Summarizing User Comments on Social Media Using Transformers

Afrodite Papagiannopoulou and Chrissanthi Angeli
School of Engineering, University of West Attica, Athens, Greece

apapagiannop@uniwa.gr
angeli@uniwa.gr
c_angeli@otenet.gr

Abstract: Social media and smart technology have invaded our daily lives. They are increasingly used to express feelings and opinions, to publish news, to support public debates on various issues and events. User comments under each post are a key factor in making economic, political and business decisions. Managing their sheer volume is an almost impossible task. Therefore, summarization seems crucial. Recent years have shown that abstractive summarization has achieved great results in the field of document summarization by producing more human-like summaries. Unlike formal documents, social media conversations face four challenges: 1) tend to be informal, consisting of slang expressions and special characters, 2) show deviations from the original theme and dependencies on previous opinions, 3) since they are short, they lack lexical richness and, 4) contain redundant and repetitive information, resulting in confusion among readers. We address these challenges by developing a system that generates abstractive summaries from pools of user comments under a specific social media post, using Transformers. Unlike previous works that do not rely on user comment pools and draw data from Reddit, Twitter or "Sina Weibo" platforms only, we use a Facebook dataset. We first reshape the raw dataset in a meaningful way for summary generation and we apply some basic pre-processing. Then, we define a task that deals with grouping comments according to the post title. A summary is generated for each group (pool) of comments. Our model is evaluated using ROUGE scores between the generated summary and each comment on the thread.

Keywords: Social media, Social comments, Abstractive summarization, Neural networks, Transformers

1. Introduction

Social media through smart technology is widespread and has become an integral part of our daily lives as it is used to communicate, express feelings, advertise, post news, political, economic and social exchanges. Social media's role in public life is very important as posts and comments shape public opinion on a variety of important issues. It has been observed that people's opinions expressed through social networks are more direct and representative than those expressed in face-to-face communication. The data exchanged in social media is a cornerstone of research because we can extract patterns of social behaviour that can be used for social, business and government decisions. Creating summaries on social media comments seems crucial for the retrieval of useful knowledge in a reasonable time period.

Summarization means converting the content of a long text into a smaller one, preserving the meaning of the original text. There are different ways to write a summary. It can be done either manually or using algorithms and artificial intelligence techniques. The latter is called "Automatic text summarization" and is an area that has attracted the interest of researchers especially in recent years (Gupta, S. and Gupta, S.K., 2019). There are two main categories of summarization techniques: a) extractive and b) abstractive (Nenkova, A. and McKeown, K., 2012) Extractive summarization finds the most important words or sentences from the original text by considering statistical and linguistic features. Then rearranges them in the correct order and produces the summary text. Abstractive summarization, on the other hand, is based on the meaning of the document and is created by either formulating new sentences or rewording existing ones with new words (Varma, V., Kurisinkel, L.J. and Radhakrishnan, P. 2017). Target text is produced by analyzing the semantic information of the original text. With deep analysis and reasoning, new sentences are created from the original text (Rachabathuni, P.K., 2017).

Initially, most summarization work was done with extractive summarization because of its simplicity. In recent years, however, this field has stagnated as it was observed that extractive summaries lacked fluency, coherence and the meaning of the original text. Producing high quality summaries that were grammatically correct, consistent, concise and informative were deemed necessary. For this reason, researchers turned their attention on abstractive summarization which, despite its complexity, still satisfies the above limitations, creating human-like summaries (Suleiman, D., and Awajan, A., 2020).

Artificial intelligence and Deep Learning techniques have been successfully applied to summarization with excellent results. By introducing the transformer model in (Vaswani et al, 2017), the traditional sequence-to-sequence model has drastically improved in terms of both accuracy and training time. Sequential processing is being replaced by parallel processing, the simple embeddings from the positional embeddings and self-

attention to multi-head attention. The attention mechanism is not a term that appeared with the emergence of transformers. But when combined together their power took off. Thus, we selected the Transformer model as our basic model for training and fine-tuning. In the field of social media summarization, there is limited research which builds abstractive text summarization systems based on transformers. Unlike these works that mainly generate summaries from Reddit, Twitter or "Sina Weibo" platforms, our approach aims to go beyond and to create a post-level summary of pools of comments and sub-comments. Our goal is to include all comments and not the most liked or those characterized by linguistic richness. Additionally, since each social media platform is unique with its own particularity on posting and commenting (Ghanem, F.A., Padma, M.C. and Alkhatib, R., 2023), we aim to expand our research to platforms other than Reddit and Twitter. The dataset we are looking at in this paper is Facebook news posts accompanied by user comments under each post.

This paper illustrates the design of our system, based on T5 pre-trained model. The rest of the paper is organized as follows: Section 2 presents a review of the research done so far on social media summarization using transformers. Section 3 illustrates our model description. Section 4 shows the experimental methodology, Section 5 discusses the results achieved so far and finally section 6 contains concluding remarks and our future work.

2. Related Work

As the internet began to conquer our daily lives in the area of news, information and opinion exchange, scientists focused on creating summaries of web posts, microblogs, and social networks. They first used probabilistic and optimization methods to generate a summary. Indicative works have been done on Twitter posts; the first is introduced by (Sharifi, B., Hutton, M-A. and Kalita, J., 2010), (Sharifi, B., Inouye, D., and Kalita, J.K., 2014) where the model automatically generates a summary by applying a Phrase Reinforcement (PR) algorithm adding the TF-IDF technique to it. The second (Chong, F., Chua, T., Asur, S., 2021) is based on two models the Decomposition Topic Model (DTM) and the Gaussian Topic Model (GDTM) exploit the temporal correlation between tweets under predicted conditions to produce a concise and information-rich summary of events.

With the development of Artificial Intelligence, NLP also started gaining attention from researchers. The main reason for turning to these techniques has been a) their abilities to support the large volume of data circulating on the internet, and b) the powerful computing power available today. Various deep learning models including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and Long-Term Memory (LSTM) RNNs have been used in the field of natural language generation. Indicative examples are research works such as of (Gao et al, 2019) which introduces a sequence-to-sequence model based on CNN and Bi-RNN creating abstractive summaries of social media. Work by (Liang, Z., Du, J. and Li, C., 2020) has created abstractive summaries of social media based on Attentional Encoder-Decoder RNNs. An additional hidden layer is used before the encoder where the model decides which information is useful and which should be removed. Another work introduced by (Wang, Q. and Ren, J., 2021) proposes a model of summary-aware attention mechanism that produces abstractive summaries. This mechanism has managed to overcome the particularity of social media content. In (Bhandarkar, P., Thomas, K. T., 2023) abstractive summaries are produced using an RNN which combines sequence-to-sequence and the attention approach. The model also uses LSTM as an encoder and decoder to build summaries. With the attention layer in this architecture the decoder is allowed to have more direct access to the input sequence.

The use of transformers and attention mechanism (Vaswani et al, 2017) has been a pivotal point in deep learning applications. The idea has been adopted by many researchers and has been extended to all NLP tasks (Gupta, A., Chugh, D. and Katarya, R., 2022). The social media summarization field is still in its infancy. We recognize, so far, the following aspects that refer to abstractive summaries on social media using transformer models: A framework using a pre-trained BERT-based model as encoder and a non-pre-trained transformer model as a decoder, had been introduced by (Li, Q. and Zhang, Q., 2020). This project produces abstractive Event Summarization on Twitter. An event topic prediction component helps the decoder to focus on more specific aspects of posts. In order to produce more coherent summaries and exceed the limit of 512 tokens taking into account only the most liked comments. In the work of (Tampe, I., Mendoza, M. and Milios, E., 2021) a multi-document text summarization scenario that includes meaningful comment detection, extending previous single-document approaches is proposed. The proposed model uses pre-trained BERT as encoder and a transformer as a decoder. This architecture is extended with an attention encoding level that is fed with user preferences. Attention encoding focuses on comments with the highest social impact. Comments are rated as

important based on user preferences. Finally (Blekanov, I. S., Tarasov N. and Bodrunova, S. S., 2022) introduces a Transformer-based abstractive summarization model which creates summaries in three languages from posts and comment pools. The datasets used have been downloaded from Reddit and Twitter. In this work T5 and LonFormer are fine-tuned and compared with BART. In addition, they apply enlarged Transformer-based models on Twitter data in three languages (English, German, and French) to discover the performance of these models on data with non-English text. In (Rawat et al, 2021) an abstractive summarization model was presented which used a transformer model to generate individual sentence summaries of review texts. Then a combination of Universal Sentence Encoder, statistical methods and graph reduction algorithm was used to select the most relevant sentences to best represent the whole text in the summary.

3. Methodology

Our system's design, based on Transformers, aims to generate abstractive summaries from pools of user comments under a specific social media post. Unlike prior works that generate summary for each comment we create a post-level summary of pools of comments and sub-comments. First, we focus on creating coherent and meaningful summaries of pools of user comments, since they present greater specificity, in terms of a) the nature of the language, b) their conceptual grouping and c) their dynamic upgrading. Second, unlike the other research works that consider only two social media platforms, Reddit and Twitter, our goal is to explore how different data sets and training models can be developed. For summarizing pools of user comments under social media posts, we used transformer model architecture as explained in (Vaswani et al, 2017). Transformers give state-of the-art results on various NLP tasks, such as sentiment analysis, language translation and text summarization. The main components of our system are:

3.1 Transformer Encoder-Decoder Architecture

Encoder-decoder architectures are widely used for developing neural networks. It can handle inputs and outputs both of which consist of variable-length sequences admirably. Therefore, these architectures, although they are applied in many fields, are best suited for problems related to natural language understanding and text generation, such as translation and summarization. In other words, they perform immensely in sequence-to-sequence modelling. The encoder takes as input a sequence of variable length and transforms it into a numerical representation that captures the important information from the input. This numerical representation constitutes the hidden state and is of fixed length. Subsequently, the decoder maps the fixed-shape coded state into a variable-length sequence to generate the output (Figure 1).

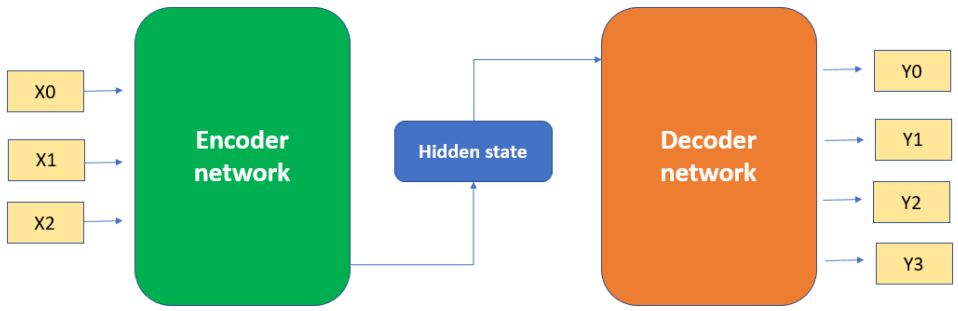

Figure 1: Encoder-Decoder architecture (Vitalflux.com/encoder-decoder-architecture-neural-network)

3.2 Model Explanation

Starting our approach, we compared three pre-trained transformer models (T5, BART and PEGASUS) for their performance in generating summaries of our dataset (Rawat A., and Singh Samant, S., 2022). What these pre-trained models have in common is that they generate abstract summaries and use the encoder-decoder transformer architecture that performs best in text generation tasks. The results showed very good behaviour

from all three models with the T5 giving the best. Without excluding the performance study of the other models, we started with T5 by feeding and training it with data from Facebook, specifically with news posts and their respective comments.

As we have already mentioned earlier, user posts and comments on social media are of particular interest when it comes to creating summaries. First of all, they are short, lack verbal and linguistic richness and tend to be informal, consisting of slang expressions and special characters. Secondly, they present deviations from the original theme and dependencies on previous views. Finally, they contain redundant and repetitive information, resulting in confusion to readers. For this reason, their pre-processing is deemed necessary in order to improve the performance of the model. Based on the topic of the post we create lists of comments ignoring parts that do not add meaning to the sentence such as short words, links, abbreviations and emoji. The proposed approach consists of the following steps:

- Data collection using Facebook.
- Preprocessing using regex libraries.
- Post-level classification of user comments using Python.
- Feed the transformer-based encoder with the lists of texts using Pytorch library.
- Feed this encoding to the Transformer decoder using Pytorch library.
- Produce the summary text
- Validate the predicted summary using ROUGE metrics.

For the creation of our system we decided to use the pre-trained models rather than building a model from scratch because pre-trained models: a) can give much better results, with careful pre-processing of the data, b) provide a learning base on which they can tune many different data sets and c) can create new models with a small change in the training session and fine-tuning, leading us to faster results (Blekanov, I. S., Tarasov N. and Bodrunova, S. S., 2022). Particularly our system creation is based on the T5 pre-trained model. T5 - stands for "Text-to-Text Transfer Transformer" (Raffel et al, 2021) and is an encoder-decoder model. It converts all problems into text-to-text format and can be trained or fine-tuned either for supervised or unsupervised data.

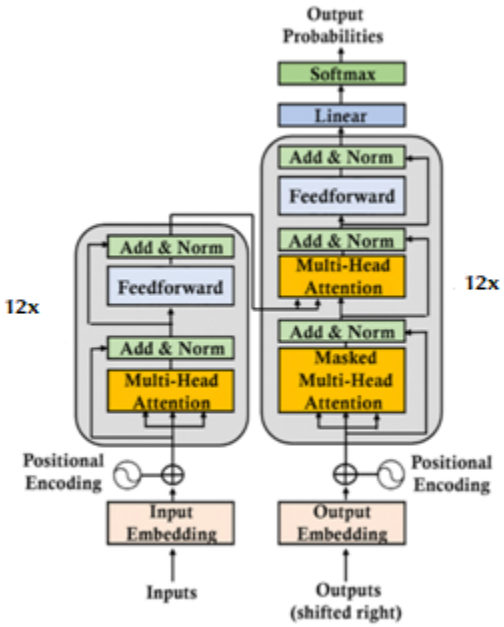

Figure 2: Transformer architecture

It is characterized as a task-agnostic model since it is suitable for any NL task, such as Translation, Language Inference, Information extraction and Summarization. Its training procedure is based on teacher forcing. It has an input sequence and a target sequence.

For the purpose of our work, it is trained and fine-tuned for text summarization from social media. The input sequence of our system is fed to the encoder which in turn the target sequence is fed to the decoder. More specifically: The input sequence is tokenized and processed to convert every word into a unique numeric id

using *body_input_ids* with *body_attention_masks.* Then in the embedding layer the transformer uses learned embeddings to transform the input tokens into vectors of 512.

Encoder part: The encoder is comprised of two major components: a multi-head attention mechanism which is followed by normalization and a feed-forward neural network. The Multi-head Attention is based on a scale dot-product attention which creates a vector of every input word. Once it is applied 8 times it generates several vectors for the same word. This helps the model to capture different representations of the words' relations in the sentence. The different attention-based matrices generated by the multiple heads are summed and passed through a linear layer to reduce the size to a single matrix.

Decoder part: The decoder side has a lot of shared components with the encoder side. It takes in two inputs: the output of the encoder (*summary_input_ids*) and the output text shifted to the right (*generated_ids*). Then it applies multi-head attention twice with one of them being "masked" (*summary_attention_masks*). The dictionary at the final layer of the decoder should have the same size as the target dictionary, our case 128. Finally, the softmax function is applied indicating the probability of each word to be present in the output.

4. Experimental Methodology

4.1 Dataset

Finding the right datasets to generate social media summaries is a really complex task. Several social media platforms place restrictions on data download. For those that are open, the data must be reformatted to meet the needs of the project. From our research we have seen that different social media platforms meet different performances. Therefore, we have used a Facebook dataset for the development of our system. The data pertains to posts of Facebook's news, each of which is accompanied by user comments and sub-comments. The raw data consists of the following 7 columns with a total of 1,781,576 rows. Essentially, our interest is limited to 3 columns: 'message', 'post_title', 'post_number' where 'post_title' and 'post_number' identify a specific post while the 'message' column represents the comments of users under the post. After the dataset was downloaded, the first step was to apply a basic pre-processing to it before entering it into the model. As we mentioned above, raw social network data contain special characters, emoticons, and referral links. Pre-processing involves cleaning the data from unnecessary and useless elements that would have led us to undesirable results. Using the python NLTK and regex libraries, we removed punctuation, special characters, emoji strings and links, as well as NULL values. The second step of pre-processing concerns the grouping of data according to the topic of discussion. Our goal is to generate a summary from a set of user comments related to a specific post-topic. Based on the post title, the data dictionary is reformatted to isolate user comments for each post into separate lists of different sizes. The new data set is the input of our model.

Experimental Setup

The Transformer model was trained on PyCharm with NVDIA GeForce 4070 GPU with 12 BG RAM. For simplicity, we used the 1/3 of our dataset (1,781,576 rows), since it is converted into lists of group-comments. We considered all sentences of each list with a summary length of 128. The dataset is split into Train, Validation and Test with 80%, 20% and 10% ratio size respectively. We trained our model based on the traditional T5-base 12-layer pre-trained model with 12 attention heads and depth of feed forward network as 3072. The dropout rate set to 0.1 and the AdamW optimizer we used is set to 1e-3. The model was trained for 12 epochs with batch size set to 64. We calculate the Train, Validation and Validation PPL loss to assess the performance of the model. Finally, we have used the Rouge score metrics to evaluate the model.

5. Results

Trying to make the right decision on which pre-trained model is more suitable for our dataset, we have started comparing three pre-trained models: BART (Lewis et al, 2019), PEGASUS (Zhang et al, 2020) and T5 (Raffel et al, 2021). Since there is a variety of pre-trained models that have achieved great results on many NLP tasks, HuggingFace hub (Wolf et al, 2020), settling on these 3 because they meet the constraints below: a) they are transformer-based, b) they have encoder-decoder architecture and c) they produce abstractive summaries. Applying these models to our data set before clustering, we obtained that T5 had the best performance. Thus, we have started our study by applying this model to our own dataset. Our main goal is to produce summaries that are accurate and coherent. In addition, T5 is a modern model that has already been successfully used in summary studies, but in the field of social media it is less explored. This is an initial estimate for our own model design, but by enriching our study with additional assumptions to optimize the results, the pre-trained models will be re-evaluated for their performance.

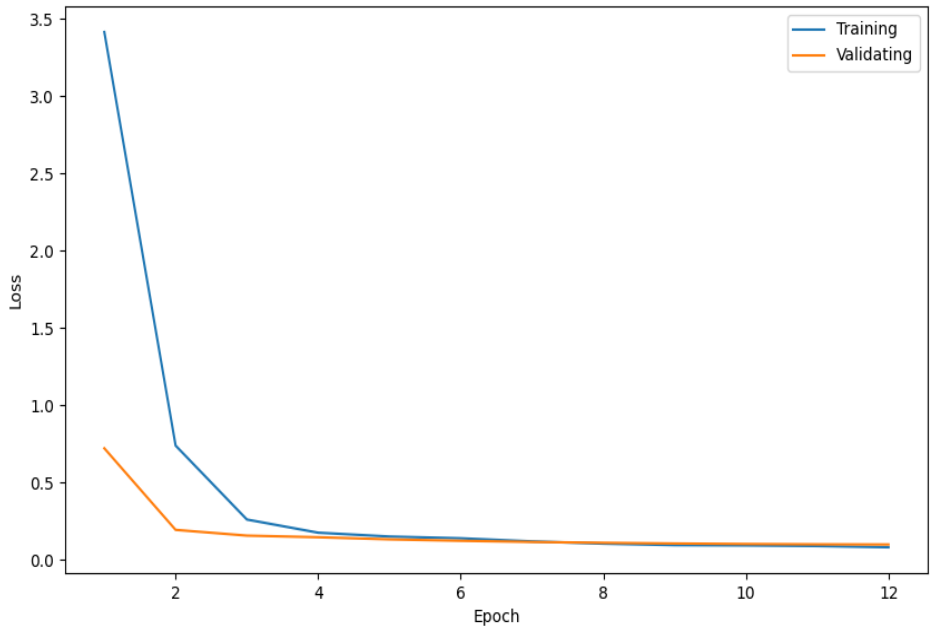

Figure 3: Train and Validation Loss

In order to assess the performance of our model, we have calculated loss. Loss quantifies the error produced by the model and it is one of the most important factors in deep learning and neural networks. After splitting the dataset into Train, Validation and Test subsets we have measured the Training and Validation loss for all the samples. As it is obvious in Figure 3 both metrics are decreasing gradually. Furthermore, the gap between validation and train loss shrinks after each epoch. This is because as the network learns the data, it also shrinks the regularization loss (model weights), leading to a minor difference between validation and train loss. Consequently, with most of our data samples it gains a good fit.

In order to evaluate the text summary produced by our system we used Rouge. The Recall-Oriented Understudy for Gisting Evaluation, ROUGE (Lin, C.-Y., 2004) is a metric for evaluating an automatically produced summary against reference summaries. It is a widely used set metrics and software package for evaluating automatic summarization. At this stage of our work Rouge compares the system-generated summary against each comment in the thread. The results are shown in Table 1. To get a better quantitative value, we compute precision, recall and F1 score. As is obvious from the results, recall is quite high which means that the reference summary has been captured by the system-generated summary. Precision gives a proportion of 0.514-0.520 of words suggested by the predicted summary that actually appears in the reference summary. On the other hand, recall gives a proportion of 0.849-0.854 of words in the reference summary captured by the predicted summary. Both of them lead to a harmonic mean f1-score of 0.640-0.645. Knowing that Rouge scores must range from 0-1 with 1 being the best value and therefore the most ideal summary, we conclude that these results look quite good for our text summarization system. It is important to notice a high Rouge score means that the system captures the most important information to put in the summary. This, on the other hand, doesn't lead us to a high-quality summary because it may contain biased text. Therefore, the evaluation of the quality of the produced text is a very complex process where the dimensions and limitations of each system must be taken into account.

Table 1: Rouge Score Results

	rouge1	rouge2	rougeL
P	0.520	0.514	0.519
R	0.854	0.849	0.854
F	0.646	0.640	0.645

6. Conclusion and Further Work

As can be seen from the measurements so far, we come to two main conclusions. First, the learning rate of the model has a good fit according to our dataset. Both the training and validation sets start with high values due to lack of learning but as they are trained the errors decrease and we have a smoothing of the curves, until in the end where they even out. Thus, the model is properly trained and can give correct results. Also, as seen from Rouge scores, the summaries generated from user comments on each post are quite satisfactory. As with all research works, however, there are areas for improvement. Earlier in this paper we pointed out that, at this stage, our model is evaluated using ROUGE scores between the generated summary and each comment in the thread. Therefore, further research would be done to compare the system-generated summary with the human-generated summary. ROUGE only works on overlays. If we only want to stick to calculated scores then a score of 1 would indicate the perfect summary. But this would only happen if both summaries had the same n-grams. Furthermore, rouge performs better on models that produce extractive summarization. Since our model focuses on paraphrasing thus producing abstractive summarization, our research will be extended to more modern tools and techniques that can achieve better results. For the purposes of this project, we have used data from facebook that correspond user comments under specific posts. A certain amount of data is manageable by the model, enabling the system to learn, train and give certain results. However, a robust mechanism can be developed to allow the desired flexibility in the model to further learn and perform on large-scale data. Most of the research work in the field of social networks has been done using data from the platforms Reddit and Twitter. Wanting to go beyond the limits and check the data of other platforms, we used Facebook data streams. However, knowing that each platform is unique and presents particularities both in the way it posts information and in the way users react, more research could be done to expand the behaviour and effectiveness of summary models for every platform.

Acknowledgements

The publication fees were totally covered by the University of West Attica.

References

Bhandarkar, P., Thomas, K. T. (2023) “Text Summarization Using Combination of Sequence-To-Sequence Model with Attention Approach”, *Springer Science and Business Media Deutschland* GmbH: 283–293, 2023, doi:10.1007/978-981-19-3035-5_22.

Blekanov, I. S., Tarasov N. and Bodrunova, S. S. (2022)“Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages”. *Future Internet 2022*, 14(3), 69. https://doi.org/10.3390/fi14030069

Chong, F., Chua, T., Asur, S. (2021) “Automatic Summarization of Events from Social Media”. *Proceedings of the International AAAI Conference on Web and Social Media*, 7(1), 81-90. https://doi.org/10.1609/icwsm.v7i1.14394.

Gao, S., Chen, X., Li, P., Ren, Z., Bing, L., Zhao, D. and Yan, R. (2019) “Abstractive Text Summarization by Incorporating Reader Comments”. *In The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 33(01):6399-6406

Ghanem, F.A., Padma, M.C. and Alkhatib, R. (2023) “Automatic Short Text Summarization Techniques in Social Media Platforms”. *Future Internet 2023*, 15(9), 311. https://doi.org/10.3390/fi15090311

Gupta, S. and Gupta, S.K. (2019) “Abstractive summarization: An overview of the state of the art”. *Expert Systems with Applications 121,* 49–65.

Gupta, A., Chugh, D. and Katarya, R. (2022) “Automated News Summarization Using Transformers”, *In Sustainable Advanced Computing,* 2022, Volume 840. ISBN: 978-981-16-9011-2.

Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang L., et al., (2021) “Pre-trained models: Past, present and future,” *AI Open*, vol. 2, pp. 225–250.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. and Zettlemoyer, L. (2019) “BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension”. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, online. Association for Computational Linguistics. https://aclanthology.org/2020.acl-main.703

Li, Q. and Zhang, Q. (2020) “Abstractive Event Summarization on Twitter”. *In The Web Conference 2020 - Companion of the World Wide Web Conference*, WWW 2020, Association for Computing Machinery: 22–23, doi:10.1145/3366424.3382678.

Liang, Z., Du, J. and Li, C. (2020) “Abstractive social media text summarization using selective reinforced Seq2Seq attention model,” *Neurocomputing*, **410**, 432–440, doi:10.1016/j.neucom.2020.04.137.

Lin, C.-Y. (2004) “ROUGE: A Package for Automatic Evaluation of Summaries”. *In Text Summarization Branches Out,* pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Nenkova, A. and McKeown, K. (2012) “A survey of text summarization techniques”, *In: Aggarwal, C., Zhai, C. (Eds) Mining Text Data.* Springer, Boston, MA. https://doi.org/10.1007/978-1-4614-3223-4_3

Rachabathuni, P. K. (2017) “A survey on abstractive summarization techniques”. *In Inventive computing and informatics (ICICI), international conference on* (pp. 762765). doi: 10.1109/ICICI.2017.8365239.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narag, S., Matena, M., Zhou, Y., Li, W. and Liu P. J. (2021) “Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer”. *In The Journal of Machine Learning Research,* Volume 21, Issue 1, 2019. ISSN: 1532-4435.

Rawat A., and Singh Samant, S. (2022) "Comparative Analysis of Transformer based Models for Question Answering". *2nd International Conference on Innovative Sustainable Computational Technologies (CISCT),* Dehradun, India, 2022, pp. 1-6, doi: 10.1109/CISCT55310.2022.10046525.

Rawat, R., Rawat, P., Elahi V. and Elahi, A. (2021) "Abstractive Summarization on Dynamically Changing Text," *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, India, 2021, pp. 1158-1163, doi: 10.1109/ICCMC51019.2021.9418438.

Sharifi, B., Hutton, M-A. and Kalita, J. (2010) “Summarizing Microblogs Automatically”. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 685–688, Los Angeles, California. Association for Computational Linguistics.

Sharifi, B., Inouye, D., and Kalita, J.K. (2014) “Summarization of Twitter Microblogs”. *The Computer Journal*, Volume 57, Issue 3, March 2014, Pages 378–402, https://doi.org/10.1093/comjnl/bxt109

Suleiman, D., and Awajan, A. (2020) “Deep Learning Based Abstractive Text Summarization: Approaches, Datasets, Evaluation Measures, and Challenges,” *Mathematical Problems in Engineering*, doi:10.1155/2020/9365340.

Tampe, I., Mendoza, M. and Milios, E. (2021) “Neural Abstractive Unsupervised Summarization of Online News Discussions”. *In: Arai, K. (Eds) Intelligent Systems and Applications. IntelliSys 2021. Lecture Notes in Networks and Systems,* vol 295. Springer, Cham.

Varma, V., Kurisinkel, L.J. and Radhakrishnan, P. (2017) “Social Media Summarization”, *In A practical Guide to Sentiment Analysis*, Chapter 7 pp.135-153.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. (2017) “Attention Is All You Need”. *In 31st Conference on Neural Information Processing Systems (NIPS 2017),* Long Beach, CA, USA., June 2017.

Wang, Q. and Ren, J. (2021) “Summary-aware attention for social media short text abstractive summarization,” Neurocomputing, **425**, 290–299, doi:10.1016/j.neucom.2020.04.136.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, M., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., et al.. (2020) “Transformers: State-of-the-Art Natural Language Processing”. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, online. Association for Computational Linguistics.

Zhang, J., Zhao, Y., Saleh, M. and Liu, P.J. (2020) “PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization”. *In ICML'20: Proceedings of the 37th International Conference on Machine Learning*, July 2020, Article No.: 1051, Pages 11328–11339.