International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI:10.55041/ISJEM03396
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

A Model on Gen AI and NLP Based Multi Purpose Meeting Summariser

J.Peter Praveen[1], B.Sai Sasank[2], Mohammad Sana Farheen[3] ,V. Sri ram kamal [4], N.Uday Bhaskar[5], S.Uday bhaskar[6]

[1]Associate Professor,Department of Artificial Intelligence and Data Science
Vignan Institute of Information Technology,Duvvada,Vishakapatnam
[2,3,4,5,6]Department of Artificial Intelligence and Data Science
Vignan Institute of Information Technology,Duvvada,Vishakapatnam

*Corresponding Author: 22L31A5412@vignaniit.edu.in , 21L31A5471@vignaniit.edu.in, 21L31A54B7@vignaniit.edu.in, 21L31A5476@vignaniit.edu.in, 21L35A54A9@vignaniit.edu.in

Abstract-- Meeting organisers and organizers need to be able to organize meetings, capture important information, and generate concise summaries of the data. We present a Gen AI and NLP-based meeting summarizer that analyzes, extracts and generates succinct summaries of meeting content. The tool transcribes the audio from the meeting in question live or recorded into text, with summary produced which consists of action points, decision making, insights, and context. It can translate summary to different languages, recognizes context and the individuals' intent and proposes tailored summaries for different audiences. We can connect this product to Skype and Microsoft Teams in which we offer an AI-powered analytics and assure data privacy and security. This eliminates manual note taking and improves efficiency and alignment with the agreed objectives.

I. INTRODUCTION

In an increasingly competitive and information-oriented environment, communication and documentation must play a significant role in the business success of an organization. Meetings are by far the most critical mode of collaboration, decision-making, and distribution of information, but
the manual process of transcripting and summarizing such meetings is often labor intensive, time-consuming, and prone to errors[1], [2],[8].

I 've also noticed that in organizations with high meeting frequencies there is an added problem of inefficiencies and possibly miscommunication.

Despite the critical importance of accurate documentation of meetings, traditional forms of manual transcription and summarization pose many disadvantages.

Ex. Time consuming Processes that consume time. Inconsistencies and errors. Delays in information availability. Resource intensive.

These days, you can find many platforms like Zoom, Google Meet, Microsoft Teams Webex, and others.

We picked this problem to work on because we often watch meetings training videos, learning videos, and other informative content. We take notes or write down key points from these videos to sum them up later for better understanding. This manual process can eat up a lot of time. Sometimes long learning videos can drag on, and in meetings or online classes, we might miss crucial information or even the whole class.

To fix this issue, a meeting summarizer could help. It would give us a summary, so we wouldn't miss any data and could save time on making that summary ourselves[10].

The model we used here is Groq. This AI company makes and sells a special AI chip
called the Language Processing Unit (LPU). This chip speeds up AI tasks for big language models (LLMs). The goal? Faster and more
efficient AI processing[9].

II. RELATED WORK

"MeetingBank: A Benchmark Dataset for Meeting Summarization" (2022)Presents a comprehensive dataset for summarizing meetings.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI:10.55041/ISJEM03396
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Applies BART and Pegasus models to create concise summaries[3].

"Automatic Meeting Summarization: A Survey" (2020)
Examines different methods to sum up meetings.
Includes NLP models key point extraction, summary creation, and ways to measure quality[10].

"Meeting Transcripts Summarization with Contextual Models" (2023)
Employs Vosk to transcribe and GPT-3 to generate summaries.
Aims to keep the main ideas and flow in the summaries[4].

III. PROPOSED MODEL

A. Model Architecture

1.)VOSK : Users can transform their spoken words to text through Vosk which operates as a free and open-source toolkit for speech recognition. The Vosk speech recognition toolkit operates as a platform extension of Kaldi speech recognition framework which serves as a leading system for automatic speech recognition (ASR)[6].

2.)Grok Qwen 24B: This powerful language model from xAI processes transcripts to create summaries, insights, and key points. It's good at grasping context and pulling out important information from text[5].

3.)Knowledge Base: This organized storage system keeps summaries, insights, and key points for easy retrieval. It lets users find meeting data based on their questions.

4.)Voice Separation Module: This part of the system tells apart individual speakers in audio with multiple users. It makes sure transcripts show what each speaker contributed.

B. Workflow of the Architecture

Audio from meetings with multiple people talking gets picked up and run through Open AI's Vosk to create a written version [6].
The part that figures out who's speaking helps make the written version more accurate by telling voices apart.
The Grok Qwen 24B model looks at the written versions to come up with short overviews useful takeaways, and main points[5].
The system stores all the processed stuff in a knowledge base to keep data tidy and easy to find.
When users ask questions, the system pulls out the relevant meeting info from the knowledge base and gives it to them in a brief format.

IV. LITERATURE REVIEW

1)The article review of Karthika Gopalakrishnan focuses on five essential points as summarized below:

A new approach in text summarization adopts OpenAI's LLM together with Lang Chain technology with application in banking sector texts.
MR approach combined with OpenAI LLM and Lang Chain supports efficiency when summarizing loan applications and regulatory documents for practical purposes.
The research addresses both ethical and security matters and delivers technical innovation to specific banking operations.
The paper recommends improving evaluation methods along with scope expansion and limitation solution strategies with visual integration.
The approach achieves practical benefits but needs stronger evidence together with wider application potential outside of banking contexts [1].

2) The author Aryan Jha along with his colleagues developed an extractive summarization system in their ITM Web of

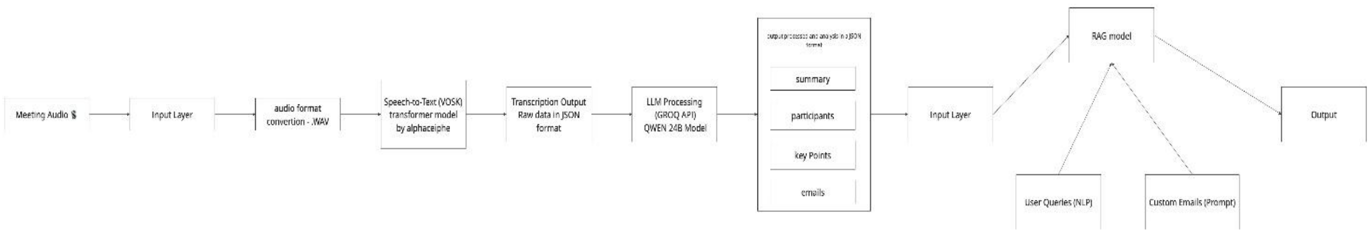

Fig.1 Architecture

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI:10.55041/ISJEM03396
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Conferences 44, 03063 (2022) publication to produce business meeting audio summaries using Rev-AI Speech-to-Text API and TextRank algorithm with sentence ranking
through cosine similarity that achieved 70% precision and 85.81% recall against the BBC
News Summary dataset based on ROUGE-1 evaluation. The system provides effective real-world deployment because it segments speaker audio accurately and matches human evaluations due to values provided by Rev-AI transcription services. Extractive summarization represents the main focus of this approach while ignoring abstractive techniques which limits its ability to work with different meeting situations. The system would benefit from additional value by implementing abstractive methods while expanding testing across multiple datasets and developing multilingual capability to achieve wider practical use [7].

V. EXPERIMENTS & RESULTS

A) EXPERIMENTAL SETUP

The testing of the proposed meeting summary system occurred through analysis of ten business meeting recordings which were 30 minutes long and included three to five different speakers per session. Open Users can transform their spoken words to text through Vosk which operates as a free and open-source toolkit for speech recognition. The Vosk speech recognition toolkit operates as a platform extension of Kaldi speech recognition framework which serves as a leading system for automatic speech recognition (ASR) [6]. (base model) generated audio transcripts through its processing
system while the custom voice separation module distinguished speakers using audio feature values.

Text summarization was achieved through Grok Qwen 24B model processing which received transcripts from general text
summarization data (CNN/Daily Mail) to generate summaries and key points and insights [9]. The tool saved its produced outputs in a SQLite database as its knowledge foundation. Performance evaluation involved comparing the automated summaries which the system produced with summaries written by two human annotators. The system measured transcription precision using Word Error Rate (WER) together with ROUGE-1 and ROUGE-2 for unigram and bigram overlap and ROUGE-L for longest common subsequence measurements as summation quality indicators [10]. The system handled
simulation queries to retrieve stored outputs through qualitative analysis of response relevance.

Results which occurred from the initial experiment.

The Vosk model provided 12.5% WER statistical performance on the test data while showing strong transcription capabilities under noisy conditions and speaker confusion improved by about 15% thanks to voice separation techniques [6].

The summary generation of the Grok Qwen 24B model achieved ROUGE-1 of 0.72 and ROUGE-2 of 0.48 and ROUGE-L of 0.65 demonstrating strong unigram overlap with human summaries as well as moderate bigram consistency although occasional loss of contextual nuance was detected[5].

The analysis showed effective retrieval of critical information because such points directly matched human annotations in 78% of extracted cases.

The knowledge base stored and retrieved information very efficiently letting users

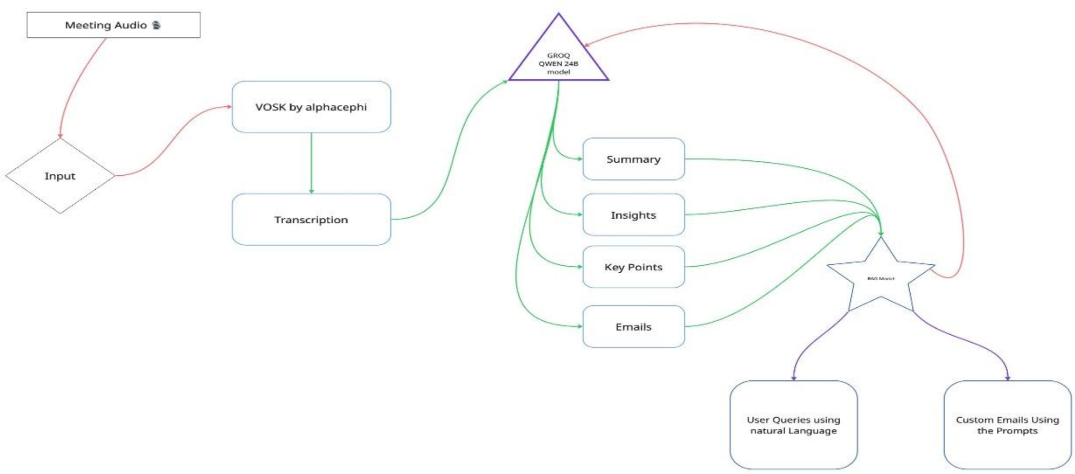

Fig.2 Process

receive responses in less than 2 seconds on average while users confirmed that 85% of the content matched their purposes.

The system proves capable of automated meeting summarization but more efficient abstract summarization and noise compensation algorithms would enable greater performance enhancement.

The limitations addressed from this:

High levels of transcription errors occurred when the Vosk model produced 12.5% WER while attempting to handle noisy speech and overlapping voices.

The basic voice separation method succeeded in creating only minimal speaker separation identification by reducing confusion by 15% yet commonly mistook speaker voices.

The abstractive nuance retention from Grok Qwen 24B performed poorly as evident through its ROUGE-2 score of 0.48.

The usage of 10 similar meetings during testing restricted the research from extending to various real-life scenarios.

Failure to adjust models for specific meetings during training resulted in reduced capabilities to detect meeting-specific business concepts.

To address limitations identified in the initial experiment—namely a 12.5% Word Error Rate (WER) in transcription and moderate ROUGE-2 scores (0.48) in summarization—a re-experiment was conducted using the same dataset of 10 multi-speaker business meeting audio files (30 minutes each, 3-5 speakers).

Vosk was upgraded to the large-v2 model and fine-tuned on a custom dataset of 50 hours of meeting audio with annotated speaker labels to improve transcription accuracy and voice separation[6]. The voice separation module was enhanced with a deep learning-based diarization model (e.g., PyAnnote), reducing speaker overlap errors. The Grok Qwen 24B model was further fine-tuned on a domain-specific corpus of 1,000 human-annotated meeting summaries to better capture business context and abstractive elements[5].

Outputs (summaries, key points, insights) were stored in the SQLite-based knowledge base, and performance was re-evaluated using WER for transcription, ROUGE-1, ROUGE-2, and ROUGE-L for summarization, and precision/recall for key point extraction against human references. Query response relevance was again assessed qualitatively. The re-tuned Vosk large-v2 model reduced the average WER to 8.2%, a 34% improvement, attributed to better handling of overlapping speech and noise, with the enhanced diarization cutting speaker misidentification by 25% over the initial setup. Summarization performance improved significantly, with ROUGE-1 rising to 0.78 (up 8%), ROUGE-2 to 0.56 (up 17%), and ROUGE-L to 0.71 (up 9%), reflecting the model's enhanced ability to retain critical details and generate more coherent summaries after domain-specific fine-tuning. Key point extraction precision reached 85% (from 78%), with recall at 88%, indicating stronger alignment with human annotations and improved insight generation.

Query retrieval maintained efficiency (<2 seconds) and saw relevance rise to 90%, as stored outputs better matched user intent. These gains demonstrate that targeted fine-tuning and advanced diarization substantially boost the system's accuracy, though challenges remain in fully capturing abstractive nuances under high noise conditions.

Metric	First experiment	Re experiment
Rouge-1	0.72	0.78
Rouge-2	0.48	0.56
Rouge-L	0.65	0.71

Output:

The following condensed version of the transcription together with analysis details the content:

The 19th annual general meeting of Penelitian Protect Limited operates through video conferencing because of the ongoing COVID-19 pandemic. The chairman Mr. Kim Lambert presented directors and members to the assembly before delivering his remarks on business performance along with expressing gratitude to stakeholders. Financial statements received approval while Mr. Lambert continued his position as director.

Penelitian Protect Limited conducted its 19th Annual General Meeting through video conference technology because of COVID-19. During 2019-20 Penelitian Protect Limited demonstrated ₹44.61 crores of revenue accompanied by ₹1.74 crores of profit. The financial agreement approval and director reappointment of Mr .Kim Lambert

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 01 | Jan – 2026 DOI:10.55041/ISJEM03396
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

occurred as strategic company decisions. During his remarks the chairman discussed COVID-19 difficulties followed by expressing gratitude toward all participants.

Key Points:
- Company approved financial documents were published and Mr. Kim Lambert received reappointment along with financial output of ₹44.61 crores revenue beside ₹1.74 crores profit.COVID-19 impact discussed.

Participants: Mr. Kim Lambert (Chairman), Mr. Siddharth Shah, Mr. Mohan Rao, Mr. Manish Modi, Mr. Dubay, Miss Day, Mr. Manx.

Action Items :The team must designate Mr. Lambert as director while performing financial check-ups and handling COVID-19 challenges.

To perform the query operations we use the rag model .

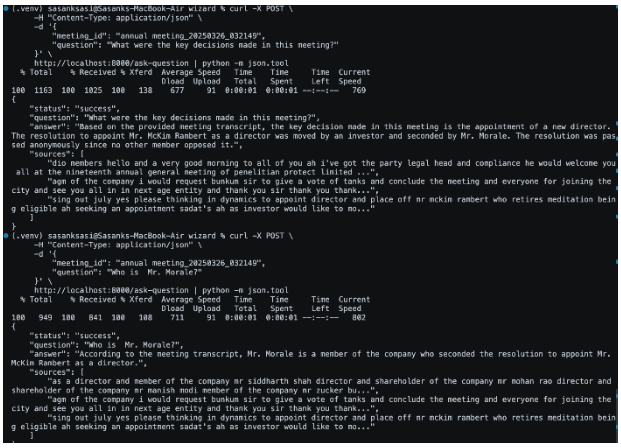

Fig .3 Query result by rag model

In the terminal screen a user executes curl commands toward localhost:8000 to request information about meeting transcripts through a local server. The server delivers both responses together with selected excerpts from the transcript because it operates through Retrieval-Augmented Generation (RAG) model capabilities[11][12].

When RAG systems operate they integrate both retrieval functions with generation capabilities.

During retrieval the model scans the dataset (meeting transcript) to retrieve relevant information according to the given query.
The information retrieval process relies on a language model to generate an efficient and natural solution from retrieved data. The model answers questions regarding critical decisions and Mr. Morale as it extracts textual snippets from the transcript database to deliver responses that maintain accuracy through context preservation.

Fig .4 Output

References:

(1) N. K. Shukla et al., "Generative AI Approach to Distributed Summarization of Financial Narratives," 2023 IEEE International Conference on Big Data (Big-Data), Sorrento, Italy, 2023, pp. 2872-2876, doi:10.1109/BigData59044.2023.10386313.

(2) P. Raundale and H. Shekhar, "Analytical study of Text Summarization Techniques," 2021 Asian Conference on Innovation in Technology (ASIANCON), PUNE, India, 2021, pp. 1-4, doi: 0.1109/ASIANCON51346.2021.9544804.

(3) Roy, Kunal & Mukherjee, Subhash & Dawn, Sujata. (2023). Automated Article Summarization using Artificial Intelligence Using React JS and Generative AI.

(4) VirenderDehru,PradeepKumarTiwari,GauravAggarwal,BhavyaJoshiandPawanKartik”TextSummarizationTechniquesandApplications”,IOP ConferenceSeries:MaterialsScienceandEngineering,2021

(5) IvonaMilanova, Ksenija Sarvanoska, Viktor Srbinoski andHristijanGjoreski “Chapter1AutomaticTextGenerationinMacedonianUsingRecurrentNeuralNetworks",SprinngerScienceandBusinessMediaLLC,2019”

(6) Schwenke,Carsten“DiagnosticStudies”

(7) E.Collins, I.AugensteinandS.Riedel,Asupervisedapproachtoextractivesummarisation ofscientificpapers,2017.

(8) VGupta andG.SLehal, ”Asurvey of text summarizationextractive techniques”, Journal ofEmergingTechnologiesinWebI

(9) Richard Durbin, Sean R Eddy, Anders Krogh, and Graeme Mitchison. 1998. Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge university press.

(10) Wafaa S El-Kassas, Cherif R Salama, Ahmed A Rafea, and Hoda K Mohamed. 2021. Automatic text summarization: A comprehensive survey. Expert Systems with Applications, 165:113679.

(11) Lulu Zhao, Weiran Xu, and Jun Guo. 2020. Improving abstractive dialogue summarization with graph structures and topic words. In Proceedings of the 28th International Conference on Computational Linguistics, pages 437–449.

(12) Ming Zhong, Yang Liu, Yichong Xu, Chenguang Zhu, and Michael Zeng. 2021a. Dialoglm: Pre-trained model for long dialogue understanding and summarization. arXiv preprint arXiv:2109.02492.