Arabic Text Summarization using transformer-based architectures

Karim Morsi*
Faculty of computer and information science, Ain-Shams University
Fatma najib†
Faculty of computer and information science, Ain-Shams University
Wedad Hussein‡
Faculty of computer and information science, Ain-Shams University
Rasha Ismail§
Faculty of computer and information science, Ain-Shams University

Abstract

Text summarizing is one of the most challenging tasks in natural language processing (NLP). This task is addressed in a large number of research projects and papers in the literature, but most of them focused on English language. Few studies are dealing with the complex Arabic language. Pre-trained Transformer-based language models have shown remarkable efficacy in addressing problems associated with text generation and natural language processing in recent times. However, there has not been much research on applying these models to Arabic text production. This study focuses on the implementation and fine-tuning pre-trained transformer-based language model structures for Arabic abstractive summarization, including AraBERT, mBERT models, and AraT5. We applied mBERT and AraBERT in the context of text summarization using a BERT2BERT-based encoder-decoder model. ROUGE measurements and manual human evaluation have been used to test the suggested models. Our models are trained and tested using XL-Sum Dataset of 46897 high-quality text-summary pairs. Their performance on out-of-domain data was also compared. We found that AraT5 outperforms AraBERT and mBERT Models, suggesting that a pre-trained Transformer with encoder-decoder functionality is more suited for text summarization. Moreover, AraT5 achieve high performance on out-of-domain dataset and received higher accuracy ratings in human evaluations compared to other models.

Key Words:Arabic natural language processing; Abstractive text Summarization; machine learning; Deep learning; Transfer learning models.

*Faculty of computer and information science, Ain-Shams University, Cario, Egypt. Email: karim.mohamed.fcis@cis.asu.edu.eg.
†Faculty of computer and information science, Ain-Shams University, Cario, Egypt. Email: fatma_mohamed@cis.asu.edu.eg.
‡Faculty of computer and information science, Ain-Shams University, Cario, Egypt. Email: wedad.hussein@cis.asu.edu.eg.
§Faculty of computer and information science, Ain-Shams University, Cario, Egypt. Email: rashaismail@cis.asu.edu.eg.

1 Introduction

It becomes more difficult to quickly and accurately extract important information from texts due to the massive volume of digital text data generated every day [1]. Moreover, automatic text summarizing is important for many applications. It improves the process of retrieving important data from digital documents through the use of advanced filtering techniques, making it easier to find embedded knowledge in these materials. Additionally, this technology helps manage the enormous amount of textual material that is accessible. Document summarizing helps to overcome the challenges caused by the huge amount of information on the Internet by reducing, organizing, and retrieving information as needed [2]. Additionally, there are several useful applications for text summarizing, such as compressing articles for online publications, optimizing search engine rankings, and making theses and research papers easier to understand. It is also helpful in creating systems for organizing and screening information sources so that only relevant information is taken out of them. Because of its flexibility, text summarization is a useful technique for increasing productivity in a variety of fields and speeding information access.

The extractive and abstractive approaches constitute the two primary types of automated text summarization [3]. The final sentences generated by summaries that only use content that has been extracted contain words or phrases that were taken from the original text. This method is called extractive summary [4], whereas abstractive summarization uses linguistic approaches to comprehend the text and compressed its essential concepts [3]. Depending on the particular needs of the assignment, these strategies are applied in different applications and accommodate alternative methodologies for producing summaries.

It's clear that the field of text summarization has focused mainly on the English language, but dealing with the Arabic language's complexity presents significant challenges. Arabic has special difficulties because of its complex morphology, diglossia, and diversity of dialects. Compared to English, automatic summarization in Arabic is a more difficult to apply because of these linguistic features. Moreover, the vast majority of text summarization systems currently in use, including those for Arabic, depend on extractive summary strategies. In particular in the Arabic context, abstractive summarization is less frequent. However, the reality that Arabic is the official language of 22 countries and is spoken by over 300 million people shows how important it is to address these issues in Arabic text summarization [2, 5]. Effective Arabic summarization systems are becoming more and more necessary in order to support the efficient processing and retrieval of information in the Arab-speaking world. The significance of developing text summarizing methods specific to the Arabic language's varied requirements and distinctive linguistic characteristics is being acknowledged by researchers and developers [6].

Text translation [7], sentiment analysis [8], text summarization, and other critical tasks have recently shown significant improvements caused by deep learning techniques [9]. Moreover, using large datasets to enhance performance is a key component of deep neural network applications [10]. The encoder-decoder model's sequence-to-sequence structure serves as the foundation for the new text summarizing techniques. The encoder and decoder are the two components of this paradigm. The encoder changes the hidden states in accordance with each new token it gets from the input sequence at each step. Regardless of the length of the input, the encoder creates a context vector representing the input sequence when it reaches the final token in the sequence. The context vector is the final hidden state to be established before the decoder. The decoder is started with ⟨SOS⟩ token, and context vector from the encoder as a first hidden state is used to start it. The decoder is taught to generate a new sequence with a predetermined length. By providing the previously created word, the device creates a new word from the vocabulary each time [10, 11]. As seen in Figure 1, the start token ⟨SOS⟩ [12] is supplied to the decoder along with the encoder's final hidden state. Numerous NLP applications, including machine translation and text summarization, have made use of this approach. A sentence in specific language is the input sequence for machine translation, while the output sequence is the same statement in a different language. In contrast, the document that has to be summarized is the input sequence in text summarization, and the summary itself is the output sequence [12, 13].

The traditional sequence-to-sequence (seq2seq) model faces a challenge in summarizing lengthy input sequences by compressing them into a single fixed-size "context vector". This method often struggles to capture all essential details, especially in longer sequences.

attention mechanism allows the model to selectively highlight important parts of the input sequence when generating each part of the output sequence as shown in Figure 2, By dynamically assigning significance scores to different segments of the input sequence during decoding, applying encoder hidden states, the attention mechanism enables the model to better understand the relevance of individual elements. This

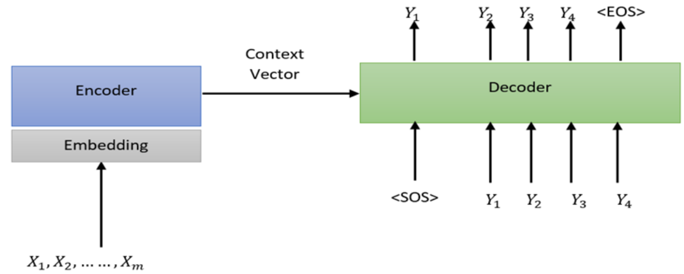

Figure 1: Sequence-to-sequence model.

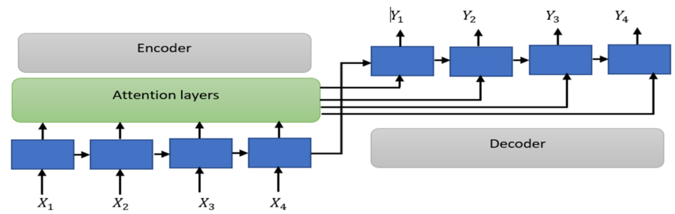

Figure 2: Sequence-to-sequence with attention.

adaptability helps in focusing on important information within longer input sequences, significantly improving the model's accuracy and performance in tasks like machine translation and summarization [13].

Attention processes come in two varieties, such as local and global attention. The context vector's derivation method determines how they vary from one another. The attended context vector in global attention is derived from all of the encoder's hidden states, whereas in local attention it is derived from a limited number of encoder hidden states [13].

The Transformer model architecture was introduced in recent years, which has allowed for performance benefits over RNN-based designs [14]. Additionally, a fresh approach known as transfer learning has surfaced and grown rapidly to take the lead in deep learning model training and application. Using a self-supervised training goal, the model is pre-trained on a vast quantity of data in the first phase of this technique. Using a supervised data set, the model is then fine-tuned on a downstream job in the second phase [15].

We used the XL-Sum dataset to fine-tune the pre-trained models mBERT, AraBERT, and AraT5 in this study. Additionally, mBERT and AraBERT have been fine-tuned using the BERT2BERT architecture. ROUGE measures were employed together with manual human evaluation to evaluate the performance of the suggested models. Their results on a test set outside of their field of study were also compared.

The rest of the paper is organized as follows. Section 2 discusses the related studies to our work. The methodology of the proposed work is presented in section 3. Section 4 presents the setup of our experiments. Results are discussed in section 5. Section 6 introduces the conclusion of the paper.

2 Related works

This section presents recent studies on Arabic abstractive text summarization. We will concentrate on the studies that used the Transformer architecture in addition to the ones that used the RNN-based sequence-to-sequence model.

It has been shown by several recent studies that transfer learning produces state-of-the-art outcomes on nearly all NLP tasks [16]. This shows that the skills acquired through neutral unsupervised learning may be effectively applied to challenges that come after. The NLP community has recently seen an increase in the use of big pre-trained models that have used this methodology as a result of its success [17, 18].

The Transformer-based pre-trained models and RNN-based architecture have not been widely used in Arabic-language works [19].[20] have improved mBERT [21], mBART-50 [22], and AraBERT [23] for cross-lingual Arabic abstractive text summarization, and found that AraBERT produces the lowest result of any of their other suggested models. AraT5 has also been improved by [24] for Arabic abstractive summarization multi-sentence.

In [19] they provided a comprehensive comparative analysis between RNN-based and Transformer-based architectures, specifically, mBERT, AraBERT, AraGPT2, and AraT5, which are well-known for their ability to understand and produce Arabic text for tasks requiring abstractive summarization. Their paper involved a sizable Arabic summarization dataset, contains 84,764 high-quality text-summary pairs, serving as both training and evaluation data. To combat potential under-fitting, an additional dataset of 280,000 examples was incorporated, leading to the improvement and enhancement of models, denoted as ”Seq2Seq-LSTM+” and ”Transformer+”. Notably, these ”+” variants, trained on the expanded dataset, shown improved performance. Achieving F-scores of 33.04 for Seq2Seq-LSTM, 32.12 for Transformer, 37.57 for Seq2Seq-LSTM+, and notably higher scores of 39.61 for Transformer+, 42.96 for mBERT2-mBERT, 40.48 for AraGPT2, 44.02 for BERT2BERT, and the highest of 46.87 for AraT5 using ROUGE-L.

A hybrid approach was developed for Arabic summarization by combining a transformer-based model with a Modified Sequence-To-Sequence (MSTS) framework [25]. (MSTS) model involves the incorporation of three encoder layers, specifically input layer, sentence layer, and named entity recognition layers, aimed at improving the summarization process. They used global attention mechanism and AraVec for embedding and building a new dictionary to cover the word that not included in AraVec. This innovative strategy involves enhancing the MSTS model. By selectively choosing and rearranging text fragments, the model generates extractive summaries. Subsequently, the transformer-based mechanism refines these extractive summaries, transforming them into abstractive summaries. The HASD (Arabic Summarization Dataset) was introduced as a novel benchmark dataset and the existing extractive EASC benchmark was modified by incorporating abstractive summaries into each text. To evaluate the quality of abstractive summaries, they proposed a new evaluation metric termed the Arabic-ROUGE measure. This metric evaluates vocabulary and structural similarity of abstractive summaries, emphasizing their coherence and linguistic essence.

In contrast, however, the study presented by [26] proposed abstractive summarization system used a sequence-to-sequence (seq2seq) model enhanced with different recurrent neural network (RNN) architectures, which are Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and Bidirectional LSTM (BiLSTM). Both the encoder and decoder components integrated global attention mechanisms, allowing the model to focus on relevant parts of the input during encoding and decoding. To enhance the understanding of Arabic words and achieve improved performance, the AraBERT preprocessing stage was incorporated into the model pipeline. Furthermore, a comparative study was conducted between two Word2Vec models, skip-gram and continuous bag of words (CBOW). The study showed that employing a Bidirectional LSTM (BiLSTM) architecture, consisting of three hidden layers, and integrating AraBERT preprocessing led to superior performance results. This finding suggests the advantage of the BiLSTM architecture in conjunction with AraBERT preprocessing for enhancing the abstractive summarization of Arabic text. They used the Arabic Headline Summary (AHS) and the Arabic Mogalad_Ndeef (AMN) datasets.

An extractive summarization system was introduced in [27] .The initial phase involved organizing an Arabic text into a graph format, where sentences act as nodes connected by edges representing similarity. Using cosine similarity, sentences exceeding a set threshold were linked, creating a highly interconnected graph. Employing the PageRank algorithm on this weighted graph assigned salience scores to each sentence, determining their significance within the network. Iteratively computed based on edge weights and damping factor, these scores identified sentences that are relevant and strongly connected. Subsequently, sentences were ranked according to their salience scores, organizing them in order of importance. This process ensured that sentences with stronger relationships and relevance stand out. Triangles within the graph were identified using De-Morgan laws, aiding in constructing a reduced graph that captured the essential elements of the text. They tested their model using EASC dataset.

An advanced text summarization model was proposed in [28] based on a sequence-to-sequence RNN architecture, specifically using LSTM units to reduce the vanishing gradient problem. Diverging from a single-layer encoder, it employed a three-layered multilayer encoder. One layer captures input text, another grasps text keywords, and the third identifies text name entities, all facilitated by word embeddings. The hidden states of the three encoder layers consist of bidirectional LSTM units. The decoder, a singular unidirectional LSTM layer, receives training input through attention to previous summary words and decoder hidden states. In testing, it depends on the previous decoder output and hidden states, initialized with "<SOS>" and the context vector. Employing global attention mechanisms enhanced prediction accuracy by adding important source text insights into the context vector. A dataset comprises 79,965 documents was used. This dataset sourced from news sites like Aljazeera, National Interest, and Financial Times along with 69,024 documents from SANAD_SUBSET, categorized into medical, finance, sports, religion, culture, politics, and technology.

In [29] they introduced finetuning the AraBART model, based on BART Base architecture, comprises 6 encoder and 6 decoders layers with 768 hidden dimensions, totaling 139M parameters. It incorporates an extra layer-normalization for stable training at FP16 precision and uses sentence piece for a 50K token vocabulary and 99.99% character coverage of the training data. This model was evaluated on datasets like Arabic Gigaword subsets and XL-Sum, it covered various abstractiveness levels in news articles and includes tasks for summary and title generation. The comparison was against baselines, AraBART is compared to C2C (a BERT2BERT-based seq2seq model), mBART25 (pretrained on 25 languages, fine-tuned for Arabic), and mT5base. They found AraBART consistently surpasses C2C and mBART25 across various datasets, Its superiority, based model. In this work, we use various Transformer-based models abstractiveness. Additionally, AraBART outperforms mT5 on the multilingual setup for XL-Sum.

In [30] an automatic and extractive method was proposed for single-document summarization in the Arabic language. The proposed method aims to create informative summaries by evaluating each sentence's importance based on a combination of statistical and semantic features like (Key-Phrases, Sentence location, Similarity with title, Sentence centrality, Sentence length, Cue words, Positive key-words, Sentence inclusion of numerical data, Occurrence of Non-essential Information). In the score-based method, important sentences were extracted based on the total scores that are assigned to them. In the machine learning approach, the extractive summarization process was modeled as a binary classification problem Then, a binary (Yes/No) classifier was trained based on a set of training documents.

In [31] they introduced SemG-TS, an innovative Arabic abstractive summarization technique based on semantic graph embeddings and a deep neural network. SemG-TS transforms text into a semantic graph, capitalizing on Arabic language nuances, followed by SemanticGraph2Vec graph embedding. The deep learning model based on a sequence-to-sequence architecture used in summarization includes LSTM in the Encoder and LSTM Basic Decoder. Using AlJazeera.net data comprising 16,770 paragraphs averaging 204 words each, SemG-TS exceeds two word2vec versions (trained and random-based) across ROUGE metrics. It achieved a 15.8% precision improvement, 29.5% in recall, and 21.4% in F-measure over the best word2vec (random-based). In human evaluation, SemG-TS shows superior relevancy, similarity, readability, and overall satisfaction compared to word2vec. The F-score for ROUGE stands at 0.047.

Previous Arabic text summarization studies focused on RNN-based model. In this work, we use various Transformer-based models. mBERT,AraBERT and AraT5 pre-trained language models were used. A high-quality dataset was used to compare the performance achieved by these models.

3 Methodology

First, as seen in figure 3, we explain the dataset preparation process that was employed in this part. Next, we go into the model architecture and training details of the several models that have been trained for the Arabic abstractive text summarizing work.

3.1 Dataset

We choose the XL-Sum dataset [32], which is appropriate for the abstractive summarization of a single document. One million texts with clear summary are included. With the use of well-designed algorithms, this dataset was generated from news articles on the BBC website. The 44 languages in the XL-Sum dataset have available ranging from low to high, with many not having public access.

3.2 Data Preprocessing

Arabic has greater difficulties than some other languages, like English, because it is a morphologically rich language. The inconsistent use of diacritical marks (Tashkil) and the omission of Hamza in Arabic texts provide difficulties for the processing of Arabic text. Modern Standard Arabic (MSA) is the Arabic language used most commonly in academic work, news, and literature, it often omits tashkil. Because the Arabic text missing the diacritical representations necessary to change a word's meaning, this omission increases ambiguity. Additionally, various dialects are spoken in Various Middle Eastern areas, each with significant differences. This diversity in dialects further complicates language processing and understanding in Arabic text analysis and summarization tasks. Each of these difficulties must be taken into consideration while processing Arabic text. The AraBERTv1-base Model [23] was used for text preprocessing. It is a powerful language model that facilitates Arabic text processing and analysis across a wide range of NLP tasks.

3.3 Data Tokenization

Tokenization is a key component of natural language processing research because it covers the gap between unprocessed textual data and the numerical input needed to build machine learning models. Tokenizer classes were strategically used to enable effective model training and evaluation, as well as efficient data preprocessing and easy integration with the corresponding models. Tokenization process was

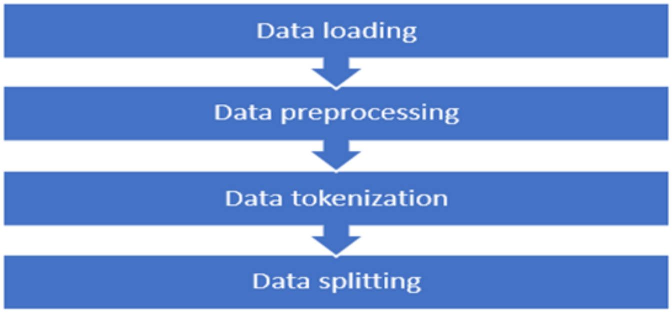

Figure 3: The steps for preparing the data.

used for improving the efficiency and adaptability of our trained models. The Hugging Face Transformers library's AutoTokenizer class was used for improving the performance of AraT5 model. Furthermore, BertTokenizer was used for optimizing the BERT2BERT and mBERT2mBERT models. In our experiments, the AutoTokenizer and BertTokenizer were used for truncating the input sequence to 512 token and padding the short sequence to it. For text summary, text was truncated to100 token and padding the short summary to it.

3.4 Data Splitting

The XL-Sum dataset includes abstractive summaries of 46897 Arabic articles that were created by humans. We split the dataset to 37519 (80% of the records) articles for training, 4689 (10%) articles for validation and 4689 (10%) articles for testing.

3.5 Building an outside domain set

We used an additional sampled dataset from the Arabic Mogalad_Ndeef Dataset (AMN) focused on single-sentence abstractive summarization [33]. Random samples were chosen about 1000 records. However, we did not train the model on this subset of the data.

4 Experiments

The architecture, experimental details, and model training process are described in detail below. Figure 4 presents the architecture of the models used in this research.

4.1 BERT2BERT

AraBERT, an encoder-only Transformer, is the Arabic version of BERT. It accepts an input of length n, called $X_{1:n}$, and generates a contextual representation based on that input, with same length $X^{-}{}_{1:n}$. AraBERT isn't appropriate for text summarization because it requires input and output lengths to match, which presents a limitation when summarizing text, as the length of the original text and the summary may differ. To apply the BERT2BERT encoder-decoder setup, we initialized

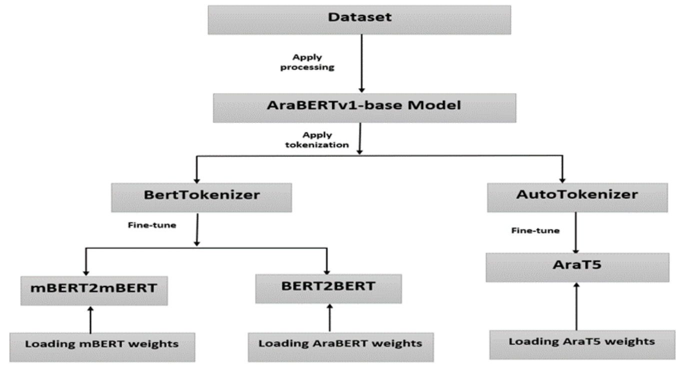

Figure 4: A diagram of the trained models.

both the encoder and decoder with AraBERT weights, enabling the utilization of AraBERT for summarization purposes [19]. To build our model, we included AraBERT parameters into the appropriate BERT2BERT layers. The decoder part was significantly modified, but the encoder component matches AraBERT without modifying its settings. In every block of the decoder, we added cross-attention layers, which had weights that were originally randomly distributed across the feed-forward and self-attention layers. Additionally, we converted bidirectional self-attention layers into unidirectional ones so that the decoder only analyzes tokens that have previously been created at each step. After the last decoder block, we added an LM head to allow for the creation of summary tokens. The final layer was initialized in the same way as the embedding layer. We suggested to reduce the total number of trainable parameters by sharing the encoder's weights with the decoder because of their close similarity. During initialization, in decoder blocks only the cross-attention layers were randomly initialized. We trained the model for 5 epochs with a batch size of 2. Before backpropagation, we collected gradients during fine-tuning for 8 steps. We used "bert-base-arabertv02" as the version of AraBERT.

4.2 mBERT2mBERT

In our approach, we used the BERT2BERT encoder-decoder architecture with mBERT initialization, a model pre-trained on a diverse corpus covering 104 languages, including Arabic sourced from Wikipedia text. During training, we implemented a strategy of accumulating gradients over 8 steps before proceeding with back-propagation, a technique aimed at saving the training process and improving convergence. Fine-tuning of the model was conducted over 5 epochs, with a relatively small batch size of 2, chosen to balance computational efficiency and model performance. To accommodate the characteristics of text summarization tasks, we restricted the length of both input sequences to 512 tokens and summaries to 100 tokens during fine-tuning. This limitation helps in managing computational resources effectively while ensuring that the model can capture essential information for summarization within the specified constraints. Overall, these training and fine-tuning strategies are designed to optimize the BERT2BERT architecture for Arabic text summarization, aiming to achieve robust performance across a range of summarization tasks and dataset.

4.3 AraT5

A modified version of the popular T5 model is the AraT5 [34], serves as an encoder-decoder framework that combines various natural language processing tasks within a single text-to-text paradigm. Because of its flexibility, AraT5 can easily handle a wide range of tasks, including text summarization, machine translation, and categorization. Notably, the model utilizes task-specific prefixes appended to input sequences to determine the kind of task, such as" translate English to Arabic" for translation or" summarize:" for summarization. During training, approximately 15% of the tokens in this model's training set are masked in Masked Language Modeling (MLM), with consecutive tokens being masked using a single sentinel token [19]. For our experimentation, we leveraged the $\text{AraT5}_{\text{Base}}$ version 8, which underwent training on a diverse dataset comprising both Modern Standard Arabic (MSA) and Twitter data, utilizing the T5_{Base} architecture. The MSA dataset, totaling 70 GB, was sourced from various Arabic repositories, while the Twitter data encompassed 1.5 billion tweets containing at least three Arabic words, randomly sampled for inclusion. The architecture of the encoder and decoder is identical to that of BERTBase, consisting of 12 layers with 12 attention heads each.

During fine-tuning, the model underwent training for 5 epochs with a batch size of 2. The input sequence length was covered at 512 tokens, with summaries restricted to 100 tokens.

In this work, several huge trained models were used. Firstly, we optimized the multilingual mBERT model, which is commonly used as a baseline in the literature. Next, improvements were implemented to the Arabic pre-trained models AraBERT and AraT5. We made use of the BERT2BERT encoder-decoder architecture to use AraBERT and mBERT for summarizing texts, where the corresponding model weights were used to warm-start the encoder and decoder . For fine-tuning the pre-trained models, we used Adam optimizer (Adaptive Moment Estimation) [35], which is an optimization algorithm commonly used in training deep learning models, including those used in natural language processing (NLP) tasks. It belongs to the family of stochastic gradient descent (SGD) optimization algorithms and is known for its efficiency and effectiveness in a wide range of applications, with a learning rate of 2e-5.

5 Experimental results and Discussion

We fine-tune three transformer models AraT5, AraBERT, and mBERT and apply improvements to better adapt them for the task of Arabic summarization. We then evaluate their performance and compare the results. Text summarization models are usually evaluated automatically using ROUGE metrics in addition to be assessed manually by human experts. While ROUGE metrics quantify overlap between generated and reference summaries, they may not fully capture qualitative aspects like coherence and readability. Manual evaluation supplements this by considering factors such as relevance, coherence, and grammaticality, providing a more detailed understanding of summary quality. By combining automated and manual evaluations, researchers gain a comprehensive view of model performance, enabling model selection and strategies for improvement. This double evaluation strategy provides an in-depth review, allowing for well-informed choices to be made in the development and research of text summarizing.

5.1 Automatic evaluation

We used the ROUGE-1, ROUGE-2, and ROUGE-L measures for automated evaluation. These metrics measure the degree to which produced summaries and reference summaries overlap in terms of unigrams, bigrams, and longest common subsequences. The model's performance in comparison to the reference summary was then measured by computing the accuracy, recall, and F-measure values for each ROUGE metric. The following formulas were used to get values for each ROUGE metric:

$$\mathbf{precision} = \frac{\left|\mathbf{grams_{refrence}} \cap \mathbf{grams_{generated}}\right|}{\mathbf{grams_{generated}}} \tag{1}$$

$$\mathbf{recall} = \frac{\left|\mathbf{grams_{refrence}} \cap \mathbf{grams_{generated}}\right|}{\mathbf{grams_{generated}}} \tag{2}$$

$$\mathbf{F-measure} = \mathbf{2.0} * \frac{\mathbf{recall} * \mathbf{precision}}{\mathbf{recall} + \mathbf{precision}} \tag{3}$$

The evaluation results, presented in terms of ROUGE F1 scores, were obtained using the rouge Python library. During summary generation, we used Beam Search algorithm which is a heuristic search algorithm commonly used in sequence generation tasks, efficiently explores the search space by maintaining a set of candidate sequences, selecting the most promising candidates at each step based on a predefined scoring criterion, and pruning less likely paths to focus on high-quality outputs [36]. The beam search algorithm was employed with a beam size of 3.

In our experimental setup, we initially trained and evaluated models using a subset of the data, consisting of 10000 records for training and 1200 records for evaluation and testing, and then we used the full dataset. Figure 5 provides ROUGE F1 scores evaluation on test set. Arat5 performed better than other models on the test set. Figure 6 shows the ROUGE F1 scores evaluation on validation set. BERT2BERT which was initialized with AraBERT weights had high performance on the validation set. Additionally, within our model comparison, AraT5 outperformed mBERT2mBERT on the validation set, showing

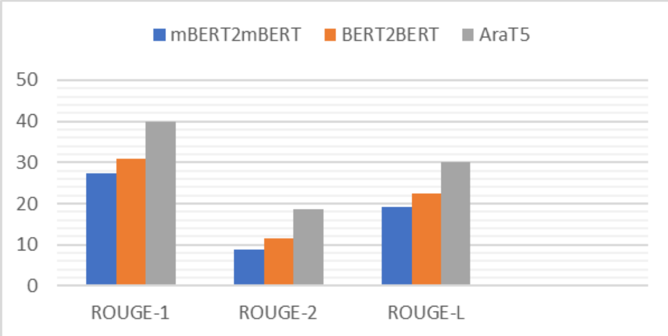

Figure 5: Test set scores based on the ROUGE F1 evaluation.

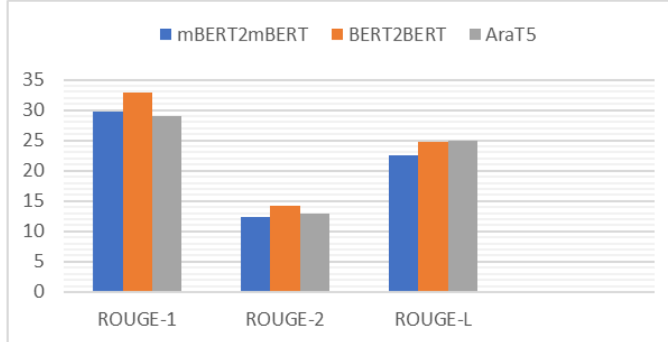

Figure 6: Validation set scores based on the ROUGE F1 evaluation.

even more the differential efficiency of different architectures in text summarizing tasks. Additionally, we evaluated our models using the 1000-record outside domain set. as shown in figure 7, and found that AraT5 achieved the highest score compared to other models. As a result, we used AraT5 model to compare our work with other comparing studies.

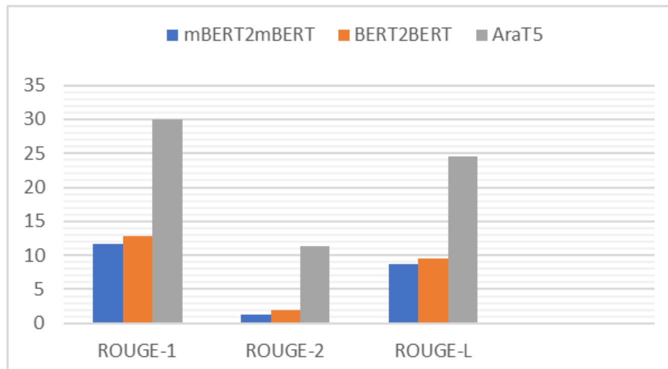

Figure 7: ROUGE F1 scores of the out-of-domain set.

5.2 Manual evaluation

As was previously mentioned, a thorough assessment of the quality and readability of produced summaries could not be possible if one only relies on ROUGE indicators [19]. Therefore, we handled human evaluations. Five fluent Arabic speakers were tasked with rating each summary on a scale of one to five based on two criteria: (1) readability, which measures grammatical accuracy and sentence structure; and (2) quality, which assesses how well the summary conveys the main ideas of the original text. We chose 20 cases at random from the test set. The human evaluation criteria are listed in Table 1. Next, we determined the reported scores' mean. The results of the manual human examination are shown in Figure 8.

The two models with the highest scores on the two measures were AraT5 and BERT2BERT, with AraT5 almost outperforming BERT2BERT. On the other hand, mBERT2mBERT scored significantly lower for both quality and readability measures. With the use of our models, we developed summaries for the two articles, which are shown in Figure 9 and 10.

5.3 Comparison with previous studies

Table 2 presents a comparison of ROUGE F1 evaluations on the XL-Sum dataset between our AraT5 model and four types of state-of-the-art baseline results [29]. The initial baseline, named C2C, is a monolingual sequence-to-sequence model [37], which is based on BERT2BERT. While the cross-attention weights are initialized at random, the encoder and decoder are initialized using CAMELBERT weights [38]. A total of 246M parameters represent C2C. The multilingual BART model mBART25 [39], pretrained on 25 languages, including Arabic, is the second baseline. mBART25 has shown successful in monolingual generative tasks like abstractive summarization, although it was initially pre-trained for neural machine translation [40]. mBART25 has 610M parameters overall. Third model is called $mT5_{base}$ model. Finally, AraBART, the fourth model, performs better than the others. As noticed our model outperforms the four compared state-of-the-art models.

5.4 Discussion and findings

ROUGE metrics indicate that fine-tuning the AraT5 model resulted in a performance increase of approximately 15.56%. In contrast, the performance of AraBERT and multilingual BERT decreased by 10.52% and 20.67%, respectively. When summarizing data from outside the domain, AraT5 performs better than AraBERT Model. Interestingly, the BERT2BERT-based model initialized using multilingual BERT shows poor performance when evaluated manually. In comparison to the models that have been recommended, AraT5 and AraBERT model consistently generate highest score summaries in terms of readability and quality, as assessed by human evaluation. Furthermore, it's observed that increasing the training data enhances the model's accuracy.

5.5 Limitations

While our research focused mostly on producing one-sentence summaries from news sources, there is still a need

Table 1: Readability and quality measures for the manual human evaluation.

Score	Quality	Readability
1	The output is irrelevant.	Inaccurate / Hard to read.
2	Key concepts are partially conveyed.	A little understood.
3	Key concepts are moderately conveyed.	Understandable in poor Arabic.
4	Key concepts are largely conveyed.	Understandable in acceptable Arabic.
5	Key concepts are completely conveyed.	Understandable in fluent Arabic.

Table 2: A comparison of different state-of-the-art models' ROUGE F1 results.

Model	ROUGE-1	ROUGE-2	ROUGE-L
C2C	26.9	8.7	23.1
mBART25	32.1	12.5	27.6
mT5base	32.8	12.7	28.7
AraBART	34.5	14.6	30.5
AraT5 (Ours)	39.78	18.77	30.21

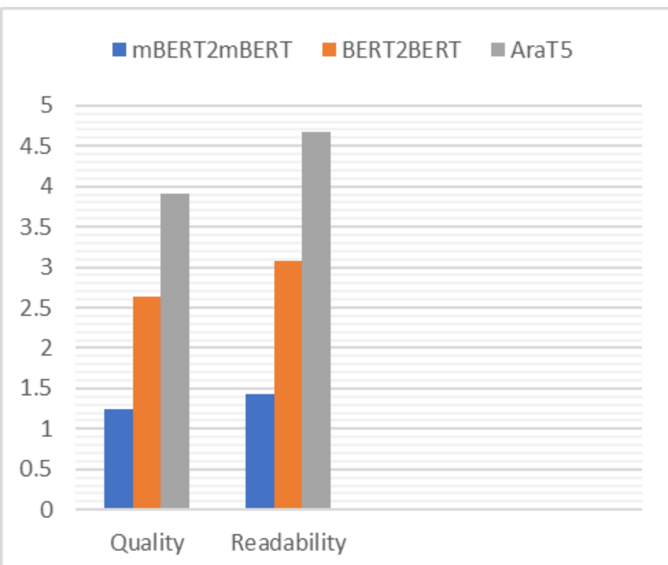

Figure 8: The manual human evaluation scores.

for further study into producing multi-sentence summaries and extending the application to other types of text sources. Further research may examine the complexity involved in summarizing information across several phrases, as well as the Arabic dialects by various text. Our models have been modified and specially designed for text summary of news. As such, we expect that without extra training data customized for particular summarization task, their performance might not be directly comparable to other models. We recognize that on sometimes, our models provide inaccurate, invalid, and grammatical outputs that could lead to general users being confused.

6 Conclusion and Future Works

For the purpose of this work, we used several pre-trained language models, such as AraT5, AraBERT, and mBERT, to summarize Arabic abstractive text. Additionally, to use encoder-only Transformer models, we used an encoder-decoder architecture based on BERT2BERT. Both manual assessment and ROUGE metrics were used to evaluate these models, and an out-of-domain dataset was used for testing. Our results show that pretrained language models perform well in Arabic text summarization tasks. According to the automated evaluation, AraT5 outperforms other models in our test set, but AraBERT outperforms other models in the validation set during training. In addition, human evaluation shows that AraT5 achieves high accuracy in terms of readability and quality. The summary generated by AraT5 is highly comparable to the reference summary. Furthermore, AraT5 outperforms other models using out-of-domain datasets. Results confirmed that the modified AraT5 performing better than other models. For future research directions, we recommend focusing on multi-sentence summarization, emphasizing grammatical correctness, understanding dialects, and incorporating semantic meaning into automatic evaluation processes. In order to improve the model's performance across multiple domains, we propose training it on multi-domain datasets. By using this method, the model's efficiency and adaptability in summarization tasks covering many topics and domains can be enhanced. While we recommend exploring fine-tuning strategies, hybrid models, and additional linguistic resources to further optimize summarization performance.

7 Future Work

In future research, we propose focusing on multi-sentence summarization with an emphasis on grammatical correctness, dialect comprehension, and the integration of semantic meaning into automatic evaluation metrics. Expanding the training data to include multi-domain datasets can further enhance the model's adaptability and performance across various topics and domains. To optimize summarization quality, future studies should explore advanced fine-tuning strategies, hybrid modeling approaches, and the incorporation of additional linguistic resources. Investigating the impact of diverse embedding techniques, reinforcement learning-based training, and transformer-based architectures could also contribute to

Source text	وأكدت وسائل الإعلام السورية أن الجيش "استعاد الأمن والاستقرار عبر حي الخالدية بالكامل". ولم ترد تأكيدات لهذا النبأ من جانب المعارضة، لكن المرصد السوري لحقوق الإنسان كان قد قال الاثنين إن الاشتباكات مستمرة في حي الخالدية. لكنه قال إن قوات الحكومة استعادت أغلب أرجاء الحي وأحكمت حصارها على المناطق المحدودة التي ما زالت متبقية تحت سيطرة المعارضة في وسط المدينة. رمز للمعارضة دبابات سورية شوهدت داخل الحي. ويعد حي الخالدية أحد الأحياء الرمزية للمقاتلين المعارضين للنظام السوري. وتعني السيطرة عليه عزل الأحياء التي يسيطر عليها المعارضون والمحاصرة منذ أكثر من عام، ويمهد هذا الطريق لسيطرة الجيش السوري على مدينة حمص برمتها. مواضيع قد تهم كنهاية وقال المرصد السوري الاثنين إن الطيران الحربي السوري نفذ غارتين على حي باب هود الواقع جنوب الخالدية. وتأتي المكاسب التي حققتها قوات الجيش بعد شهر من بدئها هجوما في حمص في إطار حملة لتكوين محور يربط بين دمشق ومناطق ساحلية على البحر المتوسط. وكان الجيش السوري قد استعاد قبل نحو شهرين السيطرة على منطقة القصير الإستراتيجية في ريف حمص، التي بقيت تحت سيطرة المعارضين لأكثر من عام. وقد لقي 100 ألف شخص حتفهم في الصراع الدائر في سوريا منذ عامين، والذي بدأ في صورة احتجاجات سلمية على حكم الرئيس السوري بشار الأسد في مارس / آذار 2011. وفر نحو مليوني سوري من الحرب. Syrian media confirmed that the army "has restored security and stability throughout the Al-Khalidiya neighborhood." There have been no confirmations of this news from the opposition, but the Syrian Observatory for Human Rights stated on Monday that clashes are ongoing in the Khalidiya neighborhood. However, he stated that government forces have regained most areas of the neighborhood and tightened their siege on the limited areas that are still under the control of the opposition in the city center. A symbol of the opposition: Syrian tanks were seen inside the neighborhood. Al-Khalidiya neighborhood is considered one of the symbolic areas for fighters opposing the Syrian regime. This means controlling the isolation of neighborhoods held by the opposition and besieged for over a year, paving the way for the Syrian army to take full control of the city of Homs. Topics that may be of interest: On Monday, the Syrian Observatory reported that the Syrian warplanes carried out two airstrikes on the Bab Hood neighborhood located south of Khalidiya. The gains achieved by the army forces come a month after they launched an attack in Homs as part of a campaign to establish a corridor linking Damascus to coastal areas along the Mediterranean Sea. The Syrian army had regained control of the strategic Qusayr area in the Homs countryside about two months ago, which had remained under the control of the opposition for more than a year. Since March 2011, the ongoing conflict in Syria, which began as peaceful protests against President Bashar al-Assad's rule, has resulted in the deaths .of 100,000 people over the past two years. About two million Syrians have fled the war.
AraT5	استعادت قوات الجيش السوري السيطرة على حي الخالدية في مدينة حمص. Syrian army forces have regained control of the Khalidiya neighborhood in the city of Hom.
BERT2BERT	"قالت الحكومة السورية إن قوات الجيش السوري على مدينة حمص، وذلك بعد يومين من سيطرة المعارضة في سوريا. ا عن مقتل 11 شخصا على الأقل بحسب ما ذكره ناشطون رسميون. مليار دولار. أنها تعرضت لوزارة الداخلية الأمريكية (داعش) الذين قتلوا خلال الساعات الأخيرة التي يسيطر عليها مسلحو تنظيم الدولة الإسلامي. "The Syrian government stated that the Syrian army has taken control of the city of Homs, just two days after the opposition gained control in Syria." About the killing of at least 11 people, according to what official activists reported. One billion dollars. She was exposed to the U.S. Department of Homeland Security (ISIS) who were killed during the last hours controlled by ISIS militants.
mBERT2mBERT	ز قالت السلطات السورية إنها قتلوا في مدينة حلب، بحسب ما ذكرته وكالة الجيش السوري للمعارضة. ا من بينهم وذلك بعد يومين شخصا على الأقل. ون إلى أن عدد من الرئيس السوري بشار الأسد التي كانت عليها " الذين يمكن أن يكون نحو 19 عاما. تاي عن مقتل 30 شخصا مع القوات الحكومية خلال البلاد. The Syrian authorities said they killed in the city of Aleppo, according to what was reported by the Syrian army opposition agency. Among them, at least one person after two days. It seems that the number of Syrian President Bashar al-Assad's years in power could be around 19 years. Reports indicate the killing of 30 individuals alongside government forces across the country.

Figure 9: A sample text that our models have summarized.

Source text	نضال حسن واعترف نضال حسن، الذي يدافع عن نفسه، بقتل الجنود، متحججا بحماية المسلمين وعناصر طالبان في أفغانستان، ولكن القاضي العسكري رفض حجته "بحماية الآخرين". وإذا أدين حسن، البالغ من العمر 42 عاما، بقتل 13 شخصا وجرح آخرين فإنه سيواجه عقوبة الإعدام. ويعتبر الحادث الأكثر دموية من بين الهجمات غير القتالية التي وقعت في قاعدة عسكرية أمريكية. وقال شهود عيان دخل في 5 نوفمبر / تشرين الثاني عام 2009 مصحة تعج بالجنود الذين كانوا ينتظرون أدوارهم إجراء فحوصات طبية أو التلقيح، ثم صعد على مكتب، وأطلق النار من سلاحين بيديه، دون توقف إلا لإعادة تعبئة السلاح. مواضيع قد تهم كنهاية وسيقدم ممثل والادعاء أدلة تفيد بأن حسن مال إلى الأفكار المتطرفة، وكان يزور المواقع بحثا عن الجهاديين "وطالبان، ساعات قبل الهجوم. وكان الرائد حسن سيلتحق بالقوات الأمريكية في أفغانستان قبل أن ينفذ هجومه. " عنف في مكان العمل "وصنفت وزارة الدفاع الأمريكية الحادث باعتباره" عنفا في مكان العمل "بدلا من تصنيفه" عملا إرهابيا "، وهو ما أغضب عائلات الضحايا، حسب ما أفاد به مراسل بي بي سي، نك براينت، في فروت هود. ويتوقع أن يدلي العديد من جرحى الحادث بشاهاداتهم أمام المحكمة. وسيواجه حسن عددا من ضحاياه في قاعة المحكمة لأنه سيتولى الدفاع عن نفسه. وهو يستخدم كرسيا متحركا لأنه أصيب بالشلل، عندما أطلق عليه شرطي في القاعدة العسكرية النار. Nidal Hassan, who is defending himself, admitted to killing the soldiers, claiming it was to protect Muslims and Taliban members in Afghanistan, but the military judge rejected his argument of "protecting others." If Hassan, 42 years old, is convicted of killing 13 people and injuring others, he will face the death penalty. It is considered the deadliest incident among the non-combat attacks that occurred at a U.S. military base. Witnesses stated that on November 5, 2009, he entered a clinic bustling with soldiers who were waiting for their turns for medical examinations or vaccinations, then climbed onto a desk and opened fire with two guns in his hands, only stopping to reload. Topics that may be of interest as a conclusion will be presented by the representative and the prosecution, providing evidence that Hassan leaned towards extremist ideas and was visiting sites in search of jihadists "and the Taliban, hours before the attack. Captain Hassan was set to join the American forces in Afghanistan before carrying out his attack. "Workplace violence" The U.S. Department of Defense classified the incident as "workplace violence" instead of labeling it as "terrorism," which angered the victims' families, according to BBC correspondent Nick Bryant in Fort Hood. Many of the accident's victims are expected to testify in court. Hassan will face several of his victims in the courtroom because he will be defending himself. He uses a wheelchair because he was paralyzed when a police officer shot him at the military base.
AraT5	قالت المحكمة الأمريكية إن مسلحا قتل 13 جنديا أمريكيا وجرح آخرين في هجوم على قاعدة عسكرية في أفغانستان. The American court stated that an armed assailant killed 13 American soldiers and injured others in an attack on a military base in Afghanistan.
BERT2BERT	"قالت وزارة الدفاع الأمريكية إنها سيواجه في هجوم شنه مسلحو حركة طالبان، وذلك بعد إدانته بأنه لم يكن إلى مقتل خمسة أشخاص على الأقل. من بينهم شخصا للجنود الباكستانيين الذين قتلوا في أفغانستان. عام 2011، حسب ما ذكره وكالة أنباء الدفاع عن شهود عيان. مليار دولار. بي بي سي إن قوات الشرطة لا يمكن أن يكون هناك. The U.S. Department of Defense stated that it will confront an attack launched by Taliban militants, following its condemnation of the incident that resulted in the deaths of at least five people. Among them was a person among the Pakistani soldiers who were killed in Afghanistan. In 2011, according to what was reported by the eyewitness news agency. One billion dollars. The BBC reports that the police forces cannot be present.
mBERT2mBERT	قالت الشرطة الأمريكية إنها قتلوا في هجوم بأنه لم يكن على الأقل، وذلك بعد يومين من بينهم. ا للرئيس الأمريكي دونالد ترامب إلى مقتل 30 شخصا. "ات مع حركة طالبان. تا عليه التي يسيطر عليها العسكرية الإسلامية. ون الذين يمكن أن تكون ما ذكره مسؤولون / كانون الماضي. ان كان قد أعلنه لا يقل عن. The American police stated that they killed in an attack that was not at least, just two days after among them. The American President Donald Trump commented on the killing of 30 people. "With the Taliban movement." Until it is controlled by the Islamic military. And those who could be what officials mentioned / last January. If he announced it, it is no less than.

Figure 10: A sample text that our models have summarized.

more accurate and contextually aware summarization models. By addressing these directions, future research can improve the efficiency, coherence, and applicability of summarization models across a broader range of real-world use cases.

References

[1] C. Slamet, A. Atmadja, D. Maylawati, R. Lestari, W. Darmalaksana, and M. A. Ramdhani, "Automated text summarization for indonesian article using vector space model," in *IOP Conference Series: Materials Science and Engineering*, vol. 288, p. 012037, IOP Publishing, 2018.

[2] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. D. Trippe, J. B. Gutierrez, and K. Kochut, "Text summarization techniques: a brief survey," *arXiv preprint arXiv:1707.02268*, 2017.

[3] L. H. Belguith, M. Ellouze, M. H. Maaloul, M. Jaoua, F. K. Jaoua, and P. Blache, "Automatic summarization," *Natural language processing of semitic languages*, pp. 371–408, 2014.

[4] A. Khan and N. Salim, "A review on abstractive summarization methods," *Journal of Theoretical and Applied Information Technology*, vol. 59, no. 1, pp. 64–72, 2014.

[5] H. N. Fejer and N. Omar, "Automatic arabic text summarization using clustering and keyphrase extraction," in *Proceedings of the 6th International Conference on Information Technology and Multimedia*, pp. 293–298, IEEE, 2014.

[6] A. Elsaid, A. Mohammed, L. F. Ibrahim, and M. M. Sakre, "A comprehensive review of arabic text summarization," *IEEE Access*, vol. 10, pp. 38012–38030, 2022.

[7] S. P. Singh, A. Kumar, H. Darbari, L. Singh, A. Rastogi, and S. Jain, "Machine translation using deep learning: An overview," in *2017 international conference on computer, communications and electronics (comptelix)*, pp. 162–167, IEEE, 2017.

[8] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis: A survey," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, no. 4, p. e1253, 2018.

[9] H. Liu, S.-G. Li, H.-X. Wang, and G.-J. Li, "Adaptive fuzzy synchronization for a class of fractional-order neural networks," *Chinese Physics B*, vol. 26, no. 3, p. 030504, 2017.

[10] M. Al-Maleh and S. Desouki, "Arabic text summarization using deep learning approach," *Journal of Big Data*, vol. 7, no. 1, p. 109, 2020.

[11] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *Advances in neural information processing systems*, vol. 27, 2014.

[12] D. Suleiman and A. Awajan, "Deep learning based abstractive text summarization: approaches, datasets, evaluation measures, and challenges," *Mathematical problems in engineering*, vol. 2020, pp. 1–29, 2020.

[13] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.

[14] V. Ashish, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, p. I, 2017.

[15] S. Ruder, *Neural transfer learning for natural language processing*. PhD thesis, NUI Galway, 2019.

[16] X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, and X. Huang, "Pre-trained models for natural language processing: A survey," *Science China Technological Sciences*, vol. 63, no. 10, pp. 1872–1897, 2020.

[17] W. Antoun, F. Baly, and H. Hajj, "Aragpt2: Pre-trained transformer for arabic language generation," *arXiv preprint arXiv:2012.15520*, 2020.

[18] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.

[19] M. Bani-Almarjeh and M.-B. Kurdy, "Arabic abstractive text summarization using rnn-based and transformer-based architectures," *Information Processing & Management*, vol. 60, no. 2, p. 103227, 2023.

[20] M. Kahla, Z. G. Yang, and A. Novák, "Cross-lingual fine-tuning for abstractive arabic text summarization," in *Proceedings of the international conference on recent advances in natural language processing (ranlp 2021)*, pp. 655–663, 2021.

[21] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[22] Y. Tang, C. Tran, X. Li, P.-J. Chen, N. Goyal, V. Chaudhary, J. Gu, and A. Fan, "Multilingual translation with extensible multilingual pretraining and finetuning," *arXiv preprint arXiv:2008.00401*, 2020.

[23] W. Antoun, F. Baly, and H. Hajj, "Arabert: Transformer-based model for arabic language understanding," *arXiv preprint arXiv:2003.00104*, 2020.

[24] E. M. B. Nagoudi, A. Elmadany, and M. Abdul-Mageed, "Arat5: Text-to-text transformers for arabic language generation," *arXiv preprint arXiv:2109.12068*, 2021.

[25] A. ELSAID, A. MOHAMMED, L. Fattouh, and M. SAKRE, "Hybrid arabic text summarization approach based on seq-to-seq and transformer," 2023.

[26] Y. M. Wazery, M. E. Saleh, A. Alharbi, A. A. Ali, *et al.*, "Abstractive arabic text summarization based on deep learning," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[27] Y. A. AL-Khassawneh and E. S. Hanandeh, "Extractive arabic text summarization-graph-based approach," *Electronics*, vol. 12, no. 2, p. 437, 2023.

[28] D. Suleiman and A. Awajan, "Multilayer encoder and single-layer decoder for abstractive arabic text summarization," *Knowledge-Based Systems*, vol. 237, p. 107791, 2022.

[29] M. K. Eddine, N. Tomeh, N. Habash, J. L. Roux, and M. Vazirgiannis, "Arabart: a pretrained arabic sequence-to-sequence model for abstractive summarization," *arXiv preprint arXiv:2203.10945*, 2022.

[30] A. Qaroush, I. A. Farha, W. Ghanem, M. Washaha, and E. Maali, "An efficient single document arabic text summarization using a combination of statistical and semantic features," *Journal of King Saud University-Computer and Information Sciences*, vol. 33, no. 6, pp. 677–692, 2021.

[31] W. Etaiwi and A. Awajan, "Semg-ts: Abstractive arabic text summarization using semantic graph embedding," *Mathematics*, vol. 10, no. 18, p. 3225, 2022.

[32] T. Hasan, A. Bhattacharjee, M. S. Islam, K. Samin, Y.-F. Li, Y.-B. Kang, M. S. Rahman, and R. Shahriyar, "Xl-sum: Large-scale multilingual abstractive summarization for 44 languages," *arXiv preprint arXiv:2106.13822*, 2021.

[33] A. M. Zaki, M. I. Khalil, and H. M. Abbas, "Deep architectures for abstractive text summarization in multiple languages," in *2019 14th International Conference on Computer Engineering and Systems (ICCES)*, pp. 22–27, IEEE, 2019.

[34] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant, A. Barua, and C. Raffel, "mt5: A massively multilingual pre-trained text-to-text transformer," *arXiv preprint arXiv:2010.11934*, 2020.

[35] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[36] A. K. Vijayakumar, M. Cogswell, R. R. Selvaraju, Q. Sun, S. Lee, D. Crandall, and D. Batra, "Diverse beam search: Decoding diverse solutions from neural sequence models," *arXiv preprint arXiv:1610.02424*, 2016.

[37] S. Rothe, S. Narayan, and A. Severyn, "Leveraging pre-trained checkpoints for sequence generation tasks," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 264–280, 2020.

[38] G. Inoue, B. Alhafni, N. Baimukan, H. Bouamor, and N. Habash, "The interplay of variant, size, and task type in arabic pre-trained language models," *arXiv preprint arXiv:2103.06678*, 2021.

[39] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, and L. Zettlemoyer, "Multilingual denoising pre-training for neural machine translation," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 726–742, 2020.

[40] M. K. Eddine, A. J.-P. Tixier, and M. Vazirgiannis, "Barthez: a skilled pretrained french sequence-to-sequence model," *arXiv preprint arXiv:2010.12321*, 2020.

Authors

Karim Morsi graduated from the Faculty of Computer and Information Science (FCIS) at Ain Shams University. Currently, I am a Master's student specializing in Natural Language Processing (NLP) at FCIS. Additionally, I work as a Teaching Assistant at FCIS. My primary area of interest is NLP, where I focus on advancing research and applications in this field.

Fatma Najib is an Assistant Professor in the Information Systems Department at the Faculty of Computer and Information Sciences, Ain Shams University. She specializes in data streams, data mining, and real-time processing, with a focus on continuous query optimization and cloud computing. Her research aims to develop innovative solutions for efficient data management and processing in dynamic and scalable environments. As an expert in these fields.

Wedad Hussein is an Assistant Professor at the Faculty of Computer and Information Sciences, Ain Shams University. Her research interests include data mining, social network analysis, and web mining. She explores innovative techniques to analyze large-scale data from social media platforms and the web, aiming to uncover valuable insights from complex datasets. With expertise in these areas, she contributes significantly to advancing research in data mining and its applications to social networks and the web.

Rasha Ismail is a Professor at the Faculty of Computer and Information Sciences, Ain Shams University. Her research covers a broad range of topics, including data science, artificial intelligence, big data analytics, and information retrieval. She is committed to advancing the application of these technologies to address complex, real-world challenges. With her extensive expertise, she plays a crucial role in shaping the direction of research and innovation within these fields at the university.