Automatic Generation of Executive Summaries for Online Meetings using NLP: A Review

Banda Sujith[1], Ramzan Shareef[2], Swathi Sowmya Bavirthi[3], Mohd. Arbaz[4]

sujithkumarb.work@gmail.com[1], mail.ramzanshareef@gmail.com[2], swathirao1254@gmail.com[3], mdarbaz3636@gmail.com[4]

Department of Information Technology [1, 2, 3, 4]

Chaitanya Bharathi Institute of Technology
Hyderabad, Telangana, India - 500075

Abstract

While meetings help make company decision-making more effective, documenting and distilling the material turns out to be a lot of time-consuming work and may also contain mistakes. The project provides an automated way of transcribing audio recording of meetings into text and applying NLP for perfect creation of useful summaries. As opposed to the existing techniques that resort to either means of human beings or platform-specific ones, our solution is a versatile way that can handle transcripts from a variety of online resources. This is a system that offers both abstractive and extractive summary techniques in the form of developed transformer models, such as BERT, to form logical summaries and TF-IDF and TextRank to focus the most important points in the summary. A wider applicability of Named Entity Recognition (NER) and Part-of-Speech (POS) tagging will allow summarization over key elements, including decisions taken and responsibilities assigned. The approach aims to make the capture of output from the meeting more efficient and reliable by automatically summarizing proceedings in meetings. User input and ROUGE scores will assess how well the system performs and guarantees quality useful summaries to stakeholders.

Keywords: Natural Language Processing (NLP), TextRank, Term Frequency-Inverse Document Frequency (TF-IDF), Named Entity Recognition (NER), Part-of-Speech (POS) Tagging.

1 Introduction

Online meetings have become essential for organizations in today's digital work environment, with platforms like Google Meet, Zoom, and Microsoft Teams facilitating discussions, decision-making, and task allocation. However, frequent sessions often lead to information overload, making it difficult for participants to retain critical details. This issue is compounded by the lack of effective automated processes for recording and summarizing conversations. Well-documented meetings enable participants to critique decisions, roles, and next steps efficiently.

Traditional note-taking methods, though useful, are prone to errors and time inefficiency. Our proposed system addresses these issues through automatic summarization of online meetings. It captures meeting audio, transcribes it into text, and applies extractive and abstractive summarization techniques using advanced NLP methods. Extractive approaches, such as TextRank and TF-IDF, identify significant information, while abstractive techniques using transformer models like BERT generate coherent summaries.

To enhance the process, the system integrates Named Entity Recognition (NER) and Part-of-Speech (POS) tagging, emphasizing key entities like names, dates, and actions. This results in concise, insightful summaries that improve the quality of meeting documentation by highlighting decisions and responsibilities.

The system's performance is evaluated using ROUGE scores for accuracy, recall, and precision, along with stakeholder feedback to refine functionality. Ultimately, this approach aims to streamline workflows, reduce manual effort, and provide instant access to meeting highlights, boosting productivity in dynamic work environments.

2 Related Work

Recent advancements in text summarization have explored various methodologies, broadly categorized into transformer-based approaches, meeting summarization techniques, and Named Entity Recognition (NER) integration. Transformer models like BERT and GPT have shown significant potential in summarization across domains such as journalism and social media, offering fine-tuned, context-preserved outputs. Asmitha et al. [4] demonstrated that transformers outperform traditional methods like TF-IDF and TextRank, particularly on large datasets like CNN/Daily Mail, due to their coherence and scalability.

The rise of online meetings has led to advancements in meeting summarization, often using hybrid models combining extractive and abstractive techniques. Muppidi et al. [3] developed an NLP-based system employing BART and T5 transformers to generate concise meeting minutes, achieving strong ROUGE scores and human evaluation results. Bhat et al. [16] proposed "Jotter," a real-time summarization tool integrating RNN-based models with audio recognition for clarity and precision. Bharti et al. [2] introduced a summarization approach tailored to video meetings, generating both text and voice summaries, making it valuable in corporate and academic settings.

NER has further enhanced summarization accuracy by classifying and prioritizing key entities like names, locations, and organizations. Tummala [13] employed BERT's contextual embeddings for unstructured data, improving summarization in domains requiring deep semantic understanding, such as legal and financial sectors. These developments highlight the progression from general-purpose models to domain-specific techniques, offering robust frameworks for future advancements in automatic summarization technologies.

2.1 Summarization Approaches

Recent advancements in text summarization have explored various methodologies, including extractive, abstractive, and hybrid approaches. These techniques aim to generate concise summaries from large amounts of data while ensuring coherence, relevance, and readability.

Extractive summarization techniques focus on ranking and selecting important sentences from a text. Islam et al. [6] compared different text summarizing methods for the Bangla language, which has limited resources for natural language processing. Their use of TextRank, cosine similarity, and advanced tokenization techniques demonstrates how well these methods handle the linguistic difficulties of Bangla, offering valuable insights into summarization. Majeed and Kala [7] compared TextRank with sentence-ranking algorithms for extractive summarization, finding that TextRank improves sentence retrieval and produces better summaries for single document inputs by maintaining document cohesion. Chaurasia et al. [8] employed the TF-IDF technique to investigate parallelization in extractive summarization, improving processing time and scalability, which is crucial for real-time applications.

Alsekait et al. [9] compared neural network models like T5 for extractive summarization with conventional machine learning, demonstrating that deep learning enhances semantic understanding and raises ROUGE scores, particularly for information-dense domains like scientific and legal writings. Similarly, Ganesh et al. [10] proposed the Test Rank method, which ranks sentences based on term frequency and similarity, making it effective for summarizing long documents while preserving coherence. RoselinKiruba et al. [12] optimized extractive summarization for scientific and technical literature by using GloVe embeddings and a Bi-GRU model, outperforming conventional techniques by a significant margin.

Abstractive summarization, in contrast, focuses on generating summaries by understanding and rephrasing the original content. Asmitha et al. [4] assessed the Hugging Face transformer models, such as BERT and GPT, for abstractive summarization, particularly in text-intensive domains like social media and journalism. Their study demonstrated that transformers are fine-tune-sensitive but excel in maintaining coherence and contextual meaning, offering an edge for summarization, especially for large datasets like CNN/Daily Mail.

Chen et al. [20] introduced the Stepwise Summary Generator (SSG), which is specifically designed for sequential data. This model dynamically adapts to new content while maintaining coherence with previous summaries, making it particularly effective for applications involving continuous streams of data, like news or social media feeds. Alqahtani and Al-Yahya [18] focused on abstractive summarization for Arabic content using the AraBART paradigm, which is fine-tuned for Arabic morphology and syntax. This approach showed promising results with ROUGE and BERTScore, particularly in headline generation tasks. Singhal et al. [14] presented a transformer-based approach for summarizing meeting discussions as dialogue data, capturing important ideas and action items for conversational contexts.

These studies collectively demonstrate the flexibility and effectiveness of transformer-based models in handling diverse languages, dialogue data, and dynamic content. However, further research is required to fine-tune these models for broader multilingual applicability and more efficient handling of real-time streaming data.

Hybrid summarization approaches, which combine both extractive and abstractive techniques, have gained attention for their potential to leverage the strengths of both methods.

This approach is more sensitive to coherence, relevance, and fluency in summaries, as it balances the retention of factual information with semantic enrichment.

Yan and Zhou [1] proposed a novel hybrid approach based on K-means clustering and text structure. They split the text into main and auxiliary parts, applying extractive methods like TextRank to the main sections and abstractive techniques like Seq2Seq to the auxiliary parts. The resulting summaries are merged in text order to improve coherence and contextual relevance. Their experiments demonstrated that this hybrid approach outperforms both purely extractive and purely abstractive methods. Muppidi et al. [3] applied a hybrid approach to automate meeting minute generation, employing both T5 and BART transformers in combination with extractive techniques to highlight relevant information. This approach produces concise summaries of meeting transcripts and audio data, making it well-suited for business environments.

Jiang et al. [19] enhanced hybrid summarization by integrating attention-based bidirectional LSTM (Bi-LSTM) into Seq2Seq models. This mechanism addresses issues like repeated phrases, out-of-vocabulary words, and semantic congruence between the source and the summary, resulting in better semantic accuracy and readability across large datasets. Hybrid approaches combine the factual precision of extractive methods with the semantic depth of abstractive methods, offering a balanced and effective summarization technique. However, there is still room for improvement in terms of scalability, real-time applications, and handling multimodal input. Future research in this area will likely focus on improving the adaptability and efficiency of hybrid summarization systems, especially in real-time scenarios and across diversified languages and text formats.

Improving Summarization by Semantic and Syntactic Linkages Tummala [13], employing NER with BERT to identify and classify things in the text improves summarization by focusing on semantic and syntactic linkages. Their work utilized deep learning techniques to enhance validation accuracy as well as to bring down training loss and proves its effectiveness to enrich domains such as legal and medical texts. In Deng et al. [15], a two-step method is presented for aspect-based meeting transcript summarization. It uses a sentence classifier with pseudo-labeling to detect the sentences relevant to a specific aspect of the meeting, and then phrases are again summarized using the transformer model. Their approach really handles mixing aspect-related sentences within long meeting transcripts.

Table 1: **Analysis of Summarization Techniques**

Technique	**ROGUE Score**	**Precision/R1**	**Recall/R2**	**F1/RL Score**	**Observations**
Sentence Ranking vs. TextRank [7]	Higher for TextRank	-	0.864	0.927	Effective in sentence ranking; TextRank performs better on short, single-document inputs.
Deep Learning-Based Comparison [9]	ROGUE-1: 0.362	0.68	0.70	0.67	Compares ML vs. DL; DL achieves better contextual relevance but requires high computation.
TextRank Algorithm [10]	Moderate	0.956	0.235	0.378	Good for news articles; balances relevance with coherence; lacks adaptability to real-time.
TF-IDF with Parallelization [8]	Processing time improved	-	-	-	Reduces computational load, suitable for large datasets; lacks semantic understanding.
Text Structure-Based Summarization [1]	ROUGE-1: 39.25	0.41	0.419	0.405	Combines K-means and Seq2Seq for coherent summaries; suitable for structured documents.
BART + T5 for Meeting Summaries [3]	Good ROGUE-1, Moderate ROGUE-2%	0.82	0.43	0.57	Generates concise summaries; effective in capturing key points in corporate settings.
Centroid-Based + BART [17]	Good ROGUE-1	0.51(R-1)	0.30(R-2)	0.33(R-L)	Centroid-based extractive summarization feeding into BART model; suited for corporate meeting summaries.
BERT and GPT Summarization [4]	ROGUE-1: Basic, ROGUE-2: Good	0.34	0.17	0.32	Effective for news and social media; requires extensive fine-tuning for domain adaptation.
Stepwise Summary Generator (SSG) [20]	Good RG1, RG2, and RG-L	0.34	0.45	0.80	Incremental updates for dynamic content; maintains coherence across updates; suited for news.

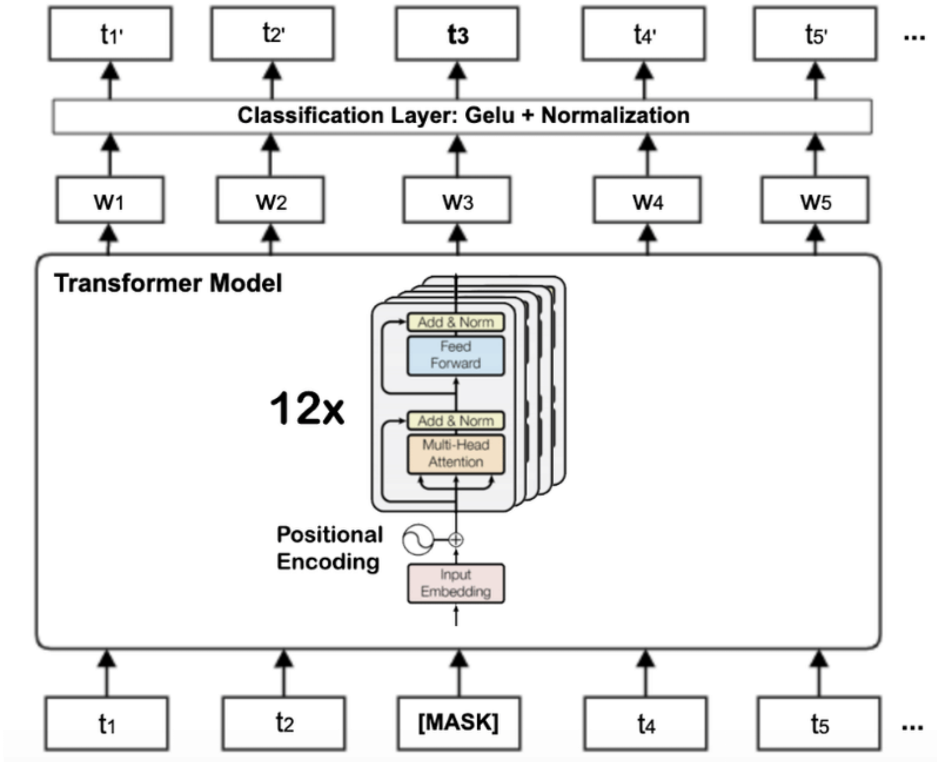

Figure 1: Architecture of Transformer-based BERT model

Precision, Recall, ROUGE, and F1 Score are basically important metrics when one measures the efficiency of a text summarization model. Precision is defined as the number of relevant phrases that are precisely identified in the summary generated out of all phrases selected by the model. High precision ensures the summary is concise, while irrelevant content is avoided in such a summary. Recall measures the ability of a model to extract all the relevant information from the source text, hence making summarization complete. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) is given by the comparison of overlapping n-grams between the generated summary and the existing reference summaries as measured through metrics such as ROUGE-1, ROUGE-2, and sequence alignment metrics like ROUGE-L. It is very suitable in relevance and coherence evaluation of abstractive summaries. The F1 score represents the balance between precision and recall, indicating how summary generation achieves accuracy without losing completeness and is essential for summary conciseness without losing the information.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}, \tag{1}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}, \tag{2}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{3}$$

2.2 Datasets

The CNN/DailyMail dataset, comprising over 300,000 news stories, is widely used for training abstractive summarization models due to its structured format. However, its focus on news articles limits representation across diverse subjects and text formats. Similarly, the BBC dataset on Kaggle, though versatile, lacks sufficient content for training long-form summaries. For Arabic summarization, datasets like EASC and Abu El-Khair, consisting of short articles from newspapers, have been employed to train hybrid summarizers. The Arabic Headline Summary (AHS) dataset, with 300K items, enhances abstractive summarization but is constrained by its reliance on headline-level summaries [11].

Advanced transformer-based models like AraBERT and AraBART have leveraged datasets such as Gigaword, XL-Sum, and WikiLingua to address domain-specific challenges, including Arabic morphology and syntax. Despite the utility of these datasets, limitations in domain variety and text length generalization remain. Additionally, smaller datasets like KALIMAT, with 20,000 items, and AMN, with 260,000 articles, have supported tasks like neural summarization but face scalability and generalization issues. This highlights the need for more extensive, varied datasets to improve model adaptability and summarization performance across broader applications [1], [12], [14].

3 Result Analysis

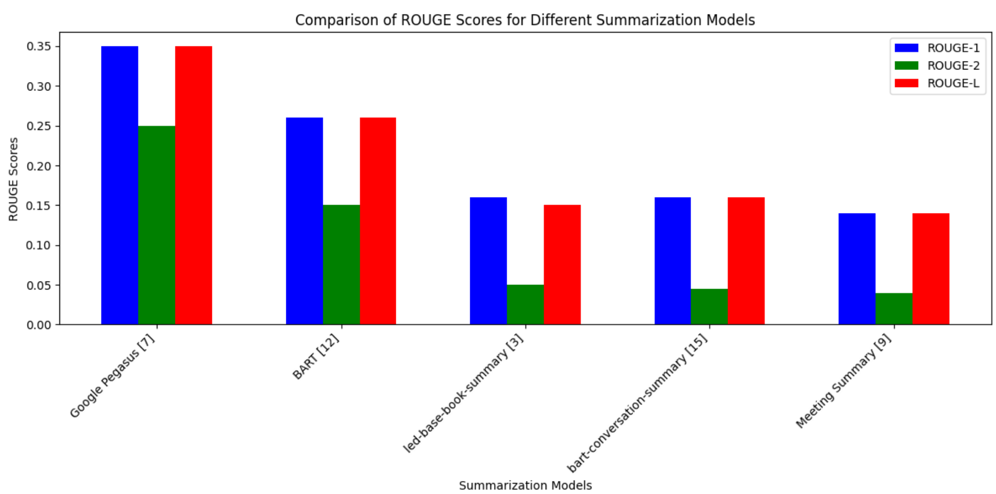

Figure 2: Comparision of ROUGE Scores for Different Summarization Models

The summarization performance of different models are compared in the bar graph comparing their ROUGE scores, which also consists of Google Pegasus [7], BART [11], led-base-book-summary [3], bart-conversation-summary [13], and Meeting Summary [9]. With this, it can be viewed that Google Pegasus [7] obtains the greatest ROUGE-1 and ROUGE-L scores as well with other models. On the other hand, BART [11] is performing good on ROUGE-2 while it lags behind Google Pegasus on ROUGE-1 and ROUGE-L. In terms of meeting summary, it has the lowest value across all the ROUGE metrics, thereby reflecting a less

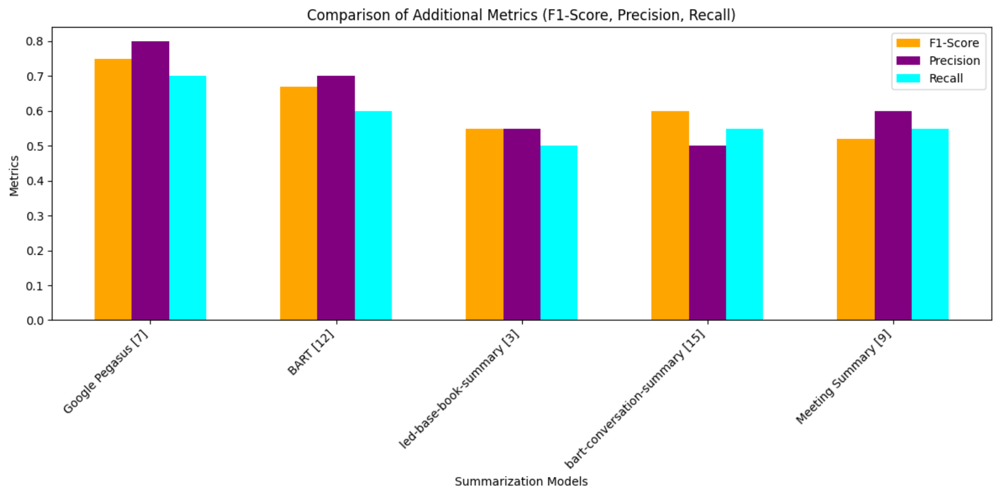

Figure 3: Comparison of Evaluation Metrics for Different Summarization Models

quality summarization compared to other models. In general, trends of performance would imply that Google Pegasus [7] seems well-set for ROUGE-1 and ROUGE-L while BART [11] places competitively for ROUGE-2.

4 Conclusion

Long meeting transcripts provide valuable information, but remote and hybrid work conditions often introduce audio issues that complicate summarization. This research proposes an automated summarization tool using a hybrid approach that combines extractive techniques like TF-IDF and TextRank with abstractive methods leveraging transformer models such as BART and T5. The system effectively highlights key meeting elements, including decisions, action items, and deadlines, enhanced by Named Entity Recognition (NER) for ranking important entities.

The system has shown promising results, receiving positive user feedback and high ROUGE scores. Its chunk-based processing pipeline efficiently handles long transcripts by grouping thematically related content, making it adaptable to various meeting types such as client calls and team discussions.

Despite these successes, challenges remain, including handling domain-specific vocabulary, improving summary accuracy for specialized sectors, and enabling multilingual support. Future enhancements will focus on real-time summarization and broader language adaptability to increase the system's utility in dynamic work environments.

In summary, this study lays the groundwork for a robust automated meeting summarization system that combines state-of-the-art NLP and AI techniques. With real-time capabilities, multilingual support, and industry-specific applications, it holds great potential to enhance productivity and decision-making in modern work settings.

References

[1] J. Yan and S. Zhou, "A Text Structure-based Extractive and Abstractive Summarization Method," in *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Xi'an, China, 2022, pp. 678-681. doi: 10.1109/ICSP54964.2022.9778497.

[2] N. Bharti et al., "An Approach for Audio/Text Summary Generation from Webinars/Online Meetings," in *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, Lima, Peru, 2021, pp. 6-10. doi: 10.1109/CICN51697.2021.9574684.

[3] S. Muppidi et al., "Automatic Meeting Minutes Generation Using Natural Language Processing," in *2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, Bengaluru, India, 2023, pp. 1-7. doi: 10.1109/EASCT59475.2023.10393102.

[4] A. M. et al., "Automation of Text Summarization Using Hugging Face NLP," in *2024 5th International Conference for Emerging Technology (INCET)*, Belgaum, India, 2024, pp. 1-7. doi: 10.1109/INCET61516.2024.10593316.

[5] G. Patil et al., "Chrome Extension for Speech-to-Text Conversion and Text Summarization Using NLP," in *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*, Bangalore, India, 2024, pp. 1-7. doi: 10.1109/ICITEICS61368.2024.10624940.

[6] T. Islam et al., "Comparative Analysis of Different Text Summarization Techniques Using Enhanced Tokenization," in *2021 3rd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, Dhaka, Bangladesh, 2021, pp. 1-6. doi: 10.1109/STI53101.2021.9732589.

[7] M. Majeed and K. M. T., "Comparative Study on Extractive Summarization Using Sentence Ranking Algorithm and Text Ranking Algorithm," in *2023 International Conference on Power, Instrumentation, Control and Computing (PICC)*, Thrissur, India, 2023, pp. 1-5. doi: 10.1109/PICC57976.2023.10142314.

[8] D. Chaurasia et al., "Enhancing Text Summarization through Parallelization: A TF-IDF Algorithm Approach," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, 2024, pp. 1503-1508. doi: 10.1109/ICoICI62503.2024.10696641.

[9] D. M. Alsekait et al., "Harnessing Deep Learning for Effective Extractive Text Summarization: A Comparative Study," in *2024 Intelligent Methods, Systems, and Applications (IMSA)*, Giza, Egypt, 2024, pp. 581-588. doi: 10.1109/IMSA61967.2024.10652647.

[10] D. Ganesh et al., "Implementation of Novel Test Rank Algorithm for Effective Text Summarization," in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, Chennai, India, 2023, pp. 1-6. doi: 10.1109/ACCAI58221.2023.10201008.

[11] N. Zade et al., "NLP Based Automated Text Summarization and Translation: A Comprehensive Analysis," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Coimbatore, India, 2024, pp. 528-531. doi: 10.1109/ICSCSS60660.2024.10624907.

[12] R. RoselinKiruba et al., "Text Summarization based on Feature Extraction using GloVe and B-GRU," in *2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, Coimbatore, India, 2024, pp. 517-522. doi: 10.1109/ICSCSS60660.2024.10625311.

[13] I. P. Tummala, "Text Summarization Based Named Entity Recognition for Certain Application Using BERT," in *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, 2024, pp. 1136-1141. doi: 10.1109/ICoICI62503.2024.10696673.

[14] D. Singhal et al., "Abstractive Summarization of Meeting Conversations," in *2020 IEEE International Conference for Innovation in Technology (INOCON)*, Bengaluru, India, 2020, pp. 1-4. doi: 10.1109/INOCON50539.2020.9298305.

[15] Z. Deng et al., "Aspect-based Meeting Transcript Summarization: A Two-Stage Approach with Weak Supervision on Sentence Classification," in *2023 IEEE International Conference on Big Data (BigData)*, 2023, pp. 636-645.doi:10.48550/arXiv.2311.04292

[16] S. S. Bhat et al., "Jotter: An Approach to Summarize the Formal Online Meeting," in *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, Ballari, India, 2023, pp. 1-6. doi: 10.1109/AIKIIE60097.2023.10390455.

[17] A. Mahadevan et al., "Minutes: Hybrid Text Summarizer for Online Meetings," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Delhi, India, 2023, pp. 1-7. doi: 10.1109/ICCCNT56998.2023.10306385.

[18] D. Alqahtani and M. Al-Yahya, "Exploring Arabic Pre-Trained Language Models for Arabic Abstractive Text Summarization," in *2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS)*, Abu Dhabi, UAE, 2023, pp. 1-7. doi: 10.1109/SNAMS60348.2023.10375464.

[19] J. Jiang et al., "Enhancements of Attention-Based Bidirectional LSTM for Hybrid Automatic Text Summarization," in *IEEE Access*, vol. 9, pp. 123660-123671, 2021. doi: 10.1109/ACCESS.2021.3110143.

[20] X. Chen et al., "Write Summary Step-by-Step: A Pilot Study of Stepwise Summarization," in *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1406-1415, 2024. doi: 10.1109/TASLP.2024.3357040.