eSystems Engineering Society EST. 2007

International Journal of Data Science and Advanced Analytics

Journal homepage: www.ijdsaa.com

AI-Driven News Summarisation for Financial Insights: Revolutionising Large Language Models

Shraddha Bhoir[1], Walaa N. Bajnaid [2] Diksha Malhotra [3]
[1]Presight AI Technologies, Abu Dhabi, UAE [2]Faculty of Communication and Media, King Abdulaziz University, Jeddah, Saudi Arabia [3]Faculty of Engineering and Technology, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, UK

[1]er.shraddhabhoir@gmail.com, [2] wnbajnaid@kau.edu.sa, [3] dikshamalhotra.dm@gmail.com

Abstract- Financial professionals rely on timely news to support decision making. However, manually reviewing large volumes of financial news is inefficient particularly when articles are complex and lengthy. Automated text summarisation using large language models (LLMs) offers a promising solution for summarising extensive textual information. This study aims to customise and optimise a text summarisation model for financial news. To achieve this, the study examines the effectiveness of transformer based LLMs by fine-tuning the FLAN-T5-XL model for this domain. Experiments were conducted using a dataset of 2,000 general news articles. Performance was assessed using ROUGE metrics and expert human evaluation. The results show that the fine-tuned FLAN-T5-XL with truncation achieved the best performance obtaining a ROUGE -1 score of 55 and 86% agreement with expert evaluation. These findings demonstrate that domain adapted LLMs can provide a practical tool for rapid information synthesis and financial decision making.

Keywords— News summarisation, Text summarisation, financial news summarisation, LLMs, financial decision making.

1. Introduction

Text summarisation refers to the process of condensing lengthy content into fewer words without changing its meaning or essential information [1]. Recently, text summarisation has become a distinctive machine-learning task that balances the need for coherence, length reduction, and the capturing of essential information. There are three approaches to automated summarisation: abstractive, extractive and hybrid. The abstractive method is built on understanding the main idea of the input text and then regenerating the content with some degree of interpretation. This process results in a human-like summary. Extractive summarisation involves selecting important parts of the input content, then merging them while maintaining the meaning. The text resulting from this method could have some grammar problems. Finally, the hybrid method combines both approaches [2].

https://doi.org/10.69511/ijdsaa.v7i1.312
Received 12 May 2025; Received in revised form 03 July 2025; Accepted 04 September 2025.
Available online 19 September 2025.

Automatically summarising text using natural language processing (NLP) has some challenges and limitations, such as ambiguities, which arise due to multiple valid summarisations of the same content. In addition, the resulting summary might be less accurate in cases of summarising multiple documents, dealing with information overload, addressing complex language features and evaluating the quality of a summary [1]. The resulting summaries could include low-quality references [3, 4]. It is costly to involve humans to evaluate the output text [5]. Thus, this study aims to introduce AI-powered summaries that are good, accurate, meaningful and that mimic human expertise.

Text summarisation in large language models (LLMs), including GPTs, combines pre-training, giant scaling and fine-tuning to perform the task. The standards of summarisation have been improved by LLMs in terms of both fluency and quality. However, LLMs also have limitations related to ethics and a tendency to produce biased, false or inaccurate information. In addition, intellectual property (IP) infringement could be an issue in the data or output of these systems. This study seeks to address these limitations by adopting recent advancements in LLMs through an abstractive approach to produce high-quality summaries.

The present study focuses on customising a text-summarisation model for financial news to assist readers in accessing key information quickly and in making timely, high-quality decisions. To achieve this, the study uses optimisation techniques, such as prompt engineering and fine-tuning, together with chunking and truncation, to improve the performance of the model.

2. Literature Review

Automated text-summarisation approaches have been developed since 1970s, varying from rule-based systems to LLMs. These models have contributed to shaping the landscape of text summarisation. In addition, the introduction of deep learning and neural networks has contributed to abstractive approaches using Sequence-to-Sequence (Seq2Seq) and transformer models, such as BERT and GPT, which are regularly improved by means of reinforcement learning (RL) [1, 6].

Various domains and applications of text summarisation have emerged. Search engines use text-summarisation techniques to summarise search results and offer quickly accessible relevant results [1]. Likewise, news summaries are displayed on news websites like CNN to help users quickly navigate to articles in which they are interested [1, 7]. Pretrained LLMs are used in summarisation tasks. [8] includes a survey of different bug-report summarisation techniques which help developers address domains specific needs.

2.1 Text-summarisation approaches

Text summarisation is classified into three types based on the output generated by the models: extractive, abstractive and hybrid. An extractive summary is formed by selecting the most significant sentences from the source text [9.10]. This approach uses different methods, including statistically based approaches, that use words and sentences to determine the relevance and important parts of the text. [11,12, 13].

Abstractive summarisation, by contrast generates a summary that captures the main idea of the source text using NLP and may generate novel sentences that differ from the source text. Due to their extensive natural language processing and complex design, abstractive approaches are slower than extractive approaches [14]. In addition, extractive models have been found to outperform abstractive and hybrid models on automated metrics, such as ROUGE and BLEU, according to recent work by [6]. Still, abstractive methods produce higher-quality summarisation that are more coherent, concise and closer to human-written summaries, and are more popular in text-summarisation tasks [9].

The hybrid approach combines both extractive and abstractive text-summarisation approaches. In this method, sentences are first selected using various extractive techniques, and then abstractive methods are applied to generate the summary. By combining extractive and abstractive methods more accurate and insightful summaries can be obtained. Commonly used methods in the hybrid approach are sentence compression, fusion, and paraphrasing [1, 12].

2.2 Text-Summarisation Models and Techniques

Recent advances in text summarisation have been driven by transformer-based pretrained language models, which have improved the quality of generated summaries [15]. These models can capture long-term dependencies among tokens and generate more coherent summaries by addressing issues such as coreference resolution [16]. Pretrained language models are widely preferred over building models from scratch because they are trained on large-scale unannotated data and can automatically learn syntactic and semantic representations. More recently,

LLMs such as GPT and T5-based systems have shown promising results in automatic text summarisation [3, 17]. Particularly when combined with prompting and fine-tuning strategies [4, 18].

Existing research highlights the importance of prompting, instruction tuning and fine-tuning in maximising summarisation performance [21]. Prior studies have found that model performance depends not only on the base architecture, but also on how the model is adapted to the task and domain [17, 22]. In addition, input context length dependency is another challenge for the performance of language models. Lengthy texts may exceed the model's token limit and affect summary quality [23]. This makes context handling strategies such as truncation and chunking relevant in practical summarisation systems.

Another important issue is evaluation. Although automated metrics such as ROUGE are widely used, previous research have shown that they do not always align with human judgements of summary quality [4,11]. For this reason, combining automatic evaluation with human assessment is highly recommended.

Finally, when selecting a summarisation model for real-world applications, factors like performance, cost, inference speed, privacy considerations, and fine-tuning capabilities significantly influence the choice of model. Although proprietary models may outperform all other models on the summarisation benchmarks, open-source models can provide a more practical balance between performance and deployment constraints [22]. This makes them suitable for domain specific applications such as financial news summarisation.

3. Materials and Methods

A transformer-based large language model was employed for this study, as transformer models have shown a strong ability to handle long-range dependencies in text and generate coherent summaries. [24, 25]. Among these models FLAN-T5-XL was selected as the base model. FLAN-T5-XL is an open-source LLM that is suitable for text summarisation tasks. Comparative methods were employed to enhance the LLM through fine-tuning and prompt-engineering techniques. The primary objective of this research was to address the limited availability of finance-news summarisation datasets with high-quality reference summaries, while identifying the most effective techniques to improve the performance of the LLM for summarisation. The study therefore compares summaries generated by both the fine-tuned and prompt-engineered versions of the model using both automated metrics and human evaluation.

3.1 Data Selection

This study used two datasets. The first dataset consisted of general news articles with high-quality reference summaries. This dataset was used to assess the performance of the selected model and to compare prompting and fine-tuning techniques. The best performing configuration from this stage was then applied to a second dataset of financial news articles obtained from publicly available Webhose financial news dataset hosted on GitHub [6].. The summaries generated from the financial dataset were reviewed by a financial expert to assess their quality.

The general news articles dataset was obtained from a recent study by [4], which contains high-quality reference summaries supported by human evaluation. This dataset contains 2,000 general news articles along with their gold-standard summaries with 500 samples drawn from each of the following sources: CNN, Dailymail, XSum and Newsroom. The dataset contains metadata information, such as the article, the reference summary used for training and reference summaries generated by four language models: Pegasus [26], GPT-3 [27, 28], T0 [29] and BRIO [30]. In the current study, we used this dataset to optimise the selected LLM for a text-summarisation task using different techniques. Specifically, the source articles and GPT-3 reference summaries were used to fine-tune the model, as GPT-3 summaries were rated highly in prior human evaluation [4]. In addition, reference summary and summaries generated by other models from the dataset were used in the multiple reference ROUGE evaluation of the generated outputs.

3.2 Data Preprocessing

Preprocessing techniques were used to improve the performance of the summarisation process [31]. Rows containing missing news-article text or reference summaries were removed from the dataset. The remaining texts were processed using the FLAN-T5-XL tokenizer. This converted the raw text into token sequences suitable for the model. The model has a limited input context length. Therefore, longer articles were handled using truncation and chunking strategies. These strategies were later compared in the experiments. Padding was applied during batch preparation. This ensured consistent sequence lengths within each batch and supported efficient computation. In

the prompt-based experiments, prompt instructions were also added to the input text. This helped guide the model to generate concise and relevant summaries.

3.3 Proposed Model

This study adopted FLAN-T5-XL, which has size of 2.85 billion parameters. Building on the findings of [32], FLAN-T5 is a language model developed by Google and made available as an open-source model for commercial use. It is trained on over 1,000 diverse tasks [33].

Three methods were tested to identify the most effective summarisation approach: zero-shot prompting, which uses instructions without training; one-shot prompting, which uses an example to guide the output; and fine-tuning, which trains the model on task-specific data. Seq2SeqTrainer was then used for efficient fine-tuning and evaluation of the best results from the three methods.

This study used a quantised version of the base model with PEFT and QLoRA to enable fine-tuning on low-cost GPUs efficiently. Quantisation improves computational efficiency by using four-bit precision while reducing memory. PEFT fine-tunes a small subset of parameters and maintains a lower computational cost. QLoRA combines quantisation with LoRA to achieve high performance with minimal resources. However, Flan-T5-XL has an input limit of 512 tokens, which creates a context length constraint. Thus, three context handling strategies were used namely truncation, recursive chunking with join and recursive chunking with rewrite.

The results were evaluated via ROUGE (1,2 and L) metrics. These metrics compare the generated summaries with reference summaries. In addition, the generated summaries were evaluated using both single-reference and multiple-reference evaluation.

3.4 Application to Finance News Articles

To verify the effectiveness of the best performing model, a case study was conducted using a financial news dataset. This dataset was previously used in research by [34] for a different financial task, namely named-entity-recognition. It contains 47,851 news articles in JSON format. It also includes metadata, such as source, country, publication date, author, title and text. Due to time and resource constraints 69 articles were randomly selected for this study. The generated summaries were evaluated by a financial expert, The evaluation was based on criteria drawn from [17] and [35], including consistency, relevance, conciseness, and coherence.

3.5 Logical Flow of the System

The flowchart in Figure 1 presents the overall workflow of the study. It shows the comparative process used to optimise the model through zero-shot prompting, one-shot prompting and fine-tuning. The process concludes with the selection of the best performing model among different approaches for further downstream application.

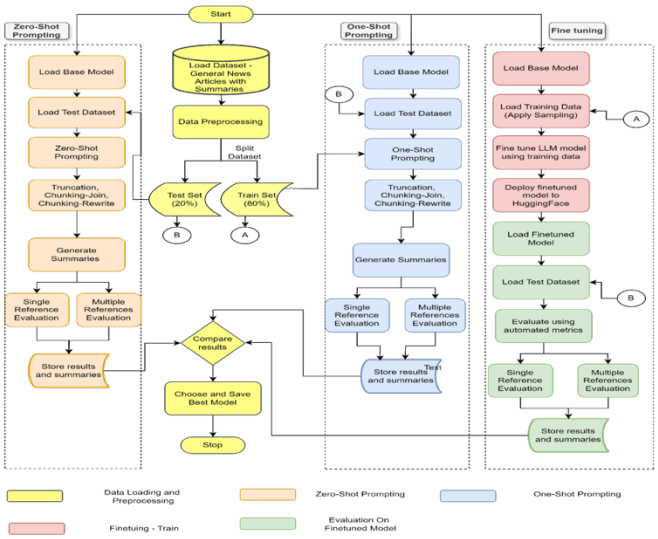

Fig. 1: Main Flowchart

4. Implementation

The implementation process began with the setup of the computational environment. Initially, the experiments were conducted using the free GPU version of Google Colab. However, due to runtime limitation after reputed runs, the training environment was switched to Lightning AI Studio. An A10G GPU machine equipped with an NVIDIA A10 Tensor Core graphics card was used for model training. The required packages were installed in the Python environment using pip.

Several libraries were used in the implementation. Pandas was used for reading and processing tabular data, JSON for loading the datasets and NumPy for numerical operations. The train_test_split function was imported from Scikit-Learn. Components from Hugging Face Transformers library were also used including T5Tokenizer, T5ForConditionalGeneration, BitsAndBytesConfig, GenerationConfig, Seq2Seq Training Arguments, Seq2SeqTrainer and DataCollatorForSeq2Seq. The Evaluate library was used to load the ROUGE metric. PyTorch was used for loading and fine-tuning the model and PEFT methods such as LoRA were applied to reduce computational and storage costs during fine-tuning.

The general news dataset was first loaded from the JSON files and prepared for the experiments. The dataset was then divided into training and test sets for repeated use across the different experimental settings. In addition, article and summary lengths were examined to better understand the input size distribution. It also guided the design of the truncation and chunking strategies.

The model was then implemented by loading the base model and applying the selected prompting and fine-tuning settings. Experiments were conducted using zero-shot, one-shot prompting and fine-tuning. Truncation and chunking were also tested to compare their effect on summarisation performance. The fine-tuned model was trained on the training set and then evaluated on the test set. All experiments were evaluated using the selected evaluation metrics.

5. Results and Analysis

The results of the three approaches examine in this study are presented below. For clarity, Table 1 summarises the ROUGE scores across all methods and techniques under both single-reference and multiple-reference evaluation settings.

Table 1 Comparison of summarisation results across methods and techniques

Approach	Technique	ROUGE(1/2/L)	
		Single Ref. Eval.	**Multiple Ref. Eval.**
Zero-Shot	**Truncation**	36.9478/15.2578/26.0821	49.7126/31.506/39.4651
	Chunking-Join	33.5017/14.7529/22.6907	41.1381/24.6133/31.1114
	Chunking-Rewrite	34.7323/12.4245/23.5366	43.1145/22.0029/32.2117
One-Shot	**Truncation**	34.0638/11.7577/22.2454	43.8525/23.8895/33.1884
	Chunking-Join	11.8686/5.8061/8.7297	13.7314/7.988/10.3833
	Chunking-Rewrite	31.1767/8.6282/20.315	36.9951/15.0746/26.0916
Finetuning	**Truncation**	44.2483/22.7619/32.9747	53.6294/35.3505/43.6055
30 epochs	**Chunking-Join**	35.6728/18.7549/25.3937	40.4719/24.799/30.3812
	Chunking-Rewrite	32.7581/13.8264/22.9311	37.0275/18.7509/27.1166

5.1 Experiment 1: Zero-Shot Prompting

The results demonstrate that multiple- reference evaluation yielded higher ROUGE scores compared to single-reference evaluations across all truncation and chunking methods. Among the chunking techniques, the Chunking-Rewrite method showed slightly better results than Chunking-Join. However, the truncation technique delivered the highest ROUGE scores overall and outperformed both chunking methods.

5.2 Experiment 2: One-Shot Prompting

Multiple-reference evaluation again produced higher ROUGE scores than single-reference evaluation across the truncation and chunking methods. As in the first experiment, the Chunking-Rewrite method performed better than Chunking-Join, while the truncation technique consistently yielded the highest scores. These findings align with those from Experiment 1.

5. 3 Experiment 3: Fine-tuning

We trained and validated the model using 15 and 30 runs. Based on performance during the training and validation phases, the model trained for 30 runs was selected as the best-performing model. The fine-tuned model was then evaluated on the test data. Consistent with the previous experiments, multiple-reference evaluation led to better ROUGE scores across all techniques. This time, the Chunking-Join method slightly outperformed Chunking-Rewrite. However, the truncation technique again achieved the highest scores overall.

We then compared the results across all three approaches. Based on this comparison, the Flan-T5-XL model fine-tuned using the truncation technique was identified as the best performing model. in addition, an expert evaluation was conducted on 17 randomly selected articles. Although this analysis was exploratory, the expert assessment indicated that the quality of the Flan-T5-XL summaries was comparable to, and in some cases better than, the GPT-3 summaries. This finding further supports the performance of the fine-tuned model.

5.4 Comparison Between Previous and Current Research

Previous research reported benchmark overlap-based ROUGE scores across four summarisation datasets. The average ROUGE-1 score reported in that study was 44.16. For comparison, the best ROUGE-1 scores from each dataset in the present study were used to calculate an overall average. The top-performing model achieved a ROUGE-1 score of 44.2483 under single-reference evaluation and 53.6294 under multiple-reference evaluation. Table 2 presents the benchmark overlap-based ROUGE-1, ROUGE-2 and ROUGE-L scores reported in previous research across the CNN, DailyMail, Xsum, and Newsroom datasets.

Table 2 Rouge Score Results from previous benchmark

Dataset	Model	Overlap-Based ROUGE(1/2/L)
CNN	PEGASUS	34.85/14.62/28.23
	BRIO	38.49/17.08/31.44
	T0	35.06/13.84/28.46
	GPT3-D2	31.86/11.31/24.71
DailyMail	PEGASUS	45.77/23.00/36.65
	BRIO	49.27/24.76/39.21
	T0	42.97/19.04/33.95
	GPT3-D2	38.68/14.24/28.08
XSum	PEGASUS	47.97/24.82/39.63
	BRIO	49.66/25.97/41.04
	T0	44.20/20.72/35.84
	GPT3-D2	28.78/7.64/20.60
Newsroom	PEGASUS	39.21/27.73/35.68
	BRIO	-
	T0	25.64/9.49/21.41
	GPT3-D2	27.44/10.67/22.18

As shown in Table 2, BRIO achieved the strongest benchmark performance on CNN, DailyMail and XSum, while PEGASUS performed best on Newsroom.

Table 3 Rouge 1 score of best performing models

Dataset	Model	ROUGE 1
CNN	BRIO	38.49
DailyMail	BRIO	49.27
Xsum	BRIO	49.66
Newsroom	PEGASUS	39.21
Average ROUGE 1 Score		**44.16**

The previous research reported an average ROUGE-1 score of 44.16 across four summarisation datasets as shown in Table 3. When the two studies were compared based on ROUGE-1, the proposed model performed slightly better than the earlier model under single-reference evaluation. Under multiple-reference evaluation, the proposed model exceeded the previous benchmarks by 9.47 points. Although BRIO and PEGASUS were identified as the top-performing models in the earlier study, GPT-3 summaries were rated highly in human evaluation. When compared with the average GPT-3 ROUGE-1 scores, the proposed model achieved improvements of 12.56 points under single- reference evaluation and 21.94 points under multiple-reference evaluation. These findings indicate that the fine-tuned Flan-T5-XL model achieved stronger summarisation performance than the previous benchmarks.

5.5 Case Study: Finance News Article Summarisation

The evaluation was conducted using both the automated ROUGE metric and expert evaluation by a finance specialist. A ROUGE-1 score of 55 points was achieved under single-reference evaluation, demonstrating that the fine-tuned model performed effectively on unseen financial news data (Table 4).

Table 4 ROUGE results on finance news articles

Dataset	Model	ROUGE(1/2/L)
Finance News Articles	flan-t5-xl finetuned	55.0923/40.4434/47.3037

Based on the expert evaluation, 55 summaries, representing 80% of the dataset, were categorised as meeting expectations. Four summaries, representing 6% were categorised as exceeding expectations. Ten summaries, accounting for 14% of the dataset, were categorised as needing improvement.

The expert evaluation indicated that the model achieved an overall accuracy of 86%, reflecting strong agreement between the generated summaries and the expert assessment. This suggests that the model is effective in producing high-quality financial summaries, on domain-specific, unseen data.

6. Conclusions

This study aimed to customise and optimise a text summarisation model for financial news. It also addressed the limited availability of finance- specific

summarisation datasets, particularly those supported by reference summaries and expert evaluation.

Overall, the proposed model demonstrated a strong performance in summarising financial news articles. The findings suggest that it can support faster access to key information while reducing the effort required for manual review. This study extends the application of automatic text summarisation to the financial domain and highlights the practical value of domain adopted large language models for knowledge aggregation and content organisation.

An important contribution of this study is the development of a finance news dataset containing both human-crafted and model-generated summaries. This dataset addresses key challenges identified in the research, such as the scarcity of finance-specific news datasets paired with reference summaries along with domain-expert evaluations. This study also has several limitations. The study was conducted only on English-language data. In addition, only 146 articles were used for testing, and 331 were used for fine-tuning because of token and resource constrains. Human evaluation was also limited to the finance news case study. Despite these limitations, the study provides a useful foundation for future research on financial news summarisation and offers directions for research, particularly in the field of domain specific summarisation using large language models.

Future work could extend the analysis to other languages, further fine-tuning the top-performing model using the finance news reference summaries developed in this study and incorporate domain specific Knowledge to improve summarisation quality.

References

1. Khan, B., Shah, Z.A., Usman, M., Khan, I., Niazi, B.: Exploring the landscape of automatic text summarization: a comprehensive survey. IEEE Access, 11, 109819–109840 (2023).
2. Kavyashree, S., Sumukha, R., Soujanya, R., Tejaswini, S. V.: Survey on automatic text summarization using NLP and Deep Learning. In: IEEE International Conference on Advances in Electronics, Communication, Computing and Intelligent Information Systems, ICAECIS 2023, pp. 523–527. Proceedings. Institute of Electrical and Electronics Engineers Inc. (2023).
3. Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., Zhong, S., Yin, B., Hu, X.: Harnessing the power of LLMs in practice: A survey on ChatGPT and beyond. In: ACM Transactions on Knowledge Discovery from Data, 186 (2024).
4. Goyal, T., Li, J.J., Durrett, G.: News summarization and evaluation in the era of GPT-3 (2023). http://arxiv.org/abs/2209.12356
5. Avramelou, L., Passalis, N., Tsoumakas, G., Tefas, A.: Domain-specific Large Language Model finetuning using a model assistant for financial text summarization. In: 2023 IEEE Symposium Series on Computational Intelligence, SSCI 2023, pp. 381–386. Institute of Electrical and Electronics Engineers Inc. (2023).
6. Webhose. (n.d.). *Financial news dataset* [Data set]. GitHub. Retrieved 2024, from https://github.com/Webhose/financial-news-dataset/tree/master/Datasets
7. Myla, S.D., Saini, E.R., Kapoor, E.N.: Auto text summarization in Natural Language Processing: Review. In: 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things, IDCIoT 2024, pp. 1258–1267. Institute of Electrical and Electronics Engineers Inc. (2024).
8. Supriyono, Wibawa, A.P., Suyono, Kurniawan, F.: A survey of text summarization: Techniques, evaluation and challenges. Natural Language Processing Journal, 7,100070 (2024).
9. Sri, S.H.B., Dutta, S.R.: A survey on automatic text summarization techniques. In: Journal of Physics: Conference Series. IOP Publishing Ltd. (2021).
10. Jin, H., Zhang, Y., Meng, D., Wang, J., Tan, J.: A comprehensive survey on process-oriented automatic text summarization with exploration of LLM-based methods (2024). http://arxiv.org/abs/2403.02901
11. Yadav, A.K., Singh, A., Dhiman, M., Vineet, Kaundal, R., Verma, A., Yadav, D.: Extractive text summarization using deep learning approach. International Journal of Information Technology (Singapore), 145, 2407–2415 (2022).
12. Watanangura, P., Vanichrudee, S., Minteer, O., Sringamdee, T., Thanngam, N., Siriborvornratanakul, T.: A comparative survey of text summarization techniques. SN Computer Science (2024).
13. Sharma, G., Sharma, D.: Automatic text summarization methods: A comprehensive review. SN Computer Science, 41 (2023).
14. Mridha, M.F., Lima, A.A., Nur, K., Das, S.C., Hasan, M., Kabir, M.M.: A survey of automatic text summarization: Progress, process and challenges. IEEE Access, 9, 156043–156070 (2021).
15. Choudhary, A., Alugubelly, M., Bhargava, R.: A comparative study on transformer-based news summarization. In: Proceedings - International Conference on Developments in eSystems Engineering, DeSE, pp. 256–261. Institute of Electrical and Electronics Engineers Inc. (2023).
16. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: 31st Conference on Neural Information Processing Systems (NIPS 2017). (2017). http://arxiv.org/abs/1706.03762
17. Kumar, S. and Solanki, A.: An abstractive text summarization technique using transformer model with self-attention mechanism. Neural Computing and Applications, 3525, 18603–18622 (2023).
18. Zhang, T., Ladhak, F., Durmus, E., Liang, P., Mckeown, K., Hashimoto, T.B.: Benchmarking Large Language Models for news summarization. In: Association for

Computational Linguistics, pp. 39–57 (2024). http://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00632/2325685/tacl_a_00632.pdf
19. Pu, X., Gao, M., Wan, X.: Summarization is (almost) dead. (2023). http://arxiv.org/abs/2309.09558
20. Gupta, A., Diksha Chugh, A., Katarya, R.: Automated news summarization using transformers. In: Aurelia, S. Hiremath, S.S., Subramanian, K. Biswas, S.K. (eds.), Sustainable Advanced Computing, pp.249–259. Springer Singapore, Singapore (2022). https://link.springer.com/chapter/10.1007/978-981-16-9012-9_21
21. Venkataramana, A., Srividya, K., Cristin, R.: Abstractive text summarization using BART. In: MysuruCon 2022 - 2022 IEEE 2nd Mysore Sub Section International Conference. Institute of Electrical and Electronics Engineers Inc. (2022).
22. Bahrainian, S.A., Feucht, S., Eickhoff, C.: NEWTS: A corpus for news topic-focused summarization. In: Findings of the Association for Computational Linguistics. pp. 493–503 (2022). https://github.com/ali-bahrainian/NEWTS
23. Tahmid, M., Laskar, R., Fu, X.-Y., Chen, C., Bhushan, S.: Building Real-world meeting summarization systems using Large Language Models: A practical perspective. In: Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track, pp. 343–352 (2023). https://openai.com/blog/chatgpt
24. Deokar, V., Shah, K.: Automated text summarization of news articles. International Research Journal of Engineering and Technology (2021). www.irjet.net
25. Rao, R., Sharma, S., Malik, N.: Automatic text summarization using transformer-based language models. International Journal of System Assurance Engineering and Management, 156, 2599–2605 (2024)..
26. Bansal, G., Chamola, V., Hussain, A., Guizani, M., Niyato, D.: Transforming conversations with AI—A comprehensive study of ChatGPT. Cognitive Computation, 165, 2487–2510 (2024).
27. Zhang, J., Zhao, Y., Saleh, M., Liu, P.J.: PEGASUS: Pre-training with extracted gap-sentences for abstractive summarization. In: Proceedings of the 37th International Conference on Machine Learning (2020).
28. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., Mccandlish, S., Radford, A., Sutskever, I., Amodei, D.: Language models are few-shot learners (2020). https://commoncrawl.org/the-data/
29. Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C.L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., Lowe, R.: Training language models to follow instructions with human feedback (2022).
30. Sanh, V., Webson, A., Raffel, C., Bach, S.H., Sutawika, L., Alyafeai, Z., Chaffin, A., Stiegler, A., Scao, T. Le, Raja, A., Dey, M., Bari, M.S., Xu, C., Thakker, U., Sharma, S.S., Szczechla, E., Kim, T., Chhablani, G., Nayak, N., Datta, D., Chang, J., Jiang, M.T.-J., Wang, H., Manica, M., Shen, S., Yong, Z.X., Pandey, H., Bawden, R., Wang, T., Neeraj, T., Rozen, J., Sharma, A., Santilli, A., Fevry, T., Fries, J.A., Teehan, R., Bers, T., Biderman, S., Gao, L., Wolf, T., Rush, A.M.: Multitask prompted training enables zero-shot task generalization. In: ICLR 2022 (2021).
31. Liu, Y., Liu, P., Radev, D., Neubig, G.: BRIO: Bringing order to abstractive summarization. In: Association for Computational Linguistics. Long Papers, pp. 2890–2903 (2022). https://github.com/yixinL7/BRIO
32. Fan, C., Chen, M., Wang, X., Wang, J., Huang, B.: A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. Frontiers in Energy Research (2021).
33. Fu, X.-Y., Laskar, M.T.R., Khasanova, E., Chen, C., Tn, S.: Tiny titans: Can smaller Large Language Models punch above their weight in the real world for meeting summarization? In: Y. Yang, A. Davani, A. Sil and A. Kumar (eds.), Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 6: Industry Track), pp. 387–394. Mexico City, Mexico: Association for Computational Linguistics (2024). https://aclanthology.org/2024.naacl-industry.33
34. Won Chung, H., Hou, L., Longpre, S., Zoph, B., Tai, Y., Fedus, W., Li, Y., Wang, X., Dehghani, M., Brahma, S., Webson, A., Shane Gu, S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Castro-Ros, A., Pellat, M., Robinson, K., Valter, D., Narang, S., Mishra, G., Yu, A., Zhao, V., Huang, Y., Dai, A., Yu, H., Petrov, S., Chi, E.H., Dean, J., Devlin, J., Roberts, A., Zhou, D., Le, Q. V, Wei, J., Salakhutdinov, R.: Scaling instruction-finetuned language models. Journal of Machine Learning Research (2024). http://jmlr.org/papers/v25/23-0870.html.
35. Shah, A., Vithani, R., Gullapalli, A., Chava, S.: FiNER: Financial Named Entity Recognition Dataset and Weak-Supervision Model (2023). http://arxiv.org/abs/2302.11157.
36. Fabbri, A.R., Kry´scí, W., Nski, K., Mccann, B., Xiong, C., Socher, R., Radev, D.: SummEval: Re-evaluating Summarization Evaluation. Association for Computational Linguistics, pp. 391–409 (2021). http://direct.mit.edu/tacl/article-pdf/doi/10.1162/tacl_a_00373/1923949/tacl_a_00373.pdf