An Overview of Sentiments Expressed Across Abstractive and Extractive Summaries

CS 7347 Final Report

Carter Mondy, Mohammed Ahsan

Abstract:

Auto summarization is one of the premier tasks in natural language processing, and as more and more literature is published digitally the need for robust summarization methods is only increasing. Our work here compares the performance and nuances of abstractive and extractive summarization methods. We analyzed summary similarity as well as summary sentiment compared to the original text. To perform this analysis, we compared two models, BERT Extractive Summarizer and GPT for both abstractive and extractive text summarization. Our dataset is the CNN/Daily Mail dataset that contains various text samples as well as their human generated summaries. We used a RoBERTa model to create the sentiment scores for the original text, human summary, and model generated summaries, and then compared the results. We also used ROGUE scores to quantify the similarity between the original text and model generated summaries. The goal of this work is to help us understand the difference between these two text summarization methods with quantifiable results to accurately measure performance. Our work found that extractive summarizations tend to preserve the original sentiment and lexical accuracy better than abstractive summarization. However, we examined the caveats with this approach as well as potential improvements to the evaluation techniques.

Introduction:

Text summarization is one of the fundamental challenges faced in the realm of natural language processing. With the volume of text data available, there is an increasing need for models that can effectively and accurately summarize text documents. There are many reasons why summarization is useful such as improved efficiency, enhanced workflows, and improved accessibility since these models can process large amounts of text information and provide accurate summaries. There are two major methods applied when summarizing text, which are abstractive and extractive. Extractive summaries function by pulling in sentences and phases directly from the corpus and arranging them into a summary. This is typically regarded as more of a classification task as the model determines whether a sentence from the corpus belongs in the summary. Abstractive summaries on the other hand, are generally created by some model or algorithm after analyzing the body of text and creating new sentences that summarize the subject material. The goal of this paper is to perform an analysis on how the sentiments of these two schemas differ. State of the art language models BERT extractive summarizer and GPT are utilized to construct summaries on the CNN-Daily mail dataset. This dataset is very well documented and provides summaries generated by human experts, which is useful when the models above need to be fine-tuned. Pretrained language models are powerful tools in this regard as the goal of this paper isn't the development of a novel summarization technique but rather an analysis on the minutiae of how the two methods differ with respect to sentiments expressed.

The summaries generated by the pretrained language models are then compared using a third pretrained sentiment analysis model, based on the RoBERTa architecture. The results of this analysis should prove insightful on the mechanisms through which the output of extractive and abstractive summarizations differ. Due to the nature of the inquiry on the minutiae of the summaries differ, pretrained models were used with little modification, to best capture the "out of the box" performance of each of these models.

Literature Survey:

Liu and Lapata [1] introduced BERTSUM, an extension of BERT [2], tailored for summarization tasks by adding sentence-level embeddings. Their model was fine-tuned for both extractive and abstractive summarization using a unique encoder-decoder structure. Similarly, Abdel-Salam and Rafea [9] compared BERT variants such as DistilBERT and SqueezeBERT, highlighting their efficiency for extractive summarization with lower computational costs. Lewis et al. [8] proposed BART, a flexible model that generalizes BERT and GPT, achieving state-of-the-art results across summarization and other NLP tasks. Hoang et al. [11] adapted pretrained transformers for summarization by adding token labels and domain-specific training, demonstrating their model's strengths and limitations on datasets like CNN/Daily Mail and XSum. Miller [12] expanded upon this field by creating the BERT extractive summarizer model, which works in a hybrid capacity expanding BERTs extractive summarization capabilities by leveraging cluster analysis. These works collectively establish a robust foundation for applying transformer-based models to summarization tasks.

Radford et al. [3] advanced generative summarization with GPT-2, leveraging large-scale web data. They showcased extractive summarization capabilities and proposed a novel abstractive approach by appending "TL;DR:" to texts. While details on fine-tuning were sparse, GPT-2's versatility and popularity make it a key model for this project. Similarly, Moratanch and Chitrakala [10] provided a comprehensive review of abstractive summarization techniques, categorizing them into structure-based and semantic-based approaches. Their emphasis on semantic understanding and linguistic modeling underscores the challenges of abstractive summarization and inspires potential improvements.

Tan et al. [5] combined RoBERTa and LSTMs to enhance sentiment analysis, demonstrating their hybrid model's ability to outperform benchmarks like Naive Bayes and GRU. This innovative integration of transformers and recurrent models serves as a guide for leveraging RoBERTa in sentiment evaluation for generated summaries. Hoang et al. [11] also explored human evaluations alongside ROUGE scores, illustrating the impact of shorter abstractive summaries on sentiment perception—a factor crucial to this project's sentiment analysis of summaries.

Further enhancing the study of summarization tools, TweetEval, introduced by Barbieri et al .[13], provides a unified benchmark for evaluating NLP tasks specific to Twitter. Spanning seven tasks—including sentiment analysis and hate speech detection—the framework consolidates datasets under a consistent preprocessing pipeline and evaluation metrics. It emphasizes the effectiveness of pretrained models like RoBERTa fine-tuned on Twitter-specific data. This focus on social media's unique linguistic patterns complements earlier summarization frameworks, particularly in expanding the domain applicability of large-scale language models and sentiment analysis.

Mihalcea and Trau [7] introduced TextRank, a graph-based ranking model for extractive summarization. By identifying key sentences or words through co-occurrence relationships, TextRank offers a unique non-transformer-based approach. This method contrasts with transformer models by relying on similarity scores rather than classification tasks, providing an alternative perspective for extractive summarization.

Methodology:

To start, we loaded the CNN/Daily Mail dataset into our Jupyter notebook. This is a well-established and publicly available dataset, which comprises of around 280k training pairs, and around 24k validation and test pairs. This dataset includes an original text article from either CNN or the Daily Mail, highlights, which were human generated summaries, and an id section which served as a unique identifier for the article. These three items are stored as strings in CSV files. For our implementation we used the Hugging Face transformers library which contains this dataset as one of its defaults. To make our process more manageable to execute, we randomly sample 100 articles with their corresponding human summary to evaluate per run. These articles are used to develop both extractive and abstractive summaries. Figure 1 depicts the overall workflow of the system

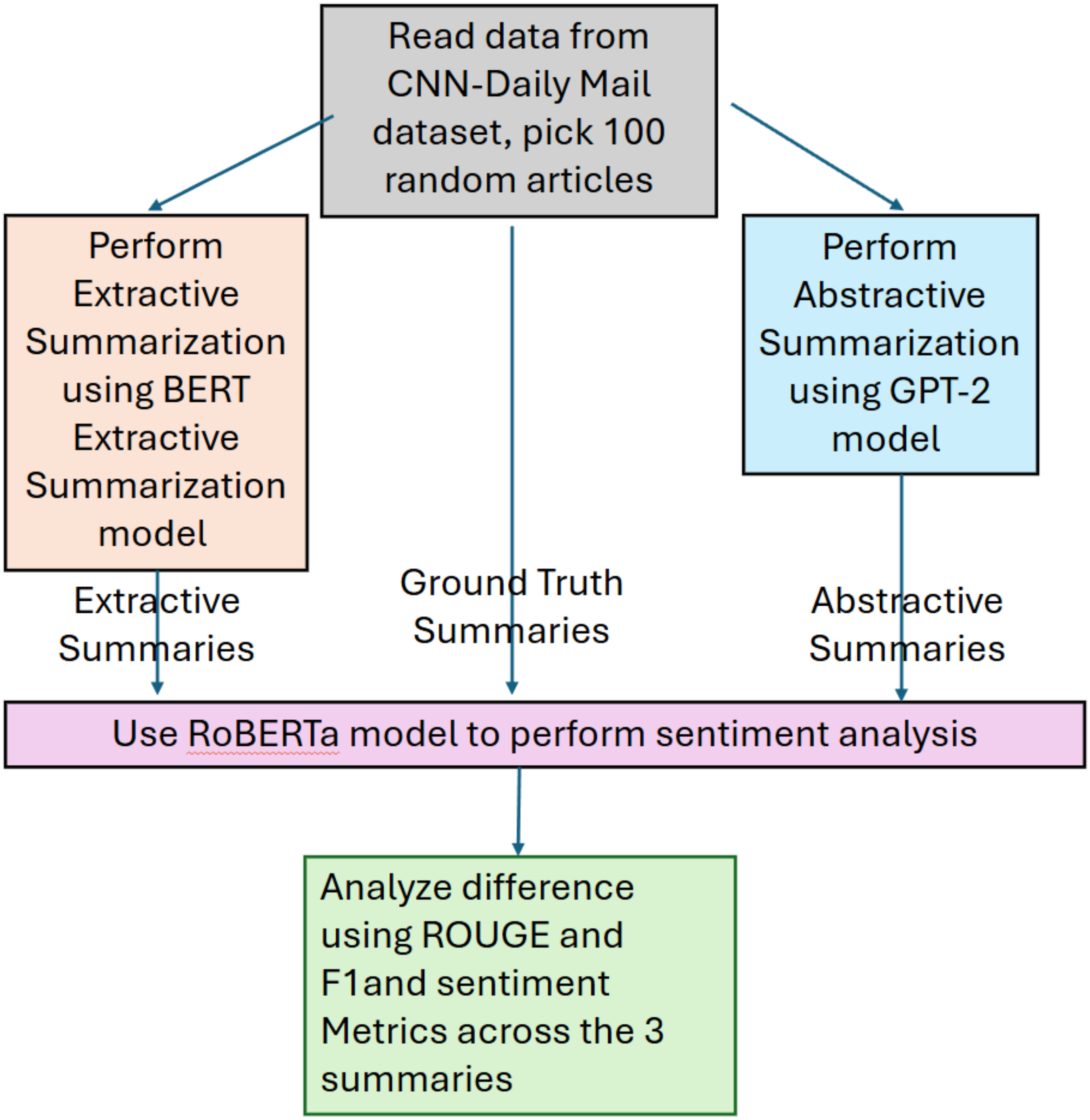

Figure 1: Layout of the summary generation and sentiment analysis process.

Once we had our dataset, two functions were created to handle the summarization tasks. For the abstractive summaries, we used the GPT2 model, where the summarization pipeline was configured with constraints on summary length. The minimum summary length was set to 30 words and the maximum length of 130 words. Sample articles were truncated to 1024 tokens to fit the model input limitations.

For extractive summaries, we used a BERT-based extractive summarization model. This model is based off Derek Miller's work in [12]. The model is publicly available as a python package on pypi and uses both unsupervised learning approaches like clustering and SOTA in NLP processes like BERT to perform extractive summarization. To evaluate the performance of these summarization models, we compared the summarizations against the human-written summaries. We used ROUGE-1 scores, which looked at unigram overlap, ROUGE-2 scores which measures bigram overlap, and ROUGE-L scores which measures the longest sequence of words that appears in the same order in both the generated summary and reference summary.

To evaluate the sentiments expressed by each of the summaries we developed a framework using the RoBERTa based model developed by Barbieri et al in [13]. The model was developed by combining the results of 3 separate roBERTa variants; the pretrained roBERTa base model developed by Liu et al, the base model but retrained on the Twitter corpus, and a roBERTa model trained on the Twitter corpus from scratch. The training of the final model took about 8-9 days on 8 NVIDIA V100 GPUs [13]. The Twitter-roBERTa-base for Sentiment Analysis model developed by Baribeiri et al is available as a model in the Hugging Face transformer library. The small modification we made in our implementation using this model was that the raw SoftMax probabilities are output as well as the final sentiment for each article analyzed. These sentiment scores are later examined in greater detail to derive some conclusions about each of the summarization methods employed for this study.

Once the sentiment analysis was completed, we analyzed the results. The JSON file containing sentiment data was flattened to ensure consistency and processed in groups of three. These groups corresponded to the abstractive, extractive, and human summaries. The key attributes were a sentiment label, which was either positive, negative, or neutral, along with the sentiment scores for each summary type. We then analyzed the summary data, by looking at the ROUGE score distributions, and looking at the sentiment data as well, such as the which summary types had differences and looking at the distribution of sentiment scores. Additionally, we used statistical tests, such as pairwise T-Test, ANOVA and a Tukey HSD post-hoc test to determine if the differences between different summary types were statistically significant.

Results:

Appendix 1 captures the articles and summaries generated both by human experts and the extractive and abstractive summarization techniques developed in this report so far. Some interesting takeaways are that the summaries generated by the abstractive summarizer seem to differ from the human gold standard quite a bit. This is potentially caused by the transformer model itself, as GPT-2's pretraining may not have been ideal for this application without modification. The extractive summarizer on the other hand places a great deal of importance in the first few sentences of the article being summarized. This again is due to the nature of the model itself. Figures 2-6 depict some of the outcomes of the analysis done on the results and discussed in much greater detail in the following section.

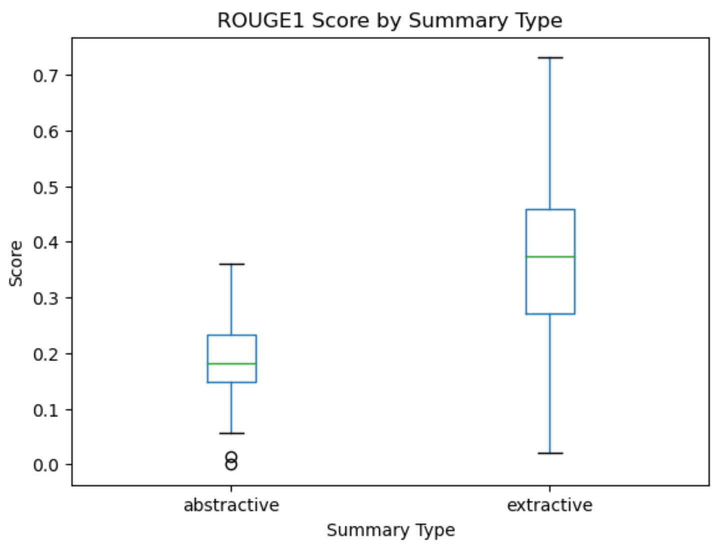

Figure 2: Distribution of ROUGE-1 scores by summarization type

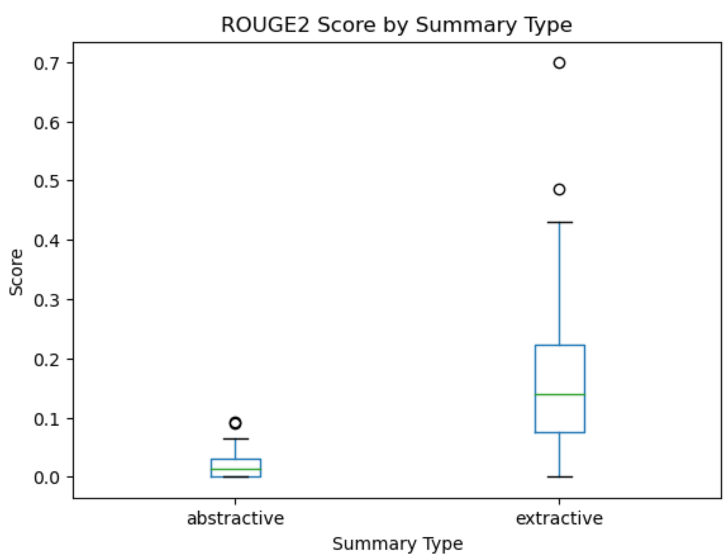

Figure 3: Distribution of ROUGE-2 scores by summarization type

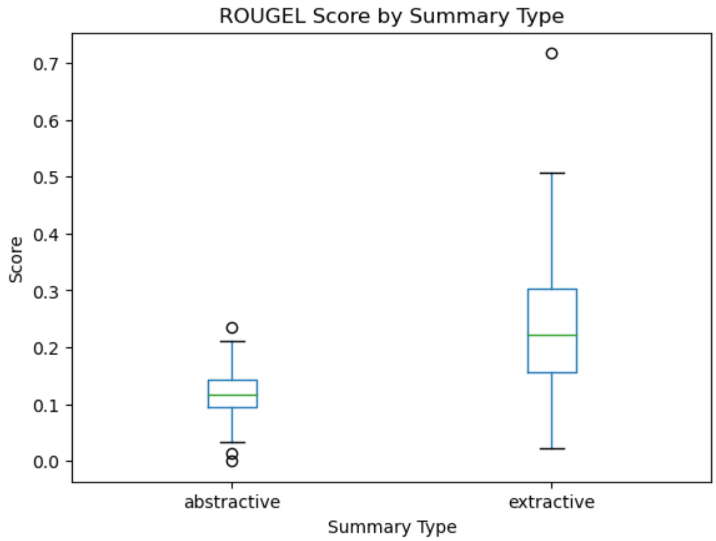

Figure 4: Distribution of ROUGE-L scores by summarization type

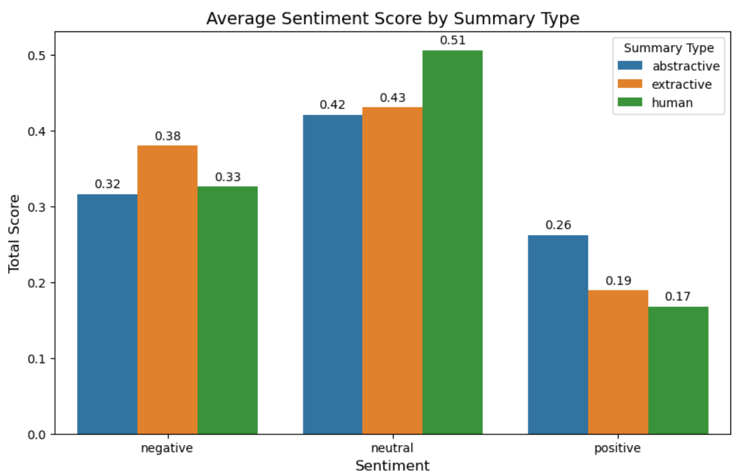

Figure 5: Distribution of sentiment scores by summarization and sentiment type

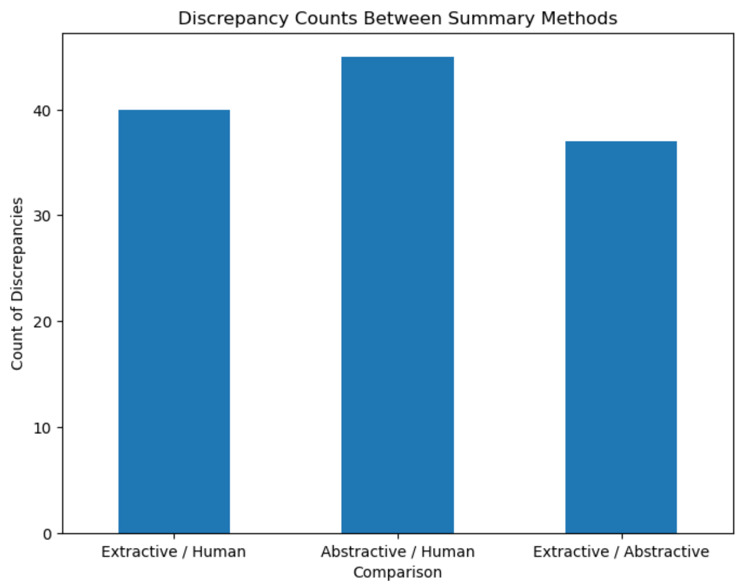

Figure 6: Count of discrepancies in sentiment label between summary types

Discussion:

The results offer valuable insights into the distinctions between abstractive and extractive summarization, specifically regarding their accuracy and sentiment representation. The distribution of ROUGE scores across summarization types indicates that extractive summarization is generally more consistent in accurately capturing the content of the original articles, as shown by the higher average ROUGE-1, ROUGE-2, and ROUGE-L scores compared to abstractive summarization. However, extractive summarization shows a larger standard deviation across all ROUGE scores, suggesting that while it tends to provide higher accuracy, the variability is higher. This means that the output quality is generally less stable for extractive summaries as opposed to abstractive summaries.

Specifically, the standard deviation for extractive summarizations were found to be 0.37 for ROUGE-1, 0.16 for ROUGE-2, and 0.11 for ROUGE-L, compared to 0.07, 0.02, and 0.04 respectively for abstractive summarizations. This significant difference in variability underscores the trade-off between accuracy and stability when choosing between summarization methods.

The statistical analysis conducted further supports these observations. A paired T-test was used to compare the average ROUGE scores between abstractive and extractive summarizations, and the results indicated that the differences in ROUGE scores were statistically significant across both types ($p < 0.05$). This suggests that extractive summarizations are statistically more accurate, despite the increased variability.

Furthermore, an ANOVA test was performed to analyze the sentiment scores across different summarization types. The ANOVA results indicated statistically significant differences in sentiment representation between human, extractive, and abstractive summaries (F-statistic =

5.62, $p < 0.01$). To further understand these differences, a Tukey HSD post-hoc test was conducted, which revealed that extractive summaries were significantly more consistent in sentiment compared to abstractive summaries, while human summaries remained the most stable.

In Figure 5, it is shown that the sentiment consistency varies significantly between abstractive and extractive summaries. Specifically, there are discrepancies in sentiment labels between the human-generated summaries and model generated ones, as indicated by Figure 6. This figure shows a bar chart that shows the frequency of discrepancies in sentiment labels between summarization types, highlighting that extractive summarization tends to maintain the original sentiment more effectively than abstractive summarization. Extractive summaries tend to adhere more closely to the sentiment of the original text, likely due to their approach of directly selecting phrases and sentences. In contrast, abstractive summaries often reinterpret the original content, which can lead to shifts in sentiment.

It is important to consider the implications of using ROUGE scores as a metric for evaluating summarization quality. While ROUGE effectively quantifies n-gram overlap between summaries, it does not inherently evaluate semantic meaning or sentiment preservation. This limitation is particularly relevant for abstractive summarization, where the generation of new content may retain the overall message but differ substantially in lexical choices. Consequently, ROUGE scores alone may not provide a complete picture of the quality of abstractive summarization. To better capture the semantic and affective accuracy of the generated summaries, future work could benefit from including other evaluation measures, such as BERTScore or human evaluations.

Conclusion:

This study adds to the existing discussion about text summarization by conducting a comparative examination of abstractive and extractive approaches with pre-trained transformer models. Our findings show that, while extractive summarization produces higher ROUGE ratings, it also has more variability, implying that it is more accurate yet inconsistent in retaining the nuances of the original text. On the other hand, abstractive summarization produces more readable text at the expense of sentiment correctness and consistency.

While the use of ROUGE scores is informative, there are limitations present. To address this, we propose using a combination of evaluation metrics to holistically assess the quality of generated summaries. This approach should take into account both lexical overlap and semantic preservation.

The results of the T-test, ANOVA, and Tukey HSD post-hoc test provide a statistical foundation for understanding the differences in performance and sentiment representation between summarization types. Extractive summaries tend to be more accurate in terms of ROUGE scores, but there is significant variability present. Overall, the insights gained from this statistical analysis have applications for helping the selection of which summarization technique, taking into consideration the trade-offs between accuracy, sentiment preservation, and interpretability.

Further considerations and future work would be to examine the effects of training set biases on the test results, GPT is very opaque with what data was used for pretraining, meaning that the CNN-Daily mail dataset could potentially have been used to develop the language model. In addition, while the BERT extractive summarizer is an excellent summarization tool it was primarily trained on lecture data collected from online learning sites. The RoBERTa model, similarly, was trained primarily on Twitter posts. Although the goal of the examination carried out in this report was to get a representative sample of the state-of-the-art approaches utilized in the summarization techniques of today, further work can be done to examine the impacts, if any, of the training sets by developing fully fledged models of comparable performance to those discussed here trained only on the CNN-Daily mail or other news datasets. Using a greater amount of data, as well as more diverse data sources, could make this analysis more robust.

References:

[1] Y. Liu and M. Lapata, “Text Summarization with Pretrained Encoders,” *arXiv:1908.08345 [cs]*, Sep. 2019, Available: https://arxiv.org/abs/1908.08345

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” *arXiv.org*, Oct. 11, 2018. https://arxiv.org/abs/1810.04805

[3] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, “Language Models are Unsupervised Multitask Learners,” 2019. Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf

[4] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, “Improving Language Understanding by Generative Pre-Training,” 2018. Available: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

[5] K. L. Tan, C. P. Lee, K. S. M. Anbananthen and K. M. Lim, "RoBERTa-LSTM: A Hybrid Model for Sentiment Analysis With Transformer and Recurrent Neural Network," in IEEE Access, vol. 10, pp. 21517-21525, 2022, doi: 10.1109/ACCESS.2022.3152828. keywords: {Sentiment analysis;Social networking (online);Training;Testing;Computational modeling;Analytical models;Blogs;Sentiment;transformer;RoBERTa;long short-term memory;LSTM;recurrent neural network;RNN}

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[7] Rada Mihalcea and Paul Tarau. 2004. TextRank: Bringing Order into Text. In Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, pages 404–411, Barcelona, Spain. Association for Computational Linguistics.

[8] Lewis, M. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*.

[9] Abdel-Salam, S.; Rafea, A. Performance Study on Extractive Text Summarization Using BERT Models. *Information* **2022**, *13*, 67. https://doi.org/10.3390/info13020067

[10] N. Moratanch and S. Chitrakala, "A survey on abstractive text summarization," *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*, Nagercoil, India, 2016, pp. 1-7, doi: 10.1109/ICCPCT.2016.7530193.

[11] A. Hoang, A. Bosselut, A. Celikyilmaz, and Y. Choi, “Efficient Adaptation of Pretrained Transformers for Abstractive Summarization,” arXiv.org, 2019. https://arxiv.org/abs/1906.00138 (accessed Oct. 14, 2024).

[12]D. Miller, “Leveraging BERT for Extractive Text Summarization on Lectures,” *arXiv:1906.04165 [cs, eess, stat]*, Jun. 2019, Available: https://arxiv.org/abs/1906.04165

[13]F. Barbieri, J. Camacho-Collados, L. Neves, and L. Espinosa-Anke, “TweetEval: Unified Benchmark and Comparative Evaluation for Tweet Classification,” Oct. 2020, doi: https://doi.org/10.48550/arxiv.2010.12421.

Appendix I: Inputs and Outputs of Summarizers

Example 1:

Article:
Bernie Ecclestone has urged Formula One's teams to start making sacrifices in a bid to save the sport from completely imploding. Ecclestone's words of warning come at the end of a miserable period for F1, notably with Caterham and Marussia entering administration. With costs spiralling out of control, and with the FIA powerless to act, three other cash-strapped teams in Lotus, Force India and Sauber have threatened to boycott the United States Grand Prix as a protest at what they perceive as a lack of action to stop the folly. Formula 1 boss Bernie Ecclestone says he is to blame for problems afflicting some of the smaller teams . Ecclestone, however, has dismissed that prospect out of hand. 'Forget all that c**p,' the F1 chief said. 'I promise they will be racing. 'They will be racing, I give you a guarantee - but I worry if they will be racing next year.' Therein lies Ecclestone's major problem - one which has surprisingly led him to concede that 'I know what's wrong, but don't know how to fix it'. Ecclestone added: 'No-one is prepared to do anything about it because they can't. The regulations have tied us up. 'If we were in a position where we could help these teams in trouble, we would do it. But we are not allowed to. Lotus are one of three teams who have threatened to boycott the United States Grand Prix . 'If one team is given 10,000 (US dollars), everyone has to have 10,000. The trouble with lots of regulations and lots of contracts is we don't think long term.' Part of the problem is the contracts Ecclestone has negotiated over the years with the teams, which includes the distribution of revenues. The smaller marques are dismayed at how the likes of Ferrari, Red Bull, McLaren and Mercedes overwhelmingly receive the lion's share of the ¬£900 million handed out. Ecclestone added: 'It makes no difference to me how the money is shared out. 'If they sat down here with me now and said they want to share out all of the money they get in a different way, I would say, 'Good, give me the bit of paper'. 'The problem is there is too much money being distributed badly - probably my fault - but, like lots of agreements people make, they seemed a good idea at the time.' Another issue Ecclestone faces is the teams now have too much say in the regulations, and, given so many vested interests, there is a lack of agreement. British-based Sauber are fuming at not having a say in the running of the sport . Ecclestone has a solution, but knows what the outcome would be. 'I would say to people getting a chunk of money that I would like to take a percentage of their performance-related payment,' Ecclestone said. 'I would put that money together to divide among the three or four we know are in trouble, and then I would put in the same amount of money. 'But there would not be one team that would think it was a bloody good idea. 'In the old days, the people sitting around a table would be the guys who could say yes or no. They would ask me to sort it out and it would be sorted. 'But none of the modern guys can agree anything, even if they wanted to. They all have to report back to somebody.' Although Ecclestone is effectively known as the commercial rights holder, it is private equity firm CVC Capital Partners which runs the sport - much to the apparent dismay of the teams. With Ecclestone no longer the man in overall charge, he said: 'If the company belonged to me, I would have done things in a different way. Smaller teams like¬†Force India¬†are struggling to make ends meet . 'That's because it would have been my money I was dealing with, but I work for people who are in the business to make money.' Another option for

Ecclestone is to 'tear up all the contracts'. He added: 'I'd take all the money, pay all the teams' debts that should be paid so people haven't suffered because of Formula One, and people haven't got credit for Formula One. 'I'd then say this is what you are going to get next year, this is the amount of money we've got, and this is what is going to happen from now on in. It's percentages.' The bottom line for Ecclestone is the big teams have to yield. 'It's not the biggest crisis I've dealt with, but I'm not happy,' he said. 'It's just you are talking about larger amounts of money, so therefore it is more difficult to solve. 'We'll have to do something about it because we can't all sit back nicely, relaxed and think the problem will go away. 'It's not like having the flu and taking a few tablets and it will disappear. 'The situation is such that if enough people want it resolved, we can resolve it. 'It's a case that the people involved in the sport will have to want to look after the sport and be prepared to make some sacrifices.'

Human Summary:

Ferrari, Red Bull, McLaren and Mercedes receive lion's share of the ¬£900million handed out . The FIA powerless to act despite costs spiralling out of control . Cash-strapped Lotus, Force India and Sauber have threatened to boycott the United States Grand Prix .

Abstractive Summary:

Ecclestone has been in the news a lot lately for his comments about the lack in action of F2 teams. He has said that the current situation is 'a bit of an anomaly' and that he has 'no idea how the situation will change'. He also said he would not be 'in the business of making decisions about F3' - a comment which is a bit like saying that you don’t want to be in business with a company that is going to make a decision about you. The F4 team has also been under fire for their lackluster performance in recent years. In a recent interview with Foto Wire, Ecclestone said: The F5 team is not doing well.

Extractive Summary:

Bernie Ecclestone has urged Formula One's teams to start making sacrifices in a bid to save the sport from completely imploding. Ecclestone's words of warning come at the end of a miserable period for F1, notably with Caterham and Marussia entering administration. With costs spiralling out of control, and with the FIA powerless to act, three other cash-strapped teams in Lotus, Force India and Sauber have threatened to boycott the United States Grand Prix as a protest

Example 2:

Article:

In an unusual opportunity a petting farm has popped up on the market, just a stone's throw away from a capital city. The Hahndorf Farm Barn is a farm and homestead in one with mountain views splayed over nine hectares of land. Located in the Adealide Hills, the sprawling property looks over Mt Lofty and the Hahndorf township - Australia's oldest surviving German township. Scroll down for video . The Hahndorf Farm Barn in South Australia is on the market . The property spans nine hectares and includes a huge homestead . The homestead overlooks the Adelaide Hills and mountains in the area . The large family home has recently undergone extensive renovations to bring it into the modern age . The farm has been in business since 1999, is a popular tourist attraction in the area and allows kids and adults alike to interact with animals and wildlife. Horse and motorbike riding, cuddles with the friendly farm animals and a chance to feed the wildlife are all offered at the the farm, which has grown by 140 percent under the current owners. And those owners are - but not for long - South Australian radio presenter Jason 'Snowy' Carter and his wife Kate, who have lived on the property for the past four years. The petting zoo is a popular tourist attraction in the area that has been up and running since 1999 . The homestead has five bedrooms with views over the vast land . There are also three bathrooms in the home . The kitchen is brand new with stone bench tops and plenty of storage room . Mr Carter works for Mix 102.3, and has four children with his wife. He grew up on a farm and wanted his children to have the same opportunity. 'The ability to be able to live on such an amount of land, and the kids be able to run around and ride horses and ride their motorcycles - it was always a dream of mine to have my family to be able to do that', Mr Carter told Arabella Hooper from Toop & Toop¬†real estate. '(And with) the Adelaide Hills being so close to the city ... it's 20 minutes in fact to get to the town without all the hustle and bustle,' he added. The home features many open plan living areas and places for the family to relax . The large number of bedrooms in the house means it would be perfect for a big family . Most rooms in the house have a spectacular view of the sprawling grounds . The living room extends into a spacious outdoor area and then on to the acreage . 'The other thing for me too was just when we saw it we thought what great value for money, and it came with a business.' The homestead has been completely renovated to 'bring the home into the new age', according to Mr Carter, and brags five bedrooms, three bathrooms, a brand new kitchen and open plan living space. There's also a kid's playroom, home office, casual dining room and floor to ceiling glass in the lounge area showing off sprawling views of the hillside. The farm is a well-established business and despite the current owners having no experience in such areas, still easily managed to run it.

Human Summary:

The Hahndrf Farm Barn has been put on the market. Located in the Adelaide Hills it's 30 minutes from the city. The property spans nine hectares .

Abstractive Summary:

The Farm is located in a small town in southern Australia. It is situated on a hill overlooking the valley of the river Adel. This is the first time that the Farm has ever been listed on Australian government's website. In fact, it is one of only two farms in Australia that have ever received a government designation. However, this is not the only farm in this country that has received government recognition. A number of other farms have also received recognition, including the Australian National University's Farm of Agriculture and Food.

Extractive Summary:

In an unusual opportunity a petting farm has popped up on the market, just a stone's throw away from a capital city. The Hahndorf Farm Barn is a farm and homestead in one with mountain views splayed over nine hectares of land. Located in the Adealide Hills, the sprawling property looks over Mt Lofty and the Hahndorf township - Australia's oldest surviving German township. Scroll down for video. The Hahndorf Farm Barn in South Australia