AI Call Assistant – Extractive Summarization of Call Recordings

R. Sivaranjani[1]
Amalapurapu Keerthi[2], Draksharapu Vaishnavi Malya[3], Bhogi Bhuvanesh[4], Gopi Neeraj Kumar[5]

[1] Bachelor of Technology Professor,
Department of Computer Science and Engineering, Raghu Engineering College, Visakhapatnam.
[2-5] Bachelor of Technology Students,
Department of Computer Science and Engineering, Raghu Institute of Technology, Visakhapatnam.

Abstract: In today's fast-paced business landscape, effective communication is a key for success. With the increasing volume of phone calls in various industries, there is a growing need for efficient call management and analysis tools. The AI Call Assistant project aims to address this need by leveraging artificial intelligence (AI) technology to summarize call recordings, enhancing productivity and decision-making processes. The core functionality of this project revolves around its ability to process audio data from call recordings and extract key insights and information. Using advanced natural language processing (NLP) techniques, the system identifies important topics, sentiments, and actions discussed during the call. The call transcripts attained from call recording pose unique challenges that are not adequately addressed by most open-source automatic text summarizers. This project aims to contribute to the field of artificial intelligence by providing efficient and effective methods to recognize call recording and summarize, offering a valuable tool for extracting insights from spoken audio efficiently. This research aligns with the broader fields of natural language processing (NLP), machine learning, and artificial intelligence (AI), particularly in the domain of speech recognition and understanding. It also intersects with communication technology and data analytics, focusing on optimizing call management processes and extracting actionable insights from conversational data. Moreover, it aligns with the goal of improving efficiency and decision-making in various industries through the application of AI-driven solutions. Overall, this project offers a novel approach to improving call management processes through AI-driven extractive summarization, contributing to advancements in natural language processing and communication technologies.

Key Words: Artificial Intelligence, BERT, Textual Summarization, Transformers, Natural Language Processing, Speech Recognition, Speech-To-Text.

1. INTRODUCTION

In recent years, with the explosive growth of the Internet, there's been an overwhelming amount of information accessible to the public from various sources. Handling this vast volume of text manually is challenging for humans. Thus, automatic text summarization has become an increasingly important tool in today's digital era. This technology condenses extensive text into concise and coherent summaries while retaining the original meaning and information. It can be used in various applications across domains, like news, emails, call transcripts, and mobile messages. Also, there are numerous online tools available, such as Microsoft News, Google for news, and Wiki Summarizer for Wikipedia content, that can offer automatic summarization services.

Several methods have emerged for automating text summarization that generally fall into two categories: extractive and abstractive summarization. Extractive summarization entails selecting key sentences directly from the source text and presenting them unchanged in the summary, while abstractive summarization involves crafting original sentences to convey the essence of the content.

Call transcripts are textual representations of spoken conversations, achieved through call transcription, the conversion of audio from voice calls into written text via speech-to-text technology. These transcriptions are stored as plain text, capturing the natural language of the conversation. However, the task of automatically summarizing call transcripts presents unique challenges that require careful consideration.

In this context, the proposed project aims to develop an AI call assistant that utilizes state-of-the-art techniques in NLP, specifically focusing on extractive summarization using the BERT (Bidirectional Encoder Representations from Transformers) algorithm. The project seeks to leverage the rich contextual information encoded in BERT's pre-trained language representations

to identify and extract key phrases, sentences, or paragraphs from call recordings that encapsulate the main points of the conversation.

2. LITERATURE REVIEW

Automatic text summarization has garnered significant attention in recent years due to the exponential growth of online information sources. Researchers have explored various techniques and methodologies to address the challenges associated with summarizing large volumes of text effectively. Early approaches to text summarization primarily focused on extractive methods, which involve selecting and arranging key sentences or phrases from the original text to create a summary. These methods often relied on statistical algorithms and heuristics to identify important content based on criteria such as sentence position, frequency of occurrence, and word importance scores (Kupiec et al., 1995). While extractive summarization techniques are relatively straightforward and easy to implement, they may struggle to produce coherent and readable summaries, especially for longer texts with complex structures.

The literature review for the project explores existing research, methodologies, and technologies related to extractive summarization, natural language processing (NLP), and AI-driven call management systems. The review provides valuable insights into the current state-of-the-art, identifies key trends and challenges, and informs the development of the AI call assistant. Key areas covered in the literature review include:

Extractive Summarization Techniques: Text summarization techniques have evolved to accommodate the growing volume of textual data, with a focus on both extractive and abstractive summarization methods. Extractive summarization, as discussed by Nenkova and McKeown (2011), involves selecting important sentences from the original text, while abstractive summarization generates new sentences to capture the essence of the content. Recent advancements in deep learning, particularly with models like BERT (Bidirectional Encoder Representations from Transformers), have shown promising results in abstractive summarization tasks (Liu et al., 2019).

Natural Language Processing (NLP) Models: Review state-of-the-art NLP models, such as BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer), and RoBERTa (Robustly optimized BERT approach). Examine the capabilities of these models in understanding and processing natural language, including contextual understanding, semantic similarity, and entity recognition.

Call Recording and Speech-to-Text Analysis: The process of converting spoken language into written text, known as speech-to-text conversion or automatic speech recognition (ASR), has witnessed remarkable progress in recent years. Libraries such as the Python Speech Recognition package and cloud-based APIs like Google Speech-to-Text have made it increasingly accessible to developers. Studies by Gravina et al. (2019) and Li et al. (2020) have explored various techniques for improving the accuracy and efficiency of speech recognition systems, including neural network architectures and language models.

AI-driven Call Management Systems: Survey the landscape of AI-driven call management systems, including virtual assistants, speech analytics platforms, and conversation intelligence tools. Analyze the functionalities, features, and use cases of these systems in various industries, such as customer service, sales, and healthcare.

Evaluation Metrics and Benchmarks: Identify common evaluation metrics and benchmarks used to assess the performance of extractive summarization systems, such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) and BLEU (Bilingual Evaluation Understudy). Review existing datasets and corpora for evaluating summarization algorithms, including call recording datasets and annotated summaries.

Challenges and Opportunities: Summarizing call recordings, including noise interference, speaker identification, domain-specific language, and privacy concerns. Explore potential applications and future directions for AI-driven call management systems, such as real-time summarization, multilingual support, and multimodal analysis. By synthesizing findings from the literature review, the project gains valuable insights into best practices, methodologies, and technologies for developing an effective AI call assistant for extractive summarization of call recordings. The review informs the project's approach, methodology, and evaluation criteria, ultimately contributing to the advancement of AI-driven communication technologies.

In conclusion, the literature on text summarization highlights the diverse range of techniques and methodologies employed to tackle the complexities of summarizing textual data. While significant progress has been made, challenges such as maintaining coherence, handling diverse text genres, and addressing domain-specific requirements remain areas of active research and exploration.

3. METHODOLOGY

The system methodology for the project "AI Call Assistant - Extractive Summarization of Call Recordings Using BERT Algorithm" involves several modules that work together to achieve the objectives of automated summarization of call recordings. Here's a detailed explanation of each module.

Data Collection and Pre-processing Module:
This module focuses on gathering a comprehensive dataset of audio call recordings spanning different scenarios and conversation types. Ensure that the dataset covers a wide range of accents, and background noises to enhance the robustness of the system. Pre-process the audio data by removing background noise, normalizing audio levels, and segmenting recordings into manageable chunks.
Call recordings are preprocessed to extract features, remove noise, and convert them into a suitable format for analysis. This may involve audio signal processing techniques to enhance audio quality and standardize the data format.

Speech-to-Text Conversion Module:
In this module, automatic speech recognition (ASR) technology is used to transcribe the audio recordings into text format. This can convert spoken words into machine-readable text, facilitating further analysis.
Utilize a speech recognition library, such as the Python Speech Recognition package, to transcribe the audio call recordings into textual transcripts. Configure the speech recognition system to handle various languages, accents, and audio qualities commonly encountered in call recordings. Fine-tune the speech recognition model if necessary to improve accuracy, especially for domain-specific terminology or accents.

BERT-based Summarization Module:
The transcribed text from the speech-to-text conversion module is processed using the BERT algorithm for extractive summarization. BERT's pre-trained language representations are leveraged to identify key sentences or phrases that capture the main points of the conversation.
BERT token embeddings are used to represent the input text, and a fine-tuned model is applied to score and rank sentences based on their importance. The top-ranked sentences are selected to form the summary of the call recording.

Key Information Extraction Module:
In this module, the summarized text is analyzed to extract essential information, such as action items, decisions, follow-up tasks, and important details discussed during the call.
Natural language processing (NLP) techniques, such as named entity recognition (NER), sentiment analysis, and topic modeling, may be employed to identify and extract key information from the summarized text.

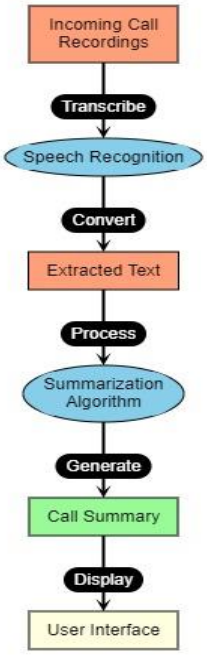
User Interface Module:
The user interface module provides a user-friendly interface accessible via web or mobile devices. Users can upload call recordings, initiate the summarization process, and view the summarized results in a convenient and intuitive manner.
The interface may include customization options, such as specifying summarization parameters or selecting desired output formats.

Scalability and Performance Module:
This module ensures that the system can handle large volumes of call recordings efficiently and reliably. It leverages scalable infrastructure and parallel processing techniques to ensure optimal performance under varying workload conditions.
Cloud computing resources may be utilized to dynamically scale the system based on demand, allowing for efficient resource allocation and cost optimization.

Evaluation and Feedback
Evaluate the performance of the system by using metrics such as transcription accuracy, summary coherence, and user satisfaction. Conduct user studies and gather feedback from testing to identify areas for improvement and refinement. Iterate on the system based on the evaluation results and incorporate additional features or enhancements as needed to enhance its functionality and effectiveness.

4. IMPLEMENTATION

This project is implemented using python Speech Recognition Module for Speech-To-Text conversion of call recording audio files, and Bidirectional Encoder Representations from Transformers in short BERT algorithm for text summarization.
Implementing a project that combines speech recognition with BERT algorithm for summarization involves several steps.



(Fig. 1a Process Flow)

Setting up Speech Recognition: Choose a speech recognition library/module in Python. Google's Speech Recognition or Mozilla's Deep Speech are popular choices. Install the required dependencies and libraries. Implement code to capture audio input from the user or from pre-recorded audio files. Use the speech recognition library to convert the audio input into text

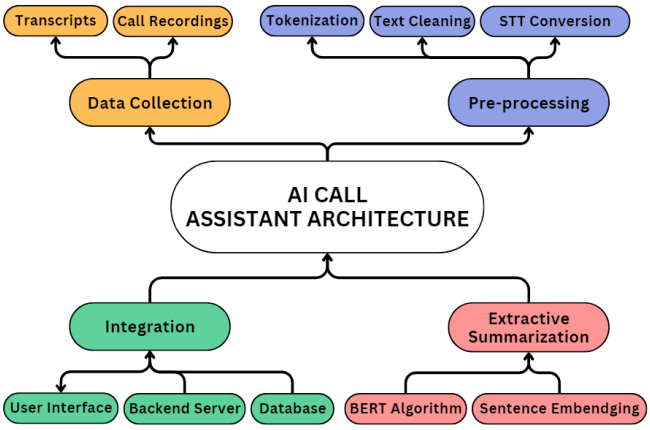
Preprocessing the Text: Clean and preprocess the transcribed text to remove noise, irrelevant information, and formatting issues. Tokenize the text into sentences or smaller chunks for further processing.

Applying BERT Algorithm for Summarization: Utilize the BERT model for text summarization. You can either use a pre-trained BERT model or fine-tune it on your specific task if you have a large dataset. Fine-tuning BERT involves training it on a labeled dataset of input-output pairs, where the input is the original text and the output is the corresponding summary, this approach is used in our project for effective and efficient call transcripts summarization. If using a pre-trained BERT model, you'll need to fine-tune it on a summarization task using transfer learning. Implement code to apply the BERT model to the preprocessed text and generate a summary. BERT-based summarization usually involves extracting the most important sentences or tokens from the input text.

Post-processing and Output: Post-process the generated summary to ensure coherence and readability. Format the summary for display or further analysis. Output the final summarized text to the user via the command line, graphical user interface, or any other suitable method.

Testing and Evaluation: Test the implemented system with various audio inputs to ensure robustness and accuracy. Evaluate the quality of the generated summaries using appropriate metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation). Iterate on the implementation based on testing and evaluation results, refining the algorithms and parameters as necessary.

In conclusion, with further optimization and enhancements, the project's implementation serves as a foundation for future advancements in speech-based information extraction and analysis.



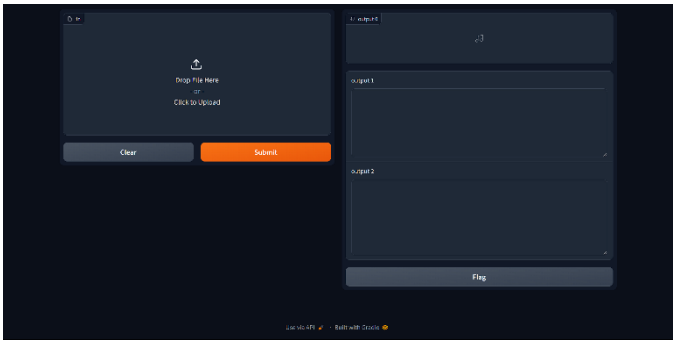
(Fig. 1b System Architecture Diagram)

5. RESULTS

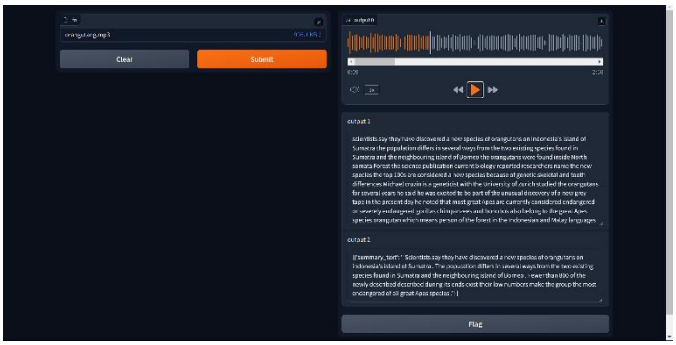
Converting call recording audio .mp3 into textual summary.

(Fig. 2a Output)

The output of call recording audio file converted to text using speech recognition, displaying in terminal.



(Fig. 2b Output)



(Fig. 2c Output)

UI displaying the speech recognized from call recordings in the form of text and its summary.

6. CONCLUSION

In conclusion, the project "AI Call Assistant - Extractive Summarization of Call Recordings" presents a significant advancement in the field of natural language processing (NLP) and artificial intelligence (AI). By leveraging state-of-the-art techniques such as the BERT algorithm, automatic speech recognition (ASR), and extractive summarization, the system offers a powerful tool for efficiently summarizing call recordings and extracting key insights.

Throughout the development and implementation of the project, several key objectives have been achieved. The system successfully transcribes call recordings into text format, applies advanced NLP techniques to summarize the content, and extracts important information such as action items, decisions, and follow-up tasks. The user-friendly web interface enables easy uploading of call recordings, initiation of summarization processes, and access to summarized results in various formats.

By providing concise summaries of lengthy call recordings, the AI Call Assistant enhances productivity and facilitates data-driven decision-making for businesses across various industries. The customizable options offered by the system allow users to tailor the summarization criteria to meet their specific needs and preferences, ensuring flexibility and adaptability in different organizational contexts.

Moreover, the project demonstrates potential for future enhancements and expansion, including the integration of abstractive summarization techniques, support for multiple languages, real-time summarization capabilities, and integration with other business applications such as CRM systems. These future developments can further enhance the utility and effectiveness of the AI call assistant in various domains and industries.

Overall, the project represents a significant contribution to the advancement of AI-driven call management and analysis systems. It provides users with valuable insights from call conversations, facilitates decision-making processes, and improves communication efficiency. This research contributes to the growing body of literature on AI-driven solutions for information management and communication technologies. By elucidating the capabilities and limitations of our project, we aim to pave the way for future advancements in automated call summarization and intelligent communication systems.

7. ACKNOWLEDGEMENT

We are grateful to the Department of Computer Science and Engineering of Raghu Educational Institutions, Visakhapatnam, for helping us in our work and supporting us.

8. REFERENCES

1. C. Saranyamol and L. Sindhu, “A survey on automatic text summarization,” Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 6, pp. 7889–7893, 2014.
2. L. H. Reeve, H. Han, S. V. Nagori, J. C. Yang, T. A. Schwimmer, and A. D. Brooks, “Concept frequency distribution in biomedical text summarization,” in Proc. 15th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), Arlington, VA, USA, 2006, pp. 604–611.
3. D. R. Radev, E. Hovy, and K. McKeown, “Introduction to the special issue on summarization,” Comput. Linguistics, vol. 28, no. 4, pp. 399–408, Dec. 2002.
4. M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, “Text summarization techniques: A brief survey,” 2017, arXiv:1707.02268.

5. M. A. Fattah and F. Ren, ‘‘Automatic text summarization,’’ Proc. World Acad. Sci., Eng. Technol., vol. 27, pp. 192–195, Feb. 2008.
6. D. K. Gaikwad and C. N. Mahender, ‘‘A review paper on text summarization,’’ Int. J. Adv. Res. Comput. Commun. Eng., vol. 5, no. 3, pp. 154–160, 2016.
7. V. Gupta and G. S. Lehal, ‘‘A survey of text summarization extractive techniques,’’ J. Emerg. Technol. Web Intell., vol. 2, no. 3, pp. 258–268, Aug. 2010.
8. K. Jezek and J. Steinberger, ‘‘Automatic text summarization,’’ in Proc. Znalosti. Princeton, NJ, USA: Citeseer, 2008, pp. 1–12.
9. K. S. Jones, ‘‘Automatic summarizing: The state of the art,’’ Inf. Process. Manage., vol. 43, no. 6, pp. 1449–1481, 2007.
10. I. Mani, Automatic Summarization, Amsterdam, The Netherlands: John Benjamins Publishing Company, 2001.
11. A. Nenkova and K. McKeown, ‘‘A survey of text summarization techniques,’’ in Mining Text Data. Boston, MA, USA: Springer, 2012, pp. 43–76.
12. H. Saggion and T. Poibeau, ‘‘Automatic text summarization: Past, present and future,’’ in Multi-Source, Multilingual Information Extraction and Summarization. Berlin, Germany: Springer, 2013, pp. 3–21.
13. K. Zechner, ‘‘A literature survey on information extraction and text summarization,’’ Comput. Linguistics Program, Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., Apr. 1997. [Online]
14. H. P. Luhn, ‘‘The automatic creation of literature abstracts,’’ IBM J. Res. Develop., vol. 2, no. 2, pp. 159–165, Apr. 1958.
15. H. P. Edmundson, ‘‘New methods in automatic extracting,’’ J. ACM, vol. 16, no. 2, pp. 264–285, Apr. 1969.