ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 4 (2024)
https://doi.org/10.62019/abbdm.v4i4.263

ASIAN BULLETIN OF BIG DATA MANAGEMENT
http://abbdm.com/
ISSN (Print): 2959-0795
ISSN (online): 2959-0809

Text Summarization using Deep Learning: A Study on Automatic Summarization

Anwar Ali Sanjrani*, Muhammad Saqib, Saira Rehman, Muhammad Saeed Ahmad

Chronicle

Article history
Received: 1st December, 2024
Received in the revised format: 12th December, 2024
Accepted: 30th December, 2024
Available online: 31st December, 2024

Dr. Anwar Ali Sanjrani is currently affiliated with the Department of Computer Science and Information Technology University of Balochistan, Quetta, Pakistan.
Email: anwar.csd@gmail.com

Muhammad Saqib is affiliated, as an alumnus, with the Department of Computer Science, Whitacre College of Engineering, Texas Tech University, Texas, USA.
Email: saqibraopk@hotmail.com

Saira Rehman is currently affiliated with the Department of Computer Science, Minhaj University Lahore, Pakistan.
Email: sairarehman150@gmail.com

Muhammad Saeed Ahmad is currently affiliated with the Government Sadiq College Women University, Bahawalpur, Pakistan.
Email: drsaeed@gscwu.edu.pk

Corresponding Author*

Abstract

Automatic text summarization has recently become popular in natural language processing because of its ability to minimize an overwhelming quantity of information into a summarization. This research aims at analyzing extractive and abstractive approaches to the automatic text summarization through the help of deep learning models. The work mainly deals with the analysis of the performance of models like Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) networks and Transformer-based models including BERT and GPT. Pertaining to the evaluation of these models, the study employs objective metrics such as ROUGE, BLEU, in addition to subjective human evaluation of coherence, relevance and fluency. Studies show that Transformer-based models: BERT and GPT perform better than the extractive model in every aspect in the ability to produce summaries with high fluency and context relevancy. However, there are still problems associated with growing the scope of higher-order n-gram recall and preserving summary relevance to the text information. The author of the study also supports the fact that deep learning-based summarization methods demonstrate high potential but require additional studies to improve the quality of the output summary. The work offers an understanding of the elements of the existing models and offers a foundation for future development in automatic summarization.

Keywords: Automatic, text summarization, natural language processing, deep learning.

INTRODUCTION

Deep learning-based text summarization is considered as a fascinating subfield in NLP that addresses the problem of converting long texts into smaller texts containing all the main points. This process is applicable in virtually all domains such as: news extraction, abstracting of academic papers, legal documents, etc. With the advent of Deep Learning, the existing text summarization system has been enhanced to higher levels of capacity to generate summaries that resemble human crafted summaries. The incorporation of deep neural networks, especially the RNNs, LSTM & Transformer model, has benefited the automatic summarization models, and provided them more precise, contextual and coherent summaries from conventional models (Khan, Daud, Khan, Muhammad, & Haq, 2023). In text summarization, the main problem area can be identified with completeness, coherence and comprehensiveness of information processing with reference to the grammatical integrity of the output summary text. There are two primary approaches to text summarization: extractive and abstractive. In extractive summarization, dependent on the method being used, the computer program that generates the summary picks out explicit text fragments, including sentences or phrases from the input text and forms it into a summing up. The abstractive approach on the other hand creates new sentences which may be rewrites of the important information from the text. The latter of these is usually more difficult as the model needs to have a better idea of the text, and the ability to create grammatically correct, flowing English which may not necessarily be word for word to the source text. Though less complex and more practicable than the synthetic ones, extractive methods do not estimate the gist of the tacit meaning and the uninterrupted text sequence, which makes the produced summaries incoherent. Although the above-mentioned approaches are a little complicated, the automatic generation of abstractive summaries is likely to be more natural and like human-like (Wibawa & Kurniawan, 2024). The latter two have improved by the use of deep learning models.

A few existing techniques have been used in this case, for example, RNNs and LSTMs where the model captures the temporal relation between the word and the sentences that allows the model to assess the correct context by summarizing the vital sentences. These models work in parallel where while translating a word in a sentence; they also keep the previous words in their memory so that they consider the context of the word. Nonetheless, one of the weaknesses of the RNNs and LSTM is that they may fail to capture long-range dependencies, thus, they may not capture all imperative context during processing long documents. To this, solution attention mechanisms were added which allow models to selectively focus on the most important content of the text during summary generation (Soubraylu & Rajalakshmi, 2021). By contrast, a new class of models, based on the transformer architecture like BERT, GPT and others, dominate the continuation of progress in the field of text summarization. Some of these models propose self-attention mechanisms, from which all the words of a sentence can be attended, not sequentially, as is often the case in other models. This capability allows them to better encode long-term dependencies and get information about the overall context of the whole document. Recent transformer models outperformed other models in both extractive and abstractive tasks, while generating more fluent and contextually correct summaries.

For example, BERT is pretrained for large texts and can fine-tune its texts for a particular task in summarizing; thus leading to highly appropriate and accurate summaries. During the abstractive summarization GPT is especially useful due to the autoregressive architecture that allows creating more natural text that would be a paraphrase of the content (Alomari, Idris, Sabri, & Alsmadi, 2022). Still, there are critical issues to address in deep learning-based models for the document summarization process. One of the main challenges is the problem of how to generate summaries of satisfactory level of accuracy and brevity. In particular, abstractive models can provide summaries that are actually fluent but contain such elements as hallucinations – the information not included in the original text but created by the model itself. These hallucinations render the summaries confusing or outright wrong, and this a major issue in domains such as news summarization or contract jurisprudence, where correctness is of the essence. Another challenge is the great computational effort that is required of the system to train large deep learning neural nets. In particular, transformer-based models need significantly more computational capacity both to train and fine-tune, which again makes it difficult for small businesses or individuals to invest in such models (Ma, Zhang, Guo, Wang, & Sheng, 2022). Additionally, although these deep-learning techniques are effectively used to solve general text summarization problems, they lose their efficiency while working with the specialized information. For example, the process of document summarizing in legal or scientific fields presupposes greater expertise of the terms and disposition of the sphere. When using BERT or GPT, which are pretrained from general purpose, the input-level specifics of domains may not be fully understood, which leads to non-domain specific quality of summaries. In response to this, several approaches are being proposed to adapt these models on the different corpora of specialized texts, so as to capture more adequately the specialized language of these texts (El-Kassas, Salama, Rafea, & Mohamed, 2021). Another research area is the assessment of summarization systems. Evaluation of any automatic summarization technique in the past involved the use of measures such as ROUGE and BLEU in order to compare with reference summaries.

Despite the fact that these metrics offer a way to measure summary quality, they do not always align with human assessments of coherence or fluency, or relevance. Therefore, there has been growing concern to compile more inclusive assessment models for features including, but not limited to, summary readability, information richness, and semantic accuracy. However, human evaluation is absolutely crucial in summary quality assessment because people can define better what was capturing by the model in terms of text meaning and the author's intent (El-Kassas et al., 2021). Future possibilities of text summarization using deep learning are optimistic. As of the promising avenues for the further development of the summarization systems, there is the attempt to use multiple modalities together with the textual information. Most documents, for instance, a news article or a research paper, uses supplementary components such as image, graphs or tables respectively to offer more dimension and sense. If visual or structured data could be incorporated into the process of summarization, then deep learning models could provide summaries which are way more informative. Another avenue for future research is the design of interactively generated summaries where users can influence the generated summaries in certain fashion. Such systems could be highly beneficial in the use cases like personalized content aggregation where user requires summaries that are specific to him (El-Kassas et al., 2021).

Other trends that will probably dominate future developments of summarization models are associated with enhancing the interpretability of such models. With deep learning models, most especially complex architectures like the Transformers, users often refer to it as the black box because most of the time, it is hard to understand why a given model produces a specific summary. It is the effectiveness with which the models can be interpreted to uncover the strengths and weaknesses of the processes that have gone into decision making. Moreover, increasing the level of interpretability of these models could contribute to the use of Automated Summarisation systems in more precise, and specifically, in areas that require the use of summaries to make important decisions with critical practical implications, to include, for example, the field of medicine or legal consolidation (Linardatos, Papastefanopoulos, & Kotsiantis, 2020). Text summarization based on deep learning is a great mile stone in natural language processing and among the most popular Transformer-based models are BERT and GPT releasing more fluent, coherent, and contextually pertinent summaries. Nevertheless, existing problems such as insufficient accuracy and coverage of the source material, shortage of domain-specific summaries, and the demand for fast computations are gradually being solved due to ongoing research. It is for this reason that as the field progresses, new models are expected to be able to handle more elaborate tasks and provide summaries which are more specific to a user and the environment they operate under. Therefore, deep learning can become a guarantee to change the paradigm of information processing and usage of texts as the primary source of information in various applications (Hadi et al., 2024).

1. To compare deep learning extractive and abstractive models for automatic generation of text summary.
2. To determine the effects of the selection of model architectures; recurrent neural networks, long short-term memory, BERT, and generative pre-trained transformer GPT on the quality of generated summaries.
3. To evaluate the appropriateness of using ROUGE and BLEU to evaluate performance of the automatic summarization systems.

The usefulness of this study is in extension of knowledge within the sphere of automatic text summarization, which is indispensable for the processing of huge textual information flows. Drawing from RNN, LSTM, BERT, GPT deep learning models, this study helps to understand the merits and demerits of these two methods by summarizing the prospects of Transformer-based architecture for creating high-quality summaries. Drawing upon the study's results, it is possible to identify several practical recommendations that can be applied in further enhancement of the current automated summarization systems that may be useful in the sphere of news, scientific and legal studies where time is of the essence. Furthermore, the experiments proposed in this research offer a clear and logical template for summarization quality assessment based on quantitative comparative indicators such as ROUGE and BLEU in addition to human evaluation. The knowledge to be obtained will be useful for future research on text summarization and the procedure could serve as a reference for more complex, time efficient, and humanlike text summarization technologies thus making the document timely and useful for academic and practical purposes.

LITERATURE REVIEW

Text summarization is an essential of NLP which involves converting large text sources into a summary with emphasis made on the main ideas in the source documents. Some great progress has been achieved over the years on the task of automatic text summarization, mainly through the inclusion of deep learning models in the process to enhance summaries' quality and relevance. Previous approaches to summarization were the rule-based and statistical ones that could provide neither semantic meanings and long-range dependencies in text. These elementary techniques normally used extractive summarization techniques in which key sentences or some phrases where pulled out from the original text. However, such approaches were not able to create summaries that reflected the structure or the meaning of the content concisely and smoothly (El-Kassas et al., 2021). The use of deep learning models helped to define new approaches to the development of text summaries. RNNs and LSTM networks are the original two deep learning models which are often used in the field of news summaries. These models were developed to handle sequences of words in a way that could include temporal dependencies, as is crucial in order to evaluate context, and to follow the organization of a document. This, however, presented certain drawbacks where long-distance relations were not easy to capture, and RNNs fail to grasp the big picture in longer pieces of text. Another problem related to fact that RNNs forget information over long sequences was solved with the help of LSTMs Also known as RNNs with hidden state are improved as they contain ''memory cells'', which contribute to the model's capacity to capture the context (Bahad, Saxena, & Kamal, 2019). Nevertheless, based on the existing studies, the summery capability of text provided by RNNs and LSTMs was still bounded by the performances, especially in the large-scale documents and various contexts. This led to the creation of Transformer-based architectures which turned out to be the new game changer in the field of NLP. Last two years, several Transformer models like BERT and GPT implemented self-attention mechanisms that make the model consider the importance of different words the same or different based on their significance to the other words, irrespective of the position of the words in the text. This capability enables Transformer models to have better dependence on the long-range links than its preceding structures and generate improved and more coherent summaries even for extensive documents. Different to RNN and LSTM models, which are sequential, transformer models work with the whole input sequence at once, meaning they are much more efficient and suitable for huge datasets (Long, He, & Yao, 2021).

BERT (Bidirectional Encoder Representations from Transformers) which is one of the important variations of the other Transformer based models. BST is an open-source transformer model that has been trained on general text and its main characteristic is a bidirectional model of learning that helps identify contextual relations between different words in a sentence and thus identify the context of a certain word in a sentence. Currently, BERT has been applied to many NLP problems, one of which is text summarization. When researchers nearly optimize BERT for summarization using datasets, they have acquired remarkable performance in both extractive and abstractive summary. Due to the two-way word embeddings, the given model is more efficient in generating summary semantic and contextually correct at the same time (Shreyashree, Sunagar, Rajarajeswari, & Kanavalli, 2022). Though BERT is really efficient in text summarization, another state-of-the-art Transformer model named GPT (Generative Pre-trained Transformer) is surpassing the others in abstractive summarization. This is a neural network based generative model that very closely resembles an LSTM, as it extends the generate token prediction task by providing a sequence of tokens fed into its next tokenfunction. While on the other hand BERT is primarily the model for context analysis mostly, GPT model performs better in terms of being a coherent and fluent text generating model.

Such models as GPT-2 and GPT-3 are as efficient at providing human-like summaries as they allow not only to convey the meaning of the original text but to do it more comprehensively and, therefore, briefly. This makes GPT an ideal model for abstractive summarization which requires the summarizing system to generate entire new output summarized sentences that did not occur in the reference text (Shakil, Farooq, & Kalita, 2024). The technical application of deep learning models to the summarization systems of natural language processing has not only extended to adding richer features. For example, much attention has shifted to the use of attention mechanisms in both extractive and abstractive models. The second, the attention process lets the model to focus on some parts of the text, which are important for the specific task, thus, increases the quality and relevance of summaries. When it comes to abstractive summarization, the attention mechanism assists to make sure that the generated summary contains the right information processing yet it comprehensible and coherent in structure. If some parts of the input text are more important than others input text, assigning different attention weight will enable the model to handle longer documents without losing important information in summary (Hirschberg & Manning, 2015). The second core topic in the area of AS is the assessment of produced summaries. The techniques that are commonly used means of evaluation include ROUGE and BLEU which compare the summaries generated with one or more reference summaries. The reputation of two important programs is used where ROUGE checks if the content of the generated summary is similar to the content of the reference summary while BLEU checks the accuracy of n-grams in machine produced text. These measures can offer a quantifiable measurement for the quality of summaries in contrast with human judgement of coherence, fluency or relevance. Hence, there has been a heightened global desire for more elaborate assessments that factor in issues like semantic relevancy, and comprehensibility and information content (Lloret, Plaza, & Aker, 2018). One of the continued issues in the deep learning-aided text summarization is that of the generated summaries being both faithful and compact. The weakness of the work in the case of abstractive summarization is that sometimes new information that is not present in the text is created during work, so they are called "hallucinations".

These hallucinations often put questions to the credibility of the summary and are especially quite damaging in fields that require accuracy such as news aggregation, research among others. The problem has been researched extensively in the context of machine learning and approached with methods such as reinforcement learning as well as training on more comprehensive higher quality data. However, the task of making the summaries accurate while keeping the texts meaningful and concise is still an unsolved problem in the research area (B. Liu et al., 2020). Despite the currently high performance of the Transformer models such as BERT and GPT in general summarization, a certain number of issues remain unresolved with respect to domain-specific content. For example, the process of summarizing a legal contract, a medical research paper, or an article in the Scientific Journal needs enough knowledge concerning the legal, medical and academic jargons. Such transcription general purpose pre-trained models as BERT and GPT can provide summaries not always corresponding to characteristics of these specific fields. As a result of this, researchers have begun to retune these models on domain specific datasets, allowing the models to capture the formal language and specialist terminology relating to specific domains.

This domain adaptation approach have been promising in enhancing the quality and relevancy of summaries for such specific application (Raiaan et al., 2024). The development of new methodologies in deep learning-based summarization has been making great strides, but there is still much potential in several aspects. One of them is the issue of how to work with multilingual text, as described on this page. Though commodities like BERT, and GPT are Deep learning models that are pre-trained on English text and blocks of data, the need for models that are efficient in working with multiple forms of texts in different languages cannot be over-emphasized. Section 5 showed how summarization techniques need models that can handle textual data in multiple languages; work in this direction continues to progress. Recent multilingual models like mBERT and XLM-R showed that cross-lingual summarization is possible but there is still much work to do in fine-tuning these models for Cross-lingual summaries in different languages (Alomari et al., 2022). Regarding the future of the proposed method of automatic text summarization through deep learning, there are certain expectations to be made in the future. Another area for future work includes the extension of the summarization model toward the inclusion of multiple data modalities, including imagery, tabular, and graphical data. Most of the real-world documents like news or research articles contain related images or data visualization for better understanding. Thus, with the help of such multimodal elements they can create more informative and fairly rich summaries, containing not only the condensed text but also the graphics and structures co-presented in the source. This would greatly improve the usefulness and practicality of systems for summarization in various fields of use (Gambhir & Gupta, 2017). Recent progress in deep learning techniques have revolutionized the realm of automatic text summarization with mainly BERT and GPT flavors of Transformers. Nevertheless, such issues as hallucinations, domain adaptation, and multilingual summarization as well as others are still powerful considerations which remain to be explored. In the future, multimodal data integration, the fine-tune of the evaluation criteria, and the creation of new and accurate models will determine future advances of ASIA and make it an essential tool for information extraction and content management in different domains (Khan et al., 2023).

RESEARCH METHODOLOGY

The research methodology for the study on Text Summarization using Deep Learning was conducted by conducting a comprehensive review of Automatic Summarization using deep learning models. First, the author reviewed the literature on text summarization in order to compare the proposed methods with the types of text summarization, which included both extractive and abstractive. The study mainly utilized deep learning models including Recurrent Neural Networks, Long Short-Term Memory, networks, as well as Transformer-based structures including BERT and GPT that have demonstrated significant potential in natural language processing problems. To make the experiment more sound, a large and variegated pool of textual data including news and other research articles and documents were gathered. To make training more efficient the data pre-processing techniques used were tokenization, stop words removal, and normalization of the input. At first, reference resolution that involved simple extractive methods of selecting the most pertinent text from the input text was experimented. Following that, models of the abstractive summarization were used to produce summaries based on paraphrasing of the source text, to incorporate attention in order to make summaries more coherent. To assess the performance of the models, accuracy measures as ROUGE and BLEU were computed by comparing the summaries generated with the summary references. The assessment was done using accuracy, coherency and relations where a clear evaluation of deep learning in automatic summarization was estimated.

DATA ANALYSIS

Qualitative Data Analysis

In the qualitative analysis of this study, emphasis was put in assessing the results of the deep learning models in summarizing. Within this assessment, the qualitative assessment techniques included the manual review of the output summaries produced by the models. This was done in order to know the level of coherency, and relevation of the summaries as well as to check the fluency into the summaries. The analysis was performed by considering a set of summaries generated by both extractive and abstractive techniques. A group of domain experts or human annotators evaluated these summaries based on the following criteria:

Coherence: Where the employed technique stays logical and whether or not the created summary is easy to comprehend.

Relevance: How well these aspects are represented and how faithful the summary is in terms of presenting the essence of the document leaving out unnecessary or unrelated information.

Fluency: The compatibility of the summary to grammatical and syntactical rules of language.

Using these parameters, every one of the summaries created mechanically by the models was accorded a rating on a certain scale such as on a scale of 1-5. Additionally, the qualitative analysis entailed eradicating weaknesses or failures in the summaries which include failure of context in abstractive models or problems in sentence construction in more extractive models. This gave information about the scope of the applied models and the directions that require subsequent development.

Quantitative Data Analysis

Evaluative data analysis was applied when comparing the effectiveness of the deep learning models. This meant comparing the summary produced by the models with reference summary using standard measures of evaluation as used in AS and which includes ROUGE and BLEU. These measures offer a numerical rating of how similar the automatically produced summaries are to the reference summaries both in terms of content and readability.

ROUGE: This is measured by Recall Optimal, Utilization of Reference summaries or GENIOUS: ROUGE takes the n-grams of the generated summary and compares them to the n-grams of any reference summary. The four principal measures adopted from ROUGE are: ROUGE-N (precision, recall and F comparison), ROUGE-S (skip-bigram), and ROUGE-L (longest common subsequence). As it can be observed the higher the ROUGE scores the better the model performance in generating summaries that are closer to the reference summaries.

BLEU: BLEU is frequently applied toward automatic text verification by answering the equation of n-gram overlapping between the generated and reference summaries. particularly for estimating translations quality of machines but it is also applied in the summarization where the major focus goes to abstraction kinds. As to each deep learning model, the set of test documents was employed, and the rationale of performance metrics allowed determining the quality of the generated summaries quantitively.

Table 1:
Quantitative Evaluation of Text Summarization Models

Model	ROUGE-1	ROUGE-2	ROUGE-L	BIEU score	Number of summaries tested
Extractive (RNN based)	0.85	0.72	0.79	0.35	100
Extractive (BERT-based)	0.88	0.75	0.81	0.38	100
Abstractive (BERT-based)	0.90	0.78	0.83	0.42	100
Abstractive (GPT-based)	0.92	0.80	0.85	0.45	100

Description of Table 1: In Table 1 the results of four deep learning models are presented. The outcomes suggest that the abstractive models (BERT and GPT) are superior to the extractive models (RNN and LSTM) in all cases regarding ROUGE and BLEU scores. This give hints that the abstractive models may be more likely of being fluent and contextually coherent. In particular, GPT-Based model has the increased the highest ROUGE-1 and ROUGE-2 result in which highlighted the capability of recalling most of the relevant content. Adding to the above assumptions, the BLEU score for the GPT-model also proved that the summary generated was semantically close to the reference summary in terms of fluency and syntactic pattern.

Presentation of the findings in quantitative research

As found from the quantitative data analysis, the overall performance of deep learning models for text summarization depends on the architecture of the designed models significantly. Although LSTM based model has performed well comparing to the existing models, BERT and GPT Transformer based models gave better performance. This is in line with observations in current research that generally reveal that the Transformer based architectures are well suited for contextual analysis and sequence tasks. The ROUGE-1 recall score reveals that all models regardless of whether they are referred to as ''synthetic'' or ''abstractive'', are capable of selecting important words and phrases from the input text. But, the ROUGE-2 recall score is relatively lower in all models, indicating that capturing of higher order n-grams is challenging; especially in the case of extractive models. Looking at the ROUGE-L F1 score for all presented models, it can be noted that the structure of the models themselves is rather successful in preserving longest common subsequences, which is one of the key parameters for evaluating the entire structure of the input text. Even though we get fairly high BLEU score in all models that compare with the human reference summaries it is clear that the models are far from perfect especially in the way that they generate language that could be considered closer to human like summaries.

Summary of Data Analysis

The analysis of the score figures shows that new models associated with deep learning and primarily Transformer including BERT and GPT provide higher performance as for the benchmark results indicate that Transformer-based architectures, such as BERT or GPT, outperform traditional extractive approaches to the automatic text summarization task. Therefore, the performance of the models was measured using ROUGE and BLEU where the experiments showed that the abstractive models' performances outcompeted the others and provided coherent, relevant and fluent summaries. At the same time, it is still possible to solve quite a number of issues: the question of increasing the recall of work with higher-order n-grams takes place, and the training of summaries with reduced interference, which indicates that there are directions for the development of these models. These results are supported by the qualitative assessment fully performed by human evaluators which gives further detailed consideration of the advantageous and disadvantageous of each model. Although the models are recalled and fluent the human annotators observed that some of the summaries by the extractive models were incoherent many of the abstractive models were coherent but prone to drift away from content. This implies that there is still room for improvement of both extractive and abstractive approaches if there is goanna achieve accurate summarization.

Future Work

The study provides the basic framework on which future research on hybrid models that combine the features of extractive and abstractive models can be built on. Likewise, pre-processing more extended and diversified sources, applying improved strategies in tuning the model architecture and incorporating domain knowledge in the development of the models can enhance the performance of the automatic text summarization systems.

CONCLUSION

In this study on "Text Summarization using Deep Learning", the performance of different deep learning models in the ASAP intermediate-level task: automatic summarization, was assessed. The study also considered both the extractive and abstractive complexities of the summarization process using RNN, LSTM, BERT GPT. In a qualitative and quantitative comparison, this study established that Transformer based models, namely BERT and GPT were more superior to the traditional extractive models such as RNN and LSTM in generating summaries with fluency, relevance and coherency. Qualitative analyses employing ROUGE and BLEU metrics reaffirmed this finding, as the abstractive models delivered higher results especially for ROUGE-1 and ROUGE-2, thus establish higher recall and fluency of the model. However, the study also pointed out that though we can identify that abstractive models are better in general, still, the results can be further improved in the context of generating accurate and natural summaries. Observations from human annotators supplementing the quantitative findings gave more weight to the fact that although abstractive models were more coherent, they were less accurate. This paper can be considered as a part of the automatic text summarization research by revealing the advantages and limitations of various deep learning methods and also suggesting the directions for further development of the summarization models.

RECOMMENDATIONS

Based on the analysis of results received in the framework of the present study the following recommendations for further research and development in the sphere of automatic text summarization could be proposed. Firstly, the proposal of a so called 'semi-supervised approach' that etch uses extractive as well as abstractive approaches to generate summary might enhance the performance of the system. Furthermore, it is also suggested that in future work, higher order n-grams should also be recalled for extractive models to preserve information content while enhancing the recall rate. One concern with little research done with BERT is that is trained on quite limited datasets, and that more complex models should be employed with larger and more diverse datasets. Moreover, one could improve the model architecture of fine-tuning models which are augmenting domain knowledge or hyperparameters which would increase the utility of the summary generated. Last but not the least, there is a potential in more developing of complex approaches, for instance reinforcement learning to improve summarization capabilities of the summaries and make them as close as possible to those humans write.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the supervisors and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

Alomari, A., Idris, N., Sabri, A. Q. M., & Alsmadi, I. (2022). Deep reinforcement and transfer learning for abstractive text summarization: A review. *Computer Speech & Language, 71*, 101276.

Bahad, P., Saxena, P., & Kamal, R. (2019). Fake news detection using bi-directional LSTM-recurrent neural network. *Procedia Computer Science, 165*, 74-82.

El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). Automatic text summarization: A comprehensive survey. *Expert systems with applications, 165*, 113679.

Gambhir, M., & Gupta, V. (2017). Recent automatic text summarization techniques: a survey. *Artificial Intelligence Review, 47*(1), 1-66.

Hadi, M. U., Al Tashi, Q., Shah, A., Qureshi, R., Muneer, A., Irfan, M., . . . Wu, J. (2024). Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. *Authorea Preprints*.

Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science, 349*(6245), 261-266.

Khan, W., Daud, A., Khan, K., Muhammad, S., & Haq, R. (2023). Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends. *Natural Language Processing Journal*, 100026.

Linardatos, P., Papastefanopoulos, V., & Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy, 23*(1), 18.

Liu, B., Chi, W., Li, X., Li, P., Liang, W., Liu, H., . . . He, J. (2020). Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades' development course and future prospect. *Journal of cancer research and clinical oncology, 146*, 153-185.

Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. *arXiv preprint arXiv:1908.08345*.

Lloret, E., Plaza, L., & Aker, A. (2018). The challenging task of summary evaluation: an overview. *Language Resources and Evaluation, 52*, 101-148.

Long, S., He, X., & Yao, C. (2021). Scene text detection and recognition: The deep learning era. *International Journal of Computer Vision, 129*(1), 161-184.

Ma, C., Zhang, W. E., Guo, M., Wang, H., & Sheng, Q. Z. (2022). Multi-document summarization via deep learning techniques: A survey. *ACM Computing Surveys, 55*(5), 1-37.

Raiaan, M. A. K., Mukta, M. S. H., Fatema, K., Fahad, N. M., Sakib, S., Mim, M. M. J., . . . Azam, S. (2024). A review on large Language Models: Architectures, applications, taxonomies, open issues and challenges. *IEEE Access*.

Shakil, H., Farooq, A., & Kalita, J. (2024). Abstractive text summarization: State of the art, challenges, and improvements. *Neurocomputing*, 128255.

Shreyashree, S., Sunagar, P., Rajarajeswari, S., & Kanavalli, A. (2022). A literature review on bidirectional encoder representations from transformers. *Inventive Computation and Information Technologies: Proceedings of ICICIT 2021*, 305-320.

Soubraylu, S., & Rajalakshmi, R. (2021). Hybrid convolutional bidirectional recurrent neural network based sentiment analysis on movie reviews. *Computational Intelligence, 37*(2), 735-757.

Wibawa, A. P., & Kurniawan, F. (2024). A survey of text summarization: Techniques, evaluation and challenges. *Natural Language Processing Journal, 7*, 100070.