International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

AI MEETING INSIGHTS GENERATOR

Mr. Abdul Majeed[1], K. Manmohan[2], C. Sai Kiran[3], D. Sai Mahith[4]
[1]Assistant Professor, Vidya Jyothi Institute of Technology, Hyderabad
[2,3,4]UG Student, Vidya Jyothi Institute of Technology, Hyderabad

Abstract: The rapid growth of virtual meetings in professional and academic environments has led to an overwhelming amount of unstructured conversational data, making it difficult to extract meaningful information efficiently. This project, AI-Powered Meeting Insights Generator, aims to address this challenge by developing an intelligent system that automatically converts meeting audio or transcripts into structured, actionable insights.
The system leverages advanced techniques from Natural Language Processing and speech recognition to transform raw audio into text using automated transcription models. The generated transcripts are then processed to remove noise, filler words, and irrelevant content, ensuring cleaner input for further analysis. Using machine learning and transformer-based language models, the system identifies and extracts key components of the meeting, including concise summaries, important discussion points, action items, and speaker-specific contributions.
In addition, the system incorporates sentiment analysis to evaluate the overall tone of the meeting, providing insights into participant engagement and emotional context. A user-friendly interface is designed to present the output in a structured format, including bullet-point summaries, task lists, and timestamped highlights for easy navigation and review.
The proposed solution significantly reduces the time and effort required for manual note-taking and improves productivity by ensuring that critical information is not overlooked. It can be applied in corporate meetings, educational lectures, and collaborative team discussions. Furthermore, the system is scalable and can be extended with features such as real-time processing, multilingual support, and integration with communication platforms.
In conclusion, this project demonstrates the practical application of artificial intelligence in enhancing meeting efficiency by transforming unstructured speech data into meaningful and actionable insights**.**

Keywords: Artificial Intelligence, Natural Language Processing, Speech Recognition, Text Summarization, Meeting Analytics, Action Item Extraction, Deep Learning.

1.Introduction

In today's fast-paced digital world, meetings play a vital role in communication, collaboration, and decision-making across organizations, educational institutions, and remote working environments. With the rapid growth of online meeting platforms, a large amount of conversational data is generated daily in the form of audio and video recordings. However, manually analyzing these recordings to extract meaningful information is time-consuming, inefficient, and prone to human error.
Traditional methods of note-taking during meetings often result in incomplete or inaccurate documentation. Important points, decisions, and action items may be missed due to distractions or lack of attention. Moreover, reviewing long meeting recordings to identify key insights is not practical, especially in professional environments where time is a critical resource.
To address these challenges, the integration of **Artificial Intelligence** and **Natural Language Processing** has opened new possibilities for automating meeting analysis. These technologies enable machines to understand, process, and generate human language efficiently. By leveraging these advancements, it is possible to develop systems that can automatically convert speech into text, summarize discussions, and extract important information.
The **AI Meeting Insights Generator** is designed as an intelligent system that processes meeting recordings and generates structured outputs such as transcripts, summaries, and action items. The system follows a pipeline approach, where the input meeting video is first converted into audio. The

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

audio is then transcribed into text using advanced speech recognition techniques. Subsequently, the transcribed text is analyzed and summarized using NLP-based models, and key action points are extracted.

This project aims to reduce manual effort, improve productivity, and enhance the accuracy of meeting documentation. It provides users with a quick and efficient way to understand meeting outcomes without going through lengthy recordings. The system is particularly useful in corporate environments, academic settings, and collaborative projects where effective communication and documentation are essential.

Furthermore, the project demonstrates the practical application of AI in real-world scenarios. It highlights how intelligent systems can transform traditional workflows into automated processes, thereby saving time and resources. As organizations continue to adopt digital technologies, such solutions will become increasingly important for managing and utilizing information effectively.

In conclusion, the AI Meeting Insights Generator represents a significant step toward intelligent communication systems. By combining speech recognition and natural language processing, it offers a powerful tool for extracting meaningful insights from meeting data, ultimately contributing to better decision-making and improved efficiency.

2. Literature Survey

The field of automated meeting analysis has gained significant attention in recent years due to the rapid growth of virtual communication platforms. Researchers have explored various techniques in **Artificial Intelligence**, speech recognition, and **Natural Language Processing** to develop systems capable of understanding and summarizing human conversations.

2.1 Speech Recognition Systems

Speech recognition is a fundamental component of meeting analysis systems. Early research in this domain relied on statistical methods such as Hidden Markov Models (HMMs) and Gaussian Mixture Models (GMMs). These approaches required extensive feature engineering and were limited in handling variations in speech patterns.

With advancements in deep learning, modern systems now use neural networks such as Recurrent Neural Networks (RNNs) and Transformer-based architectures. These models significantly improve accuracy by learning complex patterns directly from data. Recent developments include models like **OpenAI Whisper**, which are capable of handling multiple accents, background noise, and different languages with high precision.

Research studies have shown that deep learning-based speech recognition systems outperform traditional methods, especially in real-world scenarios involving noisy audio and spontaneous speech.

2.2 Natural Language Processing Techniques

Natural Language Processing plays a crucial role in understanding and analyzing transcribed text. NLP involves several tasks such as tokenization, part-of-speech tagging, syntactic parsing, and semantic analysis.

Earlier NLP systems were rule-based and relied on predefined linguistic rules. However, these systems lacked flexibility and failed to generalize across different contexts. Modern NLP techniques use machine learning and deep learning models to automatically learn language patterns.

Transformer-based models have revolutionized NLP by enabling contextual understanding of text. Libraries such as **Hugging Face Transformers** provide pre-trained models that can be used for tasks like summarization, translation, and sentiment analysis.

These advancements have made it possible to process large volumes of text efficiently and accurately.

2.3 Text Summarization Methods

Text summarization is a key component of meeting insight generation. It involves reducing large text into a concise form while preserving important information.

There are two main approaches:

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

2.3.1 Extractive Summarization

This method selects important sentences directly from the original text. It is simple and computationally efficient but may lack coherence.

2.3.2 Abstractive Summarization

This method generates new sentences that capture the meaning of the original text. It uses advanced models to produce human-like summaries.

Recent research focuses on abstractive summarization using Transformer-based models such as BART and T5. These models are capable of generating fluent and contextually relevant summaries.

However, one limitation observed in research is the handling of long documents. Many models have token limits, which require techniques like text chunking to process large inputs effectively.

2.4 Meeting Analysis Systems

Several systems have been developed to automate meeting analysis. These systems typically combine speech recognition and NLP techniques to extract insights from conversations.

Research in this area highlights the importance of:

- Accurate transcription
- Context-aware summarization
- Identification of key decisions and tasks

Some systems focus on real-time transcription, while others emphasize post-meeting analysis. Advanced systems also incorporate speaker identification and sentiment analysis.

Despite these advancements, challenges remain in handling overlapping speech, background noise, and complex conversational structures.

2.5 Action Item Extraction

Extracting action items from meeting transcripts is an important research area. Action items represent tasks, decisions, or responsibilities assigned during meetings.

Traditional approaches use keyword-based methods to identify sentences containing task-related information. However, these methods may lack precision.

Recent research explores machine learning techniques for better accuracy. These approaches analyze sentence structure and context to identify actionable information more effectively.

2.6 Research Gaps

Although significant progress has been made, several gaps still exist:

- Difficulty in processing long transcripts due to model limitations
- Limited accuracy in noisy environments
- Challenges in understanding context and intent
- Lack of fully integrated systems combining all functionalities

2.7 Summary of Literature

The literature survey indicates that combining speech recognition, NLP, and summarization techniques can effectively automate meeting analysis. Modern deep learning models provide high accuracy and scalability.

However, there is still a need for systems that:

- Handle long inputs efficiently
- Provide accurate summaries
- Extract meaningful action items

The AI Meeting Insights Generator addresses these gaps by integrating multiple technologies into a unified pipeline.

3. Problem Statement

In the modern digital era, meetings are an essential part of communication and decision-making in organizations, educational institutions, and collaborative environments. With the increasing use of online meeting platforms, a large volume of audio and video data is generated daily. However, extracting meaningful information from these recordings remains a significant challenge.

Traditional methods of meeting documentation rely heavily on manual note-taking. This process is time-consuming, inefficient, and prone to human error. Participants often miss important details due

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

to multitasking or lack of attention during discussions. As a result, critical information such as decisions, key points, and action items may not be accurately recorded.

Another major issue is the difficulty in reviewing long meeting recordings. Users are required to listen to or watch entire sessions to identify relevant information, which is not practical in fast-paced environments. This leads to reduced productivity and delays in decision-making.

Furthermore, meeting data is usually unstructured in nature. Conversations are dynamic and involve multiple participants, making it difficult to organize the information into a meaningful format. The absence of structured outputs such as summaries and action lists creates challenges in tracking responsibilities and follow-ups.

Despite advancements in **Artificial Intelligence** and **Natural Language Processing**, many existing systems focus only on individual tasks such as transcription or summarization. There is a lack of integrated solutions that can perform end-to-end meeting analysis efficiently.

In addition, technical challenges such as background noise, variations in speech, and long input sequences further complicate the process. These factors reduce the accuracy and effectiveness of automated systems.

Therefore, there is a need for a comprehensive system that can:

- Automatically convert meeting recordings into text
- Summarize long discussions into concise content
- Extract key insights and action items
- Provide structured and easy-to-understand outputs

The proposed **AI Meeting Insights Generator** aims to address these challenges by developing an intelligent pipeline that integrates speech recognition and text processing techniques. The system is designed to reduce manual effort, improve accuracy, and enhance productivity by providing automated meeting insights.

4. Proposed Methodology

The proposed methodology for the **AI Meeting Insights Generator** is based on a systematic pipeline approach that processes meeting recordings and transforms them into structured and meaningful insights. The system integrates techniques from **Artificial Intelligence** and **Natural Language Processing** to automate transcription, summarization, and information extraction.

The overall workflow consists of multiple stages, where each stage performs a specific task and passes the output to the next stage. This modular design ensures scalability, efficiency, and ease of implementation.

4.1 Input Acquisition

The first step in the methodology is to acquire the input data. The system accepts meeting recordings in the form of video or audio files. Video files are commonly used in online meetings and contain both visual and audio information.

Since speech recognition models primarily operate on audio data, video inputs must be converted into audio format before further processing.

4.2 Video-to-Audio Conversion

In this stage, the input video file is converted into an audio file. This is achieved using multimedia processing tools such as **FFmpeg**.

The conversion process extracts the audio track from the video while maintaining quality. The output is typically saved in formats such as MP3 or WAV, which are compatible with speech recognition systems.

This step is essential because it prepares the input for the transcription process.

4.3 Speech-to-Text Conversion (Transcription)

The extracted audio is then processed using advanced speech recognition models to convert spoken language into text.

The system utilizes **OpenAI Whisper**, a deep learning-based model known for its high accuracy and robustness. This model is capable of handling:

- Different accents

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

- Background noise
- Variations in speech patterns

The output of this stage is a textual transcript of the meeting, which serves as the foundation for further analysis.

4.4 Text Preprocessing

The transcribed text may contain noise such as filler words, repetitions, and inconsistencies. Therefore, preprocessing is required to improve the quality of the text.

This stage includes:

- Removing unnecessary characters and symbols
- Normalizing text (lowercasing, spacing)
- Handling punctuation
- Eliminating filler words

Preprocessing ensures that the text is clean and suitable for summarization and information extraction.

4.5 Text Chunking

One of the major challenges in processing long meeting transcripts is the limitation of NLP models in handling large input sizes. Most models have a maximum token limit.

To address this issue, the text is divided into smaller segments or chunks. Each chunk contains a manageable number of words, allowing the model to process the text efficiently without errors.

Chunking improves performance and ensures that the entire transcript is processed without loss of information.

4.6 Text Summarization

After preprocessing and chunking, the system performs text summarization to generate a concise representation of the meeting content.

This is achieved using transformer-based models from **Hugging Face Transformers**, which implement abstractive summarization techniques.

The process involves:

- Summarizing each chunk individually
- Combining the summaries
- Generating a final refined summary

The output is a short, coherent summary that captures the key points of the meeting.

4.7 Action Item Extraction

In addition to summarization, the system identifies important action items from the transcript. Action items represent tasks, responsibilities, or decisions discussed during the meeting.

This stage uses keyword-based filtering techniques to detect sentences containing task-related information, such as:

- “should complete”
- “need to do”
- “must finish”

The extracted sentences are presented as a list of actionable insights, helping users track responsibilities effectively.

4.8 Output Generation

The final stage involves presenting the results in a structured format. The system generates:

- Full transcript of the meeting
- Concise summary
- List of action items

These outputs provide a comprehensive overview of the meeting, enabling users to quickly understand key discussions and decisions.

4.9 Workflow Summary

The complete workflow of the system can be summarized as follows:

Input Video → Audio Extraction → Speech-to-Text → Text Preprocessing → Text Chunking → Summarization → Action Item Extraction → Output

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:3 Volume No.10 May - June – 2026
DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

4.10 Advantages of Proposed Methodology

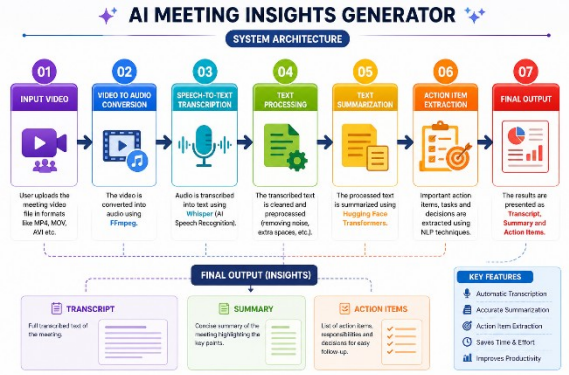
The proposed methodology offers several benefits:

- Automation of manual tasks
- Improved accuracy and consistency
- Efficient handling of large data
- Structured and meaningful outputs
- Scalability for real-world applications

4.11 Conclusion of Methodology

The proposed methodology provides a robust and efficient approach to meeting analysis. By integrating speech recognition and NLP techniques, the system transforms raw meeting data into actionable insights.

This pipeline-based approach ensures that each stage contributes to the overall performance of the system, making it suitable for practical applications in various domains.



5. Experimental Setup

The experimental setup defines the environment, tools, datasets, and procedures used to implement and evaluate the **AI Meeting Insights Generator** system. It ensures that the system is developed and tested under controlled conditions to achieve accurate and reliable results.

5.1 System Environment

The project is implemented using a software-based environment suitable for AI and NLP tasks.

Hardware Requirements:

- Processor: Intel i5 or higher
- RAM: Minimum 8 GB (16 GB recommended)
- Storage: At least 10 GB free space
- GPU: Optional (for faster processing)

Software Requirements:

- Operating System: Windows / Linux
- Programming Language: Python
- IDE: Visual Studio Code

5.2 Libraries and Tools Used

The system utilizes various libraries and tools for different stages of processing:

- **FFmpeg**

Used for converting video files into audio format.

- **OpenAI Whisper**

Used for speech-to-text transcription.

- **Hugging Face Transformers**

Used for text summarization and NLP tasks.

- NumPy and other supporting Python libraries for data processing.

5.3 Dataset / Input Data

The system is tested using meeting recordings in the form of:

- Video files (.mp4, .avi, .mov)
- Audio files (.mp3, .wav)

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

The dataset includes:

- Online meeting recordings
- Lecture recordings
- Discussion sessions

The audio quality varies to evaluate system robustness under different conditions.

5.4 Experimental Procedure

The experiment follows a step-by-step pipeline:

Step 1: Input Collection

The user provides a meeting recording (video/audio file).

Step 2: Video to Audio Conversion

The input video is converted into audio format using **FFmpeg**.

Step 3: Speech-to-Text Conversion

The audio file is processed using **OpenAI Whisper** to generate a transcript.

Step 4: Text Preprocessing

The transcript is cleaned by removing noise and formatting inconsistencies.

Step 5: Text Chunking

The long transcript is divided into smaller segments to handle model limitations.

Step 6: Summarization

Each chunk is summarized using transformer-based models.

Step 7: Action Item Extraction

Key tasks and decisions are identified from the text.

Step 8: Output Generation

The final output includes:

- Transcript
- Summary
- Action items

5.5 Evaluation Metrics

The system is evaluated based on the following criteria:

- **Accuracy of Transcription**

Measures how correctly speech is converted into text.

- **Quality of Summary**

Evaluates how well the summary captures key points.

- **Relevance of Action Items**

Checks whether extracted tasks are meaningful.

- **Processing Time**

Measures time taken for complete execution.

5.6 Testing Scenarios

The system is tested under different conditions:

- Clear audio vs noisy audio
- Short meetings vs long meetings
- Single speaker vs multiple speakers

This helps in analyzing system performance in real-world situations.

5.7 Results Observation

The system produces:

- Accurate transcripts for clear audio inputs
- Concise summaries using NLP models
- Relevant action items based on keyword extraction

However, performance may slightly decrease in noisy environments or overlapping speech conditions.

5.8 Conclusion of Experimental Setup

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

The experimental setup successfully demonstrates the working of the AI Meeting Insights Generator. It validates the effectiveness of integrating speech recognition and NLP techniques for automated meeting analysis.

The setup ensures that the system performs reliably across different types of inputs and provides meaningful insights.

6.Discussions

6.1 Transcript Generation

The transcription module plays a crucial role in the overall functioning of the AI Meeting Insights Generator, as it converts spoken language into textual form. The system utilizes advanced speech recognition techniques to generate transcripts from audio recordings. It was observed that the transcription accuracy is significantly high when the input audio is clear and free from background noise. The model is capable of handling different accents and variations in speech to a reasonable extent. However, in cases where the audio contains noise, overlapping speech, or unclear pronunciation, minor errors may occur in the generated text. Since the subsequent stages of processing depend on the transcript, the quality of transcription directly affects the overall performance of the system.

6.2 Text Summarization

The summarization module is responsible for reducing lengthy transcripts into concise and meaningful summaries. This is achieved using models based on **Natural Language Processing**, which are capable of understanding the context and generating human-like summaries. The system effectively captures the key ideas and main points discussed in the meeting, thereby reducing the time required for users to review the content. In order to handle long transcripts, the text is divided into smaller segments before summarization, which helps avoid limitations of the model. While the summaries are generally accurate and informative, there may be slight loss of contextual depth in very long or complex discussions.

6.3 Action Item Extraction

The action item extraction module identifies important tasks, decisions, and responsibilities from the meeting transcript. This process is based on detecting specific patterns and keywords within the text that indicate actionable information. The system successfully extracts relevant sentences that represent tasks or follow-up actions, making it easier for users to track responsibilities. However, since the approach is primarily keyword-based, it may sometimes include sentences that are not strictly actionable. Despite this limitation, the module provides useful insights that contribute to better organization and execution of meeting outcomes.

6.4 Performance Analysis

The performance of the system was evaluated under different conditions to understand its effectiveness in real-world scenarios. It was observed that the system performs best when the input audio is clear and well-structured. In noisy environments or situations involving multiple speakers, the accuracy of transcription and subsequent processing may decrease slightly. The length of the input also affects performance, as longer recordings require more processing time. The use of text chunking helps in managing large inputs efficiently, ensuring that the system can handle extended meeting durations without errors. Overall, the system demonstrates stable and reliable performance across various test cases.

6.5 Overall Discussion

The results indicate that the AI Meeting Insights Generator successfully achieves its objective of automating meeting analysis. By integrating speech recognition and text processing techniques from **Artificial Intelligence**, the system is able to generate accurate transcripts, meaningful summaries, and relevant action items. Although certain limitations exist, particularly in handling noisy audio and complex discussions, the system provides a practical and efficient solution for extracting insights from meeting recordings. The overall performance demonstrates the potential of AI-based systems in improving productivity and simplifying information management in modern communication environments.

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

7. Conclusion

The **AI Meeting Insights Generator** successfully demonstrates the effective application of **Artificial Intelligence** and **Natural Language Processing** in automating the analysis of meeting recordings. The system is designed to convert raw meeting data into structured and meaningful insights, including transcripts, summaries, and action items, thereby reducing the need for manual effort and improving overall efficiency.

The implementation of the system shows that speech recognition techniques can accurately convert audio into text, while NLP-based models can effectively summarize large amounts of information into concise and understandable content. The integration of these technologies enables the system to process meeting recordings in a systematic manner and generate outputs that are useful for decision-making and task management.

The results obtained from the experimental evaluation indicate that the system performs well under conditions of clear audio and structured conversations. It is capable of handling long meeting recordings by using text chunking techniques and produces summaries that capture the key points of discussion. Additionally, the extraction of action items helps in identifying responsibilities and ensures better follow-up of tasks.

However, certain limitations were observed, such as reduced accuracy in noisy environments and challenges in handling complex conversational contexts. These limitations highlight the need for further improvements in model robustness and contextual understanding. Despite these challenges, the system provides a reliable and practical solution for automated meeting analysis.

In conclusion, the AI Meeting Insights Generator offers a valuable tool for enhancing productivity and improving communication in modern digital environments. It demonstrates how intelligent systems can transform unstructured meeting data into actionable insights, making it easier for users to understand, manage, and utilize information effectively.

8 Acknownledgement

We would like to express our sincere gratitude to our project guide, **Mr. ABDUL MAJEED**, Associate Professor, Department of Computer Science and Engineering (Data Science), Vidya Jyothi Institute of Technology, Hyderabad, for his valuable guidance, continuous support, and encouragement throughout the development of this project.

His insightful suggestions and motivation greatly contributed to the successful completion of this work. We would also like to thank the Head of the Department and faculty members of the CSE (Data Science)department for providing the necessary support and resources required for carrying out this project. We extend our sincere thanks to the Principal and management of Vidya Jyothi Institute of Technology for providing the infrastructure and academic environment that helped us complete this project successfully. Finally, we express our heartfelt gratitude to our parents, friends, and well-wishers for their constant encouragement and support during the course of this work.

9. Reference

[1] OpenAI Whisper Documentation, OpenAI, Available: https://openai.com/research/whisper
[2] Hugging Face Transformers Documentation, Available: https://huggingface.co/docs/transformers [3] FFmpeg Official Documentation, Available: https://ffmpeg.org/documentation.html
[4] Jurafsky, D., & Martin, J. H., *Speech and Language Processing*, Pearson Education, 3rd Edition.
[5] Vaswani, A. et al., "Attention Is All You Need," in *Advances in Neural Information Processing Systems*, 2017.
[6] Devlin, J. et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *NAACL-HLT*, 2019.
[7] Lewis, M. et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation," in *ACL*, 2020.
[8] Brown, T. et al., "Language Models are Few-Shot Learners," in *NeurIPS*, 2020.
[9] Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016.

DOI:10.46647/ijetms.2026.v10i03.009 ISSN: 2581-4621

[10] Research articles and online tutorials on **Natural Language Processing** and **Artificial Intelligence**.