Towards Human-Centered Summarization: A Case Study on Financial News

Tatiana Passali[1], Alexios Gidiotis[1], Efstathios Chatzikyriakidis[2], and Grigorios Tsoumakas[1,2]

[1]School of Computer Science, Aristotle University of Thessaloniki
{scpassali,gidiotis,greg}@csd.auth.gr

[2]Medoid AI
{stathis.chatzikyriakidis,greg}@medoid.ai

Abstract

Recent Deep Learning (DL) summarization models greatly outperform traditional summarization methodologies, generating high-quality summaries. Despite their success, there are still important open issues, such as the limited engagement and trust of users in the whole process. In order to overcome these issues, we reconsider the task of summarization from a human-centered perspective. We propose to integrate a user interface with an underlying DL model, instead of tackling summarization as an isolated task from the end user. We present a novel system, where the user can actively participate in the whole summarization process. We also enable the user to gather insights into the causative factors that drive the model's behavior, exploiting the self-attention mechanism. We focus on the financial domain, in order to demonstrate the efficiency of generic DL models for domain-specific applications. Our work takes a first step towards a model-interface co-design approach, where DL models evolve along user needs, paving the way towards human-computer text summarization interfaces.

1 Introduction

The ever increasing amount of online text documents, such as blog posts, newswire articles and academic publications, during the last decades, has created the urgent need for appropriate natural language understanding tools. Summarization, i.e., shortening an initial text document by keeping only the most important information, plays a key role in addressing this information overload.

A lot of sophisticated summarization models have been proposed in the past, with a recent focus on Deep Learning (DL) architectures. DL models (See et al., 2017; Kryściński et al., 2018; Celikyilmaz et al., 2018; Chen and Bansal, 2018; Liu and Lapata, 2019; Song et al., 2019; Zhang et al., 2020) achieve great results in the task of summarization, outperforming most of the previously used methods. Typical DL models involve sequence to sequence architectures with RNNs (Nallapati et al., 2016; See et al., 2017) often combined with attention mechanisms (Luong et al., 2015; Bahdanau et al., 2015), as well as Transformers (Vaswani et al., 2017; Lewis et al., 2020; Raffel et al., 2020a).

Despite the success of DL models, some significant challenges remain. The low interpretability of these models (Brunner et al., 2020; Vig and Belinkov, 2019; Serrano and Smith, 2019; Vashishth et al., 2019) is a major drawback that limits significantly the trust of users in the whole process.

In addition, existing pipelines do not adequately engage the human in the summarization process (Trivedi et al., 2018; Shapira et al., 2017), providing isolated and static predictions. The engagement of users and their feedback in the whole process can be a key factor in creating high-quality models and improving the quality of existing models (Stiennon et al., 2020; Ghandeharioun et al., 2019).

To overcome the above limitations, we revisit the task of neural summarization from a human-centered perspective and with a unifying view of user interfaces and underlying summarization models. More specifically, we present a system that allows the active involvement of the user, setting the basis for human-computer text summarization interfaces. Our system allows users to choose over different decoding strategies and control the number of alternative summaries that are generated. Users can give their feedback by combining parts of the different generated summaries as a target summary for the corresponding input. These summaries are recorded, and can then be used as additional training examples, which in turn will improve the performance of the model and customize it to the preferences of the users.

Proceedings of the First Workshop on Bridging Human–Computer Interaction and Natural Language Processing, pages 21–27
April 20, 2021. ©2021 Association for Computational Linguistics

In addition, our system provides useful insights about the inner workings of the model, based on the self-attention mechanism of Transformers. Knowing which parts of the source document are most important for the generation of the final summary, can build up the trust between users and the machine.

We present a case study of the proposed system on the challenging, domain-specific task of financial articles summarization, to demonstrate the ability of the suggested approach to successfully employ generic DL models for domain-specific applications that often have different requirements. Indeed, domain-focused summarization models (Kan et al., 2001; Reeve et al., 2007) are generally more challenging, as they require deeper knowledge of the specific domain intricacies in order to generate salient summaries with logical entailment. To this end, we compiled a novel financial-focused dataset, which consists exclusively of financial articles from Bloomberg[1].

The rest of this paper is structured as follows. The main features of the proposed human-centered system are detailed in Section 2. The case study on financial news summarization is presented in Section 3. Finally, conclusions and interesting future research directions are discussed in Section 4.

2 HCI meets Summarization

In this section we will introduce the main features of our human-centered summarization system. We first present the approach used for interpreting the summaries generated by the model. Then we present the different decoding strategies we employ during inference. Finally, we explain how users can interact with our system.

2.1 Peeking into the Black Box

Our interface assumes the existence of a Transformer-based model with self-attention (Vaswani et al., 2017), which are the backbone of most modern summarization approaches. To provide insights into the produced summaries, we exploit the fact that the self-attention mechanism offers an implicit explanation about the factors that drive the behavior of the model. In particular, it helps the model identify input-output text dependencies by focusing on different parts of the input in order to generate the final sequence representation. This mechanism is typically combined with multiple attention heads. The attention weights of each head are concatenated with each other to compute the final weights.

Extracting the weights of each encoder layer separately, gives us useful insights about the model's behavior. In particular, we observe that different layers give us different types of insights regarding the way that the model perceives natural language. The first layers tend to focus on named entities and phrases taking a whole picture of the text, while the last layers attend additionally prepositions and articles in order to learn the language structure. In order to provide an overview of the model, we average all the self-attention layers along with all their attention heads, giving the user an overall picture regarding the model's learning process.

Assuming that a word which is attended by many words is more salient for the final decision of the model, we highlight the words according to their self-attention weights. Thus, high-weight words are strongly highlighted, while lower-weight words are faintly highlighted. This allows users to get a glimpse of where the model focuses on to generate the final summary.

2.2 Decoding Strategies

The selection of the right decoding strategy during inference can play a critical role in the whole process as it greatly affects the quality of a model's predictions (Holtzman et al., 2020), with different decoding strategies exhibiting different behaviors (Ippolito et al., 2019). Some decoding strategies, such as greedy search, suffer from redundancy issues (Shao et al., 2017), while others, such as beam search, might generate almost identical hypotheses among the different generated beams (Gimpel et al., 2013). Beam search is widely used in generative models, but there are also attempts that utilize other decoding mechanisms, such as top-k sampling (Fan et al., 2018b).

Our system allows for the active involvement of users into the underlying summarization process, by offering them the opportunity to select among the following decoding strategies:

- **Random sampling** selects randomly a token out of the word probability distribution. Often combined with a temperature parameter to control the entropy of the distribution (Ficler and Goldberg, 2017; Fan et al., 2018b; Caccia et al., 2020).
- **Top-k sampling** limits the space of possible

[1] https://www.bloomberg.com

next tokens to the top-k higher-ranked tokens of the distribution (Fan et al., 2018b) .

- **Top-p or nucleus sampling** selects the next token from a subset of tokens with cumulative probability up from a predefined threshold p (Holtzman et al., 2020). It can also be combined with top-k sampling.
- **Greedy search** selects the token with the highest probability at each time step.
- **Beam search** selects not only the token with the highest probability at each time step, but also a number of tokens with the highest probability according to the beam width. The number of the final generated beams is equal to the beam width. Beam search with beam width set to 1 degenerates to greedy search.
- **Diverse beam search** follows the beam search algorithm, but also adds a diversity penalty to enhance the diversity between the top most probable generated beams (Vijayakumar et al., 2016).

2.3 User Interaction

The interaction of a user with our system consists of the following steps. It starts with the user entering the source text into a text box. Then users have the option to view the visualization of the attention weights, as well as choose a particular visualization color. Next users can select among the available decoding strategies, which also gives them the opportunity to change the default hyperparameters of each decoding strategy. Finally, they can click on a button to obtain the summaries. It is also possible for users to mix and match sentences from the alternative produced summaries, as well as enter their own text, in order to create a personalized summary. This summary can then be saved, and later be used for further fine-tuning of the model.

3 Case Study: Financial Summarization

In this section, we detail our experiments with the case study of financial summarization. We first describe the data collection process and the preprocessing steps we followed. Then we discuss the models that we constructed and their evaluation. Finally we discuss concrete examples of the user experience. The code and instructions for this case study of our system is publicly available[2].

Table 1: Dataset Statistics

	Initial	**Preprocessed**
min. document length (words)	20	79
max. document length (words)	3758	2537
avg. document length (words)	676	669
avg. summary length (words)	23	23
# single-sentence summaries	21	0
# total documents	2120	2096

3.1 Dataset

We compiled a novel collection of financial news articles along with human-written summaries using the Bloomberg Market and Financial News API by RapidAPI[3]. The articles concern different financial and business categories, such as stocks, markets, currency, rates, cryptocurrencies and industries.

We removed outlier documents, i.e., relatively small (up to 70 tokens) and very large (over 3,000 tokens) ones. As most of the summaries consist of two sentences, we also removed single-sentence summaries to maintain a consistent target structure. Table 1 presents some basic statistics about our dataset before and after this simple pre-processing pipeline.

3.2 Models

We use the recently proposed PEGASUS model (Zhang et al., 2020), which is based on the transformer encoder-decoder architecture. It features 16 layers for both the encoder and the decoder, each of them with 16 attention heads. PEGASUS is already pre-trained on two large corpora, C4 (Raffel et al., 2020b) and HugeNews, and fine-tuned on 12 different downstream datasets. The model uses SentencePiece, a subword tokenizer (Kudo and Richardson, 2018), which divides rare tokens into known subword units allowing for the efficient handling of unknown words.

We experimented with two models fine-tuned on two different newswire datasets respectively, namely Extreme Summarization (XSum) (Narayan et al., 2018) and CNN/Daily Mail (Hermann et al., 2015). We used the open-sourced weights of these models to initialize our summarizers, and then further fine-tuned them on the collected financial dataset.

We observed that both model variants quickly adapted to the new dataset, and after only a few

[2] https://bit.ly/human-centered-summarization-notebook

[3] https://rapidapi.com/marketplace

Reference Summary: Local bonds a good buy despite political noise, fund head says. Rising copper prices will underpin demand for government debt.
Before Fine-tuning: All images are copyrighted.
After Fine-tuning: Peru's president set to win impeachment vote Friday. Rising copper prices will underpin demand for local debt: Credicorp.

Reference Summary: Damac will not bring new supply to the market, chairman says. Dubai home prices have slumped 30% since 2014 amid supply glut.
Before Fine-tuning: Dubai's property market will be in for a bumpy ride in the next few years, according to one of the city's biggest developers.
After Fine-tuning: Damac's Sajwani wants builders to stop building. Dubai home prices have fallen more than 30% since 2014.

Reference Summary: Kotak Mutual Fund sold part of Future bond holding at discount. A court adjourned hearings this week on the feud until Nov. 19
Before Fine-tuning: India's retail sector may be in for a rude awakening.
After Fine-tuning: Kotak sells Future bonds at 25% discount. Dispute over the e-commerce group's sale to RIL is dragging on.

Figure 1: Examples of generated summaries before and after fine-tuning.

training epochs they were capable of generating salient, non-redundant financially-focused summaries, which target explicit economic and business issues. Examples of the generated summaries before and after fine-tuning are shown in Figure 1. Fine-tuning on our dataset, leads to an improvement in performance by approximately 10 ROUGE-1 (F1 score) points (Lin, 2004) for the XSum model, which is eventually used in our system. The evaluation results are shown in Table 2.

Table 2: Evaluation Results. We measure the F1 scores for ROUGE-1, ROUGE-2, ROUGE-L and ROUGE-S.

	CNN/Daily Mail model				XSum model			
Fine-tuning	R-1	R-2	R-L	R-S	R-1	R-2	R-L	R-S
No	20.00	4.86	15.04	16.97	13.80	2.40	10.63	12.03
Yes	23.34	6.30	17.98	21.04	**23.55**	**6.99**	**18.14**	**21.36**

3.3 Samples from the User Experience

An example of the visualized self-attention weights is shown in Figure 2. The model focuses on basic named entities of the source text, which are indeed important for the final generation. We also observe that different layer depths provide different insights regarding the model's learning process as shown in Figure 3. For example, the first layers attempt to focus on every word of the input document in order to capture phrases and sentences, while the last layers pay close attention to prepositions and articles attempting to learn language structure.

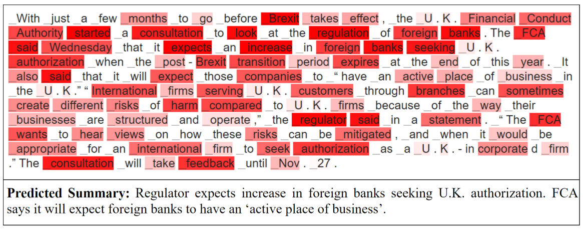
Predicted Summary: Regulator expects increase in foreign banks seeking U.K. authorization. FCA says it will expect foreign banks to have an 'active place of business'.

Figure 2: Visualization of the encoder self-attention weights. The underscore before a token indicates the start of the token according to the subword tokenizer.

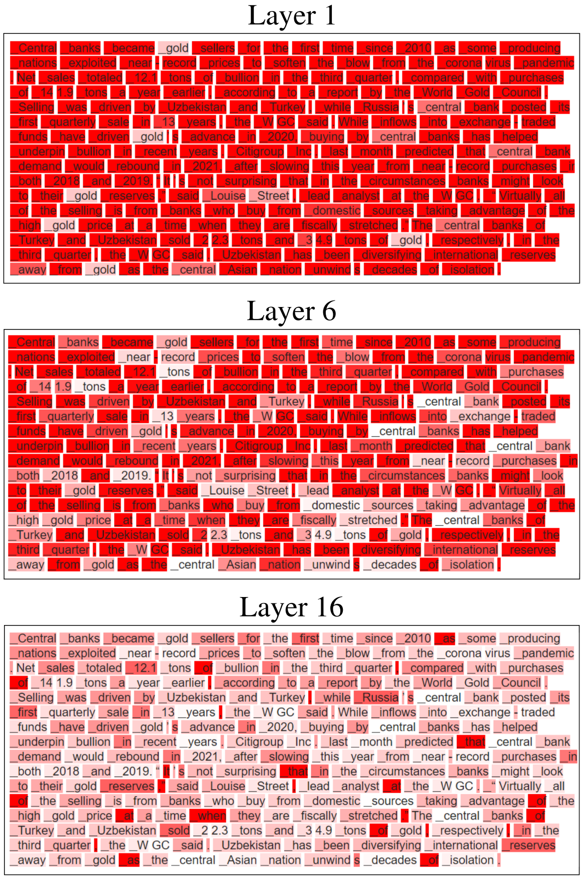

Figure 3: Self-attention weights of layers 1, 6 and 16.

An example of the output differentiation between different decoding strategies for the same input text is shown in Figure 4. The different summaries that are generated by the model, demonstrate the value of selecting an appropriate decoding strategy for the final generation.

4 Conclusions and Future Work

We presented a novel system for human-centered summarization that actively engages the user into

Reference Summary: Turkish and Uzbek central banks lead selling in third quarter. Overall gold demand fell 19% year-on-year: World Gold Council.

Greedy Search: Turkey, Russia sold gold in third quarter. Central banks bought gold at near-record pace in recent years.

Beam Search: Turkey, Russia sold gold in third quarter. Central banks bought gold in recent years to cushion blow from pandemic.

Random Sampling: Turkey, Russia sold gold in third quarter. Central banks bought gold at record pace in recent years.

Top-*k* Sampling: Turkey, Russia sold gold in third quarter. Central banks bought gold in recent years to cushion blow from pandemic.

Top-*p* Sampling: Turkey, Uzbekistan, Russia sold gold in third quarter. Central banks bought gold at near-record pace in recent years.

Diverse Beam Search:
1. Turkey, Russia sold gold in third quarter. Central banks bought gold in recent years to cushion blow from pandemic.
2. Turkey, Russia sold gold in third quarter. Central banks have been buying the metal to cushion blow from pandemic.
3. Central banks sold 12.1 tons of bullion in third quarter. Turkey, Russia posted first quarterly sales in 13 years.
4. Russia posts first quarterly sales since 2002. Turkey, Uzbekistan among nations to sell in third quarter.
5. Russia posts first quarterly sales since 2002. Turkey, Uzbekistan among nations to sell gold reserves.

Figure 4: Output for different decoding strategies.

the whole process of summarization, enabling personalized summaries and models. The users interact with the model, by entering a source text, selecting different decoding strategies, viewing a visualization of the model's attention and synthesizing a final summary from parts of multiple summaries, which can be used for further fine-tuning. We also presented a case study of our work, along with a novel dataset, on summarizing financial news. We observed that pre-trained PEGASUS models adapt quickly to our dataset, generating salient financially-focused summaries. Our work aims to inspire future research in human-centered techniques for neural summarization systems.

In future work, human involvement in the summarization process could be enhanced by using approaches that allow users to control different aspects of the generated summaries, such as length (Kikuchi et al., 2016; Liu et al., 2018; Takase and Okazaki, 2019; Fan et al., 2018a), style (Fan et al., 2018a) or generation based on a specific entity of the text (He et al., 2020; Fan et al., 2018a).

The interface we designed can be also further extended, allowing the user to evaluate the generated summaries, assessing different aspects of the text, such as salience, readability and coherence. Finally, more advanced approaches can be explored for leveraging the user submitted feedback in order to further improve the underlying model (Lertvittayakumjorn et al., 2020; Li et al., 2016).

References

Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 2015 International Conference on Learning Representations*.

Gino Brunner, Yang Liu, Damian Pascual, Oliver Richter, Massimiliano Ciaramita, and Roger Wattenhofer. 2020. On identifiability in transformers. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Massimo Caccia, Lucas Caccia, William Fedus, Hugo Larochelle, Joelle Pineau, and Laurent Charlin. 2020. Language gans falling short. In *International Conference on Learning Representations*.

Asli Celikyilmaz, Antoine Bosselut, Xiaodong He, and Yejin Choi. 2018. Deep Communicating Agents for Abstractive Summarization. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1662–1675, New Orleans, Louisiana. Association for Computational Linguistics.

Yen Chun Chen and Mohit Bansal. 2018. Fast abstractive summarization with reinforce-selected sentence rewriting. In *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*.

Angela Fan, David Grangier, and Michael Auli. 2018a. Controllable abstractive summarization. In *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*, pages 45–54, Melbourne, Australia. Association for Computational Linguistics.

Angela Fan, Mike Lewis, and Yann Dauphin. 2018b. Hierarchical neural story generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia. Association for Computational Linguistics.

Jessica Ficler and Yoav Goldberg. 2017. Controlling linguistic style aspects in neural language generation. In *Proceedings of the Workshop on Stylistic Variation*, pages 94–104, Copenhagen, Denmark. Association for Computational Linguistics.

Asma Ghandeharioun, Judy Hanwen Shen, Natasha Jaques, Craig Ferguson, Noah Jones, Àgata Lapedriza, and Rosalind W. Picard. 2019. Approximating interactive human evaluation with self-play for open-domain dialog systems. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pages 13658–13669.

Kevin Gimpel, Dhruv Batra, Chris Dyer, and Gregory Shakhnarovich. 2013. A systematic exploration of diversity in machine translation. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1100–1111.

Junxian He, Wojciech Kryściński, Bryan McCann, Nazneen Rajani, and Caiming Xiong. 2020. Ctrlsum: Towards generic controllable text summarization. *arXiv preprint arXiv:2012.04281*.

Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. 2015. Teaching machines to read and comprehend. *Advances in neural information processing systems*, 28:1693–1701.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2020. The curious case of neural text degeneration. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Daphne Ippolito, Reno Kriz, João Sedoc, Maria Kustikova, and Chris Callison-Burch. 2019. Comparison of diverse decoding methods from conditional language models. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3752–3762, Florence, Italy. Association for Computational Linguistics.

Min-Yen Kan, Kathleen R McKeown, and Judith L Klavans. 2001. Domain-specific informative and indicative summarization for information retrieval. In *In: Workshop on text summarization (DUC 2001*. Citeseer.

Yuta Kikuchi, Graham Neubig, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura. 2016. Controlling output length in neural encoder-decoders. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1328–1338, Austin, Texas. Association for Computational Linguistics.

Wojciech Kryściński, Romain Paulus, Caiming Xiong, and Richard Socher. 2018. Improving abstraction in text summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1808–1817, Brussels, Belgium. Association for Computational Linguistics.

Taku Kudo and John Richardson. 2018. SentencePiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 66–71, Brussels, Belgium. Association for Computational Linguistics.

Piyawat Lertvittayakumjorn, Lucia Specia, and Francesca Toni. 2020. FIND: human-in-the-loop debugging deep text classifiers. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 332–348. Association for Computational Linguistics.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pretraining for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Jiwei Li, Alexander H. Miller, Sumit Chopra, Marc'Aurelio Ranzato, and Jason Weston. 2016. Dialogue learning with human-in-the-loop. *CoRR*, abs/1611.09823.

Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out*, pages 74–81.

Yang Liu and Mirella Lapata. 2019. 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)*, pages 3721–3731.

Yizhu Liu, Zhiyi Luo, and Kenny Zhu. 2018. Controlling length in abstractive summarization using a convolutional neural network. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4110–4119, Brussels, Belgium. Association for Computational Linguistics.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421, Lisbon, Portugal. Association for Computational Linguistics.

Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gülçehre, and Bing Xiang. 2016. Abstractive Text Summarization using Sequence-to-sequence RNNs and Beyond. In *Proceedings of The*

20th SIGNLL Conference on Computational Natural Language Learning, pages 280–290, Stroudsburg, PA, USA. Association for Computational Linguistics.

Shashi Narayan, Shay B Cohen, and Mirella Lapata. 2018. Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. *arXiv preprint arXiv:1808.08745*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020a. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.*, 21:140:1–140:67.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020b. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, 21(140):1–67.

Lawrence H Reeve, Hyoil Han, and Ari D Brooks. 2007. The use of domain-specific concepts in biomedical text summarization. *Information Processing & Management*, 43(6):1765–1776.

Abigail See, Peter J Liu, and Christopher D Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In *Proceedings of the 2017 Annual Meeting of the Association for Computational Linguistics*, pages 1073–1083.

Sofia Serrano and Noah A. Smith. 2019. Is attention interpretable? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2931–2951, Florence, Italy. Association for Computational Linguistics.

Yuanlong Shao, Stephan Gouws, Denny Britz, Anna Goldie, Brian Strope, and Ray Kurzweil. 2017. Generating high-quality and informative conversation responses with sequence-to-sequence models. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2210–2219, Copenhagen, Denmark. Association for Computational Linguistics.

Ori Shapira, Hadar Ronen, Meni Adler, Yael Amsterdamer, Judit Bar-Ilan, and Ido Dagan. 2017. Interactive abstractive summarization for event news tweets. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 109–114, Copenhagen, Denmark. Association for Computational Linguistics.

Shengli Song, Haitao Huang, and Tongxiao Ruan. 2019. Abstractive text summarization using lstm-cnn based deep learning. *Multimedia Tools and Applications*, 78.

Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul Christiano. 2020. Learning to summarize from human feedback. *arXiv preprint arXiv:2009.01325*.

Sho Takase and Naoaki Okazaki. 2019. Positional encoding to control output sequence length. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3999–4004, Minneapolis, Minnesota. Association for Computational Linguistics.

Gaurav Trivedi, Phuong Pham, Wendy W. Chapman, Rebecca Hwa, Janyce Wiebe, and Harry Hochheiser. 2018. Nlpreviz: an interactive tool for natural language processing on clinical text. *J. Am. Medical Informatics Assoc.*, 25(1):81–87.

Shikhar Vashishth, Shyam Upadhyay, Gaurav Singh Tomar, and Manaal Faruqui. 2019. Attention interpretability across nlp tasks. *arXiv*.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

Jesse Vig and Yonatan Belinkov. 2019. Analyzing the structure of attention in a transformer language model. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 63–76, Florence, Italy. Association for Computational Linguistics.

Ashwin K Vijayakumar, Michael Cogswell, Ramprasath R Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. 2016. Diverse beam search: Decoding diverse solutions from neural sequence models. *arXiv preprint arXiv:1610.02424*.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter Liu. 2020. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR.