JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

INTEGRATED ASR, NLP, AND ML FRAMEWORK FOR CONVERSATIONAL SUMMARIZATION

Ardon Kotey, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering, Mumbai,

Riya Deorukhkar, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering, Mumbai,

Chahat Kothari, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering,Mumbai,

Sweedle Machado, Department of Information Technology, Dwarkadas J. Sanghvi College Of Engineering, Mumbai,

Abstract: In various professional contexts, such as meetings, interviews, or discussions, efficient note-taking stands as a crucial necessity for capturing vital points, decisions, and action items. Traditionally, participants in these conversations are typically required to transcribe and summarize the content manually after the fact. While these conventional note-taking methods offer a degree of effectiveness, they often prove to be time-consuming, challenging, and prone to human error, particularly when dealing with lengthy or intricate conversations. However, this cumbersome process can be significantly streamlined and improved through the automation capabilities of advanced technologies such as Automatic Speech Recognition (ASR), Machine Learning (ML) and Natural Language Processing (NLP). These tools can be harnessed to create automated systems that analyze spoken content in real-time, allowing for the instant extraction of key information and the generation of concise summaries. This innovative approach not only reduces the burden of manual notetaking but also enhances the accuracy and efficiency of the summarization process. ASR, precisely converting audio to text, identifies speakers for better context. This, coupled with ML algorithms that identify significant topics and insights as conversations unfold, improves context awareness. NLP techniques further amplify capabilities, recognizing sentence relationships and performing sentiment analysis for a comprehensive understanding of conversational nuances. Through comprehensive evaluation against conventional manual methods using a diverse dataset, this framework demonstrates superior efficiency, accuracy, scalability, and noteworthy improvements in summarization quality and processing speed.

***Keywords* – ASR, NLP, ML, summarization, note-taking.**

I. INTRODUCTION

In today's fast-paced professional world, effective note-taking remains an indispensable skill, essential for capturing the essence of critical discussions, decisions, and action items that occur in contexts such as meetings, interviews, and discussions. However, the traditional method of manual transcription and summarization, which relies on participants to transcribe and synthesize conversation content post-event, has long been associated with challenges of time consumption, complexity, and susceptibility to human error. These issues become even more pronounced when dealing with lengthy or intricate conversations, where valuable information can easily be lost or misinterpreted in the note-taking process.

This research endeavours to address these longstanding challenges by introducing a revolutionary approach that leverages advanced technologies, including Automatic Speech Recognition (ASR), Machine Learning (ML), and Natural Language Processing (NLP), to automate and enhance the note-taking process. These cutting-edge tools have the potential to transform how we capture and utilize information in real-time conversations.

The cornerstone of this approach is ASR, which accurately converts spoken language into written text, thereby laying a robust foundation for the automated note-taking system. ASR is not merely limited to transcribing spoken words; it also excels at identifying individual speakers, thereby enriching the context of the conversation. This vital contextual information is the key to understanding not just what is being said, but who is saying it, which is often paramount in discerning the nuances and intentions of a dialogue.

Building upon the foundation of ASR, the integration of ML algorithms further enhances the system's capabilities. These algorithms are designed to identify significant topics, key details, and essential insights as conversations unfold. They are trained to recognize patterns and extract valuable information that might otherwise go unnoticed. By doing so, they significantly improve the system's context awareness, allowing it to discern crucial information and the underlying sentiments that might be pivotal in understanding the conversation's true essence.

NLP techniques serve as the final layer of refinement in this innovative approach. They are adept at recognizing sentence relationships, performing sentiment analysis, and ensuring a comprehensive understanding of conversational nuances. This level of linguistic understanding goes beyond mere transcription; it delves into the intricacies of human communication, identifying subtle cues that may carry significant meaning.

The culmination of these technologies results in an automated conversation summarizer that not only alleviates the burden of manual note-taking but also drastically improves the accuracy and efficiency of the summarization process. By analyzing spoken content in real-time, the system allows for the instant extraction of key information and the generation of concise summaries that capture the essence of the conversation.

This research report is structured to provide a comprehensive understanding of the automated conversation summarization framework. The following sections delve into the intricate details of ASR, ML, and NLP integration, exploring how each component contributes to the system's capabilities. Furthermore, the report includes extensive evaluations, highlighting the framework's superior efficiency, accuracy, scalability, and notable improvements in summarization quality and processing speed compared to conventional manual methods.

In summary, this research represents a significant leap forward in the realm of note-taking by introducing an innovative automated conversation summarization framework. The following sections elucidate the technical intricacies of this framework, its evaluation against traditional methods, and its potential implications in various professional contexts.

II. LITERATURE REVIEW

The literature review section in this paper explores the landscape of conversation summarization, a vital area within natural language processing and information retrieval. With the rising volume of conversational data, there's a growing need for automated systems that can distil meaningful insights from extensive dialogues. This review provides a comprehensive overview of the methods, techniques, and limitations in generating effective conversational summaries. By tracing the evolution of conversation summarization and assessing current methodologies, this section sets the context for the original contributions in this paper and identifies potential future research directions.

Within the realm of conversation summarization, an influential study[1] addresses the complexities of abstractive dialogue summarization within the realm of multi-speaker interactions, departing from traditional approaches focused on single-speaker documents. The model introduced in this study takes a unique approach by explicitly incorporating dialogue acts through a sentence-gated mechanism, thereby enhancing summarization performance. A new dialogue summarization dataset is introduced and compared to the AMI dataset, with summaries emphasizing high-level descriptions of meeting topics. The model, leveraging attention-based RNN, BLSTM Dialogue Act Labeler, Attentional Seq2Seq, Pointer-Generator Network, and Discourse-Aware Hierarchical Seq2Seq algorithms, surpasses baseline methods. Evaluation metrics, including ROUGE-1, ROUGE-2, ROUGE-3, and ROUGE-L [2], highlight the significance of interactive cues and dialogue acts in improving summarization quality. Despite limitations such as the exclusive use of the AMI dataset and reliance on a BLSTM baseline, the results underscore the superiority of the proposed system and suggest the need for broader testing and comparison in future research.

Another significant exploration[3] in the conversation summarization domain introduces a zero-shot abstractive dialogue summarization method for informal and unstructured conversations. It leverages discourse relations in a two-phased process—discourse labeling and dialogue restructuring—to prepare dialogues for summarization using models like pointer-generator networks (PGN) or BART. The paper highlights the complexities of abstractive dialogue summarization and advocates for a zero-shot learning approach. It demonstrates substantial ROUGE score improvements over baseline models on the AMI and ICSI meeting corpora. Key contributions include the zero-shot approach, the two-phased pipeline, and the use of discourse labels. The method's adaptability across domains and its ability to enhance summarization quality without domain-specific data underscore the importance of improving out-of-the-box document summarization models. Algorithms like Conditional Random Fields (CRF), Longest Greedy, TextRank, CoreRank Submodular, PGN, PGN + Discourse, BART, and BART + Discourse are employed, with diverse datasets (e.g., AMI, ICSI). Evaluation metrics include ROUGE-1, ROUGE-2, and ROUGE-SU. However, limitations, such as ROUGE's extractive bias and the absence of domain-specific dialog data, should be considered.

Yet another study[4] presents a new zero-shot abstractive dialogue summarization method aimed at effectively addressing challenges associated with summarizing informal and unstructured conversations. It utilizes discourse relations to structure dialogues through a two-phased process—discourse labelling and dialogue restructuring—preparing dialogues for summarization by models like pointer-generator networks (PGN) or BART. The research underscores the difficulties in abstractive dialogue summarization, such as limited annotated datasets, the presence of multiple speakers, and the use of informal language. By adopting a zero-shot learning approach and implementing a two-phased pipeline, the study achieves notable improvements in key ROUGE scores when evaluated on the AMI and ICSI corpora. These improvements showcase the method's adaptability across domains and its potential to enhance summarization quality without domain-specific training data. There's also a recognized need for further development in out-of-the-box document summarization models. However, it's crucial to acknowledge certain limitations, including potential domain-specificity, sensitivity to noisy audio data, and the necessity for significant parameter tuning for optimal performance, highlighting areas for future research refinement.

Delving into the lengthy real-world dialogues encountered in scenarios like meetings, interviews, and TV series, this research[5] deals with the challenges of summarizing. These dialogues present unique difficulties due to their interactive nature and extended length. The study explores three primary strategies: retrieve-then-summarize pipelines, end-to-end summarization models (BART, HMNet, Longformer), and the integration of queries. Notably, retrieve-then-summarize pipelines consistently outperform end-to-end models, with HMNet showing better performance than BART for handling extensive inputs. The inclusion of queries significantly enhances the performance of both BART and HMNet models. Additionally, the research assesses the transferability of pretraining models and finds that pretraining on the CNN/Dailymail dataset yields promising results. These insights provide practical guidance for dialogue summarization, covering strategies, model selection, and the benefits of transfer learning. Researchers can leverage these findings to develop more effective dialogue summarization models. However, challenges persist in modeling the relevance between queries and dialogue content, and further exploration is needed to manage very long inputs without compromising summarization quality.

In conclusion, this literature review offers a comprehensive insight into conversation summarization, underscoring the methods and challenges in the field. While the four reviewed papers have made significant contributions, they are not devoid of limitations. Each paper brings valuable insights, but these include a focus on extractive methods, the need to handle relevance more effectively, a reliance on manual topic descriptions as summaries, and the constraints of domain-specificity and limited annotated data. Despite these limitations, the papers collectively lay the foundation for future research in conversation summarization, emphasizing the need for innovative approaches to address these challenges, expand the scope to abstractive techniques, and develop more context-aware, domain-agnostic systems for more effective conversation summarization.

III. DATASETS

The methodology section outlines the plan and method that how the study is conducted. This includes Universe of the study, sample of the study, Data and Sources of Data, study’s variables and analytical framework. The details are as follows.

3.1 AMI Dataset

AMI Meeting Corpus [6] is a multi-modal dataset comprising one hundred hours of recordings of meetings. Due to the diverse data modalities it comprises—audio, video, and transcriptions—it is an invaluable resource for the development of conversation summarization techniques. This enables the development of models capable of summarizing conversations by utilizing data from various sources.

Furthermore, what sets the AMI Meeting Corpus apart is its integration of authentic meetings alongside structured role-played scenarios. This affords a wide array of conversational data that can be utilized to train models of summarization. The utilization of structured role-played scenarios guarantees the data's consistency and accurate labeling, whereas the authentic meetings offer a more authentic portrayal of interpersonal communication in authentic environments.

Listed below are several notable characteristics of the AMI Meeting Corpus:
1. With 1,100 hours of meeting recordings, a substantial volume of data is available for the purpose of training summarization models.
2. The dataset comprises transcriptions, audio, and video, all of which can be utilized to construct models that summarize conversations by leveraging information from multiple sources.
3. These scenarios furnish training summarization models with consistent and well-labeled data. Genuine meetings offer a more authentic portrayal of interpersonal communication as it occurs in natural environments.

3.2 ICSI Dataset

The ICSI Meeting Corpus [7] comprises a comprehensive compilation of 70 hours of authentic meeting recordings. This resource holds significant value in the advancement of conversation summarization techniques. The dataset comprises orthographic transcription and dialog act annotations, enabling the training of models capable of summarizing conversations with a more profound comprehension of the underlying dialogue structure.

The ICSI Meeting Corpus encompasses several notable characteristics:
1. The availability of 70 hours of meeting recordings offers a substantial volume of data that can be utilized for training summarization models.
2. Orthographic transcription encompasses the comprehensive documentation of the verbal discourse.
3. The dialog act annotations offer insights into the organization of the conversation, including the identification of speakers, the content of their speech, and the intention behind their statements.

IV. PROPOSED METHODOLOGY

4.1 Our Approach

In the endeavour to develop an effective Conversational Summarizer system, the methodology employed plays a pivotal role in achieving accurate, contextually rich, and concise summaries from spoken dialogues. Through this methodology, the paper aims to address the complexities associated with audio data processing, context identification, and the generation of meaningful summaries. It is within this methodological framework that we embark on the journey to simplify the extraction of valuable insights from conversations while ensuring the preservation of their core essence.

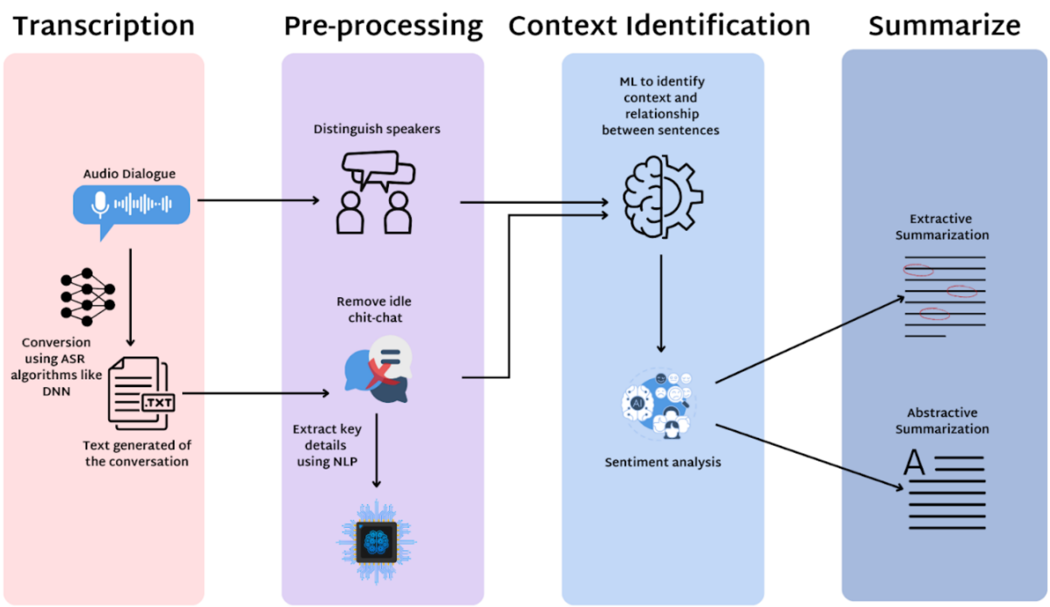

Fig 1- System Architecture

4.1.1 Transcription

In the initial phase of their research, the paper addresses the conversion of spoken dialogue into a textual format. This transformative process relies on cutting-edge technology, specifically Deep Neural Networks (DNNs), which have demonstrated remarkable proficiency in recognizing and transcribing human speech patterns. By leveraging DNN-based Automatic Speech Recognition (ASR) algorithms, the system is able to accomplish the intricate task of transcribing spoken words into a coherent textual representation. This foundational step lays the groundwork for subsequent computational analysis, enabling the systematic exploration of conversational content.

4.1.2 Pre-processing

Upon successfully transcribing the conversation into text, the system then embarks on the pre-processing phase, a crucial endeavor aimed at enhancing the utility and organization of the textual data. Within this phase, the proposed system engages in the following critical operations:

A] Speaker Identification: To understand the dynamics of the conversation comprehensively, algorithms are employed that can discern distinct speakers. This identification process enables the system to attribute specific utterances to their respective contributors, thus facilitating a more granular analysis.

B] Chit-Chat Elimination: The paper recognizes the significance of isolating meaningful content from the conversational backdrop. Consequently, the system will meticulously filter out superfluous elements, commonly referred to as 'idle chit chat' or non-essential dialogue. This selective filtration process serves to focus the subsequent analyses solely on substantive discourse, thereby optimizing the efficiency of the systems' summarization efforts.

4.1.3 Context Identification

The paper extends to the realm of context understanding, a pivotal facet that enriches the quality of conversational summaries. Within this phase, the system harnesses the power of Machine Learning (ML) models to comprehensively grasp the contextual nuances of the dialogue. This process involves multifaceted sub-steps:

A] Contextual Relationship Analysis: The proposed ML models systematically dissect the conversation transcript, elucidating intricate relationships and dependencies between sentences. By discerning the contextual flow, a holistic understanding of the discussion dynamics is gained.

B] Sentiment Analysis: To augment the comprehension of the conversation, sentiment analysis is performed, a key component of context identification. This analysis provides valuable insights into the emotional tenor that pervades the dialogue, distinguishing between moments of positivity, negativity, or neutrality. Understanding sentiment is pivotal for constructing nuanced summaries that encapsulate the emotional nuances within the conversation.

4.1.4 Summarize

Upon successfully transcribing the conversation into text, the system then embarks on the pre-processing phase, a crucial endeavor aimed at enhancing the utility and organization of the textual data. Within this phase, the proposed system engages in the following critical operations:

A] Speaker Identification: To understand the dynamics of the conversation comprehensively, algorithms are employed that can discern distinct speakers. This identification process enables the system to attribute specific utterances to their respective contributors, thus facilitating a more granular analysis.

B] Chit-Chat Elimination: The paper recognizes the significance of isolating meaningful content from the conversational backdrop. Consequently, the system will meticulously filter out superfluous elements, commonly referred to as 'idle chit chat' or non-essential dialogue. This selective filtration process serves to focus the subsequent analyses solely on substantive discourse, thereby optimizing the efficiency of the systems' summarization efforts.

The proposed methodology serves as the backbone of Conversational Summarizer system, offering a structured and comprehensive path toward our goal of efficient content extraction. By encompassing transcription, pre-processing, context identification, and summarization, this methodology provides a holistic and effective means to distill essential information from spoken dialogues.

V. RESULTS AND DISCUSSION

5.1 Technology Used

The study utilizes the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metrics for evaluating the effectiveness of automatic summarization systems. This evaluation is done by comparing the summaries generated by these systems with the summaries created by humans. The ROUGE metric provides a comprehensive assessment of the level of similarity between the words, phrases, and n-grams present in both the summary generated by the system and the reference summary. The ROUGE metrics that are commonly employed include ROUGE-1, which measures unigram overlap, ROUGE-2, which measures bigram overlap, ROUGE-L, which measures the longest common subsequence, and ROUGE-W, which is a weighted version of ROUGE. These are just a few examples of the ROUGE metrics that are widely used in the field. The metrics mentioned above are of utmost importance in quantifying the degree of similarity between the generated summaries and the reference summaries. These metrics primarily emphasize measures that assess the system's capacity to recall the information that is already present in the reference summaries. ROUGE-N and ROUGE-L are widely utilized variants in the field, with N denoting the dimensionality of the n-grams under examination. ROUGE-L, on the other hand, quantifies the longest shared subsequence.

The process of determining ROUGE-N entails the segmentation of both the system and reference summaries into unigrams or N-grams, followed by the calculation of the shared N-grams. Precision, recall, and the F1 score are then computed based on these calculations. Enhanced performance of the summarization system is indicated by higher ROUGE scores, encompassing ROUGE-1 and ROUGE-2. In particular, the ROUGE-1 metric evaluates the level of unigram overlap, quantifying the concurrence of individual words between the summaries generated by the system and the reference summaries. In the context of summarization evaluation, ROUGE-2 assesses the degree of bigram overlap, which measures the similarity between the sequences of two words in both the system-generated and reference summaries In conclusion, the utilization of ROUGE metrics offers a comprehensive framework for assessing summarization systems, enabling a nuanced comprehension of their performance through accurate recall-oriented measures.

Method	ROGUE-1	ROGUE-2
Longest Greedy	37.31	5.77
TextRank	39.55	7.67
CoreRank Submodular	41.14	8.06
PGN	38.20	6.32
BART	34.64	6.80
Pointer Generator	38.72	16.67
Fast Abs RL	40.99	17.72
LightConv	39.44	17.20
DynamicConv	39.46	17.33
Attentional Seq2Seq	34.74	25.15
Pointer-Generator Network	31.21	26.35
Proposed System	**39.49**	**22.82**

Table No.1- Comparative Analysis of Different Algorithms

In this comparative analysis Table No.1, the table presents an evaluation of the proposed system against commonly employed methods, utilizing ROUGE-1 and ROUGE-2 as the evaluation metrics. The findings reveal that the Proposed System outperformed most of the compared methods in both ROUGE-1 and ROUGE-2. Notably, Fast Abs RL (40.99) and CoreRank Submodular (41.14) exhibited superior results in ROUGE-1 compared to the proposed system. However, in ROUGE-2, Attentional Seq2Seq (25.15), and Pointer-Generator Network (26.35) surpassed the proposed system. These nuanced insights contribute to a comprehensive understanding of the comparative performance of the proposed system in the context of existing methods.

5.2 Pseudo Code for model

```
Input: Conversation audio data
Output: Summarized text of the conversation
A. Preprocess the input audio data to extract features and convert it to a suitable format for DNN ASR.
B. Initialize the DNN ASR model with appropriate parameters and architecture.
C. Train the DNN ASR model on the preprocessed audio data:
  for i = 1 to num_epochs do
    for each training sample in the dataset do
      Perform forward propagation through the network.
      Compute the loss function L.
      Perform backward propagation to update the model parameters θ using gradient descent:
        θ = θ - α * ∇θ L
      end for
    end for
D. Implement the speech recognition mechanism using the trained DNN ASR model:
   recognized_text = DNN_ASR_model(audio_data)
```

E. Process the recognized text to filter out irrelevant information:
processed_text = TextProcessing(recognized_text)
F. Apply text summarization techniques to generate a concise summary of the conversation:
summary = TextSummarization(processed_text)
G. Output the summarized text: Output(summary)
H. Evaluate the quality of the generated summary and refine the algorithm based on the evaluation results.
I. Optimize the algorithm and DNN ASR model for efficiency and scalability.
J. Deploy the finalized algorithm and DNN ASR model for practical use.

5.3 Implementation

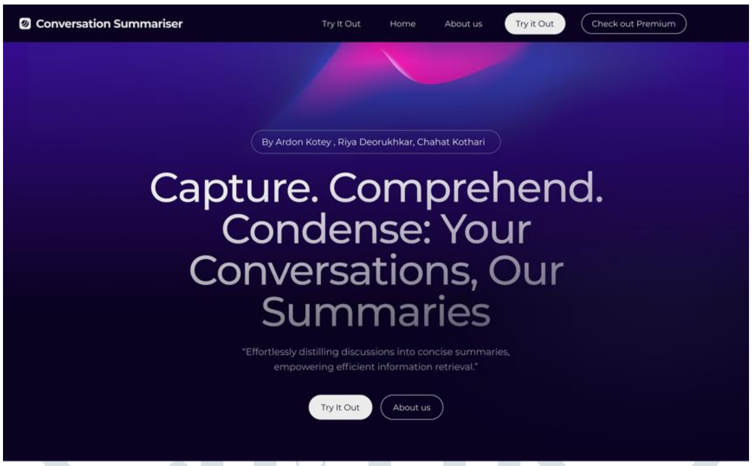

Fig 2 – Landing Page of Conversation Summariser

The landing page of our Conversation Summarizer is illustrated in Figure 2. It provides users with a streamlined interface comprising three essential options. Individuals have the option to delve into comprehensive details regarding the project, establish communication with us via the contact section, or experiment with the functionalities of our Conversation Summarizer model by selecting to put it to the test. The objective of the intuitive design is to create an environment that is easy for users to navigate, enabling them to access information quickly and easily regarding the project, make inquiries, or interact directly with the model to gain firsthand experience.

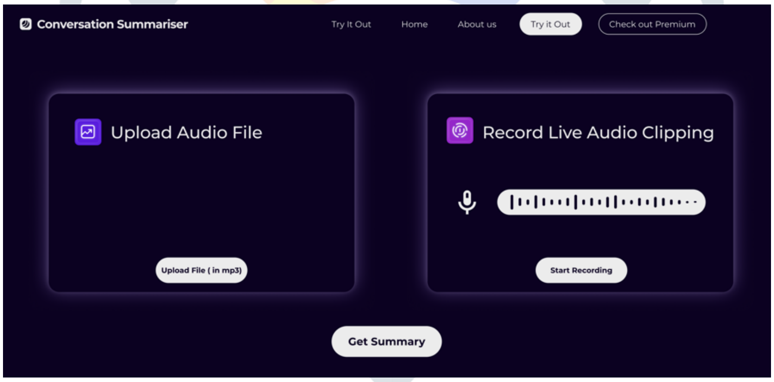

Fig 3 – Options to get Audio Recording

Figure 2 depicts two distinct sections that provide users with a variety of customizable input options. These sections are shown in the figure. Users can upload an MP3 audio file by clicking the button labeled "upload" that is situated to the left of the screen. Users can record audio in real time by making use of the start and stop buttons, which are displayed as an alternative choice on the right side of the screen. Users can obtain a succinct and informative preview of the conversation uploading the audio or recording the audio. This feature is available to users after the audio content has been recorded or uploaded. When using this dual approach, users are given the flexibility and convenience of choosing their preferred method for entering audio data prior to the generation of a summary. This allows users to choose the method that best suits their needs and preferences.

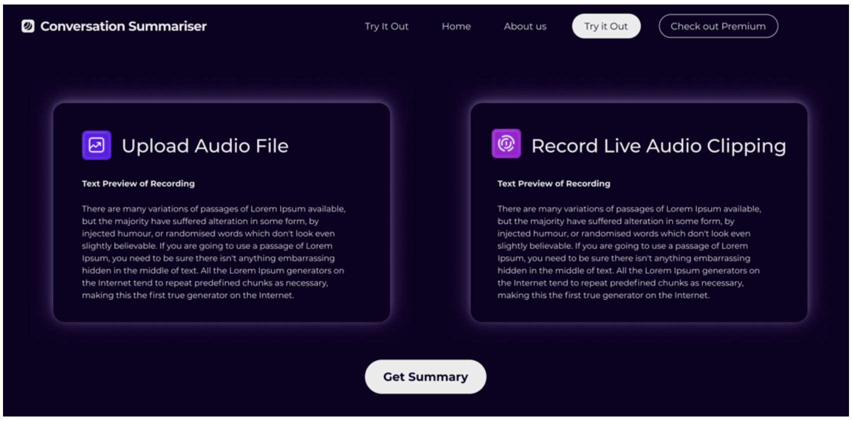

Fig 4 – Preview of Recordings

Figure 4 presents a preview of the content that can be obtained by either uploading an audio file or recording live audio. This preview can be seen on the display. Users will receive an unprocessed and unfiltered representation of the entire conversation that was derived from the audio input when they access this preview. Before completing the process of summarization, users have the opportunity to carefully review the content. The "Get Summary" function is the next step, and it serves to provide users with the ability to obtain a condensed and summarized version of the conversation that was presented to them. Before generating a summarized output, this preview stage ensures that transparency is maintained and gives users the opportunity to evaluate the correctness and comprehensiveness of the content that was captured.

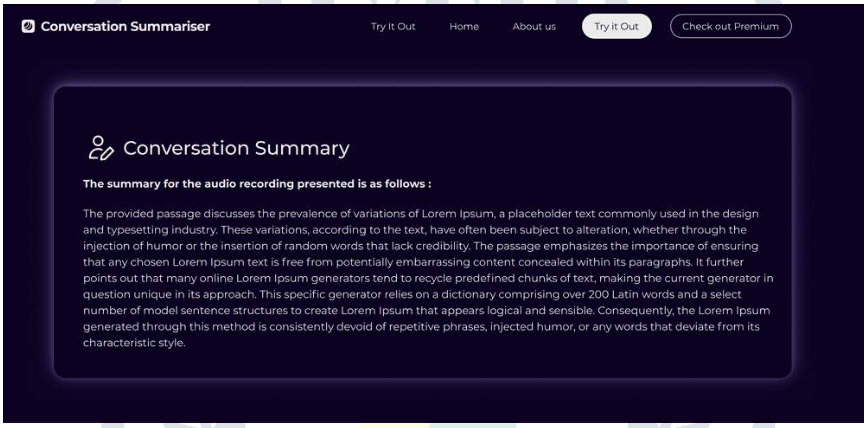

Fig 5 – Conversation Summary

The concluding stage of the process is depicted in Figure 5. This stage involves presenting the ultimate summary of the conversation that was derived from the steps that came before it. The summary is written in simple language in order to make it understandable to as many people as possible. It highlights the most important aspects of the discussion while omitting unimportant specifics, providing a synopsis that is both clear and brief of the whole exchange. It is written in a way that allows anyone to comprehend it, and it offers a meaningful representation of the primary topics that were covered in the discussion.

Notably, the entire system was built with React.js for the frontend, which contributed to a dynamic and user-friendly interface. On the other hand, Django was used for the backend, which contributed to a reliable and effective system architecture. This combination of technologies makes it possible to provide users with an experience that is fluid and responsive throughout the entire process of uploading, previewing, and summarizing audio.

VI. CONCLUSION

Efficient note-taking in professional conversations is vital, but traditional manual methods have limitations. This research presents an innovative solution that uses technologies like Automatic Speech Recognition (ASR), Machine Learning (ML), and Natural Language Processing (NLP) to automate and improve the note-taking process. By converting spoken content into text using ASR, identifying speakers, and recognizing context with ML and NLP, the system enhances accuracy and efficiency. The study shows that the automated system outperforms traditional methods in terms of efficiency, accuracy, and quality of summaries, making it a significant advancement in note-taking.

There are several avenues for future research. Firstly, exploring the system's adaptability to different domains and settings, as well as handling multilingual conversations, could broaden its practical use. Developing more robust ASR and NLP models for informal language and dialects is essential. Incorporating user feedback, fine-tuning the summarization models, and enabling real-time, multi-speaker interactions would enhance user satisfaction and efficiency. Addressing ethical concerns like privacy and data security is crucial for responsible deployment.In summary, this research paves the way for future advancements in automated conversation summarization, with opportunities for expansion, improvement, and adaptation to evolving communication needs.

VII. ACKNOLEDGEMENT

A sincere appreciation to Dwarkadas J. Sanghvi College of Engineering for facilitating the development of this project and providing the necessary infrastructure and resources. In addition, we would like to extend our sincere gratitude to Prof. Sweedle Machado (Dept. Of Information Technology) for her invaluable collaboration, support, and assistance throughout the successful culmination of this endeavor. The constructive discussions that were facilitated by my colleagues and fellow researchers were immensely valuable in refining my concepts. For this, we are extremely grateful. Without the aforementioned individuals' contributions, this undertaking would have been unattainable.

REFERENCES

[1] Goo, C.-W., & Chen, Y.-N. (2018). Abstractive Dialogue Summarization with Sentence-Gated Modeling Optimized By Dialogue Acts. National Taiwan University, Taipei, Taiwan. MediaTek Inc., Hsinchu, Taiwan, September 29, 2018.

[2] Lin, C.-Y. (2004). ROUGE: A Package for Automatic Evaluation of Summaries. In Proceedings of the ACL Workshop on Text Summarization Branches Out (WAS 2004), Barcelona, Spain.

[3] Ganesh, P., & Dingliwal, S. (2020). Restructuring Conversations using Discourse Relations for Zero-shot Abstractive Dialogue Summarization, October 13, 2020.

[4] Shang, G., Ding, W., Zhang, Z., Tixier, A. J.-P., Meladianos, P., Vazirgiannis, M., & Lorre´, J.-P. (2018). Unsupervised Abstractive Meeting Summarization with Multi-Sentence Compression and Budgeted Submodular Maximization. Ecole Polytechnique, ´ Linagora, AUEB, November 14, 2018.

[5] Yusen Zhang, Ansong Ni, Tao Yu, Rui Zhang, Chenguang Zhu, Budhaditya Deb, Asli Celikyilmaz, Ahmed Hassan Awadallah, and Dragomir Radev. 2021. An Exploratory Study on Long Dialogue Summarization: What Works and What's Next. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4426–4433, Punta Cana, Dominican Republic. Association for Computational Linguistics.

[6] Carletta, J., Ashby, S., Bourban, S., Flynn, M., Guillemot, M., Hain, T., ... & Versley, Y. (2005).
The AMI Meeting Corpus: A Pre-announcement. In Proceedings of the International Workshop on Machine Learning for Multimodal Interaction (MLMI), Edinburgh, UK.

[7] Janin, A., Baron, D., Edwards, D., Daniele, M., Ellis, D., Gelbart, D., ... & Shriberg, E. (2003).
The ICSI Meeting Corpus. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Hong Kong, China.