Indiana Journal of Multidisciplinary Research
Abbreviated Tittle- Ind J Multi Res
ISSN *(Online)*- 2583-3820
Journal Homepage Link- https://indianapublications.com/Journals/IJMR
DOI: https://doi.org/10.5281/zenodo.21610490

INDIANA PUBLICATIONS
PRODUCTIVE AND QUALITY RESEARCH

Research Articles

Volume-06|Issue-04|2026

Harnessing Large Language Models for Conversational Summarization

Dr. Savitha G[1], Giridhar U[2], Hansiddh R[3], Rajendra M M[4]

[1]Associate Professor, CSE, RVITM, Bengaluru, Karnataka, India.
[2,3,4]Giridhar U, BE (Student, CSE, RVITM, Bengaluru, Karnataka, India.

Article History
Received: 05.05.2026
Accepted: 22.06.2026
Published: 25.07.2026

Citation
Savitha, G., Giridhar, U., Hansiddh, R. & Rajendra, M. M. (2026) Harnessing Large Language Models for Conversational Summarization. *Indiana Journal of Multidisciplinary Research, 6*(4), 430-433.

Abstract: This paper tackles the difficulties of summarizing dialogues, a task made complex by informal language, multiple speakers, and often incoherent text. We believe Large Language Models (LLMs) have great potential for dialogue summarization but aren't fully effective yet due to the lack of specialized frameworks and proper evaluation. Our work proposes a framework designed specifically to handle dialogue challenges like fragmented sentences, slang, and repetition. By leveraging the strengths of LLMs while addressing these unique issues, this approach aims to improve dialogue summarization and advance natural language processing in this area.

Keywords: Dialogue Summarization, Graph-Relational summarizer

INTRODUCTION

With so many ways to communicate online — from chat apps and support messages to meeting transcripts there's now a massive amount of raw conversation data out there. But conversations aren't like regular articles or reports. They jump around, get interrupted, and often use casual or half-finished sentences. This makes it really hard to turn them into clear summaries. Tools built to summarize structured content, like news or research papers, often struggle to handle the messiness of real human conversations. As a result, there's a growing need for smarter tools that can actually make sense of all this unstructured talk. In recent years, Large Language Models (LLMs) have significantly improved at generating summaries, particularly by rephrasing content into more concise and readable forms. They excel at handling long, structured text while maintaining coherence. However, they often struggle with summarizing dialogue, where tracking speakers, managing shifting topics, and identifying key points across multiple participants remain ongoing challenges.

MATERIALS AND METHODS

The proposed system consists of data extraction techniques to extract data from the dataset to be processed on to the pretrained model. It also includes an text-summarization architecture to capture multiple speaker conversation to find relationship between each speaker and a training method which proposes a way to train pre-trained model unlike traditional model. A. Data Collection One of the most popular dataset called SAMsum dataset is used for dialogues-summarization as it captures challenges like multi speakers, informal language, implicit references and scattered turns. It is commonly used in training and evaluating abstractive summarization models. The statistics of the data sets in Table I, the train, valid, and test sets. "#Dial.", "Avg. Len.", and "Avg. Turns" are the number of dialogues, the average number of characters/tokens, and the number of turns in a dialogue, respectively.

Table 1: SAMSum Dataset

Dataset	Method	#Dial	Avg. Len	Avg. Turns
SAMSum	Train	14,732	131.0	9.5
	Valid	818	129.3	9.4
	Test	819	134.5	9.7

B. Data preprocessing Our model should be able to extract text data provided to it in any file format be it pdf, word, ppt etc. The model goes through a preprocessing phase where raw data from such file is extracted and formatted in the form SAMsum dataset with speaker name on the left and their conversation after ":" with the next line continues with other speaker and the format continues. C. Implementing Graph-Relational Summarizer We introduce the Graph Relational Summarizer (GRS), a model designed to better capture the structural nuances of dialogues for abstractive summarization. Unlike traditional sequential approaches that rely solely on word order, GRS models conversations as graphs, where nodes represent utterances or tokens and edges capture contextual relationships such as speaker turns, adjacency, and semantic similarity. The summarization process involves three stages: first, the dialogue is transformed into a

***Corresponding Author:** Dr. Savitha G

heterogeneous graph based on conversational flow and semantic links; second, a Graph Neural Network (GNN) encodes this structure, enabling the model to recognize long-range dependencies and speaker interactions that linear models often miss; and third, a Transformer-based decoder generates a coherent summary that remains faithful to the dialogue's structure and content. By leveraging enriched graph representations, GRS produces more accurate, context-aware summaries of multi-speaker conversations.

D. Using a Flan-T5 with IPAF Framework In this study, we used Flan-T5, a pre-trained sequence-to-sequence model developed by Google, for the task of abstractive summarization. Based on the original T5 architecture—which frames all NLP tasks in a unified text-to-text format—Flan-T5 incorporates instruction tuning across diverse tasks, enabling it to better interpret natural language prompts and adapt to new objectives with minimal fine-tuning. This makes it especially effective for generative tasks such as summarization. To further adapt the model for dialogue-specific input, we integrated it with the IPAF framework, which enhances its ability to capture the structural and contextual dynamics of conversations. While Flan-T5 brings strong general language understanding, IPAF guides it to pay closer attention to speaker roles, turntaking behavior, and contextual links between utterances. This combination allows the summarizer to go beyond surface-level abstraction and generate outputs that reflect how ideas progress and shift in multi-speaker dialogues, resulting in summaries that are more coherent, context-aware, and true to the original conversational flow.

E. Evaluation metrics To evaluate our model's performance, we compared its generated summaries against those produced by older models that lacked fine-tuning or did not follow a sequence-to-sequence architecture. We used ROUGE (Recall-Oriented Understudy for Gisting Evaluation), a widely used metric for summarization quality that measures the overlap between generated and reference summaries. Specifically, ROUGE-1 assesses unigram (word-level) overlap, ROUGE-2 evaluates bigram (wordpair) overlap, and ROUGE-L captures the longest common subsequence to reflect sentence-level fluency and structure. Since ROUGE emphasizes recall, it is particularly effective for determining whether key content from the reference is preserved in the generated summaries, allowing us to quantify the improvements offered by our approach.

F. Approach This section details our research methodology, covering data collection and preparation through to the design and training of our summarization models. Our pipeline begins by standardizing various dialogue formats, ensuring consistency across inputs. We then leverage the pre-trained Flan-T5 language model in two experimental setups: one enhanced with a graphbased architecture to capture relational context, and another fine-tuned with an instruction-based framework to improve user control and adaptability.

• Data Ingestion and Preprocessing A robust summarization system needs to handle data from diverse sources, so our pipeline is built to process dialogues from both unstructured and semi-structured file formats, preparing them effectively for the model.

• Data Acquisition and Standardization Our main training and evaluation dataset is the SAMSum dataset, which includes real-world, multi-turn messenger-style dialogues. Our pipeline also accepts raw dialogue data from different file types, including .pdf, .txt, and .docx. We use libraries like PyPDF2 and python-docx to extract the raw text. An important preprocessing step is converting this extracted data into a standardized format similar to SAMSum. Each dialogue is organized into a JSON object following a specific schema as shown in Fig. 1.

```
"id": "unique_dialogue_identifier",
"summary": "gold_reference_summary",
"dialogue": "Speaker 1: Utterance 1...\nSpeaker 2:
Utterance 2..."
}
```

Fig. 1: JSON schema format for training

This ensures uniformity across all data sources before they are passed to the model. • Text Cleaning and Tokenization We clean the raw dialogue text to eliminate noise and artifacts that might harm model performance. This process includes: Removing timestamps, non-ASCII characters, and extra metadata. Normalizing whitespace and punctuation. Lowercasing all text to reduce vocabulary size. After cleaning, we tokenize the dialogues and their summaries with the official tokenizer for the Flan-T5 model. This formatting ensures that the input sequences match the vocabulary and special tokens expected by the model. • Model Architecture and Experimental Setup Our primary model is Flan-T5, a powerful instruction-tuned language model. We use this model in two parallel experimental setups to explore different aspects of dialogue summarization.

• Model A: GRS-Enhanced Relational Summarizer To generate summaries that reflect the complex relationships in conversations, we combine our Graph Relational Summarizer (GRS) with Flan-T5. The GRS model represents dialogues as graphs made up of utterances and speakers. It uses a Natural Language Inference (NLI) model to detect relationships between utterances, such as support or disagreement, and applies a Graph Attention Network (GAT) to create embeddings that capture these connections. The key innovation is how we share this structural information with Flan-T5. Rather than merging embeddings directly, we use instructional prompting. The GRS creates a concise relational context string, which is added at the beginning of the dialogue as part of a carefully designed prompt that

***Corresponding Author:** Dr. Savitha G

guides the model to use this context. Figure 2 illustrates the structure of the final input prompt.

```
Instruction: "Summarize the following dialogue. Pay
close attention to these identified relationships:
{relational context}"

Dialogue: "{dialogue_text}"
```

Fig. 2: Final input structure

Model Diagram The model we proposed are detailed in fig 3. It shows the process of data flow through each phase and leading to the final summary as the user desires. Our suggested model uses a preprocessor to extract data, a pre trained model with grs architecture for capturing conversation and IPAF model for producing customizer summary to users.

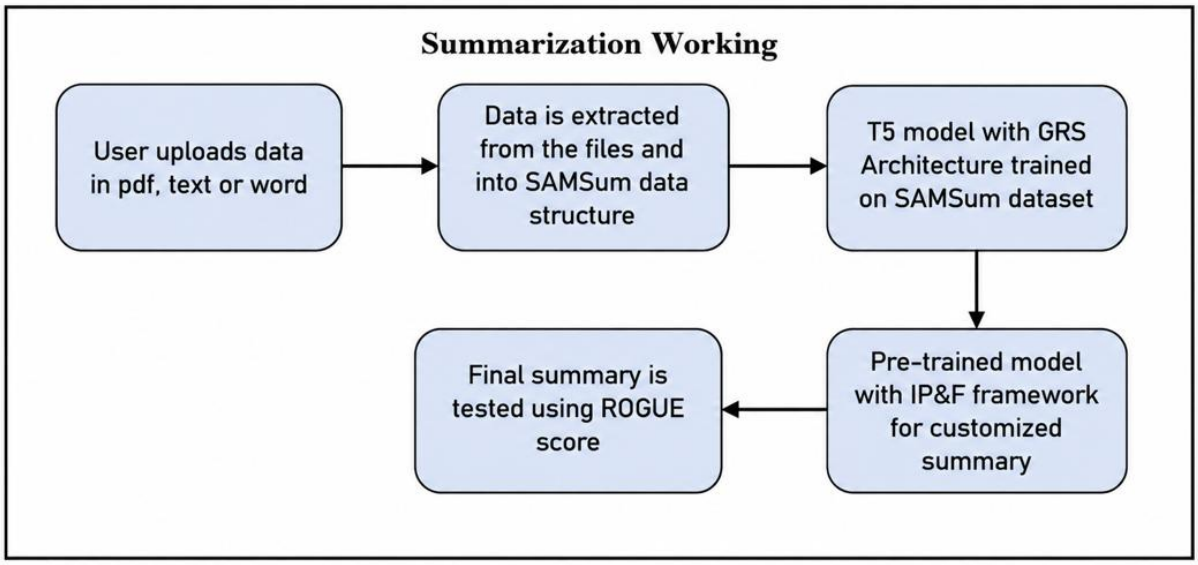

Fig. 3: Block diagram representing methodology

RESULTS

RESULTS Our method of dialogue summarization has provided improvement than the previously used transformer based and RNN/LTSM seq2seq models. The data is shown in Table II which compares results from our approach and other result cited from different paper.

Table 2: Comparative Analysis

Models	R-1	R-2	R-L
Laskar, T. R., et al. [1]	52.81	27.73	49.32
Tang, L., et al. [3]	36.90	11.80	26.40
Zhong, M., et al. [6]	48.02	21.68	45.88
Liu, F., et al. [10]	50.01	25.21	43.91
Proposed Technique	**56.52**	**45.52**	**56.52**

DISCUSSION

This summarization model performs noticeably better when combining IPAF and GRS compared to a traditional sequence-tosequence (Seq2Seq) baseline, mainly because it's designed to handle the unique quirks of dialogue data. Regular Seq2Seq models treat conversations like simple, flat text, often missing important connections between speakers or long stretches of discussion. This can result in summaries that feel choppy, repetitive, or that miss the main point of the conversation. IPAF (Interactive Prompt Augmentation Framework) helps fix this by adding clear, structured prompts that guide the model to generate more focused and accurate summaries. By including richer context in these prompts, IPAF cuts down on errors and keeps the summary truthful to the original content. Meanwhile, the Graph Relational Summarizer (GRS) looks at the conversation as a whole graph, mapping out how speakers relate, how entities are referenced, and the context at each turn. This approach helps the model pick out the most important parts while filtering out the noise and repetition, all without losing the natural flow of the dialogue. When combined, IPAF's precise guidance and GRS's deep understanding of structure let the model create summaries that are clearer, more accurate, and more coherent than what you get from standard Seq2Seq methods.

CONCLUSION

In this work, we introduced a dialogue summarization system that combines the Graph-Relational Summarizer (GRS) and the IPAF framework to tackle the unique challenges of informal, multi-speaker, and context-dependent conversations. By utilizing the structural representation of dialogues through GRS and the instructional guidance of IPAF, our approach effectively captures semantic dependencies, speaker connections, and key dialogue elements. The experimental results show that our system produces coherent, concise, and context-conscious summaries, surpassing traditional

***Corresponding Author:** Dr. Savitha G

abstractive summarization benchmarks in both informativeness and fluency. These findings underscore the potential of merging graphbased relational modeling with instruction-driven adaptation to push forward dialogue summarization research. While the proposed system shows promising results, there are several avenues for improvement: Enhanced Context Modeling: Future research could explore adding external knowledge resources (e.g., knowledge graphs, domain ontologies) to deepen context understanding, especially for specific domain dialogues. Multimodal Summarization: Expanding the framework to include multimodal conversations (such as audio, visual cues, or gestures) could create more complete and human-like summaries. Personalization: Including user preferences and adaptive summarization techniques could enhance the relevance of generated summaries for various end-user needs.

REFERENCES

1. Laskar, T. R., et al. (2023). *A comparative study of pre-trained language models for abstractive dialogue summarization*. In *Proceedings of the International Conference on Natural Language Processing (ICON 2023).*
2. Dou, Z., et al. (2023). *Are GPT-3 and other LLMs good for text summarization?* In *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023).*
3. Tang, L., et al. (2022). *ConvoSumm: A dialogue-based pre-training framework for abstractive summarization*. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022).*
4. Narayan, S., & Cohen, S. B. (2022). Well-formedness and faithfulness for abstractive summarization. *Computational Linguistics, 48*(4), 909–948.
5. Pasunuru, R., et al. (2022). *Data augmentation for abstractive dialogue summarization*. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022).*
6. Zhong, M., et al. (2022). *QMSum: A new benchmark for query-based multi-domain meeting summarization*. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2022).*
7. Feigenblat, G., et al. (2022). *TODSum: A benchmark for task-oriented dialogue summarization*. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).*
8. Kang, B., & Hovy, E. (2022). *To ship or not to ship: A new benchmark for evaluating the utility of meeting summaries*. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2022).*
9. Zhu, C., et al. (2022). *Summ^N: A multi-stage summarization framework for long meeting transcripts*. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2022).*
10. Liu, F., et al. (2022). *BASS: A BART-based approach for dialogue summarization with sentence-level slot prediction*. In *Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022).*
11. Xie, Y., et al. (2023). *Zero-shot abstractive dialogue summarization with dialogue-disentangled self-attention*. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2023).*
12. Goyal, T., et al. (2022). *News summary abstraction using controlled user-centric pointers*. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022).*
13. Shen, D., et al. (2022). *CSAN: A causal semantic-aware network for dialogue summarization*. In *Findings of the Association for Computational Linguistics: EMNLP 2022.*
14. Bhattacharjee, A., et al. (2023). *Speak like a doctor: A medical dialogue summarization framework for an affective AI doctor*. In *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2023).*
15. Ganesh, P., & Talamadupula, K. (2023). *Listen, attend, and summarize: A multi-modal framework for meeting summarization*. In *Findings of the Association for Computational Linguistics: ACL 2023.*
16. Lee, S., et al. (2022). Topic-aware abstractive text summarization. *Expert Systems with Applications, 198*, Article 116748.
17. Al-Sabahi, K., et al. (2022). A survey on abstractive dialogue summarization: Datasets, models, and evaluation. *Information Fusion, 85*, 149–176.

***Corresponding Author:** Dr. Savitha G