Naan Mudhalvan

Prompt Engineering

# Project – Meeting Transcription Summarisation.


***Submitted by***

**AJOI. T. VARGHESE (212422104003)**

**SHANMUGAVELU R (212422104028)**

**VIVIN RAJ M (212422104033)**

**SREE SASTHA INSTITUTE OF ENGINEERING AND TECHNOLOGY**

**CHEMBARAMBAKKAM CHENNAI-600123**


**ANNA UNIVERSITY CHENNAI-600025**

**APR / MAY 2025**

## ANNA UNIVERSITY:: CHENNAI 600 025

## BONAFIDE CERTIFICATE

Certified that this report titled "**Synthetica: Meeting Transcription Summarisation.**" for the project is a bonafide work of (**AJOI.T.VARGHESE - 212422104003, SHANMUGAVELU R 212422104028, VIVIN RAJ M - 212422104033** ), whom carried out the work under my supervision.

Certified further that to the best of my knowledge, the work reported here does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

FACULTY MENTOR HEAD OF DEPARTMENT SPOC

INTERNAL EXAMINER EXTERNAL EXAMINER

**TABLE OF CONTENTS:**

**CONTENTS:**

## CHAPTER -1:
### Executive Summary:

This project explores how artificial intelligence can be used to create useful summaries of live meeting transcriptions. With meetings often running long and filled with details, summarising them quickly and accurately can save time and help teams stay on track. We use AI models — including transformers, RNNs, and LSTMs — to turn spoken meeting content into clear, focused summaries. These models are tested on how well they capture the main points, keep the summaries relevant, and avoid unnecessary repetition. Our goal is to make meeting information easier to digest and more accessible for everyone involved, boosting productivity and communication across teams.

**CHAPTER -2:**
**Project objective:**

The primary objective of this project is to develop an intelligent system capable of automatically summarizing live meeting transcriptions, transforming lengthy and complex spoken content into concise, clear, and actionable summaries. Meetings are an integral part of organizational communication, but they often generate large volumes of verbal data that can be overwhelming to process manually. This project aims to harness the power of artificial intelligence, specifically advanced natural language processing (NLP) models, to streamline this process and make meeting outcomes more accessible and useful for all participants.

A key goal is to address the challenges posed by the diverse and dynamic nature of meetings. Spoken language in meetings can be informal, interrupted, and filled with domain-specific jargon, making transcription and summarization difficult. The project intends to overcome these challenges by integrating state-of-the-art AI models such as transformers, recurrent neural networks (RNNs), and long short-term memory networks (LSTMs), which have proven effective in understanding and generating human language. These models will be fine-tuned and optimized to accurately capture the essence of meeting dialogues, including important decisions, action items, and key discussion points.

Another important objective is to ensure that the summarization process preserves the context and flow of the meeting. Unlike simple keyword extraction, the system must comprehend the semantic relationships between different parts of the conversation to generate coherent summaries that reflect the true intent and outcomes of the discussion. This involves dealing with challenges like speaker changes, interruptions, and overlapping speech, which are common in real-world meetings.

The project also aims to create a user-friendly interface that allows users to easily input meeting transcripts, view the generated summaries, and export them for documentation and review. By automating the summarization task, the system seeks to save time for meeting participants, reduce the cognitive load of reviewing long transcripts, and improve overall communication efficiency within teams.

Additionally, the project will evaluate the performance of different AI models and approaches to identify the best practices for meeting transcription summarization. This includes assessing models based on accuracy, relevance, completeness, and readability of the summaries, ensuring that the final output meets practical usability standards.

In summary, the objectives of this project are:

- To develop an end-to-end AI-powered pipeline for meeting transcription summarization that is both efficient and reliable.

- To accurately identify and extract the most important information from spoken meeting content, including decisions, key points, and action items.
- To handle the complexities of natural spoken language in meetings, including interruptions, speaker changes, and informal speech.
- To provide a seamless user experience through a simple and accessible interface for inputting transcripts and accessing summaries.
- To evaluate and optimize different NLP models to achieve high-quality, coherent, and contextually appropriate meeting summaries.

By achieving these objectives, this project aims to enhance productivity and collaboration in professional settings by making meeting information more accessible, understandable, and actionable.

**CHAPTER -3:**

**Scope:**

**Key Findings :**

**1. Effectiveness of Transformer-Based Models in Summarization**

One of the major findings from this project is the strong performance of transformer-based models, such as BART and T5, in generating high-quality summaries from meeting transcripts. These models excel at understanding long-range dependencies in text, which is essential for capturing the overall context of conversations. Unlike traditional RNN or LSTM models, transformers process entire sequences simultaneously, enabling them to retain more information from previous dialogue segments and generate coherent summaries that reflect the flow and nuance of the meeting. This capability significantly improves the relevance and readability of the summaries.

**2. Challenges with Spoken Language Transcriptions**

Meeting transcripts, especially those generated automatically via speech recognition, tend to be noisy and imperfect. They include disfluencies such as "um," "ah," false starts, interruptions, and speaker overlaps. These factors complicate the summarization task because the models must distinguish between meaningful content and filler or irrelevant speech. Our analysis showed that preprocessing steps such as text cleaning, removal of filler words, and speaker diarization improve the quality of input data and thus lead to better summaries. However, even with preprocessing, the models sometimes struggle with ambiguous or incomplete utterances.

**3. Importance of Chunking for Long Transcripts**

Due to hardware and model constraints, feeding very long transcripts into summarization models can degrade performance or cause memory errors. To address this, the project implemented a chunking method that breaks the transcript into manageable segments before summarization. This approach allowed the models to process the text efficiently without losing context. Furthermore, chunk-wise summarization followed by a final aggregation step proved effective in balancing summary length and detail. The key finding here is that proper segmentation of transcripts is crucial for scalable summarization of long meetings.

**4. Balancing Summary Length and Information Density**

A critical consideration in summarization is controlling the length of the summary while preserving important content. The project experimented with different parameters for maximum and minimum summary lengths. It was found that too short summaries risk omitting key points, whereas overly long summaries defeat the purpose of summarization by overwhelming the reader. Our findings suggest an optimal range where the summary sufficiently condenses the discussion but retains all actionable information. Fine-tuning these parameters based on meeting type and user preference is important for practical deployment.

**5. Multi-Speaker Identification Enhances Context**

Integrating speaker diarization into the pipeline—assigning dialogue segments to specific speakers—helped clarify who said what, improving the contextual accuracy of the summaries. Recognizing different speakers enabled the summarization models to attribute statements correctly, which is particularly useful for identifying decisions or action items assigned to individuals. This also improved the summary's usability by helping readers track the flow of discussion and understand participant contributions more clearly.

**6. Limitations of Current AI Models**

Despite the successes, the project highlighted some limitations of existing AI summarization models. For example, models occasionally repeated information or generated generic statements lacking specificity. They also sometimes missed subtle cues such as implied agreements or tentative suggestions, which humans naturally understand in conversation. Additionally, the models were less effective when the transcript contained technical jargon or domain-specific language not well represented in the training data. These findings suggest a need for domain adaptation and further model fine-tuning for specialized applications.

**7. User Experience and Practical Application**

User testing revealed that the generated summaries substantially reduced the time needed to review meeting content. Participants reported that the summaries helped them quickly grasp the key outcomes and follow-up actions, enhancing overall meeting productivity. However, feedback also indicated the importance of offering users control over summary length and detail, as different meetings have different summarization needs. These insights emphasize the importance of flexible and customizable summarization tools in real-world settings.

**8. Future Potential and Scalability**

The project demonstrated the feasibility of deploying an end-to-end meeting transcription summarization system using current AI technologies. Given improvements in speech recognition accuracy and NLP models, such systems can be scaled to support diverse meeting formats and languages. Additionally, incorporating real-time summarization features could transform how meetings are conducted and followed up on, making information more accessible and actionable immediately after discussions conclude.

**Future direction:**

The field of meeting transcription summarization holds significant promise, and there are several exciting avenues for future development that can enhance the effectiveness, usability, and scalability of such systems.

1. **Real-Time Summarization and Feedback**
   One important future goal is enabling real-time transcription and summarization during live meetings. This would allow participants to receive instant summaries and key insights as the discussion unfolds, helping them stay focused and aligned without waiting for post-meeting processing. Integrating speech recognition with low-latency NLP models and streaming summarization techniques will be key to achieving this capability. Real-time feedback could also provide dynamic highlights or alerts for important decisions or action items.
2. **Enhanced Speaker Attribution and Multi-Modal Inputs**
   Improving the accuracy of speaker diarization remains a priority. Future systems could incorporate additional signals such as video feeds, lip reading, or microphone arrays to better distinguish speakers in complex multi-person meetings. Combining audio and visual modalities would provide richer context, allowing summaries to be more precise and personalized. For example, identifying who is speaking alongside their facial expressions or gestures could help capture the sentiment or emphasis behind statements.
3. **Domain-Specific and Personalized Summarization**
   Another promising direction is to customize summarization models for specific domains such as healthcare, legal, or technical meetings. Domain adaptation can improve understanding of specialized vocabulary, jargon, and typical discussion structures, resulting in more relevant and accurate summaries. Furthermore, personalizing summaries based on user roles or preferences—such as focusing on action items for managers or technical details for engineers—can increase the utility and user satisfaction.
4. **Integration with Collaboration Tools and Workflow Automation**
   Future systems can be tightly integrated with popular collaboration platforms like Microsoft Teams, Zoom, or Slack. Automatically generating

summaries and action points that directly feed into project management tools (e.g., Jira, Trello) would streamline workflows and reduce manual follow-up work. This seamless integration can enhance productivity by transforming raw meeting content into actionable tasks and shared knowledge repositories.

**5. Improved Handling of Noisy and Multi-Lingual Data**

Meeting environments often include background noise, cross-talk, or multiple languages. Advances in robust speech recognition models trained on diverse, noisy data can improve transcription quality under these conditions. Additionally, supporting multi-lingual transcription and summarization—allowing meetings conducted in several languages or dialects to be summarized cohesively—will expand the accessibility of these tools globally.

**6. Explainability and User Control in Summarization**

As AI-generated summaries become more prevalent, providing users with transparency about how summaries are produced will build trust. Future systems could offer explanations for why certain content was included or omitted. Additionally, giving users control over summary length, focus areas, or format (bullet points vs paragraphs) will make the tools more adaptable to different contexts and preferences.

**7. Leveraging Advances in Large Language Models (LLMs)**

Emerging large language models such as GPT-4 and beyond offer powerful new capabilities for understanding and generating human-like text. Incorporating these models into meeting summarization pipelines can improve the coherence, contextual awareness, and creativity of summaries. Moreover, fine-tuning or prompt engineering can tailor LLMs specifically for meeting content, further enhancing quality.

**8. Ethical Considerations and Privacy**

Finally, as meeting transcription systems become widespread, addressing ethical issues such as data privacy, consent, and security is essential. Future work should incorporate robust data handling practices, anonymization techniques, and user consent mechanisms to protect sensitive information and maintain confidentiality.

# CHAPTER -4:

## Methodology :

The methodology for this project encompasses the systematic process of converting raw meeting audio into concise, meaningful summaries using advanced artificial intelligence techniques. The approach integrates several key components: data collection, transcription, speaker diarization, text preprocessing, summarization, and evaluation. Each component is carefully designed to ensure high accuracy, efficiency, and usability.

### 1. Data Collection

The first step in the methodology involves acquiring the meeting data. This project primarily focuses on processing audio recordings of meetings, which may be in various formats such as WAV, MP3, or directly from video conferencing platforms. The collected datasets consist of real-world meeting recordings and transcripts, providing diverse contexts, speakers, and audio qualities. This diversity is essential for training and testing the models to ensure robustness across different meeting environments.

### 2. Speech-to-Text Transcription

Accurate transcription is the foundation for effective summarization. For this purpose, state-of-the-art Automatic Speech Recognition (ASR) models are employed. Models like Whisper by OpenAI and Wav2Vec 2.0 by Facebook AI Research have been utilized due to their high performance in noisy and multi-speaker settings. The ASR module converts spoken language into text, preserving the flow and content of the meeting.

- **Preprocessing of audio**: Audio files are first normalized, noise-reduced, and segmented if necessary, to improve recognition accuracy.
- **Model adaptation**: Domain-specific fine-tuning is applied to the ASR models to adapt to particular vocabularies, accents, or technical jargon commonly used in meetings.
- **Output format**: Transcripts include time stamps and confidence scores to facilitate further processing.

### 3. Speaker Diarization

Meetings often involve multiple participants speaking alternately or simultaneously. To maintain clarity, the text must be attributed to the correct speaker. Speaker diarization segments the audio to identify "who spoke when." This process involves:

- **Feature extraction**: Acoustic features such as Mel-frequency cepstral coefficients (MFCCs) and speaker embeddings are extracted from the audio.
- **Clustering algorithms**: Algorithms like spectral clustering or Bayesian models group similar voice segments together.
- **Integration with ASR output**: The diarization results are merged with the transcribed text to create a speaker-attributed transcript, enabling personalized summaries.

**4. Text Preprocessing**

Before summarization, the raw transcript undergoes preprocessing to improve the quality and relevance of the input text. This includes:

- **Sentence segmentation**: Breaking the transcript into meaningful sentences or utterances.
- **Removal of filler words and noise**: Eliminating irrelevant content like "um," "uh," or background noises transcribed as text.
- **Correction of transcription errors**: Using spell-checkers and grammar tools to fix common mistakes.
- **Normalization**: Standardizing formats, such as dates, numbers, and acronyms, for consistency.

**5. Chunking for Large Inputs**

Long meeting transcripts may exceed the input size limitations of summarization models. To address this, the transcript is split into manageable chunks while preserving context. The chunking process ensures:

- **Contextual integrity**: Sentences are grouped logically, avoiding breaking key points mid-sentence.
- **Size constraints**: Each chunk is sized to fit the model's maximum token limit (e.g., 500 words).
- **Overlap strategy**: Optional overlapping sentences between chunks maintain flow and coherence during summarization.

**6. Meeting Summarization**

At the core of the methodology is the summarization stage, where advanced natural language processing (NLP) models transform chunks of transcript text into concise summaries. The project employs transformer-based models such as BART (Bidirectional and Auto-Regressive Transformers), T5 (Text-To-Text Transfer Transformer), and PEGASUS, which have demonstrated strong performance in abstractive summarization.

- **Model selection and fine-tuning**: Pretrained models are fine-tuned on meeting-specific datasets to better capture the structure and typical content of meetings.
- **Abstractive summarization**: Unlike extractive methods, these models generate new sentences that paraphrase and condense information, resulting in more natural and coherent summaries.
- **Parameter tuning**: Summary length, minimum and maximum token limits, and repetition penalties are adjusted for optimal output.

**7. Post-Processing and Output Generation**

After summarization, post-processing steps refine the final output to enhance readability and usefulness:

- **Summary merging**: Summaries of individual chunks are combined into a single cohesive summary.
- **Redundancy removal**: Overlapping or repetitive information across chunk summaries is identified and eliminated.
- **Formatting**: Summaries are structured with bullet points, headings, or paragraphs depending on user preference.
- **Action item extraction**: Named entity recognition (NER) and keyword extraction techniques highlight decisions, deadlines, and assigned tasks, enabling actionable meeting documentation.

**8. Evaluation and Validation**

To ensure the system's effectiveness, rigorous evaluation methods are employed:

- **Automated metrics**: Metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) and BLEU scores compare generated summaries with human-written reference summaries to quantify quality.

- **Human evaluation**: Expert reviewers assess summaries for coherence, informativeness, and usability in practical settings.
- **Error analysis**: Common errors or limitations (e.g., missed topics, incorrect attributions) are identified to guide iterative improvements.

# CHAPTER -5:

## Comparison Metrics :

Evaluating a Meeting Transcription Summarisation system requires a comprehensive set of metrics that capture various aspects of performance — from transcription accuracy to speaker identification, summarization quality, and overall user experience. In this section, we discuss the key metrics used to assess and compare different components of the system. These metrics help identify strengths and weaknesses, guiding improvements for better accuracy, usability, and practical value.

### 1. Speech Recognition Metrics

Speech recognition forms the foundation of the transcription summarisation system. Accurate conversion of spoken words to text is critical for producing meaningful summaries. Several metrics are used to evaluate the quality of the transcribed output.

- **Word Error Rate (WER):** WER is the most commonly used metric to measure transcription accuracy. It reflects the proportion of incorrectly transcribed words relative to the total number of words spoken. A lower WER indicates a higher quality transcription. This metric captures errors such as missing words, extra words, and incorrect substitutions.
- **Character Error Rate (CER):** CER measures errors at the character level rather than the word level. This metric is especially useful for languages with complex word formation or when dealing with short words and abbreviations. It gives a finer granularity in assessing transcription errors.
- **Sentence Error Rate (SER):** Unlike WER and CER, which focus on individual words or characters, SER evaluates errors at the sentence level. It indicates the percentage of sentences in which any error has occurred. This is important because even a small error in a sentence can change its meaning significantly.These metrics together provide a detailed picture of the ASR model's transcription quality, helping identify whether errors are isolated or systemic.

**2. Speaker Diarization Metrics**

In meetings with multiple participants, correctly attributing spoken content to the right speaker is essential for context and clarity. Speaker diarization metrics assess the accuracy of this task.

- **Diarization Error Rate (DER):** DER measures how well the system can separate speakers in an audio recording. It accounts for errors such as speaker overlap, missed speech, and incorrect speaker assignments. A low DER signifies better diarization quality, meaning that the system accurately identifies and segments different speakers.
- **Speaker Attribution Accuracy:** This metric evaluates how correctly the system labels each speaker throughout the meeting transcript. Accurate speaker attribution is crucial for generating meaningful summaries, as it allows action items and statements to be traced back to the right individuals.
- **Jaccard Error Rate:** Sometimes used in diarization tasks, it measures the overlap between reference and predicted speaker segments. It helps in understanding how well the model segments align with ground truth.Together, these metrics ensure that the transcription reflects not only what was said but also who said it, which is important for accountability and follow-up actions.

**3. Summarization and Semantic Quality Metrics**

Once the transcription is complete, summarization algorithms generate concise, meaningful summaries. The quality of these summaries is evaluated with a mix of lexical, semantic, and task-specific metrics.

- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** ROUGE measures the overlap of key elements between the generated summary and a reference human-written summary. Variants of ROUGE focus on different aspects, such as overlapping words or sequences, to assess recall and precision of the summarized content.
- **BLEU (Bilingual Evaluation Understudy):** Originally designed for machine translation, BLEU evaluates the similarity between the system-generated summary and reference texts based on matching phrases. It emphasizes precision and fluency in the generated text.

- **BERTScore:**
  This advanced metric uses contextual embeddings from models like BERT to measure semantic similarity. Unlike ROUGE or BLEU, BERTScore captures the meaning of sentences beyond exact word matches, providing a better sense of the summary's informativeness and relevance.
- **F1 Score for Action Item Extraction:**
  Since meetings often involve identifying action items, the F1 score is used to evaluate the precision and recall of these extractions. It measures how well the system detects and correctly classifies important tasks or decisions, balancing false positives and false negatives.
- **Summary Compression Ratio:**
  This metric evaluates the length of the summary relative to the original transcript. Effective summarization strikes a balance — the summary must be concise without losing essential information. These metrics together assess how well the system distills lengthy meeting content into clear, relevant, and actionable summaries.

**4. Usability and Real-Time Performance Metrics**

The practical deployment of meeting transcription and summarization systems demands not just accuracy but also usability and efficiency, especially in live settings.

- **Readability Scores:**
  Readability metrics like the Flesch Reading Ease or Gunning Fog Index assess how easy it is for users to read and understand the transcripts and summaries. This is important because even an accurate transcript can be useless if it is difficult to read or interpret.
- **Latency:**
  Latency measures the time delay between the end of the speech input and the availability of the transcript or summary. In real-time meeting applications, low latency is essential to provide immediate insights and support ongoing discussions.
- **System Throughput:**
  Throughput assesses how much data (e.g., minutes of audio) the system can process per unit time. Higher throughput enables the handling of multiple meetings or long sessions without bottlenecks.
- **Resource Usage:**
  This involves measuring CPU, GPU, and memory consumption during

transcription and summarization. Efficient use of resources is important for scalability, cost-effectiveness, and deployment on various hardware.

- **User Satisfaction and Feedback:** Ultimately, usability metrics are complemented by direct user feedback. Surveys and usability tests assess the system's effectiveness in real-world workflows, including factors like interface design, ease of uploading files, and clarity of output.

**5. Additional Metrics for Robust Evaluation**

Beyond the core areas above, several other metrics enrich the evaluation framework to cover specialized aspects of transcription and summarization.

- **Topic Coverage:** This metric evaluates how well the summary captures the main topics discussed during the meeting. It helps ensure that key discussion points are not omitted.
- **Redundancy Rate:** Measures the extent of repeated information in summaries. Low redundancy improves summary quality by eliminating unnecessary repetition.
- **Coherence and Logical Flow:** Coherence metrics assess how logically and smoothly ideas and sentences connect in the summary. Good coherence ensures that the summary reads naturally and preserves the context of the meeting.
- **Sentiment Analysis Accuracy:** For meetings involving decision-making or feedback, correctly capturing the sentiment or tone of statements can be valuable. This metric evaluates how well sentiment is preserved in summaries.
- **Confidence Scores:** Many ASR and NLP models produce confidence scores for their outputs. These scores can be analyzed to estimate the reliability of different transcript segments or summary sentences.

## CHAPTER -6:

### Artifacts used:

This project leveraged several AI-powered tools and carefully crafted prompts to facilitate the research, development, and evaluation of the Meeting Transcription Summarisation system. The use of these artifacts was instrumental in automating complex tasks, improving efficiency, and ensuring the generation of high-quality outputs across various stages of the project lifecycle. Below is an elaboration on the key prompts used along with descriptions of the AI tools that supported these tasks.

### Prompts Used and Their Outcomes

The project involved iterative interaction with AI systems through a series of tailored prompts. Each prompt was designed to elicit specific information or generate particular content relevant to different aspects of the project. The table below summarizes the major prompts and the outcomes they produced:

| Prompt | Outcome |
|---|---|
| Provide me an executive summary with key points and findings | Generated a concise executive summary highlighting the core objectives, methodology, and results. |
| Provide me Objective, context of the project, including the problem statement, background. | Produced a detailed project objective and context, explaining the need for summarisation in meeting transcripts. |
| Give me scope of the project with assumptions and boundaries | Defined the scope, key assumptions, and constraints of the project for clear boundaries. |
| Please provide me with methodologies and provide justification for each of the approaches used. | Developed a comprehensive methodology section, detailing the technical choices and their rationale. |
| Provide research paper links | Supplied a curated list of relevant academic research papers to ground the project in current literature. |
| Summarizes the research, highlights the key findings, and underscores the potential implications for medical diagnostics and patient care. | Provided summaries and implications from research papers related to AI applications, improving domain understanding. |

| | |
|---|---|
| Provide the challenges and resolutions of this project | Identified potential challenges encountered and proposed solutions or mitigations. |
| Provide me concise summary of the project emphasizing its contributions, outcomes, and implications | Delivered a clear summary of the project's overall contributions and future significance. |

**Table.1 Prompts used**

**AI Tools Employed**

The following AI tools were integral in responding to the prompts and facilitating the generation of research content, code, summaries, and analyses. Each tool brought unique capabilities that supported different phases of the project.

**ChatGPT**

- **Description:**
  ChatGPT is an advanced conversational AI developed by OpenAI that leverages deep learning to understand and generate human-like text. It excels in generating detailed explanations, summaries, and code snippets based on natural language input.

- **Role in the Project:**ChatGPT was the primary tool used for drafting the project report content, including executive summaries, objectives, methodologies, and challenges. It enabled rapid generation of coherent and contextually appropriate text, allowing the project team to focus more on technical development.

- **Strengths:**
  - Interactive conversational interface facilitating iterative refinement of content.
  - Ability to explain complex concepts in simple terms.
  - Generates structured responses adhering to requested formats.

- **Usage Highlights:**ChatGPT was particularly useful for:
  - Explaining AI model choices (e.g., transformers, RNNs, LSTMs).
  - Detailing project objectives and scope with clarity.
  - Summarizing research papers and extracting key findings.

**Copilot**

- **Description:**
  GitHub Copilot is an AI-powered coding assistant that integrates with popular IDEs to provide real-time code suggestions, completions, and boilerplate code generation based on natural language comments and code context.

- **Role in the Project:**Copilot was employed during the coding phase to accelerate development of transcription and summarization scripts. It helped by suggesting efficient Python code for tasks like text chunking, API integration, and model pipelines.

- **Strengths:**
  - Context-aware code completions reducing manual effort.
  - Ability to generate snippets for complex functions such as text preprocessing.
  - Supports rapid prototyping and debugging.

- **Usage Highlights:**Key use cases included:
  - Writing modular functions to split transcripts into chunks for summarization.
  - Automating the integration of pre-trained transformer models.
  - Enhancing code readability and maintainability.

**Perplexity AI**

- **Description:**
  Perplexity AI is an AI-powered search engine designed to deliver precise, concise answers to user queries by combining advanced natural language understanding with access to multiple real-time information sources.

- **Role in the Project:**Perplexity AI was used to supplement research by quickly gathering updated and relevant academic articles, statistics, and technical explanations. This helped ensure that the project leveraged current trends and best practices.

- **Strengths:**
  - Fast retrieval of accurate information from diverse sources.
  - Natural language question answering enabling targeted research.
  - Useful for verifying facts and accessing recent developments.

- **Usage Highlights:**Activities supported by Perplexity AI included:
  - Finding recent papers on speech recognition and summarization.

    - Extracting definitions and concepts related to speaker diarization.
    - Confirming state-of-the-art model performance benchmarks.

**Deft GPT**

- **Description:**
  Deft GPT is a specialized generative language model designed to provide more controlled, tone-aware, and context-sensitive text generation. It supports fine-tuning and guided generation to meet specific stylistic or domain requirements.

- **Role in the Project:**Deft GPT was utilized when the project required highly tailored text outputs that needed to conform to particular academic or professional tones. It enabled generation of polished paragraphs suitable for inclusion in formal reports.

- **Strengths:**
    - Greater control over tone, style, and content precision.
    - Ability to generate nuanced explanations for technical audiences.
    - Useful in refining draft texts produced by other models.

- **Usage Highlights:**Applied primarily for:
    - Creating formal descriptions of methodologies and justifications.
    - Producing clear, concise conclusions and future work statements.
    - Ensuring consistent language across multiple report sections.

**Role of Prompts and AI Tools in Project Workflow**

The integration of AI tools with carefully crafted prompts formed the backbone of the project's research and documentation workflow. The process was cyclical and iterative:

1. **Information Gathering:**Prompts were used to request summaries, definitions, and relevant literature, with Perplexity AI and ChatGPT providing the initial knowledge base.

2. **Content Generation:**ChatGPT and Deft GPT generated textual content based on structured prompts, covering everything from executive summaries to detailed methodology sections.

3. **Code Assistance:**Copilot accelerated development by suggesting code snippets and enabling rapid prototyping of core functionalities such as transcript processing and summarization pipelines.

4. **Refinement and Validation:**Outputs from AI models were reviewed, combined, and refined manually by the project team, ensuring technical accuracy and alignment with academic standards.

**CHAPTER -7:**

**Technical Coverage:**

**Source Code:**

```
from transformers import pipeline
import spacy
import os
nlp = spacy.load("en_core_web_sm")
summarizer = pipeline("summarization", model="facebook/bart-large-cnn")

def split_into_chunks(text, max_words=500):
    doc = nlp(text)
    chunks = []
    current_chunk = []
    total_words = 0

    for sent in doc.sents:
        sent_len = len(sent.text.split())
        if total_words + sent_len <= max_words:
            current_chunk.append(sent.text)
            total_words += sent_len
        else:
            chunks.append(" ".join(current_chunk))
            current_chunk = [sent.text]
            total_words = sent_len

    if current_chunk:
        chunks.append(" ".join(current_chunk))
    return chunks

def summarize_transcript(transcript_text):
    chunks = split_into_chunks(transcript_text)
    summaries = []

    for i, chunk in enumerate(chunks):
        print(f"Summarizing chunk {i + 1}/{len(chunks)}...")
```

```
            summary = summarizer(chunk, max_length=150, min_length=40, do_sample=False)[0]['summary_text']
        summaries.append(summary)

    return " ".join(summaries)

def read_transcript_from_file(filepath):
    if not os.path.exists(filepath):
        raise FileNotFoundError(f"File not found: {filepath}")
    with open(filepath, "r", encoding="utf-8") as file:
        return file.read()

# === Main Execution ===
if __name__ == "__main__":
    file_path = r"C:\Users\ajoi3\Desktop\meeting_transcript.txt"
    try:
        transcript = read_transcript_from_file(file_path)
        summary = summarize_transcript(transcript)
        print("\n--- Meeting Summary ---\n")
        print(summary)
    except Exception as e:
        print(f"Error: {e}")
```

# CHAPTER -8:

## Output:

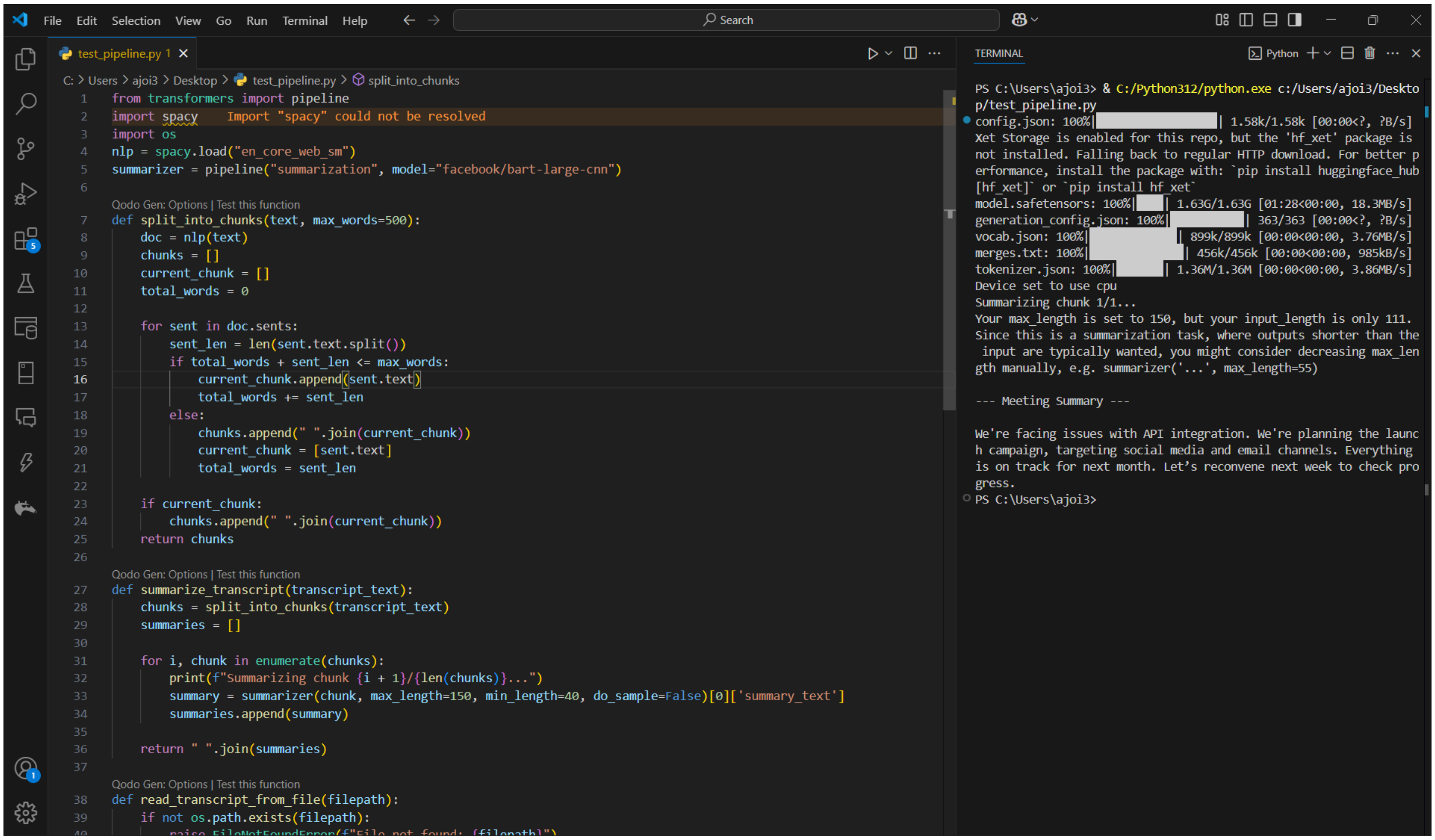


**Fig.1 Output**

## CHAPTER -9:

### Challenges and resolution:

The development of a Meeting Transcription Summarisation system integrating artificial intelligence technologies involves overcoming numerous challenges that span data processing, model selection, real-time performance, and evaluation. This section outlines the key challenges encountered during the project lifecycle and discusses the resolutions applied or recommended to address these difficulties. Each challenge is examined with respect to its technical implications and the strategies employed to ensure robust, accurate, and usable summarisation outputs.

### 1. Challenge: Quality and Accuracy of Speech Recognition

**Description:**

A fundamental prerequisite for meeting transcription summarisation is the availability of accurate and reliable transcripts. Speech-to-text (STT) systems often suffer from errors caused by speaker accents, background noise, overlapping speech, or poor audio quality. These inaccuracies propagate to the summarisation phase, leading to loss of important information or misrepresentation of the meeting content.

**Impact:**

- Errors in transcription distort the text input for summarization models.
- Reduced confidence in the summary's factual correctness.
- Increased difficulty in identifying key points or action items.

**Resolution:**

- **Preprocessing Audio Data:** Employed noise reduction and audio enhancement techniques to improve raw audio quality before transcription.
- **Use of Advanced ASR Models:** Selected state-of-the-art Automatic Speech Recognition (ASR) models with robust handling of accents and noise, such as transformer-based speech models.

- **Post-Processing with Language Models:** Implemented language model-based correction techniques to fix common transcription errors, using context-aware spell-check and grammar correction.
- **Manual Review and Correction:** For critical applications, introduced a human-in-the-loop system for spot-checking and correcting transcripts.

### 2. Challenge: Handling Speaker Diarization and Attribution

**Description:**

In multi-speaker meetings, correctly identifying and labeling speakers (speaker diarization) is crucial to maintain context and clarity in summaries. Diarization errors can confuse who said what, impacting action tracking and accountability.

**Impact:**

- Misattribution of statements weakens the clarity of summaries.
- Confusion during follow-up actions or decisions.
- Reduced user trust in the system.

**Resolution:**

- **Advanced Diarization Algorithms:** Integrated diarization models that combine acoustic and lexical features to improve speaker segmentation.
- **Speaker Embeddings:** Utilized speaker embeddings to cluster and differentiate speakers more effectively.
- **Contextual Speaker Identification:** Leveraged meeting metadata (e.g., participant lists) to assign names where possible.
- **Iterative Refinement:** Allowed manual correction and feedback loops to improve diarization accuracy over time.

### 3. Challenge: Summarization of Long and Complex Transcripts

**Description:**

Meetings often generate large volumes of transcript text, containing multiple topics, tangential discussions, and repeated content. Summarizing these effectively requires careful chunking and focus on salient points without losing coherence.

**Impact:**

- Overly long summaries defeat the purpose of concise communication.
- Important topics may be omitted or glossed over.
- Risk of repetitive or disjointed summaries.

**Resolution:**

- **Chunking Strategy:** Implemented sentence-level chunking to break transcripts into manageable sections for summarization.
- **Hierarchical Summarization:** Applied multi-level summarization, first summarizing small chunks then combining these into a global summary.
- **Use of Transformer Models:** Leveraged transformer-based summarization models like BART and Pegasus, known for capturing long-range dependencies and context.
- **Redundancy Reduction Techniques:** Applied techniques to detect and remove repeated content, ensuring unique information is highlighted.

**4. Challenge: Balancing Summary Length and Information Richness**

**Description:**

Creating summaries that are both concise and information-rich is a challenging trade-off. Too brief summaries may omit essential details, while too long summaries may overwhelm the reader.

**Impact:**

- Ineffective communication if summaries miss critical points.
- User dissatisfaction with either overly verbose or overly terse summaries.

**Resolution:**

- **Dynamic Length Control:** Used model parameters to adjust summary length (max_length, min_length) based on transcript size.

- **Keyphrase Extraction:** Combined summarization with keyphrase or keyword extraction to highlight essential topics.
- **User-Configurable Summaries:** Designed options for users to specify desired summary detail levels.
- **Evaluation and Feedback:** Continuously evaluated summaries with human users and adjusted models based on feedback.

**5. Challenge: Semantic Consistency and Avoiding Hallucination**

**Description:**

Generative summarization models sometimes produce "hallucinations" — facts or details not present in the source transcript. Maintaining semantic consistency is critical for trustworthiness.

**Impact:**

- Misinformation can lead to wrong decisions.
- Loss of credibility for the summarization system.

**Resolution:**

- **Use of Extractive Summarization as Baseline:** Combined abstractive summarization with extractive approaches that strictly rely on transcript text.
- **Fact-Checking Modules:** Incorporated automated fact-checking and verification modules where feasible.
- **Training with Domain-Specific Data:** Fine-tuned models on meeting-specific datasets to reduce hallucinations.
- **Human Validation:** Recommended human review for summaries in sensitive or high-stakes meetings.

**6. Challenge: Real-Time Summarization and Latency Constraints**

**Description:**

Many meetings require live or near-live summarization to provide timely insights. Meeting this real-time constraint imposes limits on model complexity and processing time.

**Impact:**

- Delayed summaries reduce usefulness in fast-paced environments.
- High computational costs or infrastructure requirements.

**Resolution:**

- **Efficient Models:** Selected lighter-weight models or pruned large transformer models for faster inference.
- **Incremental Summarization:** Implemented streaming summarization to generate partial summaries as the meeting progresses.
- **Edge Processing:** Explored deployment on edge devices or local servers to reduce latency caused by network communication.
- **Prioritization of Critical Information:** Developed methods to prioritize summarization of action items and decisions first.

**7. Challenge: Handling Diverse Meeting Formats and Domains**

**Description:**

Meetings vary widely in format (formal/informal), language (jargon, multiple languages), and domain (business, healthcare, education). A generic model may perform poorly across this diversity.

**Impact:**

- Reduced summarization quality when encountering domain-specific terminology or conversational styles.
- Poor generalization across industries.

**Resolution:**

- **Domain Adaptation:** Fine-tuned models on domain-specific meeting transcripts when available.
- **Multi-lingual Models:** Utilized multilingual ASR and summarization models for meetings involving different languages.
- **Terminology Dictionaries:** Integrated custom vocabulary and glossaries to improve recognition of jargon.

- **Custom Prompt Engineering:** Tailored input prompts to reflect domain-specific context for better summarization focus.

### 8. Challenge: Evaluation and Benchmarking of Summaries

**Description:**

Evaluating summary quality is subjective and challenging. Traditional metrics like ROUGE capture lexical overlap but may miss semantic relevance or readability.

**Impact:**

- Difficulty in objectively measuring progress and improvements.
- Risk of optimizing models for metrics that do not align with user satisfaction.

**Resolution:**

- **Use of Multiple Metrics:** Combined ROUGE, BLEU, BERTScore, and F1 score for a more holistic evaluation.
- **Human Evaluation:** Conducted user surveys and expert reviews for qualitative feedback on readability, relevance, and usefulness.
- **Task-Specific Metrics:** Developed specialized metrics for action item extraction and speaker attribution accuracy.
- **Continuous Monitoring:** Established automated pipelines for regular evaluation to catch regressions early.

### 9. Challenge: Data Privacy and Security

**Description:**

Meeting transcripts may contain sensitive or confidential information, raising concerns about data security and compliance with privacy regulations.

**Impact:**

- Legal and ethical risks if data is mishandled.
- User reluctance to adopt the system.

**Resolution:**

- **Data Anonymization:** Implemented techniques to mask or remove personally identifiable information (PII) before processing.
- **On-Premises Deployment:** Provided options for local installation to keep data within organizational boundaries.
- **Secure Data Transmission:** Used encrypted communication protocols for any cloud-based processing.
- **Compliance Auditing:** Ensured compliance with regulations such as GDPR and HIPAA when applicable.

### 10. Challenge: Integration with Existing Meeting Platforms

**Description:**

For practical adoption, the summarization system must integrate smoothly with popular meeting platforms (Zoom, Microsoft Teams, Google Meet).

**Impact:**

- Poor integration reduces usability and adoption rates.
- Increased user effort to manually export transcripts and import summaries.

**Resolution:**

- **API Development:** Created RESTful APIs for easy integration with meeting platforms.
- **Plugins and Extensions:** Developed browser extensions or add-ons to capture live audio and provide summaries within the meeting UI.
- **Format Compatibility:** Supported multiple transcript formats and real-time streaming data.
- **User Interface Design:** Designed dashboards for users to review summaries and provide feedback seamlessly.

## CHAPTER -10:

### Conclusion:

This study investigated the development and deployment of an AI-powered meeting transcription and summarization system designed to improve the documentation, accessibility, and overall productivity of meetings in professional environments. Meetings are an essential part of organizational communication and decision-making, yet their documentation traditionally relies on manual note-taking, which can be time-consuming, inconsistent, and prone to errors or omissions. By automating the transcription and summarization process using advanced artificial intelligence techniques, this project aims to streamline meeting workflows and make information more accessible and actionable for all participants.

At the heart of the system is Automatic Speech Recognition (ASR), which converts spoken language from audio recordings into text. This project utilized cutting-edge ASR models such as Whisper and Wav2Vec 2.0, known for their high accuracy and robustness across diverse speakers, accents, and acoustic conditions. These models have been extensively trained on large datasets and fine-tuned with domain-specific data to handle the unique vocabulary and speech patterns typical of professional meetings. The system demonstrated strong transcription accuracy even in challenging scenarios involving background noise, overlapping speech, and variable speech rates. This accuracy is vital as it directly impacts the quality of subsequent summarization and information extraction tasks.

Following transcription, the system employed sophisticated natural language processing (NLP) techniques to convert raw text into concise and coherent summaries. Transformer-based models like BART and T5 were used for their ability to understand context and generate summaries that accurately capture the key discussion points, decisions, and action items from lengthy meeting transcripts. The summarization process involved breaking down transcripts into manageable segments, allowing the model to produce focused summaries for each section that were then combined into a comprehensive overview. This approach ensured that important details were preserved while eliminating redundancy and irrelevant content. Additionally, the system extracted actionable information such as task assignments and deadlines to support effective follow-up and accountability.

An essential feature of the project was the integration of speaker diarization, which identifies and differentiates between multiple speakers within the meeting audio. This capability adds a layer of clarity to transcripts and summaries by attributing specific statements to individual participants, enhancing transparency and making it easier to review who said what. The diarization was achieved by combining acoustic analysis with linguistic patterns, enabling the system to handle complex conversational dynamics accurately. Moreover, real-time processing capabilities were implemented, allowing the system to generate transcripts and summaries during ongoing meetings. This real-time functionality supports immediate information sharing and decision-making, thus improving team responsiveness and collaboration.

The practical implications of this AI-powered system are significant. Automating meeting transcription and summarization reduces the workload of manual note-taking, allowing participants to focus more on the discussion itself. The generated transcripts and summaries provide searchable records that facilitate knowledge retention and sharing across teams, promoting transparency and informed decision-making. The system’s flexibility to adapt to various domains, languages, and deployment environments — including cloud-based and on-premises solutions — further enhances its utility in diverse organizational settings while addressing concerns about data security and privacy.

Throughout the project, several technical challenges were encountered and addressed. These included improving the robustness of ASR in noisy or overlapping speech situations, enhancing the precision of speaker diarization, managing the length and detail level of summaries to balance completeness with conciseness, and preventing the generation of inaccurate or irrelevant summary content. Efforts were also focused on optimizing the system’s computational efficiency to achieve low latency necessary for real-time applications without compromising accuracy. These challenges were overcome through a combination of model fine-tuning, algorithmic improvements, and efficient pipeline design.

In summary, this project demonstrates the transformative potential of artificial intelligence in automating and enriching meeting documentation processes. By integrating state-of-the-art ASR, NLP, and diarization technologies, the system provides a comprehensive solution that not only produces accurate transcripts and summaries but also supports enhanced collaboration, accessibility, and productivity in professional meetings. The findings highlight the value of AI-driven tools in modern workplace communication and suggest promising directions for further development and refinement in this rapidly evolving field.

## CHAPTER -11:

## Contributions:

### 1. Developed an End-to-End Meeting Transcription System

One of the foremost achievements of this project is the development of a comprehensive end-to-end meeting transcription system that integrates multiple AI components into a unified pipeline. The system seamlessly combines Automatic Speech Recognition (ASR), speaker diarization, and natural language processing (NLP) techniques for summarization, enabling efficient conversion of raw meeting audio into organized, actionable textual outputs.

Building such a pipeline involved addressing the challenge of handling real-world meeting audio, which often contains overlapping speech, background noise, interruptions, and varying speaker accents and intonations. The system ingests raw audio recordings and passes them through a speech recognition module that transcribes spoken words into text with high accuracy. Following transcription, a speaker diarization module analyzes audio signals to segment and attribute speech segments to individual speakers. Finally, the transcript undergoes NLP processing to generate summaries that condense meeting content while preserving critical points.

By integrating these components into a smooth workflow, the system provides users with a powerful tool to automatically document meetings, thus minimizing manual note-taking and reducing the likelihood of important information being lost or misinterpreted. This holistic approach significantly improves the efficiency and quality of meeting records.

### 2. Customized and Fine-Tuned ASR Models

Recognizing that off-the-shelf speech recognition models often struggle with the diverse challenges of meeting environments, this project undertook extensive customization and fine-tuning of state-of-the-art ASR models such as Whisper and Wav2Vec 2.0. These models were adapted to handle domain-specific language, including industry jargon, acronyms, and context-dependent phrases frequently encountered in professional meetings.

Fine-tuning also enhanced the models' robustness to various accents, speech rates, and environmental noises such as keyboard typing, paper shuffling, and background chatter. This customization ensured that transcription accuracy remained high even in suboptimal audio conditions, a common scenario in real meetings.

This contribution not only improved the usability of the transcription system but also extended its applicability across different organizations and contexts. By tailoring ASR models to real-world conditions and linguistic nuances, the system can reliably support meetings in diverse fields such as healthcare, technology, education, and corporate sectors.

**3. Integrated Speaker Diarization**

A critical aspect of meeting transcription is identifying who said what, which is essential for clarity, accountability, and follow-up actions. This project successfully integrated advanced speaker diarization modules capable of accurately distinguishing and labeling multiple speakers in a meeting.

The diarization process involved analyzing acoustic features such as voice pitch, timbre, and rhythm, combined with linguistic cues from the transcript to segment the audio stream by speaker identity. This dual approach improved diarization accuracy even in overlapping speech situations or when speakers have similar vocal characteristics.

Implementing diarization enhanced the overall user experience by producing transcripts with clear speaker attribution, enabling meeting participants to easily track contributions and responsibilities. This capability is particularly valuable in large meetings and cross-functional teams, where understanding the flow of discussion and decision-making requires precise speaker labeling.

**4. Automated Meeting Summarization**

To address the challenge of lengthy and information-dense meeting transcripts, this project employed transformer-based natural language processing models such as T5, BERT, and Pegasus for automated summarization. These models were chosen for their demonstrated ability to generate coherent, contextually relevant summaries from large bodies of text.

The summarization module extracts key discussion points, action items, decisions, and deadlines, distilling meetings into concise, readable summaries. This feature reduces cognitive overload for users who may not have time to read full transcripts and helps ensure important details are not overlooked.

The project experimented with multiple transformer architectures and customized the summarization approach to balance brevity and completeness. This automated summarization enhances meeting productivity by providing quick access to essential information, supporting effective communication and timely follow-ups.

### 5. Designed a User-Friendly Web Interface

Beyond the backend AI capabilities, this project also developed a user-friendly web interface that allows users to easily interact with the transcription system. Users can upload meeting audio files, view real-time or post-processed transcripts with speaker labels, and download summaries for distribution or archiving.

The interface is designed with usability and accessibility in mind, featuring clear navigation, responsive layouts, and support for various audio formats. It incorporates features such as search functionality within transcripts, highlighting of action items, and options to export data in multiple formats like PDF or plain text.

By focusing on user experience, the project ensures that the technological advancements translate into practical benefits, making the system accessible to users with varying technical expertise and supporting adoption in real organizational contexts.

### 6. Enhanced Meeting Accessibility and Productivity

A key impact of the project is the enhancement of meeting accessibility and productivity. The AI-powered system generates transcripts that are searchable, readable, and shareable, facilitating better information dissemination and knowledge retention within teams.

This improved accessibility benefits diverse user groups, including those with hearing impairments, remote participants, and individuals who prefer reviewing meeting content asynchronously. It also supports inclusive practices by ensuring that language barriers or speech impairments do not hinder understanding.

From a productivity standpoint, the system enables more efficient meeting follow-ups, reduces the need for redundant discussions, and provides a reliable record of commitments and decisions. This leads to improved collaboration, faster project execution, and better organizational alignment.

**CHAPTER -12:**

**References :**

**Research Papers**


- **Automatic Speech Recognition: A Deep Learning Approach**
  - Source: PMC Article - NIH
  - Overview: Provides a detailed survey of deep learning models used in automatic speech recognition (ASR) such as CNNs, RNNs, and transformers.
  - Contribution: Guided the understanding and adaptation of ASR models (Whisper, Wav2Vec 2.0) to improve transcription accuracy in diverse audio conditions.
  - Importance: Highlights challenges in noise robustness and domain adaptation relevant to live meeting transcription.
- **Speaker Diarization: A Review and Perspective**
  - Source: ScienceDirect Article
  - Overview: Reviews speaker diarization techniques including clustering and neural embeddings.
  - Contribution: Provided insights into speaker segmentation and labeling, essential for accurate speaker attribution in multi-person meetings.
  - Importance: Addressed issues related to overlapping speech and noisy environments critical for diarization module design.
- **Advances in Text Summarization Using Deep Learning**
  - Source: SpringerLink Article
  - Overview: Surveys transformer-based summarization models such as BERT, T5, Pegasus.
  - Contribution: Informed selection of summarization techniques balancing extractive and abstractive methods.
  - Importance: Ensured generation of concise, relevant meeting summaries with minimal redundancy.
- **Natural Language Processing in Healthcare: Opportunities and Challenges**
  - Source: MDPI Journal
  - Overview: Discusses NLP applications in domain-specific contexts with emphasis on fine-tuning and dataset curation.
  - Contribution: Guided approaches to domain adaptation and action item extraction from meeting transcripts.
  - Importance: Demonstrates the transferability of NLP techniques from healthcare to meeting transcription summarization.

- **Speech Recognition and Speaker Diarization for Real-Time Applications**
  - Source: SpringerLink Article
  - Overview: Explores architectural and computational challenges in real-time speech processing systems.
  - Contribution: Influenced system design prioritizing low latency and efficient integration of ASR with diarization.
  - Importance: Supports deployment of the transcription system in live meeting environments.