Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. CIST 117
DOI: 10.11159/cist23.117

Fine-Tuned PEGASUS: Exploring the Performance of the Transformer-Based Model on a Diverse Text Summarization Dataset

Mohammed Alsuhaibani
Department of Computer Science, College of Computer, Qassim University,
Buraydah 52571, Saudi Arabia
m.suhibani@qu.edu.sa

Abstract **-** Text summarization is the art of succinctly capturing the essence of a lengthy text document through a concise summary. The intricate craft of text summarization involves distilling a voluminous text document into a brief and elegant summation that conveys its core message. Ultimately, the goal of text summarization is to help on grasping the essence of a text without having to wade through its entire length. In this research, we propose to fine-tune and explore the quality performance of the deep learning and transformer-based PEGASUS model for abstractive text summarization on a diverse dataset. The diversity of the dataset is expected to challenge the model and test its capabilities in generating summaries for a wide range of text types and styles. Our experimental results indeed indicate that the model's performance varies based on the topic and category of the text reaching as high as 88.03 F1 (ROUGE-1) score with some topics and as low as 81.22 with others. This is crucial as texts, such as political, economic, literary, legal, and medical, have distinctive writing conventions and styles, and a model that performs well on a diverse dataset is more likely to adapt to other text types.

Keywords**:** Text summarization, Transformers, Diverse dataset, Text processing, Deep learning

1. Introduction

Text summarization is a topic of extensive investigation in Natural Language Processing (NLP), aimed at reducing lengthy text to its core essence, while still maintaining the important information and overall meaning of the original text [1]. The discipline of text summarization utilizes two primary strategies: extractive and abstractive [2]. Extractive summarization entails selecting key sentences or phrases from the source text and composing them to create the summary. In contrast, abstractive summarization involves constructing a summary that is not an exact replica of the source text, but rather a new and concise version of it, written in the model's own words. Abstractive text summarization is a more challenging task than extractive summarization, as it requires the model to have a deeper understanding of the input text and to be able to generate coherent and grammatically correct summaries in natural language [3]. This task is important for a variety of applications, such as summarizing long documents, generating summaries of news articles, and condensing social media posts.

Recently, transformer-based models have been used extensively in various NLP tasks and have shown promising results [4, 5, 6]. Transformer-based models are a type of Neural Network (NN) architecture that uses self-attention mechanisms to process input sequences and generate output sequences. They have been shown to be particularly effective in tasks such as machine translation [7], language modelling [8], and text classification [9]. Pre-training with Extracted Gap-sentences for Abstractive Summarization (PEGASUS) [10] model is a SOTA abstractive text summarization. It is based on transformer architecture and uses a combination of techniques to generate high-quality summaries of long documents.

In this research, we propose to fine-tune and explore the PEGASUS model for abstractive text summarization on a diverse dataset. The diversity of the dataset is expected to challenge the model and test its capabilities in generating summaries for a wide range of text types and styles. This is important because different types of text, such as politics, economics, literature, laws, and medicine, have different writing styles and language conventions, and a model that can perform well on a diverse dataset is more likely to generalize well to other text types. The goal of this study is to evaluate the performance of the transformer- based model on this diverse dataset and to identify any potential improvements that can be made to the model to improve its summarization abilities.

To the best of our knowledge, this is the first study to evaluate the performance of a transformer-based model on a diverse text summarization dataset. Previous research has primarily focused on using transformer-based models on specific types of text, such as news articles or scientific papers. By evaluating the model on a diverse dataset, we aim to provide a more comprehensive assessment of its performance and to identify any potential challenges that the model may face when summarizing different types of text. Overall, this research has the potential to contribute to the field of text summarization by providing a detailed evaluation of a transformer-based model on a diverse dataset and by identifying potential areas for improvement in the model's performance.

The remainder of the paper is structured as follows: first, we will review related studies in section 2. Next, in section 3, we will detail the methodology used in our research. Section 4 will describe the experimental setup and procedures along with the obtained results. Finally, section 5 will summarize the key contributions and limitations of the study, and suggest possibilities for future research.

2. Related Work

In the early days of ATS, most approaches relied on statistical models to select and copy the most relevant parts of the original text [11]. For instance, using statistical techniques to calculate the relative significance of words based on their frequency and distribution, and generated a summary by using the most significant sentences [12]. However, these early statistical approaches were unable to produce abstractive summaries because they lacked the ability to understand and analyse the meaning of natural language. This led to the development of more advanced systems that could address the limitations of these methods by incorporating a deeper understanding of the semantics of natural language.

Abstractive ATS systems aim to generate a condensation of the original text created by grasping its central concepts, rather than simply copying and pasting sentences [13]. These systems utilize NLP approaches to create a summary that is different from the original written content. Abstractive methods can be broadly classified into three categories: structure-based, semantic-based approaches and deep learning-based approaches [3]. Deep NNs, specifically Sequence-to-Sequence (Seq2Seq) [14] learning models, have been effectively used in automated text summarization (ATS) systems. These approaches often employ a Recurrent Neural Network (RNN). For example, a study by Shi et al. [12] conducted a thorough review of various seq2seq models for abstractive text summarization, focusing on summary generation methods, training techniques and network architecture.

Kouris et al. [15] has put forth a novel approach to enhance abstractive text summarization by integrating deep learning techniques and semantic data trans- formations. Firstly, they introduce a theoretical framework for semantic-driven text generalization, which is combined with a deep encoder-decoder architecture to produce a summary in a generalized form. Then, they present a methodology that converts the generalized summary into a human-readable format, while pre- serving important information from the source text and overcoming the problem of infrequent or out-of-vocabulary words.

ATS systems has been greatly influenced by the introduction of the transformer architecture, thanks to the pioneering work of Vaswani et al. [16]. This paradigm-shifting mechanism focuses on capturing input and output representations without having to depend on sequence recurrence or convolution, making it a game-changer in the field. The real marvel that came next, building upon this architecture, were the pre-trained Transformer Language Models (TLMs) such as BERT[17]. These have quickly risen to be the go-to models in NLP because of their ability to achieve state-of-the-art results in numerous tasks.

The structure of these TLMs allows them to be initially trained on a general task, then finely tuned to accommodate specific tasks with little modification needed in the model architecture. For example, BERT can easily be tweaked to handle ATS by adding a single output layer. This flexibility is achieved by introducing inputs and outputs specific to the task at hand into the BERT model, and then tweaking the parameters to best suit that particular task. Consequently, a wide array of pre-trained models has been constructed and meticulously fine-tuned for ATS tasks, encompassing diverse languages. Renowned models like mBART [18], PEGASUS, and mT5 [19] stand as testament to this, showcasing the ability to generate high-quality summarizations.

Abdel-Salam and Rafea [20] introduced a study whose aim is to conduct an investigation into the efficacy of diverse iterations of BERT-based models on the task of text summarization by means of a series of empirical analyses. Furthermore, Furthermore, this study proposes a new summarization model which has been trained and fine-tuned utilizing the SqueezeBERT encoder variant. Similarly, Grail et al. [21] presents a hierarchical propagation layer that facilitates the dissemination of information across numerous transformer windows. Their methodology employs a hierarchical framework framework wherein the input is partitioned into multiple independent blocks, each processed by scaled dot-attentions, and and subsequently merged across successive layers. They substantiate the efficacy of their approach by conducting experiments on three diverse extractive summarization corpora encompassing lengthy scientific papers and news articles. In the same vein, Doan et al. [22] have ingeniously leveraged the power of Heterogeneous Graph Neural Networks (HeterGNN) [23] to tackle the challenging task of processing long documents. To this end, they have devised an approach that involves constructing a homogeneous graph consisting of sentence-level nodes, which is then integrated with HeterGNN to capture both inter and intra-sentence connections and ex- tract valuable semantic information. The efficacy of their method has been extensively validated through rigorous experimentation on two widely-used bench- mark datasets, namely PubMed and ArXiv.

The DeepSumm approach, as proposed by Joshi et al. [24], ingeniously employs the concepts of topic modelling and word embeddings to distill the crux of single documents through an extractive technique. Each sentence in a document is encoded through a pair of distinct RNNs, one leaning on the probabilistic topic distributions and the other on word embeddings. This encoding is then passed through a Seq2Seq network. In a subsequent step, the outputs from the encoder and decoder in the Seq2Seq networks are converged. This integration process is guided by an attention mechanism, which helps weigh the significance of each element. After this, a multi-layer perceptron network steps in to transform these weights into a score, which then helps in producing the summary. Tsvigun et al. [25], on the other hand, focus on addressing the ATS challenge through a novel query strategy for Active Learning (AL) rooted in diversity principles. Their research has revealed notable outcomes. In particular, they discovered that the employment of a unique strategy in AL annotation markedly elevates the model's performance concerning ROUGE and consistency scores. This enhancement is observed while staying within a specified annotation budget, further demonstrating the efficiency of their approach.

Most recently, Chouikhi and Alsuhaibani [26] present study endeavours to delve into cutting-edge model advancements in the realm of ATS. Specifically, the focus is on exploring and scrutinizing five SOTA-deep TLMs. In order to accomplish this objective, various text summarization datasets, specially curated for the Arabic language, have been meticulously scrutinized and adapted for the purpose of evaluation. To thoroughly comprehend the capabilities of these models, they have been compared against the conventional deep learning and machine learning-based baseline models. The primary intention of this comparative analysis is to evaluate the performance of the state-of-the-art models with respect to the existing conventional models in the task of text summarization. The study is a comprehensive and rigorous investigation into the advancements in ATS and aims to provide an in-depth understanding of the latest TLMs developments in this field. However, its lack of diversity in the datasets used in the study raises the concern that the findings and conclusions may not be generalizable to a broader range of text summarization tasks or to other languages.

3. Methodology

PEGASUS stands as a ground-breaking pre-training approach, specially crafted for abstractive summarization tasks. Its design intends to enhance the performance of models by harnessing the power of extensive unannotated text data. The approach kicks off by first training a language model on a voluminous text corpus. This trained model then steps into the role of a miner, digging into the corpus to extract sentences that harbor a “gap”, a missing nugget of information that begs to be inferred or summarized. These “gap” filled sentences are then repurposed as a specialized training dataset for a summarization model. The model is then fine-tuned on this task-specific data, leading to an upswing in its performance. The heart of PEGASUS lies in its use of a transformer-based language model that is pre-trained on an expansive text corpus. This model is then set to the task of unearthing sentences with “gaps”. What sets PEGASUS apart is its ability to guide the model towards generating more coherent and fluent summaries. This is made possible by training the model on these extracted "gap" sentences. As these sentences are more inclined to hold the key information that begs summarization, the model is schooled in the art of identifying and drawing out this vital information with enhanced effectiveness. Thus, PEGASUS serves as an innovative approach in the field of abstractive text summarization.

As such, PEGASUS presents a new pre-training objective called Gap Sentences Generation (GSG) and compares the masked language model objective of BERT through individual and combined evaluations. The idea behind as clearly stated in the paper, is the belief that if the pre-training objective is more similar to the end task, the fine-tuning performance will be better and faster. With the goal of using the model for abstraction summary, a pre-training objective generates text from an input document is introduced. To take advantage of a large text corpus for pre-training, they created a self-supervised seq2seq objective that does not require abstractive summaries.

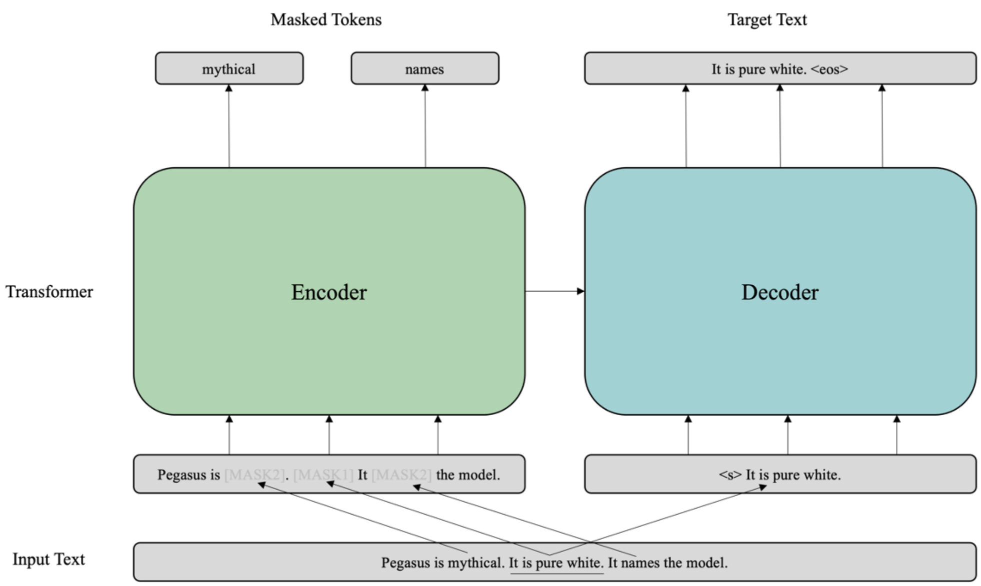

Figure 1: Overall architecture of the PEGASUS model [32].

The methodology employed by PEGASUS, as outlined in [10], builds upon the triumph of masking individual words and clusters of words and opts to obscure complete sentences from the document, then collates the masked sentences to form a simulacrum of a summary. This procedure transposes each selected sentence with a marker symbol [MASK1], serving as a notification to the model. The Gap Sentence Ratio (GSR), defined as the proportion of selected gap sentences relative to the overall number of sentences within the document, is analogous to the mask rate that has been explored in prior studies. To make the pre-training objective more similar to the summary, it selects sentences that seem crucial or fundamental to the text. This modus operandi confers the dual advantages of masking and adheres to the downstream task's format, thus further enhancing its efficacy.

The model uses Masked Language Modelling (MLM) to train a Transformer encoder, following BERT. This involves selecting 15% of the tokens in the input text and then randomly replacing them with a mask token, a random token, or leaving them unchanged. It also uses the GSG in conjunction with MLM. However, MLM might not significantly improve performance on downstream tasks when used for a substantial number of training steps and ultimately, it was chosen not to be included in the large model [10].

4. Experiments and Results

In order to fine-tune the model for the downstream task, the text summarization, we just followed and adopted the original model to perform the summarization, in which relying solely on publicly available abstractive summarization datasets, such as XSum dataset [27]. XSum dataset features a collection of 227,000 articles from the BBC, covering a wide range of subjects, for a period of 7 years. It also includes one-sentence summaries which are written by professionals.

Table 1: Trained model's settings.

Layers	Parameters	Vocab Size	Epochs	Batch Size
16	568M	96K	3	6

Transformers library from HuggingFace[1] was utilized for both the experiments and the model training. Following [26], the input documents were limited to 200 tokens, and the generated summaries had a maximum of 12 tokens. Beam search with a value of 4 for num-beam was employed and the batch size = 6. The experimental settings are shown in Table 1 as adopted by [26].

For the diverse dataset, we opted to use the Arabic diverse NADA dataset [28]. It is an amalgamation of two pre-existing datasets, the DAA [29] and the OSAC [30]. The former boasts nine distinct categories, each containing 400 meticulously pre- processed documents. Each classification is designated its own directory, harbouring all relevant files. The latter dataset encompasses six categories, each housing raw documents. Both datasets underwent a refining process and were seamlessly joined to create the comprehensive and diverse NADA dataset. Table 2 shows the details of the dataset in regard to the number of instances of each category/topic.

Table 2: Statistics of the diverse used text summarization NADA dataset.

Topics	Number of Instances
Literature	400
Economy	1307
Politics	400
Law	1644
Sports	1416
Art	400
Religions	515
Pure Sciences	400
Health Sciences	428
Astronomy	400

After the pre-processing steps (e.g. tokenization) are applied to the dataset, it is saved in a CSV file format. Then, tokenization is performed to obtain special tokens which are used as inputs for the transformer models (either encoder or decoder). The output generated will be a summary, and AdamW [31] optimizer is utilized with a maximum summary length of 150.

Assessing the merit of a summary can prove to be a challenging task, as a single document or compendium of documents may possess a multitude of viable summarizations [26, 32]. The well-known and utilized evaluation method is program-based assessment like recall, accuracy, and F1-score. A commonly used automated measure in text summarization is ROUGE [26, 33], which compares the generated summaries to the corresponding references. ROUGE-1 gauges the congruence between unigrams, while ROUGE-2 appraises the agreement between bigrams. ROUGE-L, on the other hand, determines the maximal continuous sequence shared between two text pairs, both with and without sentence segmentation into distinct lines. The model has undergone a rigorous evaluation on the diverse NADA dataset, a corpus that encompasses a diverse array of subjects, including Literature, Economy, Politics, Law, Sports, Art, Religious, Sciences, Health, and Astronomy The

[1] https://huggingface.co/docs/transformers/index

efficacy of the model was assessed utilizing the ROUGE-1, ROUGE-2 and ROUGE-L metrics. The evaluation outcome was depicted in Table 3 as the F1- score for each metric, with the best results prominently emphasized in bold

The reported results in Table 3 portray that the PEGASUS model has achieved its best performance in the diverse in the economy domain with 88.03 F1-score in ROUGE-1, with the second-best performance being witnessed in the of law. This pattern is also witnessed with ROUGE-2 and ROUGE-3. This could be attributed to the abundance of and documents available for the model to draw knowledge from in these two topics. On the contrary, it is plausible that model faced difficulties in the other topics due to the paucity of data available. We have also incorporated Fig. 2 of 1 results from Table 3 for enhanced readability. The figure showcases a visually appealing and intuitive illustration of performance of the PEGASUS model on the diverse NADA dataset, making it easier for the reader to grasp and the results.

Table 3: Reported F1-score results on the NADA dataset topics.

Topics	**ROUGE-1**	**ROUGE-2**	**ROUGE-L**
Literature	81.22	75.96	77.52
Economy	**88.03**	81.06	**84.37**
Politics	84.67	79.46	81.45
Law	86.94	**81.19**	82.33
Sports	85.08	80.16	82.03
Art	83.65	77.60	80.42
Religions	81.77	76.95	78.10
Pure Sciences	83.12	76.99	79.19
Health Sciences	84.33	78.22	80.59
Astronomy	82.98	76.97	79.22

To advance our understanding of the effectiveness of the model in the context of diverse text summarization, in Fig. 3, we conducted a meticulous qualitative analysis on a select corpus of summaries that were generated by the model using the NADA dataset. Our inquiry was particularly focused on summaries generated in the fields of economics and law, which exhibited comparable patterns to those in other topics. As a measure to enhance accessibility, we translated the outcomes from Arabic to English. Our scrutiny revealed that the chosen summaries amply exemplify the model's excellent output quality, as they precisely captured the principal themes of the original text while preserving syntactic accuracy and coherence. For example, on the topic of the economy, the original text was discussing a piece of news regarding the warning by international economists about the unbearable oil prices with further details about the particular oil prices. The model was able to accurately summarise it with few words delivering the core idea. Taken in totality, the qualitative analysis corroborates the potential of the PEGASUS model in the realm of diverse text summarization tasks.

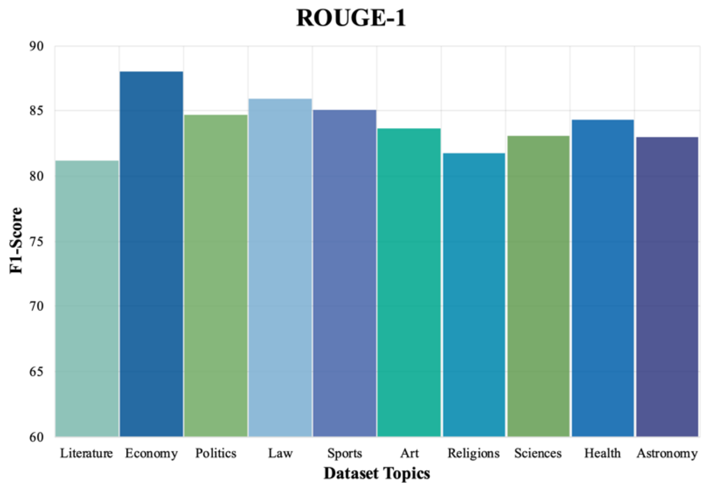

Figure 2: Reported F1-score on the diverse NADA dataset topics using ROUGE-1 scores.

Topic	**Economy**
Original Text	An international economist warned that the global economy is unable to bear more in oil prices. The British newspaper Financial Times quoted "the chief economist of the International Energy Agency, which allows Birol that the rise of oil prices and their survival above the level of $70 a barrel may prevent the global economy from recovering from the stagnation he is going through. Birol said, in an interview with the newspaper Friday and published on Tuesday: "If we take one forward and we see more height, this may cause the economic recovery to slow down." Last Friday, the return of the price of the American crude barrel witnessed a level of $70, and the two exceeded the price of 73.75 dollars in the European markets, which is the highest This year, driven by industry data issued by China, and those related to the construction sector in the United States of America. Related links to future oil contracts are thrived on the impact of the dollar and market gains, during the Group of Eight Group in Italy last month, was invited to a better audit in the energy markets, while organizational parties in the United Kingdom are viewed if the work markets are It is subject to appropriate control.
Summarised Text	High oil prices may hinder the recovery of the global economy
Topic	**Law**
Original Text	Dubai Real Estate Foundation, provided that the profits achieved from the investment and management of these clubs and stadiums will be transferred to the Dubai Foundation for the Management and Organization of Events. The Foundation also informs all the rights and obligations related to the previous agreements and contracts on the issuance of this law and concluded in the name of the government or in the name of Dubai Golf Foundation, the Emirates Golf Club Authority and the Dubai Creek Club related to the activities included in the jurisdiction of the institution or transferred to it according to the provisions of this law. The law stated that all the employees of the Dubai Golf Foundation, the Dubai Creek Club Foundation and the Emirates Golf Club Authority be transferred to the Foundation without prejudice to their acquired rights. According to Article (26) of the Dubai Foundation for Event Management Law, Decree No. (2) of 1989 will be cancelled by the establishment of a body called the Emirates Golf Club and Decree No. (13) for the year 1992 regarding the establishment of a public institution known as the Dubai Creek Club and Decree No. (17) for the year 2001 Establishing the Dubai Golf Corporation, and any text in any other legislation is cancelled to the extent that contradicts the provisions of this law. This law is published in the Official Gazette and it shall be established from the date of its publication.
Summarised Text	A law for the Dubai Foundation for Events Management

Figure 3: Selected summaries generated by the model on two topics, economy and law.

It is imperative to acknowledge that the results represented in the table are merely a glimpse into the performance of the PEGASUS model on the diverse NADA dataset. Further comprehensive evaluations and testing are imperative to attain a more robust understanding of the model's proficiency. Moreover, it would be advantageous to compare the results of PEGASUS with other algorithms in the field using the diverse dataset to gauge its competitiveness.

5. Conclusion

In conclusion, this study has proposed the fine-tuning and evaluation of the PEGASUS model for abstractive text summarization using a diverse dataset. The multifaceted nature of the dataset presents a formidable test for the model, demanding that it exhibits proficiency in generating summaries for a vast spectrum of text types and styles. In fact, the experimental findings unequivocally substantiate that the model's efficacy is contingent upon the specific topic and category of the text in question. While the model's performance scales to an impressive 88.03 F1 score for certain topics, it can decrease significantly to as low as 81.22 for others. Nevertheless, the results of the experiments demonstrate the model's capability in generating summaries for a wide range of text types and styles while showing variations in performance based on the topic and category of the text. This highlights the importance of training models on diverse datasets for improved generalization to unseen texts. The findings of this study contribute to the advancement of the field of text summarization and have potential implications for various NLP applications.

Future work could extend this study by incorporating other SOTA models for comparison and exploring additional strategies for improving the performance on diverse text genres.

References

[1] Yogan, J.K., Goh, O.S., Halizah, B., Ngo, H.C., Puspalata, C.: A review on automatic text summarization approaches. Journal of Computer Science 12(4), 178–190 (2016).

[2] Nenkova, A., McKeown, K.: A survey of text summarization techniques. Mining text data pp. 43–76 (2012).

[3] Song, S., Huang, H., Ruan, T.: Abstractive text summarization using lstm-cnn based deep learning. Multimedia Tools and Applications 78, 857–875 (2019).

[4] Gaanoun, K., & Alsuhaibani, M. Fabricated Hadith Detection: A Novel Matn-Based Approach With Transformer Language Models. IEEE Access, 10, 113330-113342 (2022).

[5] Myagmar, B., Li, J., Kimura, S.: Cross-domain sentiment classification with bidirectional contextualized transformer language models. IEEE Access 7, 163219–163230 (2019).

[6] Pilault, J., Li, R., Subramanian, S., Pal, C.: On extractive and abstractive neural document summarization with transformer language models. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 9308–9319 (2020).

[7] Huang, F., Zhou, H., Liu, Y., Li, H., Huang, M.: Directed acyclic transformer for non-autoregressive machine translation. In: International Conference on Machine Learning. pp. 9410–9428. PMLR (2022).

[8] Borgeaud, S., Mensch, A., Hoffmann, J., Cai, T., Rutherford, E., Millican, K., Van Den Driessche, G.B., Lespiau, J.B., Damoc, B., Clark, A. and de Las Casas, D.: Improving language models by retrieving from trillions of tokens. In: International conference on machine learning. pp. 2206–2240. PMLR (2022).

[9] Terechshenko, Z., Linder, F., Padmakumar, V., Liu, M., Nagler, J., Tucker, J.A., Bonneau, R.: A comparison of methods in political science text classification: Transfer learning language models for politics. Available at SSRN 3724644 (2020).

[10] Zhang, J., Zhao, Y., Saleh, M., Liu, P.: Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In: International Conference on Machine Learning. pp. 11328–11339. PMLR (2020).

[11] Saggion, H., Poibeau, T.: Automatic text summarization: Past, present and future. Multi-source, multilingual information extraction and summarization pp. 3–21 (2013).

[12] Shi, T., Keneshloo, Y., Ramakrishnan, N., Reddy, C.K.: Neural abstractive text summarization with sequence-to-sequence models. ACM Transactions on Data Science 2(1), 1–37 (2021).

[13] Yang, M., Li, C., Shen, Y., Wu, Q., Zhao, Z., Chen, X.: Hierarchical human-like deep neural networks for abstractive text summarization. IEEE Transactions on Neural Networks and Learning Systems 32(6), 2744–2757 (2020).
[14] Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. pp. 3104–3112 (2014).
[15] Kouris, P., Alexandridis, G., Stafylopatis, A.: Abstractive text summarization based on deep learning and semantic content generalization. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. pp. 5082–5092 (2019).
[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems 30 (2017).
[17] Kenton, J.D.M.W.C., Toutanova, L.K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT. pp.4171–4186 (2019).
[18] Liu, Y., Gu, J., Goyal, N., Li, X., Edunov, S., Ghazvininejad, M., Lewis, M., Zettlemoyer, L.: Multilingual denoising pre training for neural machine translation. Transactions of the Association for Computational Linguistics 8, 726–742 (2020).
[19] Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., Raffel, C.: mt5: A massively multilingual pre-trained text-to-text transformer. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 483–498 (2021).
[20] Abdel-Salam, S., Rafea, A.: Performance study on extractive text summarization using bert models. Information 13(2), 67 (2022).
[21] Grail, Q., Perez, J., Gaussier, E.: Globalizing bert-based transformer architectures for long document summarization. In: Proceedings of the 16th conference of the European chapter of the Association for Computational Linguistics: Main volume. pp. 1792–1810 (2021).
[22] Doan, X.D., Le Nguyen, M., Bui, K.H.N.: Multi graph neural network for extractive long document summarization. In: Proceedings of the 29th International Conference on Computational Linguistics. pp. 5870–5875 (2022).
[23] Zhang, C., Song, D., Huang, C., Swami, A., Chawla, N.V.: Heterogeneous graph neural network. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 793–803 (2019).
[24] Joshi, A., Fidalgo, E., Alegre, E., Fern´andez-Robles, L.: Deepsumm: Exploiting topic models and sequence to sequence networks for extractive text summarization. Expert Systems with Applications 211, 118442 (2023).
[25] Tsvigun, A., Lysenko, I., Sedashov, D., Lazichny, I., Damirov, E., Karlov, V., Belousov, A., Sanochkin, L., Panov, M., Panchenko, A. and Burtsev, M.: Active learning for abstractive text summarization. arXiv preprint arXiv:2301.03252 (2023)
[26] Chouikhi, H., Alsuhaibani, M.: Deep transformer language models for arabic text summarization: A comparison study. Applied Sciences 12(23), 11944 (2022).
[27] Narayan, S., Cohen, S.B., Lapata, M.: Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. pp. 1797–1807 (2018).
[28] Alalyani, N., Marie-Sainte, S.L.: Nada: New arabic dataset for text classification. International Journal of Advanced Computer Science and Applications 9(9) (2018).
[29] Abuaiadah, D., El Sana, J., Abusalah, W.: On the impact of dataset characteristics on arabic document classification. International Journal of Computer Applications 101(7) (2014).
[30] Saad, M.K., Ashour, W.: Osac: Open source arabic corpora. In: 6th ArchEng Int. Symposiums, EEECS. vol. 10 (2010)
[31] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
[32] Zhang, M., Zhou, G., Yu, W., Huang, N., Liu, W.: A comprehensive survey of abstractive text summarization based on deep learning. Computational Intelligence and Neuroscience 2022 (2022).
[33] ROUGE, L.C.: A package for automatic evaluation of summaries. In: Proceedings of Workshop on Text Summarization of ACL, Spain (2004).