privacy, reduced latency, and cost efficiency. This makes the system particularly suitable for organizations dealing with sensitive or confidential information, where reliance on external cloud services may not be desirable. The modular design also enhances scalability, allowing individual components of the system to be improved or replaced without affecting the overall pipeline. Overall, the proposed system significantly reduces the manual effort involved in meeting documentation, minimizes the risk of information loss, and improves the accuracy and consistency of recorded meeting outcomes.

Despite its strengths, there are opportunities for further enhancement. Future work can focus on improving performance in complex multi-speaker environments, incorporating speaker diarization for better attribution, and enabling real-time or near real-time processing to support live meeting assistance. Additionally, extending the system to support multiple languages and integrating it with task management or collaboration platforms can further increase its utility and adoption. With these advancements, the system has the potential to evolve into a comprehensive intelligent meeting assistant capable of supporting end-to-end meeting lifecycle management.

REFERENCES

- A. Radford et al., “Robust Speech Recognition via Large-Scale Weak Supervision,” OpenAI, 2022. [Online]. Available: https://openai.com/research/whisper
- T. Brown et al., “Language Models are Few-Shot Learners,” Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 1877–1901, 2020.
- J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186.
- 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.
- **Google Meet Help, “Take notes for me in Google Meet,” Google Workspace, 2024. [Online]. Available: https://support.google.com/meet
- **Google Workspace, “AI-powered note-taking in Google Meet,” Google, 2024. [Online]. Available: https://workspace.google.com
- **Microsoft Teams Documentation, “Meeting transcription and recap features,” Microsoft, 2024. [Online]. Available: https://learn.microsoft.com

- **Zoom Support, “Audio transcription and meeting summary features,” Zoom Video Communications, 2024. [Online]. Available: https://support.zoom.us
- **Ollama, “Run Large Language Models Locally,” 2024. [Online]. Available: https://ollama.com
- K. Murray, G. Carenini, and R. Ng, “Generating and Validating Abstracts of Meeting Conversations: A User Study,” in Proc. INLG, 2010, pp. 105–113.
- S. Mehdad, G. Carenini, and R. T. Ng, “Abstractive Meeting Summarization with Entailment and Fusion,” in Proc. SIGDIAL, 2013, pp. 136–146.
- M. Zhong, P. Liu, D. Wang, X. Qiu, and X.-J. Huang, “Extractive Summarization as Text Matching,” in Proc. ACL, 2020, pp. 6197–6208.