



Article

Transformer-Based Abstractive Summarization for Reddit and Twitter: Single Posts vs. Comment Pools in Three Languages

Ivan S. Blekanov ^{1,2,*} , Nikita Tarasov ² and Svetlana S. Bodrunova ³ ¹ School of Mathematics and Computer Science, Yan'an University, Yan'an 716000, China² Faculty of Applied Mathematics and Control Processes, St. Petersburg State University, 199034 St. Petersburg, Russia; nkt.tarasov@yandex.ru³ School of Journalism and Mass Communications, St. Petersburg State University, 199034 St. Petersburg, Russia; s.bodrunova@spbu.ru

* Correspondence: i.blekanov@spbu.ru

Abstract: Abstractive summarization is a technique that allows for extracting condensed meanings from long texts, with a variety of potential practical applications. Nonetheless, today's abstractive summarization research is limited to testing the models on various types of data, which brings only marginal improvements and does not lead to massive practical employment of the method. In particular, abstractive summarization is not used for social media research, where it would be very useful for opinion and topic mining due to the complications that social media data create for other methods of textual analysis. Of all social media, Reddit is most frequently used for testing new neural models of text summarization on large-scale datasets in English, without further testing on real-world smaller-size data in various languages or from various other platforms. Moreover, for social media, summarizing pools of texts (one-author posts, comment threads, discussion cascades, etc.) may bring crucial results relevant for social studies, which have not yet been tested. However, the existing methods of abstractive summarization are not fine-tuned for social media data and have next-to-never been applied to data from platforms beyond Reddit, nor for comments or non-English user texts. We address these research gaps by fine-tuning the newest Transformer-based neural network models LongFormer and T5 and testing them against BART, and on real-world data from Reddit, with improvements of up to 2%. Then, we apply the best model (fine-tuned T5) to pools of comments from Reddit and assess the similarity of post and comment summarizations. Further, to overcome the 500-token limitation of T5 for analyzing social media pools that are usually bigger, we apply LongFormer Large and T5 Large to pools of tweets from a large-scale discussion on the *Charlie Hebdo* massacre in three languages and prove that pool summarizations may be used for detecting micro-shifts in agendas of networked discussions. Our results show, however, that additional learning is definitely needed for German and French, as the results for these languages are non-satisfactory, and more fine-tuning is needed even in English for Twitter data. Thus, we show that a 'one-for-all' neural-network summarization model is still impossible to reach, while fine-tuning for platform affordances works well. We also show that fine-tuned T5 works best for small-scale social media data, but LongFormer is helpful for larger-scale pool summarizations.

Keywords: natural language processing; deep learning models; transformer models; abstractive summarization; social networks; opinion mining; Reddit; Twitter; pool summarization



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

Academic Editor: Daniel Gutiérrez Reina

Received: 11 January 2022

Accepted: 18 February 2022

Published: 23 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays, social networks play an increasingly big role in social and political life, with prominent examples of user discussions shaping public opinion on a variety of important issues. Extraction of public views from social media datasets is a goal that unites sociolinguistic tasks such as sentiment analysis, emotion, and topic extraction, finding product-related recommendations, political polarization analysis, or detection of hidden

user communities may be performed via text classification, fuzzy or non-fuzzy clustering, text summarization, text generation, and other techniques of information retrieval. At their intersection, study areas such as sentiment analysis or topic modeling studies have formed.

In particular, topicality extraction is a task that most often employs probabilistic clustering such as topic modeling [1–3], its extensions, or combinations with sentiment analysis [4] or other classes of tasks. However, topic modeling, despite its wide area of application, has many well-known disadvantages. Among them is topic instability [5–7], difficulties with interpretation [8], artifacts of topic extraction [9], and low quality of topicality extraction for very short texts that need to be pooled to receive distinguishable topics [10]. However, an even bigger problem lies in the fact that, for fast enough assessment of topicality, topic modeling may not fit, as it provides only hints to topics via top words but does not summarize the meaning of the texts relevant to a given topic. Nor does it help outline the themes of specific elements of networked discussions, such as comment threads or ‘echo chamber’ modules. Moreover, topic modeling takes the dynamics of discussions for statics of a collected dataset, and even topic evolution studies do not allow for clear extraction of particular meanings of discussion fragments and cannot trace how exactly the discussion themes change [11]. This is just one example of how text summarization could have helped social media studies. Thus, it is regrettable that text summarization techniques are not developed for it.

Their lack is even more regrettable as, today, a range of new-generation deep learning models for text summarization are available that provide for significantly better results for summarization of structurally diverse and noisy texts. Of them, transfer-learning Transformer-based models have produced state-of-the-art results (see Section 2.2). They, though, need to be adapted for tasks relevant for social media datasets and fine-tuned with regard to the specific features of the data such as use of specific language and platform affordances that shape the data properties, e.g., short/varying length, high level of noise, presence of visual data, etc. Here, the advantage is that, for neural-network text summarization models, problems similar to those in topic modeling may be resolved somewhat easier, by pre-training and fine-tuning instead of development of specialized models of model extensions.

Despite this, we still have no clear understanding which Transformer-based model(s) work best with social media data, due to the scarcity of comparative research. Each of these models has notable drawbacks and need pre-testing and fine-tuning for social media data; only after such fine-tuning may these models be compared in quality. This is the first research gap, which we identify and address regarding these new models that may be used for text summarization, such as those based on BART and Transformer architectures, which have not yet been compared after their fine-tuning for particular types of data.

For some time already, text summarization studies have used text collections from Reddit. Among different social media platforms, Reddit is undeniably one of the most prominent platforms. First, it differs from others thanks to higher structural heterogeneity of discussion groups (subreddits), which also implies opinion diversity. Second, it has many times sparked large-scale cross-platform discussions leading to cross-national outrage (e.g., on the Boston Marathon Bombing suspects in 2013) and shaped trends in developed economies (e.g., AMC and Gamestop stock investments changing financial strategies of Wall Street). At one point, it used to be the third most popular website in the USA. Third, and most important for text summarization studies, it has platform affordances that demand users create author text summarizations that may be used as target texts.

Compared to other discussion platforms such as Twitter or Facebook, Reddit has been under-researched for application of automated text analysis in general, and of transfer learning in particular. Examples of the recent Reddit-based opinion explorations include studying COVID-19 publications using statistical techniques (with Reddit being just one of many datasets) [12], community detection using topic information [13], and applying sentiment analysis to Reddit data [14], but such studies do not seem to form a notable field within opinion mining from social media. On the other hand, for abstractive text

summarization studies, Reddit, in effect, is today the only social media platform that has provided multiple datasets of user-generated content—or, more precisely, users' self-authored summaries, thanks to its "Too Long; Didn't Read" (TL;DR) feature that allows discussion participants leave human summarizations for posts. Text summarization studies, thus, employ Reddit datasets to test models—but next-to-never are applied to answer sociological or communication science questions with the help of these models, which leaves the question of their applicability open. This is the second research gap that we address. The most recent works that involve Reddit [15,16] compare deep-learning extractive and abstractive summarization, and abstractive models, especially BART and its extensions, are shown to have the best results in comparison with ground truth. However, newer models such as T5 and LongFormer have not yet been tested on Reddit, even if T5 has already been used in summarization studies [17].

Out of it, one can clearly see that text summarization studies for social media are still at the very early stage of development, and the highly advanced neural-network models that produce good results for other types of textual data are still to be applied to various social media platforms beyond Reddit. We see this as the third research gap. Today, the scarcity of labeled datasets for text summarization studies highly limits the development of text summarization models for other platforms; we will perform a step further here and apply the models to real-world data on Twitter.

Moreover, as we have shown above, social media data have a complicated nature. They are not only noisy, but also structurally diverse. What matters sociologically is not only the meaning of single posts but also that of pools of user posts and/or comments that correspond to elements of networked discussion structure. Summarization of text pools exists, but not for social media data; this is the fourth gap that we address.

Thus, this study aims at several consecutive steps in expanding today's abstractive summarization studies in their application to social media data. First, it is the fine-tuning and comparative assessment of Transformer-based abstractive summarization models for the Reddit data. As per other scholars, we use Reddit as an exemplary platform, but add T5 and LongFormer models to BERTSum and BART already tested on Reddit. We perform this by utilizing a research pipeline that includes data collection and preprocessing; model loading, fine-tuning, and application; and comparative assessment of the results and their visualization. As stated above, the chosen models are first fine-tuned on a standard open-source-labeled Reddit dataset, to receive the large-scale pre-trained models adapted to social media analysis. Second, the chosen fine-tuned model is additionally tested on posts collected from a specific subreddit of 150,000 posts, to define how well the fine-tuned models work against the original versions. Third, we apply the fine-tuned model to pools of the most popular comments and juxtapose the respective summarizations, to see by human assessment whether the model can, in principle, be applied to pooled social media data. However, the fine-tuned models, due to the known limitation of their capacity (they work best with texts or text pools limited to 512 tokens) do not provide for solving real-world analysis of comment threads or discussion outbursts. This is why, fourth, we apply enlarged Transformer-based models to Twitter data in three languages, to discover how enlarged models would work with pooled data and with non-English texts. In particular, we test the selected models on Twitter data of 2015, using the discussion on the *Charlie Hebdo* massacre on three languages (English, German, and French), and show that, despite the fine-tuning and its satisfactory results for Reddit in English, the models need more work for German and French, while it works well for short English texts.

Thus, the novelty of this paper lies in three aspects. First, we fine-tune the models and choose between the LongFormer, Bart, and T5 ones. Second, we test the fine-tuned models on the new Reddit dataset, which can be considered testing on raw data. Third, we test our models on Twitter data on three languages and show that they need additional work for the languages other than English, while short texts provide fewer complications than expected.

The remainder of this paper is organized as follows. Section 2 provides a review on deep learning in text summarization studies, including how text summarizations have been applied to Reddit data until now. Section 3 provides the fine-tuning methodology, including the research pipeline and evaluation metrics, as described above. Section 4 provides the results of comparative assessment of the fine-tuned models. Section 5 describes the new real-world dataset, provides examples the resulting summaries for posts and pools of comments, as well as and their human assessment. In the absence of labeled target collections for comment pools, we use post summaries as targets for summarizations of the respective pools of comments. In Section 6, to overcome the limitations of the original models, we apply the enlarged models to Twitter data on three languages. We conclude by putting our results at the background of today's leading research.

2. Literature Review

2.1. Deep Learning Models in Application to Text Summarizations

Summarization is a task mainly used for dimensionality reduction in large text corpora for better and faster discovery of meaning in the data for specialists in different fields (sociologists, economists, etc.) [18]. While serving a role similar to topic modeling, summarization models are different, mainly in the final result's representation. These models can produce concise and fluent representations of a large text document or collection of documents by either choosing the most important sentences or creating new ones, briefly summarizing the original data. Summarization techniques can, in general, be split into two distinct categories: extractive and abstractive. They differ in their architecture, core principles, and form of the resulting summary.

Extractive summarization can be described as a simple classification task where the main objective is to identify the most important sentences or messages in a group of underlying texts. Deep learning approaches typically focus on the task of language modeling for a downstream classification task [19]. SUMMARUNNER [20] was one of the first algorithms to use this principle with recurrent neural network architecture. BERTSUM [21,22] and HIBERT [23] are examples of more recent algorithms, based on the ideas of transfer learning and a much more refined transformer architecture.

Abstractive summarization, on the other hand, has the potential to train models capable of truly understanding text semantics and thus generating a summary in a form of short phrases—an idea closer to manual summarization. It defines this task as a sequence-to-sequence problem (seq2seq). In these types of models, there are two main parts: an encoder, which transforms initial tokens into their corresponding encodings, and a decoder that generates target summaries using the autoregression method. There are many classic implementations of this idea, including PTGEN [24], which improves word generation in summaries; DRM, which uses reinforcement learning techniques; TCONVS2S [25], which uses supplementary topic distributions to improve summarization quality, and many others. Generally, abstractive models can provide concise representations by generating readable summaries, as opposed to blocks of text, combined by extractive model.

Currently, the rise of models based on the Transformer architecture and usage of transfer learning methods has allowed researchers to achieve high-quality text representations and, thus, obtain state-of-the-art results in different text analysis tasks. The Transformer architecture has allowed for creating general better-quality models, and has brought in one of the first scalable neural text models. While traditional models used recurrent neural networks and their main modification was called Long-Short Term Memory (LSTM) [26], Transformer [27] relied entirely on the self-attention mechanism. This allowed the model to solve the long-term problem of RNNs, namely the inability to model longer sequences without data loss. At the same time, it has enabled parallel training, which was impossible previously, with sequential recurrent models.

Transfer learning [28], on the other hand, allows for model reuse in different subsequent tasks. This idea allows to train a model from a certain starting point, without the need to train all models from scratch. Not only does this process save resources, it also

allows for training models on much smaller datasets, since general linguistic interactions are modeled at the initial training step. Fine-tuning is the next optional step in this process, which allows for, given the same objective, fitting new data into the existing pre-trained model. This ensures better domain modeling, e.g., tuning a model initially developed for working with Wiki data for working with Twitter data instead, which is different in its nature.

2.2. The Newest Text Summarization Models

In this section, we describe some of the more recent Transformer-based models able to perform abstractive text summarization.

2.2.1. BERTsum and BART

As for BERT [29] sentence encoding, this model was one of the first to introduce pre-trained encoders to the field of text summarization. This includes both extractive and abstractive approaches. In this architecture, the authors employ the BERT encoder with a randomly initialized initial Transformer decoder. BART [30] builds on the ideas of BERT in a different way to allow the model to generate text data. It combines BERT's bidirectional encoder with an autoregressive encoder allowing for text generation, while still using core principles of the transformer architecture. This allows for usage of this model in machine translation, summarization, and other text generation tasks without significant architectural changes. Authors have shown that this model outperforms BERTsum on the summarization task, especially with additional fine-tuning.

2.2.2. The T5 Model

The main motivation behind the paper, 'Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer' [31], which introduces the T5 model, lies in the idea of using true sequence-to-sequence modeling. This algorithm solves a variety of text processing problems (question answering, summarization, and translation) by presenting the initial problem as a task of transforming one text to another. It is important to note that text-to-text translation architecture allows to directly apply the same architecture, target, training, and inference steps for different subtasks, which is an important property when dealing with full analysis of a given dataset. At its core, this model follows the originally proposed Transformer model very closely, in particular in its encoder/decoder structure.

In this structure, firstly, the input sequence of tokens is encoded by a set of embeddings, which are then passed to the encoder. The encoder consists of a stack of "blocks", each consisting of two components: a self-attention layer and a small feedforward neural network. Layer normalization [32] is applied to the input of each component. The authors use a simplified version of layer normalization, in which only activations are scaled, without applying additive displacement. After the normalization layer, the Residual skip connection [33] adds the input of each subcomponent to its output. Dropout [34] is used in feedforward networks to skip part of the connections in attention layer weights, and also at the input and output of the entire structure. The decoder is similar in structure to the encoder, with a notable exception of including a standard attention mechanism after each layer of self-attention. The self-attention mechanism in the decoder also uses a form of autoregression, or causal self-attention, which allows the model to pay attention only to past results. The output of the last decoder block is passed to a dense layer with a softmax output, the weights of which are shared with the input set of embeddings. One of the major contributions of the T5 model's authors was the process of collecting and preprocessing data based on the CommonCrawl initial dataset. As a result of multiple steps of processing, including duplicate, stopwords, and special symbols removal, English-language text filtering, etc., this model was able to achieve state-of-the-art results for a variety of different tasks, particularly on the SuperGlue benchmark [35] where it was able to obtain the results closest to those of a human.

2.2.3. The LongFormer Model

Transformer-based models usually cannot handle long sequences due to their self-attention, an algorithm that scales quadratically with the length of the input sequence. This can be an important detriment for one of the variants of comments summarization task, as it will be shown in the experiment section. LongFormer [36] circumvents these limitations by using an attention mechanism that scales linearly with sequence length, making it easier to process texts consisting of thousands of tokens or more. The LongFormer attention mechanism is a direct replacement for standard self-attention and combines local window attention with global attention motivated by individual tasks. Given the importance of local context, this model's attention pattern uses a fixed-size attention window surrounding each token. The use of multiple composite layers of such windowed attention results in a large receptive field where the upper layers have access to all input locations and have the ability to create views that include information throughout the input, similar to CNN. The resulting modification of the Transformer architecture allows to effectively capture the order and meaning of words throughout large documents.

2.2.4. Conclusions on the Transformer Models

The described models provide state-of-the-art results for a variety of text analysis tasks, including abstractive summarization. Advantages of Transformer architecture include parallelization capabilities, bidirectional encoding introduced by BERT, and a generalized approach for different tasks and models. The T5 model presents a unified seq2seq approach, providing for potential model co-training (e.g., sentiment analysis and summarization). LongFormer, in its turn, allows for encoding and summarizing longer or aggregated texts, which is commonly used in case of social media texts. This drastically reduces the GPU memory requirements, thus relieving the researchers from a major drawback of the Transformer architecture.

2.3. Text Summarization Types for Opinion Mining

Besides the differences between extractive, abstractive, hybrid, and compressive [37] summarization stated above, several other model features need to be discussed. Thus, in 1997, three task features and the respective summarization typology were suggested [38]:

- by intent: indicative/informative;
- by focus: generic/user-directed;
- by coverage: single-document/multi-document.

For opinion mining on social media, abstractive informative generic multi-document summarization fits best, as we need to assess what many people say nearly simultaneously. Further on, we will mostly orient our literature review to abstractive multi-document summarizations (for a recent review of 55 single-document summarization approaches, see [39]).

2.4. Previous Applications of Deep-Learning Summarization Models to Datasets from Reddit

To our best knowledge, deep learning was first applied to text summarization for Reddit datasets not later than 2017 [40]. In as late as 2019, authors [41] were still noting that, 'as most recent models ha[d] been evaluated exclusively on news corpora, our knowledge of their full capabilities [was] still superficial' [39: 524]. Until today, Reddit remains the only large social network that provides for large-scale data for text summarizations, and only very rare works challenge the quality of human summarizations [42].

So far, three major Reddit-based datasets have been developed for testing summarization models. In all of them, the unique feature of Reddit posts mentioned above and called 'Too Long; Didn't Read' (TL;DR) is utilized: the Reddit users provide their own summaries to longer posts, and their summaries are seen as benchmark human summarizations.

In particular, in 2017, the Webis-TLDR-17 was collected and labeled. This first-ever English-language summarization dataset contains over 3 million content-summary pairs. It has been used for creating a method for collecting author summaries for long texts

from social media [40], as well as for a 2019 text summarization ‘TL;DR challenge’ [41], despite it being considered ‘slightly noisier than the other datasets’ [42]. By 2018, a labeled dataset Reddit TIFU (from the TIFU subreddit) had been developed [43]. It comprises 123,000 posts from Reddit with the same TL;DR summaries. For the purposes of so-called extreme summarization of very high levels of abstraction, the TLDR9+ dataset of over 9 million Reddit content-summary pairs was developed in 2021; from it, the authors [15] extracted high-quality summaries, to form another, smaller TldrHQ sub-dataset. These datasets, in effect, remain the only ones for abstractive summarization of social media data.

The authors of the Reddit TIFU dataset suggested their own multi-level memory networks [43] that performed better than basic seq2seq and extractive models available by 2019. Within the 2019 TL;DR challenge, authors [44] compared LSTM, LSTM + copy, Transformer, Transformer + copy, and Transformer + pretrained models for n-gram abstractiveness in summarizations. The results have shown that the pretrained Transformer model unnecessarily outperformed the ground truth for news summarizations, but worked best for the Reddit data. Thus, Transformer models were shown to be efficient without changes to the model architecture; only pretraining and fine-tuning was recommended, and more extractive approaches were recommended for less noisy data. The authors, though, claimed that none of the models could “learn true semantic natural language compression” (p. 520). Other works [45] combined extractive and abstractive summarization (without a focus on application) and, by this, helped the challenge organizers to identify influential summary aspects that affect summarization, including sufficiency and formal text quality [41].

By 2021, though, several works had shown that BART-based models worked best for the Reddit datasets [46]. Among them, very recently, low-resource abstractive summarization methods based on a combination of transfer learning and meta learning [47] and perfection of pre-training by using extractive-type text masks [17] have been suggested. However, they have brought mixed results in terms of model applicability to news and social media data. The authors in [47] have shown the best improvement for the Reddit TIFU dataset in comparison with eight other datasets, which demonstrates the potential of social media data to be well summarizable, despite their higher level of noise and lexical diversity. This corresponds to the earlier results by [46] who showed that abstractive models worked better than extractive ones on the XSum and Reddit TIFU datasets. Contrary to this, the authors [17] have shown that, on news datasets such as CNN/DailyMail and long-text XSum, the pre-training enhancer worked better than on Reddit TIFU. This may be explained by the fact that their model was more extractive and did not correspond to the abstractive nature of the Reddit summarizations.

To this, we need to add that, till today, fine-tuning summarization models for Reddit has not aimed much to real-world applications; moreover, in current review papers, Reddit gains very low attention, featuring literally one paper on Reddit each [48,49]. Most works use the Reddit datasets along with many others such as in [46], rather than focus on them. Our goal, thus, is to finally apply the models not only to Reddit but to other social media as well. Neural-network summarization models are expected to work alike on datasets with similar limitations that come from comparable platform affordances, including pre-set maximum or average text length. Different factors that affect the quality of summarization on different platforms are yet to be explored.

In this study, we make several steps forward from the existing models for Reddit and try the models for another platform, another type of social media data, and two other languages. First, we deal with the models for Reddit: we fine-tune them, apply to real-world non-labeled Reddit data, and compare their performance, to select the best one for the Reddit posts. Second, we apply the model to comments, pooled by popularity. Third, we apply the models to other pooled data, namely the tweet pools from Twitter, to assess real-world applicability and detect the shortcomings of the models for another platform and other languages.

3. The Experimental Methodology

3.1. The Proposed Pipeline

The proposed approach is comprised of the following steps:

1. Data collection using Reddit API—Praw library.
2. Preprocessing using general regex and nltk methods.
3. Model loading using the huggingface Transformers library.
4. Fine-tuning (on Reddit data) using the Pytorch library.
5. Visualization using standard Python tools.

Data collection for Reddit is performed using the official API, and pre-processed and grouped as will be shown in the next section. Models are loaded in pre-trained state from Hugging Face and tuned using a standardized Pytorch Lightning framework.

The main advantages of this approach are:

- Additional training and tuning for other social networks, with advantages provided by transfer learning.
- Fast and reliable inference provided by the transformer models.
- Robust preprocessing ensures less noise in the data and as such more effective fine-tuning.
- Visualization module capable of presenting basic statistics, results of summarization and sentiment analysis.
- Standardized fine-tuning approach provides an opportunity to add new models and modify training structure with minimal code changes.

Figure 1 demonstrates key steps of the process. Overall, the pipeline follows established training and fine-tuning techniques, with the focus, though, on their application to the social discussions on social networks.

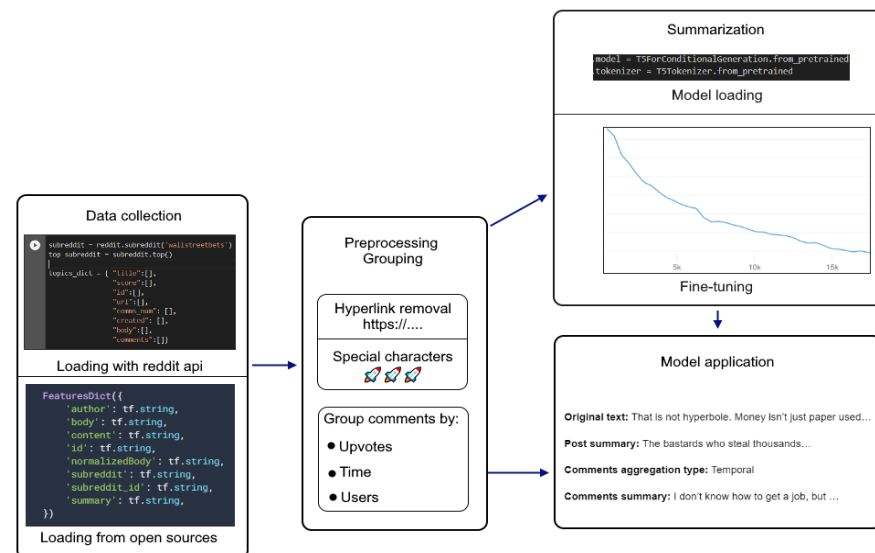


Figure 1. Key steps of the research pipeline.

3.2. Model Selection

To find the model most suited for the task of social media posts summarization, the following pipeline is proposed:

- Loading the initial model states from their basic pre-trained options, setting the parameters for fine-tuning;
- Preprocessing labeled data, splitting them into training, validation, and test subsets by removing hyperlinks, special characters, and bot messages (in case of comments);
- Fine-tuning the models on labeled data, with the model training from the loaded state continued;

- Comparing the results using automatic evaluation metrics (on them, see below);
- Inspecting summarization results on new data, namely the real-world unlabeled datasets from individual topical subreddits.

3.3. The Datasets

The models described above are tested on a subsection of the TLDR Reddit dataset [40], collected specifically for the task of abstractive summarization. These data take advantage of Reddit users leaving shortened versions of longer messages. This allows for obtaining the initial text and its author's own summarization as labelled data for the task of abstractive summarization. Table 1 presents the split distribution for the data.

Table 1. The data splits.

| Design Steps | No. of Posts | Average Post Length | Average Summary Length |
|------------------|--------------|---------------------|------------------------|
| Training (80%) | 307,866 | 224.9 | 21.2 |
| Validation (10%) | 38,483 | 225.8 | 21.3 |
| Test (10%) | 38,484 | 222.7 | 21.1 |

The data have the following post length and summary length distributions, presented in Figure 2.

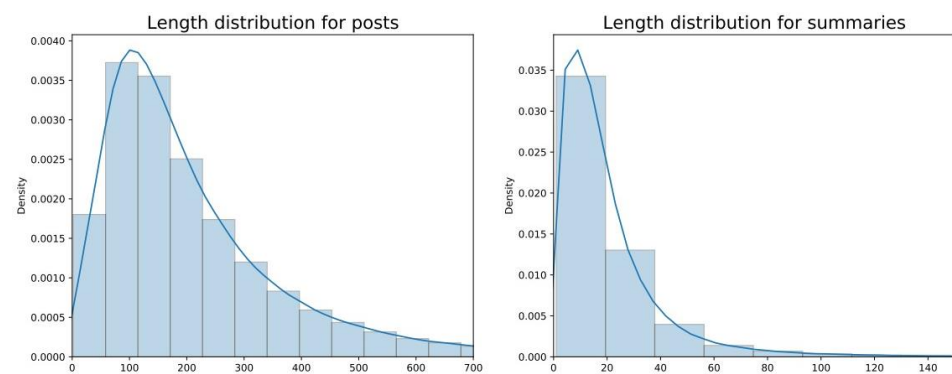


Figure 2. Length distributions for posts and their summaries, by the number of tokens.

The distributions show that, generally, initial texts do not exceed the 500 token limit and summaries do not exceed 150 tokens. As such, standard values for these parameters can be used for all models described in the next section.

3.4. Fine-Tuning Parameters

The models were fine-tuned on the Reddit data using the Pytorch Lightning framework. The initial model names and parameters used for tuning were the following: for BART, distilbart-cnn-12-6 (305 M parameters); for T5, t5-base (220 M parameters); for LongFormer, led-base (161 M parameters). The parameters were chosen empirically close to their standard values. Thus, weight is the regularization parameter, with weight decay fixed at 1×10^{-6} . Adam epsilon determines the number of weight updates and, as such, regulates the learning rate, fixed at 1×10^{-8} . Low number of epochs, namely, two, was chosen to use more training data instead of selecting a portion of the dataset and, thus, create a more general solution.

3.5. Metrics

As the ROUGE complex of metrics is used most often in this research area, including for non-English-language datasets [50], we also employ it. However, we are aware of its criticism, as there is repeated evidence of its low interpretability and lack of relevance to human judgment [51,52].

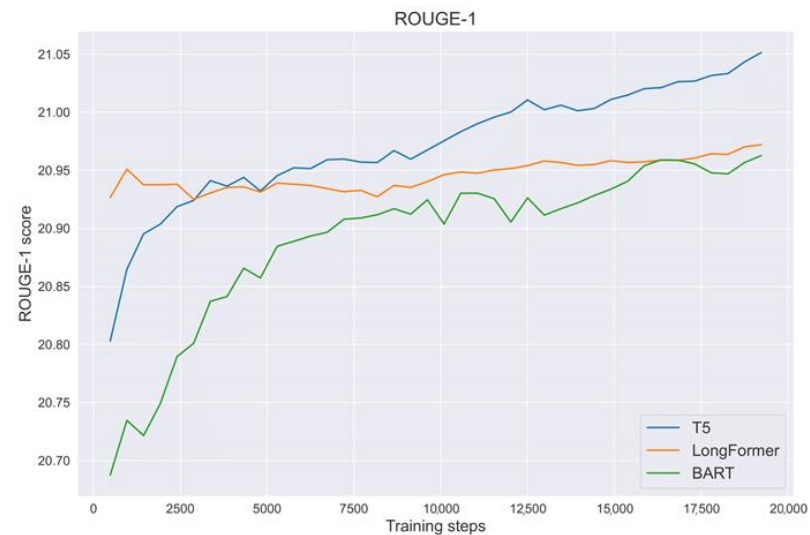
Rouge-N [53] is a set of metrics originally proposed in 2004, but to this day, is widely regarded as one of the most accurate, and mostly resembling human judgement. Similar to the machine translation metric BLEU [54], different variations of ROUGE use the idea of calculating the percentage of overlapping elements in the output, produced by a summarization model and baseline target summary. ROUGE-1 and ROUGE-2 refer to the overlap of unigrams and bigrams respectively. ROUGE-L uses the Longest common subsequence, taking sentence structure into account.

4. Results of Fine-Tuning and Model Selection

After text preprocessing, model fine-tuning, and metric calculations at different steps, the following results were obtained.

Figure 3 demonstrates core metrics, obtained during the fine-tuning of each model for the Reddit case. The metrics were calculated on a validation set at different training steps. Table 2 shows the average, minimum, and maximum values of metrics for BART, T5, and LongFormer, the highest reached values highlighted.

(a) Scores for the ROUGE-1 metric:



(b) Scores for the ROUGE-2 metric:

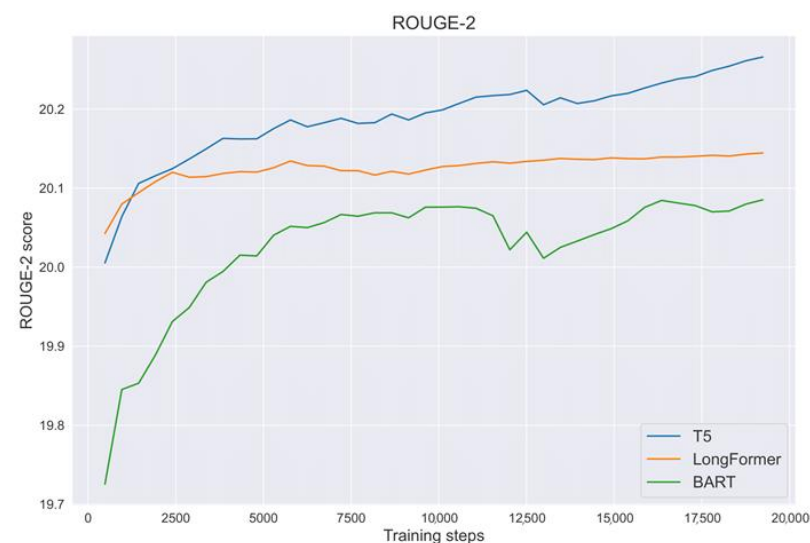


Figure 3. Cont.

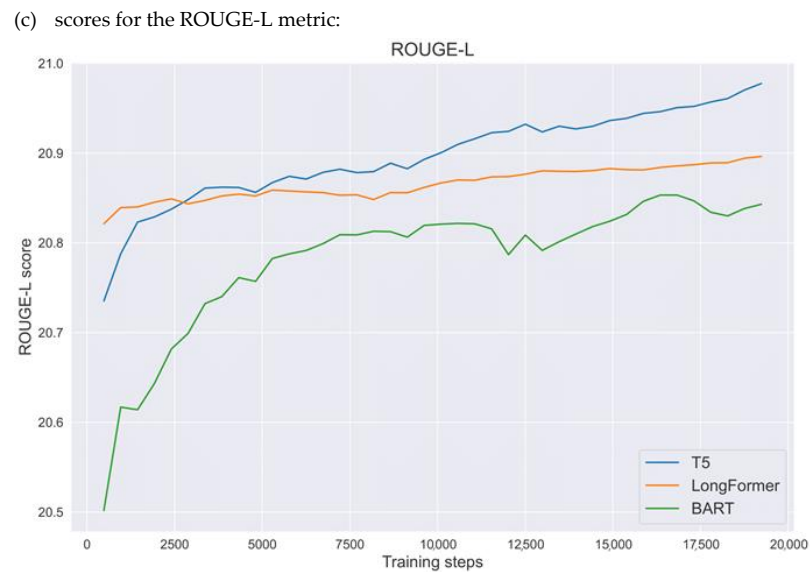


Figure 3. The results of fine-tuning, as evaluated by the ROUGE metrics.

Table 2. The results of fine-tuning, as measured by the ROUGE metrics.

| Metrics | | BART | T5 | LongFormer |
|---------|---------|--------|---------------|------------|
| ROUGE-1 | average | 20.891 | 20.972 | 20.946 |
| | minimum | 20.687 | 20.802 | 20.925 |
| | maximum | 20.962 | 21.051 | 20.972 |
| ROUGE-2 | average | 20.025 | 20.189 | 20.125 |
| | minimum | 19.725 | 20.005 | 20.042 |
| | maximum | 20.085 | 20.265 | 20.144 |
| ROUGE-L | average | 20.781 | 20.896 | 20.865 |
| | minimum | 20.501 | 20.735 | 20.821 |
| | maximum | 20.853 | 20.977 | 20.896 |

It can be discerned that the LongFormer model shows the best performance on each metric, having weaker initial values, but bigger increase in quality during each iteration, with the resulting increase in quality of around 1.2% compared to the original version. This is comparable to other studies related to fine-tuning Transformer-based models for other types of textual data; however, social media data are much more noisy and less structured than other types of texts. This is why our result may be considered to be even better than the baseline results obtained by other research teams for other types of data.

The T5 model shows even stronger initial metrics and comparable resulting quality, thus, though, showing somewhat lesser impact of fine-tuning, and the highest maximum values—a 1% to 2% increase in comparison with the BERT-, BART-, and LongFormer-based models, in both their baseline and fine-tuned forms. This is why we recommend the fine-tuned T5 model for further use on the Reddit data.

5. The Model Application to Real-World Reddit Posts and Comments

5.1. Applying the Model to the Dataset of Unlabeled Reddit Posts

The chosen model has been tested on new data to show its efficiency on the previously unseen data. In our previous practice of working with automated text analysis, we have seen multiple times that the models that are believed to work well by automatic metrics do not really provide for satisfactory results when the result is human assessed. This is why we have applied the fine-tuned T5 model to the unlabeled data, to assess by human eye the user summaries (TL;DRs) vs. machine summaries and reveal possible shortcomings. We especially underline here that the Transformer-based models especially fine-tuned for

a certain type of data (Reddit) have not been previously tested on the raw data from the platform; before our tests, only non-fine-tuned models were applied to massive Reddit datasets, and the results were compared to data of a different nature, which did not lead to better performance for opinion mining from social media.

We have obtained the data from 'r/wallstreetbets' subreddit, one of the most popular on Reddit (of over 10 million users). Recently, it has been the origin point of a major social-economic discussion involving stock trades. We have collected this subreddit in order to test the model on the data that do not belong to any existing dataset specifically prepared for summarization tests; thus, the collected data may be called real-world, unlabeled, or raw data. The top posts were collected, most of them dated circa mid-2020 (around the time of the highest user engagement). A total of 150,000 posts were collected.

First, we received the automated summaries and assessed them per se (as sentences) in the working group that contained experienced Internet researchers, communication scientists, and linguists. Here, we need to note that not many posts on Reddit have the TL;DR addition; this is why it is important to assess the automated summaries vs. both the posts themselves, not only vs. TL;DRs. We have assessed both.

When juxtaposed to the posts themselves, the summaries were coherent short sentences with clear relation to post title (see Table 3); although, the extent of this connection requires further automated testing. The summaries correctly reflected the nature of posts, including the first-person speech ('I hate the market . . . '), user's opinion/position ('Weird price is now around where the volumes dropped off'), argumentation structure ('This post is not about the fundamentals in the first place'), and emotions ('The bastards who steal thousands of years off the regular man's lives . . . '). There were some shortcomings in logic ('I hate the market, and I don't think we need to talk to each other') and grammar, e.g., some rare statements with no predicate. We have also spotted some hint of extraction, rather than abstraction, for shorter texts where key short phrases were added to abstractive parts of summaries ('There are a lot of factors out here. It is not a meme.'). In addition, when the posts were changed many times (e.g., had 'Edit 5', 'Edit 6', etc.) or had many multi-vector reasons the author gave to argue his/her position, the machine tended to summarize with excessive generalization ('There are a lot of factors here'; 'There is a good reason for GME'). However, the summaries, in general, mirrored the contents of the Reddit posts correctly and to the point. For examples of posts with no TL;DR and their summaries, see Appendix A.


Table 3. Examples of post summaries without TL;DRs.

| Post Title | Automated Summary |
|---|---|
| Daily reminder that buying GME at 300 before 28/1 was a great play at the time | This post is not about the fundamentals in the first place, it was a risky bet based on the Incoming short squeeze. |
| Why should any American company ever act responsibly again? | I hate the market, and I don't think we need to talk to each other |
| At these levels, it's NOT about the roe of the stock. tis the number of shares in the hedge funds' possession. | Weird price is now around where the volume dropped off. I think we're at the bottom, as long as we hold |
| People are risking their lives to wage war against the suits and it brings tears to my eyes to watch them do it | The bastards who steal thousands of years of the regular man's lives every single day. |
| Pledge to Delete RobinHood After The Great Rise. Who's With Me? | There are a lot of factors out there. It's not a meme. |

More surprising results were brought when we have tried to compare the automated summaries to user TL;DRs (see Table 4). The number of the posts with TL;DRs was very small and, thus, could be closely assessed by the experts. The machine summaries

corresponded to the user summaries in only circa 50% of posts. However, this was not due to the low quality of post summarization, but because the Reddit users did not properly use the TL;DR function, making sarcastic, aggressive, or non-sensical claims under ‘TL;DR’. Where user summaries truly reflected the essence of the posts, the machine summaries closely resembled the user TL;DRs. However, where the authors used TL;DR for conveying some emotional message, swearing, or slang-based claims, the machine summary could not create a similar one, as it was oriented to the post, not to TL;DR (see Table 4, line 4). In case when the posts consisted of links and its TL;DR had words, the machine returned unreliable results (see Table 4, line 5).

Table 4. Examples of post summaries with TL;DRs.

| Post Title | User Summary | Automated Summary |
|--|--|---|
| A dark part of WallStreetBets history, and why it’s important for you to know and remember | Founder bad, greedy, got banned for being greedy. Being greedy again with new spotlight on the sub. | Founder bad, greedy, got banned for being greedy. Being greedy again with new spotlight on the sub. tlak is a good mods. u/Speaks and post posts. |
| I suspect the hedgies are illegally covering their short positions | **Melvin and gang hasn’t covered shit. They’ve been illegally “closing out” their short positions and if we hold they will 100% get fucked. There is far more nefarious shit at play.** | There is far more nefarious shit at play.** So this morning I saw the S3 and put into the SEC has been for a long time, but it’s not worth it. |
| The real reason Wall Street is terrified of the GME situation | We probably own way more of GME than we think and that is freaking out Wall Street because it could prove they’ve been up to some extremely illegal shit and the whole system could implode as a result. | There’s a good reason for GME. |
| Can you say “CLASS ACTION LAWSUIT”. Fuck you RobinHood | GME to 1000 EOD, GME 5000 EONW. THE SUQEEZE HAS NOT SQUOZEN YET, BUY AND HOLD  | I hate autism. I think it’s a bad idea. |
| Upvote to ban all of Canada from the internet | #CanaBan2017 | CanaBan20172017 2017 2017201720172017.com 2016s 2011–2007 and 2012–2007’s last year. |

This revelation problematizes the Reddit-based summarization research, as it poses a fundamental question: Do we teach the machine to ‘ideally’ summarize the post, or do we teach it to produce human-like summaries (with all the ironic and unpredictable uses of TL;DRs), as this is not the same?

All in all, the fine-tuned T5 model showed its high capacity for summarizing posts correctly and close to human summaries and may be recommended for further practical of the data from Reddit—however, with no orientation to the TL;DR section of the posts, as this section itself might be highly misleading.

5.2. The Model Application to Pools of Comments

As stated above, for social media analysis, it is crucial to have a possibility to summarize pools of user posts and/or comments. Thus, we have conducted tests on a collection of comments from the abovementioned collection of posts.

The test sub-dataset of comments was collected on the basis of their popularity. As the model we applied had a limitation of working best with the 512-token texts, we had to limit the length of each pool; this limitation was not relevant for posts, most of which were shorter, but it was crucial for the comment pools (see Figure 4 for the length distribution in the sub-datasets of posts and comment pools).

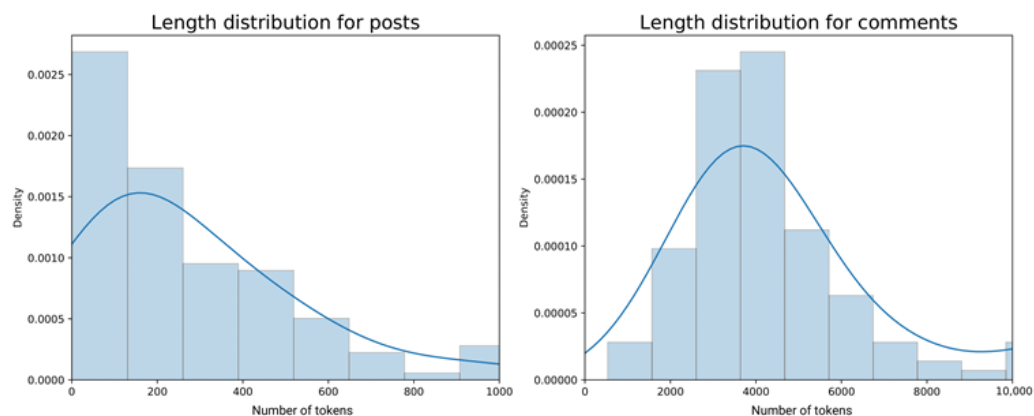


Figure 4. Length distributions of tokens for posts and their comments in additional data, by the number of tokens.

Due to their high volume and the model limitations, the collection of comments had the following logic. First, we uploaded the top popular 30% of all comments per post, without additional subsequent answers to comments. Even after this, as Figure 4 demonstrates, the average length of comment pools subjected to summarization was ~4000 words, instead of ~150+ words for the posts. Thus, only the most popular comments of the 30% pools that fitted to the 512-token limitation were summarized.

On the comments, the fine-tuned model has, in general, shown stronger results than the non-fine-tuned model, and the baseline threshold of 512 tokens was, technically, not an issue. This means that abstractive summarization can be directly applied to summarize pools of user texts, e.g., the posts/comments by one author, discussion threads, or topic-related pools previously discovered by topic modeling techniques.

Human assessment has shown that some comment summaries are capable of capturing both the topicality of the post and the general mood of commenters, e.g., their opposition to the author (see Table 5, line 1, the post on prices of shares), just as their shared sentiment with the author of the post (see Table 5, line 2, the post on wage wars). The comments also had some grammar structures typical for argumentation ('This is not about the market, it's about how much you are willing to pay attention to your own product', see Table 5, line 1). Examples of comment summaries may also be found in Appendix A.

However, this was the case in less than 50% cases. In many enough cases, the summaries conveyed only the basic opinion (see Table 5, line 3) or the basic emotion, often quite a rude one (see Table 5, line 4). There are also cases where the comments are unanimously supportive of the author, and the post, its TL;DR, and its automated post summary are clear, but the comments are not summarized well and only convey minor elements of meaning (see Table 5, line 5).

All in all, summarizing the comments looks less successful than that of posts, which is expected, given that there are up to dozens of texts that are summarized. They all have varying user emotions, micro-topics, and argumentation structures. Our advice here would be to deal with sub-pools of comments delineated technically (e.g., comment threads), topically (e.g., via topic modeling), or via clustering (e.g., via agglomerative clustering with the Markov moment [55]).

One more reason for smaller success could be that the comments to be summarized were of highly varying length, not only of different emotionality or topicality. This is why

we have decided to apply the Transformer-based methods to pools of texts where the length does not vary that significantly. Such datasets can be found on Twitter.

Table 5. Examples of summaries for comment pools.

| Post Title | Post Summary | Comment Summary |
|--|---|--|
| At these levels, it is NOT about the price of the stock. It is the number of shares in the hedge funds' possession. | Weird price is now around where the volume dropped off. I think we're at the bottom, as long as we hold. | This is not about the market, it's about how much you are willing to pay attention to your own product. |
| People are risking their lives to wage war against the suits and it brings tears to my eyes to watch them do it. | The bastards who steal thousands of years off the regular man's lives every single day. | I don't know how to get a job, but I think that's why I hate the f*ck out of it. |
| I suspect the hedgies are illegally covering their short positions | There is far more nefarious shit at play.** So this morning I saw the S3 and put into the SEC has been for a long time, but it's not worth it. | I agree with you, but I don't believe in this. |
| 30 Seconds From Triggering Market Nuclear Bomb | this is a big deal. | I'm a nutjob. |
| Analysis on Why Hedge Funds Didn't Reposition Last Thursday, Why They Didn't Cover on Friday, and Why They Want You to Think They Did. (GME) | Yes, hedge funds are pretty good for hedge fund management. But they can't be managed with the same amount of foreclosing on the 28th RH is not an actual number of hedge fund companies that have a very rare net SI due to lack of knowledge and lack of understanding of what it takes to find out how much money they need to put in between \$350 and \$400 M positions. | this is not a great thing. EDIT: I just wrote this on my phone, but it's all about the money you're looking for! |

6. Applying the Transformer-Based Models to Twitter Data

6.1. Twitter and Abstractive Summarization

Social media platforms today provide a high variety of platform affordances [56], which implies high variability in length and structure of user posts. For instance, Twitter limits the post/comment volume at 280 characters, which affects how users shape their speech.

Previous studies that used Twitter data for topic modeling to solve communication science tasks, e.g., for agenda detection purposes or finding pivotal moments in networked discussions [11], have had only moderate success. Moreover, social media data, and especially the noisy, unstructured, and extremely short texts from microblogs such as Twitter, have created multiple complications for topicality, sentiment, and opinion detection [57]. This implies that other methods, such as text summarization, the success of which does not depend that much on text length, may be applied to discussion evolution studies with better results.

Nevertheless, the studies of text summarizations beyond Reddit are scarce; of all platforms, Twitter is next to attract the researchers' attention. Neural-network abstractive summarization for Twitter has been explored in [58], and Reddit and Twitter were compared in [59]. These studies have shown the principal possibility of applying abstractive summarizations to Twitter data, but several important gaps have remained unresolved.

Of them, we address two that relate to the real-world features of Twitter discussions on the global scale. First, when such discussions erupt, they quickly become multilingual [60]. However, despite Twitter being extremely multi-lingual both within and beyond such discussions, there are no studies that apply abstractive summarization to tweets on multiple languages. Second, tweets are too short to summarize, but their accumulation into author, thread, or sub-topic pools, in line with the concept of cumulative deliberation [61], is what matters within heated discussions for public opinion formation. However, it has not been tested how the neural-network summarization models would work on the pools of tweets.

Thus, addressing these two gaps, we have conducted tests of abstractive summarization for a discussion on Twitter dedicated to the *Charlie Hebdo* massacre in Paris.

6.2. The Twitter Case, the Dataset, and Pooling Tweets

We have previously used the *Charlie Hebdo* dataset of over 2 million tweets in several studies; we collected it ourselves, with the help of a patented web crawler, in 2016 [60], and continue to use it instead of more recent data for a number of reasons. First, it is not as bot-infested as the modern datasets are, and we do not need to develop a separate methodology for cleaning it from algorithmic noise. Second, the way it is collected is independent from the Twitter recommendation system and, thus, allows for adequate representation of the discussion dynamic, especially the growth of the number of tweets, disregarding the final popularity of individual tweets. Third, we know the data, as we had worked with them for several years and can judge qualitatively whether the summarizations convey the messages that mirror the discussion dynamics.

For this paper, we have used a part of the dataset selected via a hashtag #jesuischarlie, as it contained more user opinions and emotions in comparison to more news-oriented #charliehebdo. The #jesuischarlie sub-dataset comprised the first three days of the discussion and, after preprocessing, had 420,080 tweets. In the language distribution, it largely reproduced the one of Twitter itself [62], with circa 50% of posts being in English [60].

We chose the three main European languages, namely, English, German, and French, for testing the models. The language of tweets was detected by using the FastText algorithm [63]. However, as the events that caused the discussion happened in France, the 'global' English-language and the 'local' Francophone segments of the discussion were similar in volume, while the German one was significantly smaller (213,558, 133,671, and 7430 tweets, respectively). That is, the part of the discussion in the German-language Twitter was much looser in time. This has posed an issue of equality of discussion representation in the three languages via summarizations; in other words, the principle of pool formation had to be chosen.

In choosing between the pool formed via time-oriented increment (e.g., all tweets for certain minutes/hours) and the pool based on equal number of tweets, we have chosen the latter, as we just needed to test the models and, thus, demanded the equal number of tweets to be gathered into the pools. This formed a very differing number of pools, but, for all the three languages, their number was sufficient for the testing purposes.

Then, we had to decide on the number of tweets in a pool. After several tests, the limitation of the model showed circa 300 tweets to be the optimal number to be summarized. Thus, we have used 300-tweet pools for testing the models.

6.3. The Model Selection

As already mentioned above, the fine-tuned T5 model has worked well for the comments of varying length on Reddit. However, the model limitation was 512 tokens per text, as fine-tuning larger-scale models would demand an unfeasible volume of resources that were as yet unavailable for the working group. This is why, to Twitter pools, we have applied non-fine-tuned models with greater capacity (up to 4096 tokens). If they are successful, the subsequent users of our models will, at least have a choice between fine-tuned models for smaller entries and non-fine-tuned models for larger-scale entries.

Pre-tests have shown that LongFormer works better with English than T5. Thus, a pretrained but not fine-tuned Longformer model, the longformer-large-4096, was used to summarize the subset in the English language. For French and German, the T5 pretrained non-fine-tuned models t5-base-fr-sum-cnndm and mt5-small-german-finetune-mlsum were used, respectively.

6.4. The Results for the Three Languages

We have received 537 pools and summarizations for English, 422 for French, and 22 for German. Examples of the summarizations are presented in Table 6.

Table 6. Examples of text summarizations for Twitter in three languages.

| English | French | German |
|--|--|--|
| 7 January 2015, 12:16 the terrorist attack on the offices of the satirical magazine Charlie Hebdo has left 12 dead. | 7 January 2015, 12:03 Le mardi, à Bruxelles, il y a une minute de silence pour | Jan 07, 12:24 RT und klingt nicht nur ähnlich . . . Du caricaturiste haïtien Bousiko |
| 7 January 2015, 12:31 this is a tribute to the 12 people who were killed in the attack on the satirical magazine. | 7 January 2015, 12:28 Le hashtag #suischarlie est le plus utilisé de toute l'histoire de | 7 January 2015, 15:20 Anschlag auf Satire-Zeitung: Dieter Nuhr rettet sich in Bunker |
| 7 January 2015, 12:33 the attack on the offices of the satirical magazine Charlie Hebdo has left 12 people dead and 12 others wounded. | 7 January 2015, 12:29 Le dessin de Cabu s'éteint en hommage aux 12 | 7 January 2015, 21:17 Der Pariser Anschlag ist auch auf Facebook und Twitter das grosse Thema. Irr |
| 7 January 2015, 12:36 we are all Charlie. | 7 January 2015, 12:32 Nouveau: hommage à Stéphane Charbonnier et | 8 January 2015, 01:27 Cartoonist Jean Jullien: "Wir trauern um Frankreichs bedeuten |

Note. Data partly mentioned in [64]. Yellow: unfinished sentences marked.

As the Table 6 demonstrates, the English-language summarizations were clearly understandable. They conveyed the main discussion message, were long enough where the main message was news-like and shorter where it was opinionated. Moreover, in English, it was possible to detect via summarizations when and how the initial outburst of news spread on Charlie Hebdo changed to a real discussion where related issues started to be mentioned and formulated (see Table 7).

Table 7. Gradual growth of issue-oriented discussion detected via text summarization.

| Time | Summarization | News/Issue |
|------------------|---|------------|
| 07-01-2015 8:51 | this is a tribute to the 12 people killed in the terrorist attack on the satirical magazine in Paris today. | News |
| 07-01-2015 12:26 | the terrorist attack on the offices of the satirical magazine Charlie Hebdo has left 12 dead. | News |
| 07-01-2015 12:28 | the attack on the satirical magazine Charlie Hebdo in the heart of the capital of France is being described as a terrorist attack on freedom of expression. | Issue |
| 07-01-2015 12:29 | freedom of speech is a powerful thing. | Issue |
| 07-01-2015 12:31 | this is a tribute to the 12 people who were killed in the attack on the satirical magazine. | News |
| 07-01-2015 12:33 | the attack on the offices of the satirical magazine Charlie Hebdo has left 12 people dead and 12 others wounded. | News |
| 07-01-2015 12:34 | this is a tribute to the 12 people killed in the terrorist attack on the satirical magazine on the 12th of January 2015. | News |

Table 7. Cont.

| Time | Summarization | News/Issue |
|------------------|---|------------|
| 07-01-2015 12:36 | we are all Charlie. | Issue |
| 07-01-2015 12:38 | this is a tribute to the 12 people who were killed in the attack on the satirical magazine Charlie Hebdo. | News |
| 07-01-2015 12:40 | this is a tribute to the 12 people who were killed in the attack on the satirical magazine Charlie Hebdo in france. | News |
| 07-01-2015 12:42 | the attack on the satirical magazine Charlie Hebdo has left 12 people dead and 12 others wounded. this is the deadliest terrorist attack in the history of the world. | News |
| 07-01-2015 12:43 | this is a tribute to the 12 people who were killed in the terrorist attack on the satirical magazine Charlie Hebdo on 8th de January 2015. | News |
| 07-01-2015 12:45 | the attack on in was 100% about religion. | Issue |
| 07-01-2015 12:47 | in the wake of the attack on the satirical magazine, the world is united in solidarity with the victims and their families. | Issue |
| 07-01-2015 12:48 | we live in a world where a man with a gun is intimidated by a man with a pen. if extremism has taught us anything, its that the pen is mightier than rocket launcher. | Issue |
| 07-01-2015 12:50 | freedom of expression is everyone's basic human right and is one that shouldn't need defending. what happened in France today is unacceptable. | Issue |
| 07-01-2015 12:51 | freedom of speech is the foundation that allows a culture to advance. | Issue |
| 07-01-2015 12:53 | freedom of expression is a gift. | Issue |

However, the summarizations for German and French returned understandable but broken sentences with loose endings (for examples, see Table 6, marked yellow). Additional runs did not bring better results. This may have been caused by the fact that Twitter breaks too-long user posts into two parts, and, thus, the model captures the broken structure of tweets. However, random checks of the subsets have not proved this assumption. This means that, most probably, the T5 model for languages other than English needs not only fine-tuning, but also additional training for proper grammar structures.

7. Conclusions

7.1. Putting Our Findings in Context

Opinion mining from social networks is one of most common and useful applications of natural language processing tools. There have been a variety of approaches, models, and datasets designed to work specifically for social networks. This includes topic modeling approaches, extractive and abstractive summarizations, sentiment analysis, and many others. In this work, we mainly focused on abstractive summarization approaches, as this class of models provides great insights for topicality detection in textual data and may be combined with other models.

One of the most interesting social media platforms for application of automatic analysis is Reddit. It is unique mainly in its community-driven structure: the website consists of different subreddits divided by topics and opinions created and curated largely by the users themselves. Additionally, its explosive growth in user base and popularity makes opinions shared on the platform a driving factor in some major real-world events. In this study, we show that abstractive summarization models can effectively be applied to analyzing such content, both in terms of original posts and their respective comments. Our contribution to abstract summarization studies is the provision of evidence on the varying quality of neural network models for summarization tasks and selection of the best models. In addition, we went beyond judging the models trained on open-source data and had them tested on real data. Experimental results prove that the LongFormer and T5 models demonstrate the best

performance as judged by the ROUGE score (around 2% increase in quality in comparison with other fine-tuned models and their baseline results) on the initial data with high quality of subsequent summarizations. This shows higher applicability of the T5 model, and more possibilities for improvement for the LongFormer model.

Our main contributions may be formulated as follows:

1. State-of-the-art Transformer-based models were fine-tuned and tested on open-source-labeled data obtained from Reddit, with an emphasis on tuning the existing models trained on large datasets, to improve their performance when working with social media data. The results for the real-world English-language data beyond the existing datasets were a highly satisfactory human assessment, namely, being clearly understandable, mirroring the contents, and capturing the posts' main points, argumentation, and emotions. They also went beyond the baseline results in terms of ROUGE-based assessment. The results were also compared to BART and were better, both in comparison to BART and the non-fine-tuned models. The difference reached 1.2% and 2% for LongFormer and T5, respectively. This is in line with the results of fine-tuning for other types of textual data; however, given that the user-generated texts are noisy and less structured, this result is well beyond what was expected.
2. A new dataset was collected to test the chosen model on a real-world pool of Reddit data for both posts and their respective comments (see Appendix A), available for other scholars on request.
3. Application of the fine-tuned T5 model to the real-world Reddit data of this new dataset has shown that the received summaries are lexically and grammatically satisfactory. They mirror the general meaning of the texts, as well as the authors' sentiment and argumentation patterns. However, the summarizations are much closer to the general meaning of the posts than to the posts' TL;DRs, which poses a fundamental question on whether at all the quality of summaries can be the target of testing, and whether deep learning needs to orient to TL;DRs from Reddit. Several more shortcomings have been noted, among them, an inclination towards extractive summarization and excessive generalization.
4. The fine-tuned model was also applied to comment pools of a significantly more complicated structure than average user posts on Reddit. For some pools, the results obtained resemble those for the posts, which opens wide possibilities for analysis of text pools from social media. However, many pools returned insufficient summarizations that did not correspond to the contents of comments. In many cases, finding a summary for a long a diverse pool created by many users just did not seem feasible, even for a human summarizer; in other cases, a major emotion or a major opinion was detected correctly, in a generalized way; in some rare cases, minor features of comments were picked up. Due to this, for comment sections, it is advised to add a proper comment pooling procedure—affordances-based or computational.
5. The varying length of Reddit comments could be a reason for limited success in comment summarization. This is proved by the fact that summaries of short text pools from microblogs have returned better results. Testing on pools of tweets in English has given satisfactory results by human assessment, allowing for fine-grained detection of agenda shifts and topicality evolution in networked discussions, as well as for an array of other possible applications. However, LongFormer has worked better than T5 for tweet pools, unlike for Reddit. This tells us that, for the same language, platform affordances that define the length and structure of social media texts may demand both different summarization models and additional fine-tuning. This also provides the first explanation of the current impossibility to reach the 'one-for-all' summarization model for all social media that is demanded by the market and some academics.
6. The results of testing T5 on German and French tweet pools, though, has revealed unsatisfactory representations of tweet pools, which might be an artifact of data collection but, with higher probability, it is the result of the models being not fine-

tuned to languages other than English or to social media data. This demands further fine-tuning of the Transformer-based models for social media studies in other languages. This provides the second explanation for unavailability of the aforementioned ‘one-for-all’ model.

In addition to this, we would like to especially underline that LongFormer is often considered to perform worse than full-fledged hard BERT-based models. However, LongFormer has appeared to be very efficient on tweets, presumably (and paradoxically) due to its major drawback of not having a fully connected structure, being outweighed by the unbalanced nature of social media texts and, especially, comments.

7.2. Future Work

After providing evidence that the T5 and LongFormer models provide new state-of-the-art results in abstractive summarization on real-world data from social media (subreddits), we would also like to set prospects for future research.

First, future work implies further model improvement, primarily for better utilization of the Reddit structure and affordances. For example, the use of information from sub-comments and of rating data can improve results, as has been shown in some multi-document summarization techniques. Another way of improving the model’s capacity is to fully utilize the LongFormer model by training it with larger-scale initial data, e.g., longer posts and blocks of comments. Human assessment of the resulting summarizations would bring additional information on what has been summarized and what has not. Another interesting opinion-mining idea is to only summarize partial comment data, either based on time of post (an idea more common for traditional social networks such as Twitter) or sentiment of posts. By only summarizing negative or positive comments, it should be possible to obtain respective key points of negatively or positively aligned comments.

Second, combination of models also looks very promising for opinion mining as additional information on text entities have been proven to improve the quality of the summarization models [65]. One suggestion is to combine T5 and LongFormer with topic-aware summarization models [66–69] that integrate topical information into sequential ones; this would be especially helpful for pools of comments with varying length. First attempts in this area are promising but need to be much enhanced for use on various social media and to put opinion mining in focus, and to render real-world discussions and assess topicality dynamics. Another suggestion is to combine topic modeling and/or sentiment analysis techniques with summarization. Text conglomerates can first be selected by probabilistic clustering based on word co-occurrence and assessed for quality [7], and then the essence of the most relevant texts can be summarized in an abstractive way for better understanding.

Third, an as yet practically unexplored line for future studies is developing datasets and training models for other social media platforms. The question that may be posed is to what extent the platforms differ in results when applied to real-world tasks; the scarce available evidence shows that we might expect varying levels of how platforms + models fit to different sociological tasks [59]. As stated above, our own attempts to apply the T5 and LongFormer text summarizations to Twitter datasets of varying volumes and languages [64] have shown that the models need much more additional fine-tuning, for both the languages and the affordances-based limitations of the collected data.

Author Contributions: Conceptualization, I.S.B. and N.T.; Funding acquisition, S.S.B.; Investigation, I.S.B., N.T. and S.S.B.; Methodology, I.S.B., N.T. and S.S.B.; Project administration, S.S.B.; Resources, S.S.B.; Software, N.T.; Supervision, I.S.B. and S.S.B.; Visualization, N.T.; Writing—original draft, N.T.; Writing—review & editing, I.S.B. and S.S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in full by Russian Science Foundation, grant 21-18-00454 (2021–2023).

Data Availability Statement: The datasets are property of Center for International Media Research of St. Petersburg State University, Russia, and may be available on request.

Conflicts of Interest: Authors declare no conflict of interest.

Appendix A. Examples of the Real Data

Here we present the examples of texts and their respective summaries. Original texts are presented and may include profanities or offensive language.

Post title: Why should any American company ever act responsibly again?

Original text: Whats the point of good corporate governance and fiscal responsibility? The companies that leveraged themselves to the moon, did stock buybacks to hyper-inflate their stock price, live on constant debt instead of good balance sheets are now being bailed out by unlimited QE. Free money to cover your mistakes. Why would anyone run a good business ever again? Just cheat and scheme and get bailed out later.

Edit: I am truly honored to be the number 1 post on WSB. To get validation from you autists and retards, the greatest American generation, is the peak moment of my life. Thank you all.

Edit 2: Many of you are saying this post is socialist. It is anti-capitalist. It is anti-wall street. It is none of that. My post is in fact about fixing capitalism so it is done the right way. Don't reward companies that are managed poorly and don't invest their profits wisely. Capitalism is about survival of the fittest and rewarding the winners not the schemers and cheaters. I'd rather have a profitable company that pays its workers livable wages, doesn't use sweat shop labor, doesn't pollute our environment, gives good quality healthcare, paid family leave, sick leave, maternity/paternity leave, reinvests in improving infrastructure, keeps low debt to equity, and has a 12 month emergency fund for a black swan event. Not companies that give all the money to the CEO and Board and nothing to the workers, do stock buy-backs with profits instead of improving infrastructure or saving for emergency funds. Let the greedy poorly run companies fail so we can invest only in good quality companies that treat their workers well. We will all make tons of profits in the market with well run companies and main street America will also be able to live a decent quality life.

Edit 3: I am not a salty bear. In fact I want the market to do well. But this is not the way. Bailing out weak companies that didn't save for a black swan event because of CEO greed is just making this bubble bigger and bigger and it will only pop worse later on.

JPow will ruin our market and the economy with this fake bubble with his printer. Let the market be free so we can shed weak companies and true capitalism can see a rise of the strong companies and the market can moon again. JPow and his printer are really helping the Wall street elite. Jpow doesn't care about you. Now the tax payers are bailing out shadow banking. Junk bonds are risky loans that private equity, hedge funds, and other shadow banking institutions give out to desperate companies that can't get loans from regular banks anymore. That's why junk bonds are shadow banking instead of traditional banking. JPow is using his unlimited printer to BAILOUT and give free money to the shadiest and greediest characters of wall street and society in general—private equity, hedge fund managers, shady billionaires.

PE, hedgies, shady billionaires were screwed because the economy just halted and companies were going to default on these risky loans since they had no revenue coming in. This is who JPow is helping. He just bailed them all out by buying these risky junk bonds on the back of the American tax payer. You may become homeless and starve, but private equity, hedge fund managers, and shady billionaires will be made whole by the fed.

Post summary: I hate the market, and I don't think we need to talk to each other.

Comments summary: Every company's goal should be to become 'too big to fail' with enough employees to be worth bailing out.

Post title: People are risking their lives to wage war against the suits and it brings tears to my eyes to watch them do it.

Original text: That is not hyperbole. Money isn't just paper used to buy cocaine and hookers. It is a physical representation of a portion of a man's life. Make 50,000 a year in a

soul sucking desk job? Taking a 50,000 position on GME is risking a year off your life to fight this good fight. The amount of years so many of you are willing to put on the line is an amazing testament to how dedicated this sub is to fucking these bastards raw. The bastards who steal thousands of years off the regular man's lives every single day. You're fighting the good fight, and your sacrifices will not be forgotten.

Post summary: The bastards who steal thousands of years off the regular man's lives every single day.

Comments summary: I don't know how to get a job, but I think that's why I hate the fuck out of it.

Post title: I Pledge to Delete Robin Hood After The Great Rise. Who's With Me?

Original text: Let's show them how many of their traders are out. These fuckers are set to IPO this year, fools. Who's trading data are you gonna sell if all of us leave? Shilling out to big money was the worst financial mistake of your pathetic wanna-be corps life. I'm out, but I'm not selling on RH until it's motherfucking done. Collect tendies and your tax information and split. Who's with me?

GME 5000 is not a meme.

I am not a financial advisor

Post summary: There are a lot of factors out there. It's not a meme.

Comments summary: The market is not as bad as you think. These examples demonstrate three types of summaries: entirely new text in the first example, a line from the original post (resulting in summary, common for extractive approach) and combination of words from original post with new ones.

References

1. Kherwa, P.; Bansal, P. Topic modeling: A comprehensive review. *EAI Endorsed Trans. Scalable Inf. Syst.* **2020**, *7*. Available online: <https://eprints.eudl.eu/id/eprint/682/1/eai.13-7-2018.159623.pdf> (accessed on 10 January 2022).
2. Potapenko, A.; Vorontsov, K. Robust PLSA performs better than LDA. In Proceedings of the 35th European Conference on Information Retrieval, Moscow, Russia, 23–24 March 2013; LNCS. Springer: Berlin/Heidelberg, Germany, 2013; Volume 7814, pp. 784–787.
3. Bodrunova, S.S. Topic Modeling in Russia: Current Approaches and Issues in Methodology. In *The Palgrave Handbook of Digital Russia Studies*; Palgrave Macmillan: Cham, Switzerland, 2021; pp. 409–426.
4. Rana, T.A.; Cheah, Y.N.; Letchmunan, S. Topic Modeling in Sentiment Analysis: A Systematic Review. *J. ICT Res. Appl.* **2016**, *10*, 76–93. [[CrossRef](#)]
5. Blekanov, I.; Tarasov, N.; Maksimov, A. Topic modeling of conflict ad hoc discussions in social networks. In Proceedings of the 3rd International Conference on Applications in Information Technology, Aizu-Wakamatsu, Japan, 1–3 November 2018; pp. 122–126.
6. Koltcov, S.; Koltsova, O.; Nikolenko, S. Latent dirichlet allocation: Stability and applications to studies of user-generated content. In Proceedings of the 2014 ACM Conference on Web Science (WebSci), Bloomington, IN, USA, 23–26 June 2014; pp. 161–165.
7. Bodrunova, S.; Koltsov, S.; Koltsova, O.; Nikolenko, S.; Shimorina, A. Interval semi-supervised LDA: Classifying needles in a haystack. In Proceedings of the 12th Mexican International Conference on Artificial Intelligence (MICAI'2013), Mexico City, Mexico, 24–30 November 2013; LNCS. Springer: Berlin/Heidelberg, Germany, 2013; Volume 8265, pp. 265–274.
8. Bodrunova, S.S.; Blekanov, I.S.; Kukarkin, M. Topics in the Russian Twitter and relations between their interpretability and sentiment. In Proceedings of the 2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS'2019), Granada, Spain, 22–25 October 2019; pp. 549–554.
9. Boyd-Graber, J.; Mimno, D.; Newman, D. Care and feeding of topic models: Problems, diagnostics, and improvements. In *Handbook of Mixed Membership Models and Their Applications*; Taylor Francis Group: New York, NY, USA, 2014; pp. 225–255.
10. Qiang, J.; Qian, Z.; Li, Y.; Yuan, Y.; Wu, X. Short text topic modeling techniques, applications, and performance: A survey. *IEEE Trans. Knowl. Data Eng.* **2020**, *34*, 1427–1445. Available online: <https://arxiv.org/pdf/1904.07695.pdf> (accessed on 10 January 2022).
11. Smoliarova, A.S.; Bodrunova, S.S.; Yakunin, A.V.; Blekanov, I.; Maksimov, A. Detecting pivotal points in social conflicts via topic modeling of Twitter content. In Proceedings of the 5th International Conference on Internet Science (INSCI'2018), St. Petersburg, Russia, 24–26 October 2018; LNCS. Springer: Cham, Switzerland, 2018; Volume 11193, pp. 61–71.
12. Kousha, K.; Thelwall, M. COVID-19 publications: Database coverage, citations, readers, tweets, news, Facebook walls, Reddit posts. *Quant. Sci. Stud.* **2020**, *1*, 1068–1091. [[CrossRef](#)]
13. Jiang, H.; Sun, L.; Ran, J.; Bai, J.; Yang, X. Community detection based on individual topics and network topology in social networks. *IEEE Access* **2020**, *8*, 124414–124423. [[CrossRef](#)]
14. He, L.; Yin, M.; Shi, Y. Love, Hate Thy Neighbour? Or Just Don't Care Much about Them: A Sentiment Analysis of China-Related Posts and Comments on Reddit. *Com. China Rep.* **2020**, *56*, 204–220. [[CrossRef](#)]

15. Sotudeh, S.; Deilamsalehy, H.; DERNONCOURT, F.; Goharian, N. TLDR9+: A Large Scale Resource for Extreme Summarization of Social Media Posts. *arXiv* **2021**, arXiv:2110.01159.
16. Liu, Y.; Jia, Q.; Zhu, K. Keyword-aware Abstractive Summarization by Extracting Set-level Intermediate Summaries. In Proceedings of the Web Conference (WWW'2021), Ljubljana, Slovenia, 19–23 April 2021; pp. 3042–3054.
17. Zhang, J.; Zhao, Y.; Saleh, M.; Liu, P. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In Proceedings of the 37th International Conference on Machine Learning (PMLR), Virtual, 13–18 July 2020; Volume 119, pp. 11328–11339.
18. Maybury, M. *Advances in Automatic Text Summarization*; MIT Press: Cambridge, MA, USA, 16 July 1999; Available online: <https://dl.acm.org/doi/book/10.5555/554275> (accessed on 10 January 2022).
19. Jing, K.; Xu, J. A Survey on Neural Network Language Models. *arXiv* **2019**, arXiv:1906.03591. Available online: <https://arxiv.org/pdf/1906.03591.pdf> (accessed on 10 January 2022).
20. Nallapati, R.; Zhai, F.; Zhou, B. Summarunner: A recurrent neural network-based sequence model for extractive summarization of documents. In Proceedings of the AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017; pp. 3075–3081.
21. Liu, Y. Fine-tune BERT for Extractive Summarization. *arXiv* **2019**, arXiv:1903.10318. Available online: <https://arxiv.org/pdf/1903.10318.pdf> (accessed on 10 January 2022).
22. Liu, Y.; Lapata, M. Text Summarization with Pretrained Encoders. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), Hong Kong, China, 3–7 November 2019; pp. 3730–3740.
23. Zhang, X.; Wei, F.; Zhou, M. HIBERT: Document Level Pre-training of Hierarchical Bidirectional Transformers for Document Summarization. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 5059–5069.
24. See, A.; Liu, P.J.; Manning, C.D. Get to the Point: Summarization with Pointer-Generator Networks. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Florence, Italy, 28 July–2 August 2017; Association for Computational Linguistics (ACL): Stroudsburg, PA, USA; pp. 1073–1083.
25. Narayan, S.; Cohen, S.B.; Lapata, M. Don't Give Me the Details, Just the Summary! Topic-Aware Convolutional Neural Networks for Extreme Summarization. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, 31 October–November 2018; pp. 1797–1807.
26. Cheng, J.; Dong, L.; Lapata, M. Long short-term memory-networks for machine reading. In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, Austin, TX, USA, 1–5 November 2016; pp. 551–561.
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 2017 Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
28. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [[CrossRef](#)]
29. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding Universal Language Model. *arXiv* **2018**, arXiv:1810.04805.
30. Lewis, M.; Liu, Y.; Goyal, N.; Ghazvininejad, M.; Mohamed, A.; Levy, O.; Stoyanov, V.; Zettlemoyer, L. 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, Virtual, 5–10 July 2020; pp. 7871–7880.
31. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *J. Mach. Learn. Res.* **2020**, *21*, 1–67.
32. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer Normalization. *arXiv* **2016**, arXiv:1607.06450.
33. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
34. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
35. Wang, A.; Pruksachatkun, Y.; Nangia, N.; Singh, A.; Michael, J.; Hill, F.; Levy, O.; Bowman, S.R. Super GLUE: A stickier benchmark for general-purpose language understanding systems. In Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–12 December 2019; pp. 3266–3280.
36. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The Long-Document Transformer. *arXiv* **2020**, arXiv:2004.05150.
37. Torres-Moreno, J.M. Automatic Text Summarization: Some Important Concepts. In *Automatic Text Summarization*; Torres-Moreno, J.M., Ed.; Wiley & Sons: London, UK, 2014; pp. 23–52.
38. Hand, T.F. A Proposal for Task-based Evaluation of Text Summarization Systems. In *Intelligent Scalable Text Summarization: Proceedings of a Workshop Sponsored by the Association for Computational Linguistics*; Mani, I., Maybury, M., Eds.; ACL: Madrid, Spain, 1997; pp. 31–38.
39. Syed, S.; Völske, M.; Lipka, N.; Stein, B.; Schütze, H.; Potthast, M. Towards summarization for social media—results of the tldr challenge. In Proceedings of the 12th International Conference on Natural Language Generation, Tokyo, Japan, 29 October–1 November 2019; pp. 523–528.

40. Völske, M.; Potthast, M.; Syed, S.; Stein, B. Tldr: Mining reddit to learn automatic summarization. In Proceedings of the Workshop on New Frontiers in Summarization, Copenhagen, Denmark, 7 September 2017; pp. 59–63.
41. Syed, S.; Yousef, T.; Al-Khatib, K.; Jänicke, S.; Potthast, M. Summary Explorer: Visualizing the State of the Art in Text Summarization. *arXiv* **2021**, arXiv:2108.01879.
42. Bommasani, R.; Cardie, C. Intrinsic evaluation of summarization datasets. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), virtual, 16–20 November 2020; pp. 8075–8096.
43. Kim, B.; Kim, H.; Kim, G. Abstractive summarization of Reddit posts with multi-level memory networks. *arXiv* **2018**, arXiv:1811.00783.
44. Gehrmann, S.; Ziegler, Z.; Rush, A.M. Generating abstractive summaries with finetuned language models. In Proceedings of the 12th International Conference on Natural Language Generation, Tokyo, Japan, 29 October–1 November 2019; pp. 516–522.
45. Choi, H.; Ravuru, L.; Dryjański, T.; Rye, S.; Lee, D.; Lee, H.; Hwang, I. VAE-PGN based Abstractive Model in Multi-stage Architecture for Text Summarization. In Proceedings of the 12th International Conference on Natural Language Generation, Tokyo, Japan, 29 October–1 November 2019; pp. 510–515.
46. Chen, Y.; Liu, P.; Zhong, M.; Dou, Z.Y.; Wang, D.; Qiu, X.; Huang, X. CDEvalSumm: An empirical study of cross-dataset evaluation for neural summarization systems. *arXiv* **2020**, arXiv:2010.05139.
47. Chen, Y.S.; Shuai, H.H. Meta-Transfer Learning for Low-Resource Abstractive Summarization. *arXiv* **2021**, arXiv:2102.09397.
48. Alomari, A.; Idris, N.; Sabri, A.Q.M.; Alsmadi, I. Deep Reinforcement and Transfer Learning for Abstractive Text Summarization: A Review. *Comput. Speech Lang.* **2022**, *71*, 101276. [\[CrossRef\]](#)
49. Shi, T.; Keneshloo, Y.; Ramakrishnan, N.; Reddy, C.K. Neural abstractive text summarization with sequence-to-sequence models. *ACM Trans. Data Sci.* **2021**, *2*, 1–37. [\[CrossRef\]](#)
50. Ertam, F.; Aydin, G. Abstractive text summarization using deep learning with a new Turkish summarization benchmark dataset. *Concurr. Comput. Pract. Exp.* **2021**, e6482. [\[CrossRef\]](#)
51. Liu, F.; Liu, Y. Exploring correlation between ROUGE and human evaluation on meeting summaries. *IEEE Trans. Audio Speech Lang. Process.* **2009**, *18*, 187–196.
52. Kryscinski, W.; Keskar, N.S.; McCann, B.; Xiong, C.; Socher, R. Neural text summarization: A critical evaluation. *arXiv* **2019**, arXiv:1908.08960.
53. Lin, C.Y. Rouge: A package for automatic evaluation of summaries. In Proceedings of the Text Summarization Branches Out: Proceedings of the ACL-04 Workshop, Barcelona, Spain, 24–25 July 2004; pp. 74–81.
54. Papineni, K.; Roukos, S.; Ward, T.; Zhu, W.J. Bleu: A method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, Philadelphia, PA, USA, 7–12 July 2002; pp. 311–318.
55. Bodrunova, S.S.; Orekhov, A.V.; Blekanov, I.S.; Lyudkevich, N.S.; Tarasov, N.A. Topic detection based on sentence embeddings and agglomerative clustering with Markov moment. *Future Internet* **2020**, *12*, 144. [\[CrossRef\]](#)
56. Bucher, T.; Helmond, A. The Affordances of Social Media Platforms. In *The SAGE Handbook of Social Media*; Burgess, J., Marwick, A., Poell, T., Eds.; Sage Publications: London, UK, 2018; pp. 233–253.
57. Blekanov, I.S.; Bodrunova, S.S.; Zhuravleva, N.; Smoliarova, A.; Tarasov, N. The ideal topic: Interdependence of topic interpretability and other quality features in topic modelling for short texts. In Proceedings of the International Conference on Human-Computer Interaction (HCI International 2020), Copenhagen, Denmark, 19–24 July 2020; LNCS. Springer: Cham, Switzerland, 2020; Volume 12194, pp. 19–26.
58. Li, Q.; Zhang, Q. Abstractive event summarization on twitter. In Proceedings of the Companion Proceedings of the Web Conference, Taipei, Taiwan, 20–24 April 2020; pp. 22–23.
59. Priya, S.; Sequeira, R.; Chandra, J.; Dandapat, S.K. Where should one get news updates: Twitter or Reddit. *Online Soc. Netw. Media* **2019**, *9*, 17–29. [\[CrossRef\]](#)
60. Bodrunova, S.S.; Smoliarova, A.S.; Blekanov, I.S.; Zhuravleva, N.N.; Danilova, Y.S. A global public sphere of compassion? #JeSuisCharlie and #JeNeSuisPasCharlie on Twitter and their language boundaries. *Monit. Obs. Mneniya Ekon. i Sotsial'nyye Peremeny* **2018**, *1*, 267–295.
61. Bodrunova, S.S. Social Media and Political Dissent in Russia and Belarus: An Introduction to the Special Issue. *Soc. Media Soc.* **2021**, *7*, 20563051211063470. [\[CrossRef\]](#)
62. Hong, L.; Ahmed, A.; Gurumurthy, S.; Smola, A.J.; Tsioutsoulouklis, K. Discovering geographical topics in the twitter stream. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; pp. 769–778.
63. Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T. Bag of tricks for efficient text classification. *arXiv* **2016**, arXiv:1607.01759.
64. Bodrunova, S.S.; Blekanov, I.S.; Tarasov, N. Global Agendas: Detection of Agenda Shifts in Cross-National Discussions Using Neural-Network Text Summarization for Twitter. In Proceedings of the International Conference on Human-Computer Interaction (HCI International 2021), Virtual, 24–29 July 2021; LNCS. Springer: Cham, Switzerland, 2021; Volume 12774, pp. 221–239.
65. Kouris, P.; Alexandridis, G.; Stafylopatis, A. Abstractive Text Summarization: Enhancing Sequence-to-Sequence Models Using Word Sense Disambiguation and Semantic Content Generalization. *Comput. Linguist.* **2021**, *47*, 813–859. [\[CrossRef\]](#)
66. Nguyen, T.; Luu, A.T.; Lu, T.; Quan, T. Enriching and controlling global semantics for text summarization. *arXiv* **2021**, arXiv:2109.10616.
67. Zheng, C.; Zhang, K.; Wang, H.J.; Fan, L. Topic-Aware Abstractive Text Summarization. *arXiv* **2020**, arXiv:2010.10323.

-
68. Wang, Z.; Duan, Z.; Zhang, H.; Wang, C.; Tian, L.; Chen, B.; Zhou, M. Friendly topic assistant for transformer based abstractive summarization. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Virtual, 16–20 November 2020; pp. 485–497.
 69. Ailem, M.; Zhang, B.; Sha, F. Topic augmented generator for abstractive summarization. *arXiv* **2020**, arXiv:1908.07026.