Effect of fine-tuning Beam Search and Length Penalty on Abstractive Text Summarization with PEGASUS

Jeyabarani Seenivasagam
UC Berkeley School of Information
ksjeyabarani@berkeley.edu

Jasdeep Basra
UC Berkeley School of Information
jbasra@berkeley.edu

Abstract

Abstractive-Summarization is the automatic generation of text to paraphrase the source text. It has an advantage over Extractive-Summarization methods because it can reduce redundancy and provide more coherent summaries similar to a summarization made by human. In this paper, we study the effect of the fine-tuning the beam search hyper-parameters during evaluation time using the recently released PEGASUS model for Abstractive Summarization (Zhang et al., 2019). We evaluate the model using ROUGE, BLEU, Length and Repetition metrics, as well as the Extractive Fragment Coverage and Density metrics as described by Grusky et al (Grusky et al., 2018).

1 Introduction

Automatic text summarization is generally categorized into two approaches abstractive and extractive. Extractive summaries identify the sentences that contain the main ideas and extract phrases from the original text. On the other hand, abstractive summaries result from the generation of new text that highlight and convey the main ideas.

Text summarization models for domain specific use are still not very common since most ”current research on abstractive summarization focuses primarily on the genre of news”, making it difficult to summarize other types of text such as social media posts (Syed et al., 2019). Additionally, now during the pandemic, there is an increasing need for models that can perform in biomedical contexts (Gigioli et al., 2018).

We use a large pre-trained model *PEGASUS* **Pre-training with Extracted Gap-sentences for Abstractive Summarization** released by Google(Liu and Zhao, 2020). We intend to study the effects of using domain specific fine-tuned models on cross-domain data-sets to see how to generalize an out-of-box model by changing the hyper-parameters at evaluation time. We have picked six of the same data-sets used in PEGASUS across different domains such as social media posts, biomedical articles, news articles, U.S. State Bills, and Emails.

We hypothesize that we could get a better quality text summary using a state-of-art pre-trained - fine-tuned architecture such as PEGASUS while setting the evaluation hyper-parameter on beam search and length-penalty to a bigger value.

2 Background

Traditionally, abstractive text summarization used Attentional Encoder-Decoder Recurrent Neural Networks. These sequence-to-sequence networks map input sequences to output sequences. While effective, they were difficult to parallelize across large amounts of data (Nallapati et al., 2016). Transformers provided an alternative approach by focusing on the attention mechanism or self-attention (Vaswani et al., 2017). Transformers are able to achieve ”state of the art performance” and have become ubquitous (Wu et al., 2020). Thus, Transformers have formed the basis of many large pre-trained language models such BERT, GPT, and others that can be used in a wide array of tasks and fine tuned via transfer learning. Pre-trained models such as BERT are of great value to researchers and enable more practical use cases for text summarizing but allowing very short training periods and moderate computing resources to achieve state-of-the-art performance in a wide array of tasks including text generation. Indeed, ”Pre-trained language models have recently emerged as a key technology for achieving impressive gains in a wide variety of natural language tasks” (Liu and Lapata, 2019). GPT-3 and PEGASUS have emerged recently as powerful pre-trained models. Though, GPT-3 is still not publicly available yet, which is why we focus on PEGASUS for our study.

PEGASUS, that is used in this study has reported state-of-art performance on several different scores for abstractive text summarization. PEGASUS uses a Transformer encoder-decoder model combined with a supervised objective called *gap-sentence generation* (Liu and Zhao, 2020). Their work primarily focuses on the pre-training objectives and they have presented their results on twelve different data-sets for text summarization. According to the Google team, prior models that used "self-supervised objectives in pre-training were somewhat agnostic to downstream applications." Since the model is trained specifically for the task of abstractive text summarization, we believe it will be more effective than other pre-trained models.

3 Methods

In our work, we focus on studying the effects of tuning the hyper-parameters during the evaluation time, namely beam search and length penalty (beam α).

Data Set and Models: We use vocabulary, pre-trained PEGASUS model (pegasus_ckpt) and fine-tuned models on domain-specific datasets: (CNN/Daily Mail News - cnn_dailymail, News newser.com - multi_news, Emails - aeslc, US State Bills - billsum, social media posts from reddit_tifu , a *subreddit* of the social media site Reddit - fu, and Scientific papers - pubmed) and their corresponding Tensorflow text-summarization data-sets for our experiments.

Design: We picked a set of 100 text articles from each domain specific data-set to evaluate. We then evaluate the test samples on:

- baseline (PEGASUS check-point)
- fine-tuned model on the same domain data-set (e.g. CNNDailyMail test data on fine-tuned CNN model checkpoint)
- fine-tuned model across domains (e.g. CNNDailyMail test data on fine-tuned multi-news model, CNNDailyMail test data on fine-tuned billsum model, etc...)

We repeat this experiment by changing the beam-size and beam α for the following combinations:

- Beam Size 1 and beam α 0.5
- Beam Size 5 and beam α 0.5
- Beam Size 10 and beam α 0.1
- Beam Size 10 and beam α 0.5
- Beam Size 10 and beam α 0.9

We investigate the effects of these values under different metrics for text summarization.

Evaluation: We evaluate our model with a robust method for automatic scoring of the results including ROUGE and other metrics which have been widely used to evaluate the quality of abstractive text summaries (Zopf et al., 2018). Comprehensively, we will use the following metrics for analysis:

- Recall scores, Precision score, f1-scores for:
  - ROUGE1
  - ROUGE2
  - ROUGEL
  - ROUGESum
- Coverage and Density
- Length score used to compare target and prediction lengths
- Repetition Evaluation of Generated Sequences

4 Results

Coverage and Density

We calculate the Extractive Fragment Coverage and Density metrics. The coverage measure is the extent to which a summary is derivative of a text. The density is how well the word sequence of a summary can be described as a series of extractions of the text.(Grusky et al., 2018)

In the general case, setting the evaluation hyper-parameters for beam search and length-penalty to higher values give the best results.

For models fine-tuned on a larger body of text such as in the news domain (cnn daily mail and multi news), government bills and published research articles, higher beam search value and higher length-penalty clearly works best in terms of coverage and density . The same is true when validating larger text on fine-tuned models on any domains.

On the other hand, for models fine-tuned on shorter text such as email (aesic) and reddit (reddit_tifu - short), shorter beam search value clearly performs better in terms of coverage when tested using data-sets across different domains.

This study shows that the hyper-parameters used for beam search and length-penalty during validation shows a correlation on the coverage metrics for shorter vs longer body of text used during fine-tuning. We suspect that the attention heads learn differently when trained on shorter vs longer length of text. This could be a future extension of research from this work.

ROUGE

Precision: Generally, the summarization task:

- performs better with higher beam search value keeping length-penalty a constant.

- performs better with lower length-penalty keeping beam search a constant

Recall:

- performs better with lower beam search value keeping length-penalty a constant

- performs better with higher length-penalty keeping beam search a constant

Repetition Evaluation of Generated Sequences

It is the ratio for n-gram based consecutive repeating sub-sequences. We consider the repetition score for predicted text. The lower the score, the better. Generally, the summarization task:

- performs better with higher beam search value keeping length-penalty a constant

- performs better with lower length-penalty keeping beam search a constant

Length scorer to compare targets and prediction lengths

char value of 10, word value of 5, and relative value of 0.1 are used.

Generally, the summarization task:

- performs better with lower beam search value keeping length-penalty a constant

- performs better with higher length-penalty keeping beam search a constant

The detailed results of our experiments and investigation are in the tables of Appendix 2.

5 Conclusion

We examined the impact of altering the hyper-paramters Beam Search and Length Penalty under various metrics including ROUGE, BLEU, Coverage and Density, Length and Repetition metrics. We notice that when the parameter Beam size increases, ROUGE-Precision, Repetition Score, and the Coverage and Density metrics all improve, while ROUGE-Recall and Length Score have diminished results. For the Length-Penalty beam α, as it increases, ROUGE-Recall, Length Score, and the Coverage and Density metrics improve, while ROUGE-Precision and Repetition Score have weaker results.

We would have liked to recruit human raters as an additional metric for a qualitative review, but due to time and resource constraints we cloud not. This could be used in further research to achieve greater confidence or variation in the results presented here. Further research could be done in other datasets and with a wider range of paramters.

References

P. Gigioli, N. Sagar, A. Rao, and J. Voyles. 2018. Domain-aware abstractive text summarization for medical documents. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2338–2343.

Max Grusky, Mor Naaman, and Yoav Artzi. 2018. Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies.

Yang Liu and Mirella Lapata. 2019. Text summarization with pretrained encoders.

Peter J. Liu and Yao Zhao. 2020. Pegasus: A state-of-the-art model for abstractive text summarization, Jun.

Ramesh Nallapati, Bowen Zhou, Cicero Nogueira dos santos, Caglar Gulcehre, and Bing Xiang. 2016. Abstractive text summarization using sequence-to-sequence rnns and beyond.

Shahbaz Syed, Michael Völske, Nedim Lipka, Benno Stein, Hinrich Schütze, and Martin Potthast. 2019. 12th international conference on natural language generation (inlg 2019). In *12th International Conference on Natural Language Generation (INLG 2019)*. Artificial Intelligence Research Center of Japan, Nov.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need.

Zhanghao Wu, Zhijian Liu, Ji Lin, Yujun Lin, and Song Han. 2020. Lite transformer with long-short range attention.

Jingqing Zhang, Yao Zhao, Mohammad Saleh, and Peter J. Liu. 2019. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization.

Markus Zopf, Eneldo Loza Mencía, and Johannes Fürnkranz. 2018. Which scores to predict in sentence regression for text 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 1782–1791.

Appendix 1 - Data sources and model check-points:

Data Set Sources:

1. https://www.tensorflow.org/datasets/catalog/cnn_dailymail
2. https://www.tensorflow.org/datasets/catalog/multi_news
3. https://www.tensorflow.org/datasets/catalog/aeslc
4. https://www.tensorflow.org/datasets/catalog/billsum
5. https://www.tensorflow.org/datasets/catalog/reddit_tifu
6. https://www.tensorflow.org/datasets/catalog/scientific_papers#scientific_paperspubmed

Model Checkpoints:

1. https://console.cloud.google.com/storage/browser/pegasus_ckpt
2. https://console.cloud.google.com/storage/browser/pegasus_ckpt/cnn_dailymail
3. https://console.cloud.google.com/storage/browser/pegasus_ckpt/multi_news
4. https://console.cloud.google.com/storage/browser/pegasus_ckpt/aeslc
5. https://console.cloud.google.com/storage/browser/pegasus_ckpt/billsum
6. https://console.cloud.google.com/storage/browser/pegasus_ckpt/reddit_tifu
7. https://console.cloud.google.com/storage/browser/pegasus_ckpt/pubmed

Appendix 2 - Tables and Figures:

Figure 1: ROUGE1: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (precision|recall|f1)	beamsearch_5_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.1 (precision|recall|f1)	beamsearch_10_α_0.9 (precision|recall|f1)
base	cnn	0.292 | 0.4269 | 0.3273	0.3064 | 0.4293 | 0.3402	0.3063 | 0.4378 | 0.3416	0.3208 | 0.4244 | 0.3433	0.2986 | 0.4419 | 0.3412
base	multi_news	0.2831 | 0.1881 | 0.2107	0.4732 | 0.2333 | 0.2952	0.4653 | 0.2272 | 0.2909	0.5161 | 0.2018 | 0.2718	0.4286 | 0.2471 | 0.298
base	billsum	0.2124 | 0.2549 | 0.2106	0.3147 | 0.2305 | 0.2253	0.3301 | 0.2189 | 0.2218	0.4428 | 0.1722 | 0.2219	0.2403 | 0.2544 | 0.2209
base	pubmed	0.2893 | 0.2197 | 0.2269	0.3743 | 0.2054 | 0.2459	0.398 | 0.2057 | 0.2513	0.442 | 0.159 | 0.2176	0.33 | 0.2531 | 0.2665
base	reddit_tifu	0.1176 | 0.3542 | 0.1657	0.1194 | 0.3523 | 0.1685	0.1212 | 0.3551 | 0.171	0.1226 | 0.3423 | 0.1703	0.1185 | 0.362 | 0.1689
base	aeslc	0.0632 | 0.2708 | 0.0959	0.0573 | 0.2204 | 0.0858	0.0559 | 0.2078 | 0.0824	0.0532 | 0.1982 | 0.0786	0.0555 | 0.2073 | 0.0822
cnn	cnn	0.4279 | 0.482 | 0.4424	0.4794 | 0.3873 | 0.4158	0.4522 | 0.3831 | 0.4033	0.5217 | 0.3213 | 0.3801	0.4191 | 0.4785 | 0.4304
cnn	multi_news	0.6111 | 0.1778 | 0.2689	0.6028 | 0.1478 | 0.2312	0.5816 | 0.1427 | 0.2238	0.622 | 0.1009 | 0.1694	0.5385 | 0.2117 | 0.2956
cnn	billsum	0.5572 | 0.2737 | 0.3277	0.6446 | 0.1333 | 0.1976	0.6565 | 0.1275 | 0.193	0.672 | 0.0958 | 0.1553	0.5224 | 0.2703 | 0.3047
cnn	pubmed	0.5283 | 0.2112 | 0.2902	0.5348 | 0.1129 | 0.1763	0.5443 | 0.1101 | 0.1718	0.5845 | 0.08 | 0.1351	0.506 | 0.2045 | 0.2754
cnn	reddit_tifu	0.1371 | 0.4096 | 0.1904	0.1609 | 0.2742 | 0.1857	0.1641 | 0.2587 | 0.1812	0.1855 | 0.1997 | 0.1756	0.1022 | 0.4268 | 0.159
cnn	aeslc	0.0827 | 0.4454 | 0.1346	0.0878 | 0.3251 | 0.1317	0.0761 | 0.294 | 0.1151	0.0859 | 0.2608 | 0.1205	0.0835 | 0.3541 | 0.1279
multi_news	cnn	0.2383 | 0.5061 | 0.3162	0.2578 | 0.5198 | 0.3369	0.249 | 0.494 | 0.3223	0.2507 | 0.4948 | 0.3238	0.2443 | 0.4998 | 0.3205
multi_news	multi_news	0.4909 | 0.4551 | 0.4656	0.5122 | 0.4361 | 0.4613	0.5213 | 0.4282 | 0.4603	0.5343 | 0.4194 | 0.4594	0.4985 | 0.4569 | 0.4679
multi_news	billsum	0.234 | 0.3869 | 0.2656	0.2542 | 0.4047 | 0.2806	0.2706 | 0.4101 | 0.2932	0.2713 | 0.4085 | 0.293	0.2625 | 0.4155 | 0.2924
multi_news	pubmed	0.2336 | 0.3573 | 0.272	0.2454 | 0.3629 | 0.2822	0.2464 | 0.3651 | 0.2852	0.2467 | 0.3648 | 0.2852	0.2432 | 0.3674 | 0.2833
multi_news	reddit_tifu	0.0828 | 0.4304 | 0.1345	0.0857 | 0.4272 | 0.1383	0.0888 | 0.4339 | 0.1421	0.0895 | 0.4344 | 0.1431	0.0882 | 0.436 | 0.1418
multi_news	aeslc	0.0225 | 0.1246 | 0.0372	0.0274 | 0.1565 | 0.045	0.0334 | 0.1794 | 0.0548	0.0333 | 0.1793 | 0.0546	0.0337 | 0.1823 | 0.0555
billsum	cnn	0.2839 | 0.3228 | 0.2821	0.3167 | 0.2624 | 0.2653	0.3234 | 0.243 | 0.257	0.313 | 0.2161 | 0.2396	0.2692 | 0.3135 | 0.2695
billsum	multi_news	0.4422 | 0.1935 | 0.2473	0.493 | 0.1365 | 0.1988	0.5041 | 0.1189 | 0.1805	0.5017 | 0.1077 | 0.1667	0.4067 | 0.1979 | 0.2363
billsum	billsum	0.566 | 0.6896 | 0.5817	0.6798 | 0.6213 | 0.5935	0.7249 | 0.5577 | 0.55	0.7491 | 0.4885 | 0.4968	0.673 | 0.6269 | 0.5816
billsum	pubmed	0.3817 | 0.1949 | 0.2341	0.3959 | 0.159 | 0.2083	0.3917 | 0.1585 | 0.2072	0.3936 | 0.152 | 0.2026	0.3218 | 0.2063 | 0.2238
billsum	reddit_tifu	0.1162 | 0.1834 | 0.1234	0.135 | 0.1607 | 0.1291	0.1493 | 0.1473 | 0.1301	0.1554 | 0.1415 | 0.1304	0.103 | 0.1977 | 0.114
billsum	aeslc	0.075 | 0.304 | 0.1117	0.0792 | 0.2793 | 0.1153	0.0832 | 0.2774 | 0.1186	0.0835 | 0.275 | 0.1178	0.0817 | 0.2826 | 0.1154
pubmed	cnn	0.1502 | 0.247 | 0.1779	0.188 | 0.2011 | 0.1808	0.1882 | 0.1975 | 0.1802	0.2025 | 0.1976 | 0.1877	0.1521 | 0.2183 | 0.1701
pubmed	multi_news	0.2955 | 0.1915 | 0.2205	0.3313 | 0.1319 | 0.1734	0.331 | 0.1238 | 0.1657	0.3816 | 0.1102 | 0.1621	0.2161 | 0.178 | 0.1879
pubmed	billsum	0.2697 | 0.29 | 0.2523	0.3054 | 0.2517 | 0.2459	0.2933 | 0.2312 | 0.2331	0.3137 | 0.2237 | 0.2365	0.2341 | 0.2598 | 0.2248
pubmed	pubmed	0.4633 | 0.4768 | 0.4571	0.4634 | 0.4398 | 0.4405	0.4649 | 0.4303 | 0.4335	0.4795 | 0.4082 | 0.4236	0.4382 | 0.4662 | 0.4419
pubmed	reddit_tifu	0.0963 | 0.3327 | 0.1418	0.1202 | 0.2847 | 0.1528	0.1191 | 0.297 | 0.1563	0.1355 | 0.2721 | 0.1678	0.0985 | 0.3696 | 0.1459
pubmed	aeslc	0.0189 | 0.0929 | 0.0305	0.0209 | 0.106 | 0.0338	0.0167 | 0.085 | 0.0266	0.0165 | 0.086 | 0.0265	0.0166 | 0.085 | 0.0265
reddit_tifu	cnn	0.4266 | 0.201 | 0.2632	0.3748 | 0.2135 | 0.2517	0.3789 | 0.2328 | 0.2681	0.27 | 0.1389 | 0.1707	0.3281 | 0.3183 | 0.2943
reddit_tifu	multi_news	0.2552 | 0.0351 | 0.0593	0.1533 | 0.0212 | 0.0364	0.1601 | 0.0236 | 0.0403	0.0821 | 0.0102 | 0.0178	N/A
reddit_tifu	billsum	0.1816 | 0.0792 | 0.0924	0.2306 | 0.0425 | 0.0611	0.1725 | 0.03 | 0.0421	0.3353 | 0.0351 | 0.0593	0.0931 | 0.0375 | 0.034
reddit_tifu	pubmed	0.567 | 0.1398 | 0.1954	0.4914 | 0.1212 | 0.1702	0.5211 | 0.1305 | 0.184	0.568 | 0.0926 | 0.1514	0.2614 | 0.2653 | 0.2482
reddit_tifu	reddit_tifu	0.3415 | 0.3077 | 0.2967	0.3538 | 0.2584 | 0.2719	0.3544 | 0.246 | 0.2628	0.3554 | 0.234 | 0.255	0.2783 | 0.299 | 0.2373
reddit_tifu	aeslc	0.054 | 0.1886 | 0.0756	0.032 | 0.0794 | 0.04	0.0324 | 0.0863 | 0.0428	0.0284 | 0.0616 | 0.0353	0.0597 | 0.1824 | 0.0816
aeslc	cnn	0.4446 | 0.0791 | 0.124	0.4933 | 0.076 | 0.1197	0.5013 | 0.0694 | 0.1095	0.4662 | 0.0377 | 0.0648	0.481 | 0.0998 | 0.143
aeslc	multi_news	0.5961 | 0.0241 | 0.0453	0.5278 | 0.0146 | 0.0279	0.5064 | 0.0131 | 0.0251	0.4877 | 0.0102 | 0.0196	0.4709 | 0.0238 | 0.0379
aeslc	billsum	0.4641 | 0.0387 | 0.067	0.5498 | 0.0278 | 0.0494	0.4985 | 0.0251 | 0.0444	0.53 | 0.0224 | 0.0394	0.4746 | 0.0339 | 0.0535
aeslc	pubmed	0.5206 | 0.0244 | 0.0459	0.508 | 0.0162 | 0.0302	0.5413 | 0.0144 | 0.0277	0.504 | 0.0126 | 0.0243	0.4399 | 0.0384 | 0.0422
aeslc	reddit_tifu	0.0925 | 0.0147 | 0.0245	0.0983 | 0.0127 | 0.0218	0.0867 | 0.0108 | 0.0185	0.07 | 0.0076 | 0.0133	0.0891 | 0.0171 | 0.0259
aeslc	aeslc	0.4473 | 0.397 | 0.391	0.4916 | 0.4244 | 0.4216	0.4738 | 0.4144 | 0.4117	0.4724 | 0.3906 | 0.3957	0.47 | 0.4213 | 0.4138

Figure 2: ROUGE2: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (precision|recall|f1)	beamsearch_5_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.1 (precision|recall|f1)	beamsearch_10_α_0.9 (precision|recall|f1)
base	cnn	0.1165 | 0.1631 | 0.1266	0.1241 | 0.1679 | 0.1337	0.1271 | 0.1746 | 0.1375	0.1312 | 0.1702 | 0.1379	0.1219 | 0.175 | 0.1355
base	multi_news	0.0824 | 0.0482 | 0.0573	0.1499 | 0.0715 | 0.0919	0.1484 | 0.0709 | 0.0918	0.1632 | 0.0638 | 0.0861	0.1342 | 0.0732 | 0.0904
base	billsum	0.0493 | 0.0597 | 0.0487	0.0996 | 0.0634 | 0.0655	0.1025 | 0.0649 | 0.0653	0.1586 | 0.06 | 0.078	0.0711 | 0.0708 | 0.0614
base	pubmed	0.0624 | 0.047 | 0.0486	0.103 | 0.0532 | 0.0653	0.1143 | 0.0543 | 0.0683	0.121 | 0.043 | 0.0592	0.0914 | 0.0633 | 0.0701
base	reddit_tifu	0.0248 | 0.0752 | 0.0358	0.0252 | 0.0735 | 0.0362	0.0252 | 0.0724 | 0.0361	0.0249 | 0.0687 | 0.0354	0.0246 | 0.0729 | 0.0356
base	aeslc	0.0213 | 0.104 | 0.0344	0.0163 | 0.0782 | 0.0261	0.0167 | 0.077 | 0.0265	0.0147 | 0.0676 | 0.0234	0.0168 | 0.0765 | 0.0265
cnn	cnn	0.1919 | 0.2167 | 0.1986	0.2227 | 0.1809 | 0.1939	0.2045 | 0.1737 | 0.1827	0.2601 | 0.1571 | 0.186	0.1928 | 0.2165 | 0.1968
cnn	multi_news	0.1988 | 0.0581 | 0.0876	0.2126 | 0.0511 | 0.0802	0.1947 | 0.0466 | 0.0732	0.2244 | 0.0342 | 0.0578	0.1696 | 0.0684 | 0.0947
cnn	billsum	0.2418 | 0.1137 | 0.1379	0.38 | 0.0673 | 0.1025	0.4015 | 0.0672 | 0.1046	0.4197 | 0.054 | 0.088	0.2566 | 0.1253 | 0.1404
cnn	pubmed	0.1946 | 0.0744 | 0.1043	0.219 | 0.0429 | 0.0685	0.226 | 0.0408 | 0.0665	0.2514 | 0.0328 | 0.0559	0.1934 | 0.0726 | 0.1006
cnn	reddit_tifu	0.0323 | 0.0893 | 0.044	0.0355 | 0.0577 | 0.0405	0.0367 | 0.0545 | 0.0395	0.0425 | 0.0467 | 0.041	0.0212 | 0.0849 | 0.0328
cnn	aeslc	0.0289 | 0.1783 | 0.0476	0.0248 | 0.1168 | 0.0387	0.0209 | 0.0981 | 0.0325	0.0313 | 0.1056 | 0.0436	0.0254 | 0.1204 | 0.0395
multi_news	cnn	0.0789 | 0.1679 | 0.1047	0.0859 | 0.1761 | 0.1127	0.0802 | 0.1633 | 0.1047	0.0811 | 0.1642 | 0.1055	0.0791 | 0.1662 | 0.1046
multi_news	multi_news	0.1753 | 0.1651 | 0.1676	0.1945 | 0.1686 | 0.1771	0.2028 | 0.1637 | 0.1775	0.2084 | 0.1622 | 0.1784	0.1935 | 0.1753 | 0.1805
multi_news	billsum	0.0406 | 0.0849 | 0.0486	0.0569 | 0.1179 | 0.0673	0.0712 | 0.1217 | 0.0796	0.0709 | 0.1214 | 0.0793	0.0707 | 0.1257 | 0.0811
multi_news	pubmed	0.0392 | 0.064 | 0.0466	0.0509 | 0.0778 | 0.059	0.0505 | 0.0757 | 0.0586	0.0509 | 0.0766 | 0.0589	0.0499 | 0.0769 | 0.0583
multi_news	reddit_tifu	0.017 | 0.0892 | 0.0275	0.0161 | 0.0793 | 0.0259	0.0172 | 0.0887 | 0.0279	0.0172 | 0.0876 | 0.0278	0.0171 | 0.0883 | 0.0279
multi_news	aeslc	0.0022 | 0.0146 | 0.0037	0.0029 | 0.0229 | 0.0049	0.0075 | 0.0493 | 0.0126	0.0076 | 0.0511 | 0.0127	0.0077 | 0.0508 | 0.013
billsum	cnn	0.0775 | 0.0888 | 0.0773	0.0974 | 0.0813 | 0.0823	0.1023 | 0.0764 | 0.0815	0.0953 | 0.065 | 0.0732	0.0864 | 0.0994 | 0.0853
billsum	multi_news	0.0963 | 0.0402 | 0.0527	0.1194 | 0.0339 | 0.0493	0.121 | 0.0278 | 0.043	0.1181 | 0.0247 | 0.0389	0.0958 | 0.0425 | 0.0531
billsum	billsum	0.3875 | 0.4763 | 0.397	0.5062 | 0.4436 | 0.4299	0.5601 | 0.4042 | 0.4056	0.5967 | 0.356 | 0.3658	0.5057 | 0.4503 | 0.4217
billsum	pubmed	0.0693 | 0.0311 | 0.0393	0.0792 | 0.0256 | 0.0356	0.0744 | 0.0242 | 0.0341	0.0714 | 0.0224 | 0.0318	0.0532 | 0.031 | 0.0353
billsum	reddit_tifu	0.0171 | 0.0179 | 0.0159	0.0204 | 0.0195 | 0.0167	0.022 | 0.0146 | 0.0161	0.0204 | 0.0128 | 0.0143	0.0146 | 0.0242 | 0.0157
billsum	aeslc	0.0266 | 0.1199 | 0.0389	0.0289 | 0.1108 | 0.042	0.0307 | 0.114 | 0.0437	0.0322 | 0.122 | 0.0457	0.031 | 0.1131 | 0.0437
pubmed	cnn	0.0206 | 0.0334 | 0.0243	0.0307 | 0.0296 | 0.028	0.0349 | 0.0316 | 0.0309	0.0364 | 0.0315 | 0.0321	0.0281 | 0.037 | 0.0299
pubmed	multi_news	0.0365 | 0.0236 | 0.027	0.0423 | 0.0154 | 0.0209	0.0472 | 0.0152 | 0.0215	0.054 | 0.0143 | 0.0217	0.0218 | 0.0189 | 0.0196
pubmed	billsum	0.043 | 0.0459 | 0.0398	0.05 | 0.0399 | 0.0398	0.046 | 0.036 | 0.0362	0.0492 | 0.0354 | 0.037	0.0359 | 0.0409 | 0.035
pubmed	pubmed	0.1965 | 0.1999 | 0.1928	0.1958 | 0.1866 | 0.1866	0.1951 | 0.1833 | 0.184	0.2005 | 0.1736 | 0.1799	0.1829 | 0.1962 | 0.1857
pubmed	reddit_tifu	0.019 | 0.0654 | 0.028	0.0262 | 0.0628 | 0.0332	0.0243 | 0.0615 | 0.0322	0.0255 | 0.0549 | 0.0322	0.0209 | 0.0775 | 0.0309
pubmed	aeslc	0.0034 | 0.0167 | 0.0056	0.0047 | 0.03 | 0.008	0.0033 | 0.0183 | 0.0055	0.0032 | 0.0183 | 0.0053	0.0033 | 0.0183 | 0.0056
reddit_tifu	cnn	0.1763 | 0.0803 | 0.107	0.1583 | 0.0863 | 0.1028	0.1641 | 0.0962 | 0.1121	0.1131 | 0.0583 | 0.0708	0.1426 | 0.1252 | 0.1212
reddit_tifu	multi_news	0.0803 | 0.01 | 0.0175	0.0609 | 0.0082 | 0.0141	0.0606 | 0.0084 | 0.0144	0.0268 | 0.0031 | 0.0055	N/A
reddit_tifu	billsum	0.0695 | 0.0261 | 0.032	0.1149 | 0.017 | 0.0256	0.0607 | 0.0116 | 0.0168	0.1821 | 0.0199 | 0.0331	0.0132 | 0.0119 | 0.0101
reddit_tifu	pubmed	0.2 | 0.0431 | 0.0621	0.1957 | 0.0404 | 0.0601	0.2286 | 0.045 | 0.0685	0.2353 | 0.0353 | 0.0591	0.0818 | 0.0814 | 0.0772
reddit_tifu	reddit_tifu	0.1176 | 0.099 | 0.0991	0.1405 | 0.0922 | 0.101	0.1413 | 0.0881 | 0.0985	0.1461 | 0.0858 | 0.098	0.1053 | 0.1004 | 0.0831
reddit_tifu	aeslc	0.0104 | 0.0431 | 0.0163	0.0106 | 0.033 | 0.0135	0.0164 | 0.0433 | 0.0205	0.0137 | 0.03 | 0.0159	0.0234 | 0.072 | 0.0314
aeslc	cnn	0.1428 | 0.02 | 0.0317	0.201 | 0.0304 | 0.0471	0.2143 | 0.027 | 0.0422	0.1904 | 0.0134 | 0.0226	0.1903 | 0.0364 | 0.053
aeslc	multi_news	0.2035 | 0.0076 | 0.0142	0.21 | 0.0048 | 0.0092	0.2035 | 0.004 | 0.0076	0.186 | 0.0028 | 0.0054	0.173 | 0.0069 | 0.0114
aeslc	billsum	0.3015 | 0.0224 | 0.0389	0.3525 | 0.0179 | 0.0321	0.3278 | 0.0168 | 0.0301	0.29 | 0.0143 | 0.0254	0.3339 | 0.0202 | 0.0338
aeslc	pubmed	0.1146 | 0.0047 | 0.0089	0.1226 | 0.004 | 0.0076	0.1264 | 0.0039 | 0.0076	0.1119 | 0.0034 | 0.0065	0.1143 | 0.0085 | 0.01
aeslc	reddit_tifu	0.0183 | 0.0016 | 0.0029	0.055 | 0.0037 | 0.0069	0.045 | 0.0025 | 0.0048	0.03 | 0.0017 | 0.0031	0.035 | 0.0031 | 0.0052
aeslc	aeslc	0.2631 | 0.2242 | 0.2256	0.2848 | 0.2466 | 0.2456	0.2877 | 0.2456 | 0.2461	0.2763 | 0.2268 | 0.2319	0.2911 | 0.2503 | 0.2496

Figure 3: ROUGEL: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (precision|recall|f1)	beamsearch_5_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.1 (precision|recall|f1)	beamsearch_10_α_0.9 (precision|recall|f1)
base	cnn	0.1945 | 0.2827 | 0.2166	0.2014 | 0.2829 | 0.2234	0.2029 | 0.2886 | 0.2247	0.2157 | 0.2844 | 0.2299	0.1981 | 0.2925 | 0.2256
base	multi_news	0.1726 | 0.1156 | 0.1283	0.2766 | 0.1355 | 0.1712	0.2675 | 0.1296 | 0.1658	0.3035 | 0.1146 | 0.1555	0.2427 | 0.1408 | 0.1686
base	billsum	0.1612 | 0.2017 | 0.1615	0.2283 | 0.1781 | 0.1666	0.2385 | 0.1671 | 0.1619	0.3164 | 0.1252 | 0.1592	0.1803 | 0.2013 | 0.1682
base	pubmed	0.1948 | 0.154 | 0.156	0.2488 | 0.1353 | 0.1617	0.2631 | 0.1335 | 0.1632	0.2968 | 0.1036 | 0.1427	0.214 | 0.1651 | 0.1719
base	reddit_tifu	0.0829 | 0.2541 | 0.1169	0.0848 | 0.253 | 0.1191	0.0857 | 0.2553 | 0.1207	0.0874 | 0.2475 | 0.1209	0.0846 | 0.2612 | 0.1202
base	aeslc	0.055 | 0.2434 | 0.0843	0.0499 | 0.1984 | 0.0749	0.0495 | 0.1914 | 0.0734	0.0469 | 0.1821 | 0.07	0.0486 | 0.1894 | 0.0724
cnn	cnn	0.302 | 0.3409 | 0.3123	0.3452 | 0.2782 | 0.2989	0.3195 | 0.2726 | 0.2865	0.3927 | 0.2393 | 0.2832	0.2849 | 0.3239 | 0.292
cnn	multi_news	0.3618 | 0.1036 | 0.1577	0.3845 | 0.0914 | 0.1436	0.3632 | 0.0866 | 0.1365	0.4172 | 0.0657 | 0.1109	0.3135 | 0.1211 | 0.17
cnn	billsum	0.3899 | 0.1873 | 0.2257	0.4881 | 0.1018 | 0.1512	0.4947 | 0.098 | 0.147	0.528 | 0.0763 | 0.1228	0.3729 | 0.1933 | 0.2148
cnn	pubmed	0.3419 | 0.1349 | 0.1864	0.4029 | 0.083 | 0.1305	0.4167 | 0.082 | 0.1298	0.4632 | 0.0622 | 0.1055	0.3306 | 0.1287 | 0.1752
cnn	reddit_tifu	0.0995 | 0.304 | 0.1387	0.1163 | 0.1993 | 0.134	0.1219 | 0.1946 | 0.1357	0.1475 | 0.1568 | 0.1394	0.0712 | 0.3085 | 0.1114
cnn	aeslc	0.0737 | 0.4036 | 0.1204	0.0709 | 0.2736 | 0.107	0.0613 | 0.2468 | 0.093	0.0745 | 0.2309 | 0.1047	0.0682 | 0.3086 | 0.1062
multi_news	cnn	0.1416 | 0.3071 | 0.1891	0.154 | 0.3189 | 0.2028	0.1477 | 0.3009 | 0.1932	0.1494 | 0.3015 | 0.1944	0.1437 | 0.3026 | 0.1906
multi_news	multi_news	0.2495 | 0.2325 | 0.2371	0.2683 | 0.2276 | 0.2415	0.2783 | 0.2243 | 0.2429	0.2859 | 0.2203 | 0.243	0.2609 | 0.2361 | 0.2433
multi_news	billsum	0.1289 | 0.2289 | 0.1488	0.146 | 0.2494 | 0.1635	0.1552 | 0.2532 | 0.1708	0.1556 | 0.2518 | 0.1705	0.1506 | 0.259 | 0.1709
multi_news	pubmed	0.1235 | 0.1949 | 0.145	0.1317 | 0.1996 | 0.1525	0.1345 | 0.2033 | 0.1567	0.1344 | 0.2022 | 0.1563	0.1333 | 0.2053 | 0.1561
multi_news	reddit_tifu	0.0582 | 0.3149 | 0.0953	0.057 | 0.2995 | 0.0928	0.0597 | 0.3093 | 0.0966	0.0599 | 0.3077 | 0.097	0.0591 | 0.3097 | 0.0962
multi_news	aeslc	0.0199 | 0.1107 | 0.0328	0.024 | 0.1425 | 0.0396	0.0303 | 0.1666 | 0.0498	0.0307 | 0.1682 | 0.0503	0.0303 | 0.1663 | 0.0498
billsum	cnn	0.186 | 0.2128 | 0.1855	0.2093 | 0.1708 | 0.1738	0.2189 | 0.1643 | 0.1741	0.2154 | 0.1478 | 0.1642	0.1863 | 0.2154 | 0.1859
billsum	multi_news	0.2567 | 0.1055 | 0.1364	0.3105 | 0.0809 | 0.1195	0.325 | 0.0725 | 0.1114	0.3315 | 0.067 | 0.1048	0.2431 | 0.1175 | 0.1397
billsum	billsum	0.4385 | 0.5405 | 0.4508	0.5742 | 0.5126 | 0.492	0.6247 | 0.4658 | 0.4627	0.6604 | 0.4119 | 0.4207	0.5573 | 0.5055 | 0.4691
billsum	pubmed	0.2433 | 0.1236 | 0.1476	0.2614 | 0.1027 | 0.1346	0.2549 | 0.1021 | 0.1336	0.2563 | 0.0966 | 0.1294	0.2083 | 0.1353 | 0.145
billsum	reddit_tifu	0.0923 | 0.1449 | 0.0966	0.1081 | 0.1309 | 0.1031	0.1188 | 0.1189 | 0.1037	0.1229 | 0.1137 | 0.103	0.0806 | 0.164 | 0.0919
billsum	aeslc	0.067 | 0.2759 | 0.1	0.0687 | 0.2483 | 0.1007	0.0728 | 0.2501 | 0.1041	0.074 | 0.25 | 0.1051	0.0717 | 0.2551 | 0.1017
pubmed	cnn	0.098 | 0.1674 | 0.118	0.123 | 0.1348 | 0.1199	0.1278 | 0.1363 | 0.1232	0.1372 | 0.1338 | 0.1274	0.1071 | 0.1553 | 0.1204
pubmed	multi_news	0.1776 | 0.1177 | 0.134	0.2085 | 0.0838 | 0.1096	0.2142 | 0.0803 | 0.1069	0.2442 | 0.0691 | 0.1012	0.1432 | 0.1172 | 0.1236
pubmed	billsum	0.1677 | 0.1887 | 0.1588	0.2021 | 0.1731 | 0.1646	0.1956 | 0.1584 | 0.156	0.2072 | 0.1513 | 0.1562	0.1588 | 0.1826 | 0.1539
pubmed	pubmed	0.2867 | 0.2922 | 0.2812	0.2841 | 0.269 | 0.2698	0.2898 | 0.2668 | 0.2696	0.2973 | 0.2507 | 0.2613	0.2691 | 0.2843 | 0.2701
pubmed	reddit_tifu	0.0704 | 0.2517 | 0.1047	0.0904 | 0.2224 | 0.1165	0.0877 | 0.2278 | 0.1168	0.1009 | 0.2068 | 0.1257	0.0711 | 0.2823 | 0.1072
pubmed	aeslc	0.0182 | 0.0916 | 0.0293	0.0192 | 0.0979 | 0.0313	0.0158 | 0.0807 | 0.0252	0.0158 | 0.0812 | 0.0253	0.0155 | 0.0798 | 0.0248
reddit_tifu	cnn	0.3159 | 0.1477 | 0.1936	0.2705 | 0.1478 | 0.1765	0.2695 | 0.1602 | 0.1863	0.1946 | 0.0939 | 0.1173	0.227 | 0.2166 | 0.2008
reddit_tifu	multi_news	0.1751 | 0.0237 | 0.0401	0.1055 | 0.0143 | 0.0246	0.1083 | 0.0155 | 0.0266	0.0556 | 0.0066 | 0.0115	N/A
reddit_tifu	billsum	0.1482 | 0.0654 | 0.076	0.2057 | 0.0359 | 0.0518	0.1577 | 0.0261 | 0.037	0.3139 | 0.0327 | 0.0552	0.0874 | 0.0336 | 0.0305
reddit_tifu	pubmed	0.3952 | 0.0955 | 0.1326	0.3561 | 0.0836 | 0.1183	0.3898 | 0.0913 | 0.131	0.4376 | 0.0686 | 0.1129	0.174 | 0.1771 | 0.1656
reddit_tifu	reddit_tifu	0.2759 | 0.2515 | 0.2411	0.3017 | 0.2191 | 0.2291	0.3042 | 0.2121 | 0.2255	0.3111 | 0.2044 | 0.2223	0.2321 | 0.2463 | 0.1971
reddit_tifu	aeslc	0.0482 | 0.1734 | 0.0677	0.0323 | 0.0776 | 0.0392	0.0313 | 0.0823 | 0.0409	0.0283 | 0.0597 | 0.0352	0.056 | 0.1655 | 0.0758
aeslc	cnn	0.3818 | 0.063 | 0.0998	0.4382 | 0.0657 | 0.1036	0.458 | 0.0593 | 0.0942	0.437 | 0.034 | 0.0585	0.4146 | 0.0809 | 0.1166
aeslc	multi_news	0.5204 | 0.0196 | 0.0369	0.4853 | 0.0124 | 0.0238	0.4597 | 0.0109 | 0.0209	0.4559 | 0.0086 | 0.0167	0.4244 | 0.0189 | 0.0303
aeslc	billsum	0.4346 | 0.0365 | 0.0632	0.5411 | 0.0273 | 0.0485	0.4878 | 0.0245 | 0.0432	0.5235 | 0.022 | 0.0386	0.4586 | 0.0304 | 0.0492
aeslc	pubmed	0.4549 | 0.0208 | 0.0391	0.4602 | 0.0143 | 0.0267	0.4908 | 0.0129 | 0.0247	0.4656 | 0.0114 | 0.022	0.4041 | 0.0325 | 0.0362
aeslc	reddit_tifu	0.0884 | 0.014 | 0.0234	0.0983 | 0.0125 | 0.0214	0.085 | 0.0107 | 0.0182	0.0683 | 0.0075 | 0.0129	0.0889 | 0.0164 | 0.0254
aeslc	aeslc	0.433 | 0.3847 | 0.3799	0.4791 | 0.4136 | 0.4107	0.4593 | 0.4037 | 0.4007	0.4573 | 0.3823 | 0.3867	0.4537 | 0.409 | 0.4019

Figure 4: ROUGELSum: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (precision|recall|f1)	beamsearch_5_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.5 (precision|recall|f1)	beamsearch_10_α_0.1 (precision|recall|f1)	beamsearch_10_α_0.9 (precision|recall|f1)
base	cnn	0.2519 | 0.3636 | 0.2812	0.2636 | 0.3673 | 0.2914	0.2643 | 0.3748 | 0.2935	0.2784 | 0.367 | 0.2977	0.2573 | 0.3798 | 0.2929
base	multi_news	0.1598 | 0.1012 | 0.1147	0.2752 | 0.1367 | 0.1719	0.2689 | 0.1318 | 0.168	0.3044 | 0.1162 | 0.1571	0.2459 | 0.1439 | 0.1721
base	billsum	0.1237 | 0.1505 | 0.1214	0.2072 | 0.1432 | 0.1403	0.2165 | 0.1376 | 0.1385	0.3233 | 0.1265 | 0.1616	0.147 | 0.1537 | 0.1323
base	pubmed	0.2439 | 0.183 | 0.1909	0.3054 | 0.1633 | 0.197	0.3249 | 0.163 | 0.2009	0.3704 | 0.1303 | 0.1799	0.268 | 0.2018 | 0.2133
base	reddit_tifu	0.0838 | 0.2566 | 0.1179	0.0849 | 0.2531 | 0.1193	0.0861 | 0.2575 | 0.1215	0.0885 | 0.2495 | 0.1226	0.085 | 0.2629 | 0.1206
base	aeslc	0.0554 | 0.2453 | 0.0847	0.0508 | 0.2023 | 0.0765	0.0492 | 0.1906 | 0.0734	0.0474 | 0.1831 | 0.0705	0.0497 | 0.1912 | 0.074
cnn	cnn	0.397 | 0.4458 | 0.4092	0.4462 | 0.359 | 0.3858	0.4212 | 0.3556 | 0.3752	0.4889 | 0.2988 | 0.3547	0.3893 | 0.4452 | 0.4
cnn	multi_news	0.4293 | 0.1245 | 0.1885	0.4367 | 0.1056 | 0.1655	0.4149 | 0.101 | 0.1586	0.4534 | 0.0738 | 0.1238	0.3737 | 0.1454 | 0.2034
cnn	billsum	0.4562 | 0.2204 | 0.2671	0.5439 | 0.1124 | 0.1675	0.5566 | 0.1072 | 0.1632	0.5782 | 0.0805 | 0.1322	0.4394 | 0.2247 | 0.2537
cnn	pubmed	0.4801 | 0.1901 | 0.2625	0.4928 | 0.1035 | 0.1622	0.5026 | 0.1009 | 0.1583	0.5347 | 0.0721 | 0.1224	0.4591 | 0.183 | 0.2481
cnn	reddit_tifu	0.1125 | 0.3471 | 0.1577	0.1218 | 0.2131 | 0.1415	0.1268 | 0.2061 | 0.1418	0.1492 | 0.1602 | 0.1417	0.0815 | 0.352 | 0.1274
cnn	aeslc	0.0747 | 0.4081 | 0.1217	0.0723 | 0.2781 | 0.1092	0.0631 | 0.2535 | 0.0963	0.0751 | 0.2325 | 0.1056	0.0695 | 0.3108 | 0.1082
multi_news	cnn	0.1855 | 0.3923 | 0.2461	0.2039 | 0.4111 | 0.2664	0.1938 | 0.3867 | 0.2514	0.1949 | 0.3867 | 0.2522	0.1894 | 0.3902 | 0.2491
multi_news	multi_news	0.2498 | 0.233 | 0.2374	0.2687 | 0.2278 | 0.2415	0.2775 | 0.2234 | 0.2423	0.2857 | 0.2193 | 0.2426	0.2615 | 0.2363 | 0.2439
multi_news	billsum	0.1506 | 0.2551 | 0.1713	0.1703 | 0.2765 | 0.1877	0.1797 | 0.2785 | 0.195	0.1794 | 0.2779 | 0.1945	0.174 | 0.2847 | 0.1946
multi_news	pubmed	0.1837 | 0.2812 | 0.214	0.193 | 0.2825 | 0.2212	0.1957 | 0.2867 | 0.2259	0.1966 | 0.2863 | 0.226	0.194 | 0.2891 | 0.2252
multi_news	reddit_tifu	0.0583 | 0.3145 | 0.0952	0.0568 | 0.2999 | 0.0926	0.0597 | 0.3086 | 0.0967	0.0599 | 0.3072 | 0.0968	0.0592 | 0.3104 | 0.0963
multi_news	aeslc	0.0198 | 0.111 | 0.0328	0.0242 | 0.1412 | 0.0398	0.0308 | 0.1678 | 0.0507	0.0308 | 0.1682 | 0.0505	0.0306 | 0.1677 | 0.0501
billsum	cnn	0.2404 | 0.2732 | 0.2391	0.2747 | 0.2252 | 0.2295	0.2828 | 0.2105 | 0.2239	0.2721 | 0.1862 | 0.208	0.235 | 0.2732 | 0.235
billsum	multi_news	0.2678 | 0.1121 | 0.1437	0.3093 | 0.0815 | 0.1201	0.3264 | 0.0732 | 0.1124	0.333 | 0.0681 | 0.1062	0.2412 | 0.1154 | 0.1374
billsum	billsum	0.4807 | 0.5854 | 0.4922	0.6046 | 0.5419 | 0.5219	0.6523 | 0.4878 | 0.4848	0.6806 | 0.4289 | 0.438	0.5933 | 0.5397 | 0.503
billsum	pubmed	0.3293 | 0.1678 | 0.202	0.3469 | 0.1368 | 0.1799	0.3422 | 0.1359 | 0.1796	0.3401 | 0.1291 | 0.1734	0.2795 | 0.1777 | 0.1938
billsum	reddit_tifu	0.0931 | 0.1458 | 0.0975	0.1089 | 0.1317 | 0.1037	0.1192 | 0.1178 | 0.1031	0.1244 | 0.1132 | 0.1033	0.0805 | 0.1633 | 0.0916
billsum	aeslc	0.0676 | 0.2795 | 0.101	0.0685 | 0.2477 | 0.1007	0.0735 | 0.2519 | 0.1043	0.0747 | 0.251 | 0.1058	0.0718 | 0.2548 | 0.1023
pubmed	cnn	0.1343 | 0.2209 | 0.1588	0.1662 | 0.1792 | 0.1608	0.169 | 0.1783 | 0.1619	0.1807 | 0.1778 | 0.1683	0.1375 | 0.1985 | 0.1543
pubmed	multi_news	0.1722 | 0.1129 | 0.129	0.2072 | 0.0779 | 0.1039	0.2086 | 0.073 | 0.0993	0.2505 | 0.0693 | 0.1024	0.1202 | 0.0958 | 0.1013
pubmed	billsum	0.2026 | 0.221 | 0.1896	0.2229 | 0.184 | 0.1788	0.2155 | 0.1689 | 0.1694	0.2314 | 0.1659 | 0.1731	0.1631 | 0.1846 | 0.1572
pubmed	pubmed	0.4148 | 0.4233 | 0.4071	0.4126 | 0.3897 | 0.3916	0.4136 | 0.3805 | 0.3844	0.4269 | 0.3608 | 0.3756	0.3902 | 0.4132 | 0.3929
pubmed	reddit_tifu	0.0761 | 0.271 | 0.1134	0.0941 | 0.2303 | 0.1211	0.0892 | 0.2308 | 0.1185	0.1029 | 0.2124 | 0.1288	0.0752 | 0.2992 | 0.1132
pubmed	aeslc	0.0182 | 0.0907 | 0.0294	0.0195 | 0.1008 | 0.0318	0.0163 | 0.0845 | 0.026	0.0161 | 0.0848 | 0.0259	0.0164 | 0.0851 | 0.0262
reddit_tifu	cnn	0.3859 | 0.1786 | 0.2353	0.3338 | 0.185 | 0.2194	0.3316 | 0.1986 | 0.2312	0.2355 | 0.1186 | 0.1471	0.2882 | 0.2744 | 0.2559
reddit_tifu	multi_news	0.1737 | 0.0236 | 0.0398	0.1041 | 0.0142 | 0.0246	0.1081 | 0.0154 | 0.0264	0.0556 | 0.0066 | 0.0116	N/A
reddit_tifu	billsum	0.156 | 0.0693 | 0.08	0.2089 | 0.035 | 0.0511	0.1607 | 0.0278 | 0.0391	0.3126 | 0.0327 | 0.0553	0.0856 | 0.0329 | 0.0293
reddit_tifu	pubmed	0.4988 | 0.1197 | 0.1683	0.4281 | 0.1019 | 0.1444	0.4575 | 0.1104 | 0.1574	0.4986 | 0.0781 | 0.1286	0.224 | 0.225 | 0.2128
reddit_tifu	reddit_tifu	0.2765 | 0.251 | 0.2411	0.3003 | 0.2196 | 0.2292	0.3041 | 0.2119 | 0.2251	0.3113 | 0.2041 | 0.222	0.2326 | 0.246 | 0.1976
reddit_tifu	aeslc	0.0478 | 0.1726 | 0.0675	0.0319 | 0.0768 | 0.0392	0.0322 | 0.0828 | 0.0414	0.0288 | 0.061 | 0.0356	0.0547 | 0.1653 | 0.074
aeslc	cnn	0.4257 | 0.0742 | 0.1167	0.4777 | 0.0733 | 0.1151	0.4918 | 0.067 | 0.1054	0.4574 | 0.0362 | 0.0625	0.4633 | 0.0944 | 0.136
aeslc	multi_news	0.521 | 0.0195 | 0.0368	0.4849 | 0.0123 | 0.0236	0.4585 | 0.0109 | 0.0209	0.4534 | 0.0086 | 0.0167	0.4264 | 0.0182 | 0.0295
aeslc	billsum	0.4408 | 0.0368 | 0.0638	0.5452 | 0.0271 | 0.0481	0.4939 | 0.0249 | 0.0439	0.5252 | 0.0224 | 0.0392	0.4611 | 0.031 | 0.0499
aeslc	pubmed	0.4785 | 0.0218 | 0.041	0.4885 | 0.0153 | 0.0287	0.5163 | 0.0139 | 0.0268	0.4848 | 0.0122 | 0.0234	0.426 | 0.0352 | 0.0396
aeslc	reddit_tifu	0.0897 | 0.0141 | 0.0235	0.1 | 0.0129 | 0.0221	0.0867 | 0.0106 | 0.0183	0.0683 | 0.0076 | 0.0131	0.0886 | 0.0164 | 0.0252
aeslc	aeslc	0.4357 | 0.3881 | 0.3826	0.4778 | 0.4144 | 0.4117	0.4591 | 0.403 | 0.4005	0.458 | 0.3829 | 0.3873	0.4534 | 0.4093 | 0.4021

Figure 5: REGS1 (Repetition Evaluation of Generated Sequences): Comparison of validation results on cross-domain trained models

A	B	C	D	E	F	G
Model	**Testset**	**beamsearch_1_α_0.5 (target|predicted)**	**beamsearch_5_α_0.5 (target|predicted)**	**beamsearch_10_α_0.5 (target|predicted)**	**beamsearch_10_α_0.1 (target|predicted)**	**beamsearch_10_α_0.9 (target|predicted)**
base	cnn	0.2705 | 0.4295	0.27 | 0.4018	0.2694 | 0.396	0.2697 | 0.3941	0.27 | 0.4078
base	multi_news	0.3857 | 0.7612	0.385 | 0.5168	0.3851 | 0.4757	0.3856 | 0.3867	0.3849 | 0.5189
base	billsum	0.4525 | 0.9073	0.4539 | 0.8387	0.4526 | 0.8184	0.4533 | 0.6925	0.453 | 0.8972
base	pubmed	0.544 | 0.7875	0.5437 | 0.6696	0.5441 | 0.613	0.5434 | 0.4576	0.5441 | 0.7647
base	reddit_tifu	0.1291 | 0.371	0.1297 | 0.3665	0.1299 | 0.353	0.1294 | 0.3361	0.1293 | 0.3587
base	aeslc	0.0079 | 0.1026	0.0079 | 0.0893	0.0084 | 0.0859	0.0085 | 0.0839	0.0079 | 0.0871
cnn	cnn	0.2695 | 0.3213	0.2693 | 0.2275	0.2697 | 0.2453	0.2701 | 0.1817	0.2699 | 0.295
cnn	multi_news	0.385 | 0.3131	0.3853 | 0.2544	0.3849 | 0.2447	0.3853 | 0.1619	0.3852 | 0.3381
cnn	billsum	0.4529 | 0.5085	0.4543 | 0.2256	0.4528 | 0.2258	0.452 | 0.1318	0.4541 | 0.4645
cnn	pubmed	0.5433 | 0.4116	0.5441 | 0.2388	0.5443 | 0.2383	0.5435 | 0.1086	0.544 | 0.3831
cnn	reddit_tifu	0.1293 | 0.4173	0.1291 | 0.2625	0.1291 | 0.2413	0.1309 | 0.1237	0.1296 | 0.3997
cnn	aeslc	0.0078 | 0.184	0.0084 | 0.1007	0.0079 | 0.1016	0.0084 | 0.061	0.0079 | 0.1207
multi_news	cnn	0.2695 | 0.3814	0.2695 | 0.3535	0.2698 | 0.3463	0.2692 | 0.3467	0.2701 | 0.3513
multi_news	multi_news	0.3854 | 0.4704	0.385 | 0.425	0.3853 | 0.4173	0.3852 | 0.4104	0.3855 | 0.4345
multi_news	billsum	0.4543 | 0.6005	0.4541 | 0.5936	0.4522 | 0.5957	0.4536 | 0.5921	0.4527 | 0.6113
multi_news	pubmed	0.5439 | 0.5754	0.5441 | 0.5779	0.5432 | 0.598	0.5441 | 0.5975	0.5446 | 0.6047
multi_news	reddit_tifu	0.1303 | 0.4359	0.1298 | 0.4211	0.1294 | 0.4153	0.1302 | 0.4159	0.1291 | 0.4197
multi_news	aeslc	0.0078 | 0.1654	0.0084 | 0.1578	0.0084 | 0.1489	0.0084 | 0.1492	0.0079 | 0.1484
billsum	cnn	0.2689 | 0.404	0.27 | 0.3246	0.2701 | 0.3335	0.2699 | 0.2724	0.2694 | 0.5687
billsum	multi_news	0.3853 | 0.4522	0.385 | 0.402	0.3849 | 0.3589	0.385 | 0.3191	0.3856 | 0.6068
billsum	billsum	0.4523 | 0.5368	0.4523 | 0.4743	0.4522 | 0.4189	0.453 | 0.3692	0.4531 | 0.4643
billsum	pubmed	0.5437 | 0.467	0.544 | 0.4262	0.5442 | 0.432	0.5441 | 0.4203	0.544 | 0.6119
billsum	reddit_tifu	0.1286 | 0.3956	0.1303 | 0.3122	0.13 | 0.2878	0.1301 | 0.2438	0.1289 | 0.545
billsum	aeslc	0.0084 | 0.2412	0.0079 | 0.1592	0.0079 | 0.1413	0.0084 | 0.1375	0.0079 | 0.1597
pubmed	cnn	0.2698 | 0.6919	0.2697 | 0.627	0.2697 | 0.6031	0.2698 | 0.5814	0.2701 | 0.7033
pubmed	multi_news	0.3849 | 0.7748	0.3853 | 0.6904	0.3853 | 0.6616	0.3849 | 0.5699	0.3854 | 0.8761
pubmed	billsum	0.452 | 0.8053	0.4528 | 0.7863	0.4531 | 0.7965	0.4528 | 0.7716	0.4522 | 0.8804
pubmed	pubmed	0.5444 | 0.6132	0.5444 | 0.5591	0.5446 | 0.5469	0.5437 | 0.5263	0.5444 | 0.5675
pubmed	reddit_tifu	0.1297 | 0.5976	0.13 | 0.4888	0.1299 | 0.4892	0.1296 | 0.4252	0.1301 | 0.601
pubmed	aeslc	0.0079 | 0.3295	0.0079 | 0.2509	0.0079 | 0.2571	0.0079 | 0.259	0.0084 | 0.258
reddit_tifu	cnn	0.2699 | 0.1744	0.2688 | 0.1574	0.2693 | 0.155	0.2696 | 0.0754	0.269 | 0.4088
reddit_tifu	multi_news	0.3849 | 0.566	0.385 | 0.2095	0.3852 | 0.2432	0.3852 | 0.0339	N/A
reddit_tifu	billsum	0.453 | 0.7803	0.4544 | 0.6456	0.4538 | 0.7495	0.4535 | 0.1928	0.4538 | 0.9722
reddit_tifu	pubmed	0.5438 | 0.3746	0.5445 | 0.4049	0.5436 | 0.3765	0.5441 | 0.2422	0.544 | 0.8442
reddit_tifu	reddit_tifu	0.1307 | 0.1676	0.1296 | 0.1111	0.1298 | 0.0913	0.1295 | 0.084	0.1291 | 0.2595
reddit_tifu	aeslc	0.0078 | 0.1191	0.0084 | 0.068	0.0079 | 0.0763	0.0079 | 0.0138	0.0079 | 0.1941
aeslc	cnn	0.2702 | 0.0377	0.2699 | 0.0448	0.2694 | 0.0664	0.2693 | 0.0417	0.2695 | 0.1154
aeslc	multi_news	0.3851 | 0.0178	0.3856 | 0.0182	0.3852 | 0.0171	0.3844 | 0.0062	0.3856 | 0.0938
aeslc	billsum	0.4534 | 0.0255	0.4535 | 0.0055	0.4539 | 0.01	0.4531 | 0.0129	0.4537 | 0.0399
aeslc	pubmed	0.5433 | 0.0231	0.5436 | 0.0321	0.5443 | 0.0108	0.5445 | 0.01	0.5443 | 0.2039
aeslc	reddit_tifu	0.1296 | 0.011	0.1305 | 0.03	0.13 | 0.03	0.13 | 0.02	0.1289 | 0.0617
aeslc	aeslc	0.0079 | 0.0086	0.0084 | 0.0061	0.0079 | 0.0076	0.0078 | 0.0061	0.0079 | 0.0081

Figure 6: REGS2 (Repetition Evaluation of Generated Sequences): Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (target|predicted)	beamsearch_5_α_0.5 (target|predicted)	beamsearch_10_α_0.5 (target|predicted)	beamsearch_10_α_0.1 (target|predicted)	beamsearch_10_α_0.9 (target|predicted)
base	cnn	0.0169 | 0.2243	0.0172 | 0.1958	0.0172 | 0.1812	0.0173 | 0.1956	0.017 | 0.1891
base	multi_news	0.0491 | 0.6723	0.049 | 0.3255	0.049 | 0.2672	0.0489 | 0.1798	0.049 | 0.3113
base	billsum	0.1202 | 0.8501	0.1202 | 0.7568	0.1205 | 0.7456	0.1206 | 0.585	0.1199 | 0.854
base	pubmed	0.1354 | 0.6575	0.1356 | 0.481	0.136 | 0.3894	0.135 | 0.2136	0.1359 | 0.6026
base	reddit_tifu	0.0115 | 0.0977	0.0121 | 0.0813	0.0119 | 0.0709	0.0121 | 0.0654	0.012 | 0.0692
base	aeslc	0.0 | 0.0127	0.0 | 0.0199	0.0 | 0.0247	0.0 | 0.0242	0.0 | 0.0255
cnn	cnn	0.0172 | 0.0454	0.0171 | 0.0186	0.017 | 0.0168	0.017 | 0.0131	0.017 | 0.028
cnn	multi_news	0.0491 | 0.049	0.0491 | 0.0232	0.0491 | 0.0188	0.0491 | 0.0057	0.0489 | 0.063
cnn	billsum	0.1204 | 0.1971	0.1207 | 0.051	0.1203 | 0.056	0.1199 | 0.0157	0.1204 | 0.1556
cnn	pubmed	0.1357 | 0.1007	0.1356 | 0.0347	0.1355 | 0.0464	0.1354 | 0.0082	0.1352 | 0.081
cnn	reddit_tifu	0.0119 | 0.094	0.0116 | 0.0495	0.0119 | 0.0461	0.0121 | 0.0096	0.0121 | 0.0674
cnn	aeslc	0.0 | 0.0346	0.0 | 0.0107	0.0 | 0.0098	0.0 | 0.0067	0.0 | 0.0099
multi_news	cnn	0.017 | 0.0608	0.017 | 0.0563	0.0171 | 0.0572	0.0172 | 0.0562	0.0172 | 0.057
multi_news	multi_news	0.0491 | 0.118	0.0489 | 0.0979	0.0491 | 0.0902	0.0491 | 0.0873	0.049 | 0.105
multi_news	billsum	0.1203 | 0.2528	0.1196 | 0.2755	0.1195 | 0.2966	0.1199 | 0.2942	0.1209 | 0.3192
multi_news	pubmed	0.1354 | 0.2071	0.1355 | 0.2513	0.135 | 0.2756	0.1348 | 0.2755	0.1354 | 0.2839
multi_news	reddit_tifu	0.0122 | 0.0978	0.012 | 0.1234	0.0121 | 0.1316	0.0123 | 0.1304	0.0118 | 0.1351
multi_news	aeslc	0.0 | 0.0138	0.0 | 0.0198	0.0 | 0.0171	0.0 | 0.0171	0.0 | 0.0171
billsum	cnn	0.0169 | 0.1858	0.017 | 0.1647	0.0172 | 0.1874	0.0172 | 0.1203	0.0171 | 0.4326
billsum	multi_news	0.049 | 0.2051	0.0489 | 0.2104	0.0489 | 0.193	0.0487 | 0.1622	0.0491 | 0.4577
billsum	billsum	0.1196 | 0.1983	0.1199 | 0.1541	0.1198 | 0.124	0.1207 | 0.1075	0.1196 | 0.1574
billsum	pubmed	0.1353 | 0.1725	0.1347 | 0.1331	0.1358 | 0.1501	0.1355 | 0.1319	0.1352 | 0.3739
billsum	reddit_tifu	0.0119 | 0.1712	0.0121 | 0.1412	0.0122 | 0.1341	0.0122 | 0.0901	0.0118 | 0.3766
billsum	aeslc	0.0 | 0.0983	0.0 | 0.0531	0.0 | 0.0376	0.0 | 0.038	0.0 | 0.0394
pubmed	cnn	0.0171 | 0.4101	0.017 | 0.3702	0.017 | 0.3894	0.0173 | 0.35	0.0171 | 0.5057
pubmed	multi_news	0.049 | 0.5693	0.049 | 0.493	0.0491 | 0.4584	0.0489 | 0.3332	0.0489 | 0.7724
pubmed	billsum	0.12 | 0.5925	0.1205 | 0.5936	0.12 | 0.6132	0.1198 | 0.5634	0.1202 | 0.7483
pubmed	pubmed	0.1353 | 0.2322	0.1355 | 0.1561	0.1351 | 0.1459	0.1353 | 0.1345	0.1349 | 0.1674
pubmed	reddit_tifu	0.0118 | 0.285	0.0121 | 0.199	0.012 | 0.226	0.012 | 0.1535	0.012 | 0.3115
pubmed	aeslc	0.0 | 0.1002	0.0 | 0.0674	0.0 | 0.0854	0.0 | 0.0857	0.0 | 0.084
reddit_tifu	cnn	0.0168 | 0.045	0.017 | 0.0489	0.017 | 0.0426	0.0169 | 0.0175	0.0171 | 0.2917
reddit_tifu	multi_news	0.0492 | 0.5006	0.0491 | 0.1819	0.0489 | 0.2133	0.0493 | 0.02	N/A
reddit_tifu	billsum	0.1205 | 0.6902	0.1198 | 0.5229	0.1195 | 0.6547	0.1201 | 0.082	0.1205 | 0.9476
reddit_tifu	pubmed	0.1355 | 0.1956	0.1355 | 0.2592	0.1353 | 0.2149	0.1351 | 0.103	0.1356 | 0.7483
reddit_tifu	reddit_tifu	0.012 | 0.0278	0.0119 | 0.0206	0.0121 | 0.0075	0.0118 | 0.0076	0.012 | 0.1568
reddit_tifu	aeslc	0.0 | 0.0512	0.0 | 0.0412	0.0 | 0.0524	0.0 | 0.0	0.0 | 0.1229
aeslc	cnn	0.017 | 0.0194	0.0171 | 0.028	0.0171 | 0.038	0.0171 | 0.0222	0.017 | 0.088
aeslc	multi_news	0.049 | 0.001	0.0491 | 0.0099	0.049 | 0.01	0.049 | 0.0	0.049 | 0.0874
aeslc	billsum	0.12 | 0.0073	0.1198 | 0.0014	0.1197 | 0.01	0.1204 | 0.01	0.1203 | 0.0293
aeslc	pubmed	0.1355 | 0.009	0.1352 | 0.0162	0.1351 | 0.0	0.1356 | 0.0	0.1349 | 0.1872
aeslc	reddit_tifu	0.0118 | 0.006	0.0117 | 0.03	0.0119 | 0.03	0.0122 | 0.02	0.0118 | 0.06
aeslc	aeslc	0.0 | 0.0	0.0 | 0.0	0.0 | 0.0	0.0 | 0.0	0.0 | 0.0

Figure 7: REGS3 (Repetition Evaluation of Generated Sequences): Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (target|predicted)	beamsearch_5_α_0.5 (target|predicted)	beamsearch_10_α_0.5 (target|predicted)	beamsearch_10_α_0.1 (target|predicted)	beamsearch_10_α_0.9 (target|predicted)
base	cnn	0.0023 | 0.1851	0.0023 | 0.1619	0.0023 | 0.1476	0.0023 | 0.1599	0.0023 | 0.1494
base	multi_news	0.0086 | 0.6547	0.0085 | 0.2953	0.0086 | 0.2414	0.0085 | 0.1529	0.0085 | 0.2836
base	billsum	0.0498 | 0.8146	0.0497 | 0.7273	0.0494 | 0.7193	0.0496 | 0.5513	0.0493 | 0.8328
base	pubmed	0.0482 | 0.6247	0.0485 | 0.4413	0.0483 | 0.356	0.0481 | 0.1768	0.0482 | 0.5696
base	reddit_tifu	0.0035 | 0.0669	0.0034 | 0.0504	0.0033 | 0.0398	0.0033 | 0.0361	0.0031 | 0.0389
base	aeslc	0.0 | 0.0036	0.0 | 0.0123	0.0 | 0.0193	0.0 | 0.0213	0.0 | 0.0193
cnn	cnn	0.0023 | 0.0111	0.0021 | 0.0027	0.0023 | 0.0023	0.0023 | 0.0025	0.0023 | 0.0054
cnn	multi_news	0.0086 | 0.0146	0.0086 | 0.0057	0.0087 | 0.0031	0.0086 | 0.0004	0.0085 | 0.0333
cnn	billsum	0.0497 | 0.1215	0.0497 | 0.0269	0.0501 | 0.0313	0.0498 | 0.0079	0.0495 | 0.0938
cnn	pubmed	0.0485 | 0.0418	0.0483 | 0.0133	0.0485 | 0.0233	0.0484 | 0.0026	0.0479 | 0.037
cnn	reddit_tifu	0.0031 | 0.0473	0.0033 | 0.0274	0.0031 | 0.026	0.0035 | 0.0017	0.0031 | 0.0275
cnn	aeslc	0.0 | 0.0159	0.0 | 0.0024	0.0 | 0.0024	0.0 | 0.0014	0.0 | 0.0031
multi_news	cnn	0.0023 | 0.0159	0.0023 | 0.0205	0.0023 | 0.0193	0.0023 | 0.0187	0.0023 | 0.0196
multi_news	multi_news	0.0087 | 0.0518	0.0087 | 0.0445	0.0086 | 0.0391	0.0086 | 0.0377	0.0086 | 0.049
multi_news	billsum	0.0498 | 0.1423	0.0494 | 0.1798	0.0496 | 0.2103	0.0493 | 0.2068	0.0498 | 0.2297
multi_news	pubmed	0.0482 | 0.1013	0.048 | 0.1489	0.0479 | 0.1717	0.048 | 0.1709	0.0485 | 0.1817
multi_news	reddit_tifu	0.0035 | 0.0386	0.0031 | 0.0754	0.0035 | 0.0827	0.0035 | 0.0837	0.0035 | 0.0868
multi_news	aeslc	0.0 | 0.003	0.0 | 0.0124	0.0 | 0.0059	0.0 | 0.0053	0.0 | 0.0051
billsum	cnn	0.0022 | 0.1337	0.0023 | 0.1344	0.0023 | 0.1563	0.0023 | 0.0922	0.0022 | 0.3902
billsum	multi_news	0.0086 | 0.142	0.0086 | 0.1705	0.0084 | 0.1557	0.0086 | 0.1281	0.0086 | 0.4145
billsum	billsum	0.0495 | 0.1037	0.0493 | 0.076	0.0495 | 0.0587	0.0494 | 0.0525	0.0499 | 0.0825
billsum	pubmed	0.0484 | 0.1179	0.048 | 0.0945	0.0487 | 0.1105	0.0479 | 0.0914	0.0484 | 0.316
billsum	reddit_tifu	0.0033 | 0.1209	0.0035 | 0.101	0.0031 | 0.1061	0.0035 | 0.0581	0.0031 | 0.3241
billsum	aeslc	0.0 | 0.059	0.0 | 0.0314	0.0 | 0.0197	0.0 | 0.0191	0.0 | 0.0202
pubmed	cnn	0.0023 | 0.3134	0.0023 | 0.2988	0.0022 | 0.3289	0.0023 | 0.285	0.0023 | 0.4483
pubmed	multi_news	0.0084 | 0.4831	0.0087 | 0.4387	0.0086 | 0.4045	0.0086 | 0.2716	0.0086 | 0.7282
pubmed	billsum	0.05 | 0.4848	0.0496 | 0.512	0.0493 | 0.5415	0.0495 | 0.4874	0.0495 | 0.6828
pubmed	pubmed	0.0479 | 0.1224	0.048 | 0.0725	0.0481 | 0.0658	0.0484 | 0.0598	0.0487 | 0.0814
pubmed	reddit_tifu	0.0031 | 0.1964	0.0034 | 0.1427	0.0031 | 0.1741	0.0035 | 0.1056	0.0031 | 0.2549
pubmed	aeslc	0.0 | 0.0451	0.0 | 0.0398	0.0 | 0.0533	0.0 | 0.0534	0.0 | 0.052
reddit_tifu	cnn	0.0024 | 0.0377	0.0023 | 0.0346	0.0023 | 0.0165	0.0023 | 0.0014	0.0023 | 0.2791
reddit_tifu	multi_news	0.0086 | 0.4986	0.0086 | 0.1797	0.0086 | 0.2103	0.0086 | 0.02	N/A
reddit_tifu	billsum	0.0498 | 0.6674	0.0495 | 0.5118	0.0494 | 0.6506	0.0498 | 0.0821	0.0492 | 0.9434
reddit_tifu	pubmed	0.0482 | 0.1605	0.0482 | 0.2362	0.0481 | 0.1943	0.0482 | 0.0734	0.0485 | 0.7254
reddit_tifu	reddit_tifu	0.0033 | 0.0126	0.0031 | 0.0168	0.0031 | 0.0012	0.0031 | 0.0012	0.0035 | 0.1459
reddit_tifu	aeslc	0.0 | 0.0377	0.0 | 0.04	0.0 | 0.05	0.0 | 0.0	0.0 | 0.1215
aeslc	cnn	0.0022 | 0.0192	0.0023 | 0.0269	0.0023 | 0.0369	0.0023 | 0.02	0.0022 | 0.085
aeslc	multi_news	0.0085 | 0.0	0.0086 | 0.0099	0.0088 | 0.01	0.0086 | 0.0	0.0085 | 0.0875
aeslc	billsum	0.0499 | 0.0029	0.0492 | 0.0	0.0493 | 0.01	0.0491 | 0.01	0.0495 | 0.0288
aeslc	pubmed	0.0482 | 0.0067	0.0482 | 0.0095	0.0483 | 0.0	0.0484 | 0.0	0.0484 | 0.1811
aeslc	reddit_tifu	0.0035 | 0.005	0.0031 | 0.02	0.0035 | 0.02	0.0033 | 0.01	0.0031 | 0.06
aeslc	aeslc	0.0 | 0.0	0.0 | 0.0	0.0 | 0.0	0.0 | 0.0	0.0 | 0.0

Figure 8: REGSTCR (Repetition Evaluation of Generated Sequences): Comparison of validation results on cross-domain trained models

A	B	C	D	E	F	G
Model	**Testset**	**beamsearch_1_α_0.5 (target|predicted)**	**beamsearch_5_α_0.5 (target|predicted)**	**beamsearch_10_α_0.5 (target|predicted)**	**beamsearch_10_α_0.1 (target|predicted)**	**beamsearch_10_α_0.9 (target|predicted)**
base	cnn	0.2705 | 0.4295	0.27 | 0.4018	0.2694 | 0.396	0.2697 | 0.3941	0.27 | 0.4078
base	multi_news	0.3857 | 0.7612	0.385 | 0.5168	0.3851 | 0.4757	0.3856 | 0.3867	0.3849 | 0.5189
base	billsum	0.4525 | 0.9073	0.4539 | 0.8387	0.4526 | 0.8184	0.4533 | 0.6925	0.453 | 0.8972
base	pubmed	0.544 | 0.7875	0.5437 | 0.6696	0.5441 | 0.613	0.5434 | 0.4576	0.5441 | 0.7647
base	reddit_tifu	0.1291 | 0.371	0.1297 | 0.3665	0.1299 | 0.353	0.1294 | 0.3361	0.1293 | 0.3587
base	aeslc	0.0079 | 0.1026	0.0079 | 0.0893	0.0084 | 0.0859	0.0085 | 0.0839	0.0079 | 0.0871
cnn	cnn	0.2695 | 0.3213	0.2693 | 0.2275	0.2697 | 0.2453	0.2701 | 0.1817	0.2699 | 0.295
cnn	multi_news	0.385 | 0.3131	0.3853 | 0.2544	0.3849 | 0.2447	0.3853 | 0.1619	0.3852 | 0.3381
cnn	billsum	0.4529 | 0.5085	0.4543 | 0.2256	0.4528 | 0.2258	0.452 | 0.1318	0.4541 | 0.4645
cnn	pubmed	0.5433 | 0.4116	0.5441 | 0.2388	0.5443 | 0.2383	0.5435 | 0.1086	0.544 | 0.3831
cnn	reddit_tifu	0.1293 | 0.4173	0.1291 | 0.2625	0.1291 | 0.2413	0.1309 | 0.1237	0.1296 | 0.3997
cnn	aeslc	0.0078 | 0.184	0.0084 | 0.1007	0.0079 | 0.1016	0.0084 | 0.061	0.0079 | 0.1207
multi_news	cnn	0.2695 | 0.3814	0.2695 | 0.3535	0.2698 | 0.3463	0.2692 | 0.3467	0.2701 | 0.3513
multi_news	multi_news	0.3854 | 0.4704	0.385 | 0.425	0.3853 | 0.4173	0.3852 | 0.4104	0.3855 | 0.4345
multi_news	billsum	0.4543 | 0.6005	0.4541 | 0.5936	0.4522 | 0.5957	0.4536 | 0.5921	0.4527 | 0.6113
multi_news	pubmed	0.5439 | 0.5754	0.5441 | 0.5779	0.5432 | 0.598	0.5441 | 0.5975	0.5446 | 0.6047
multi_news	reddit_tifu	0.1303 | 0.4359	0.1298 | 0.4211	0.1294 | 0.4153	0.1302 | 0.4159	0.1291 | 0.4197
multi_news	aeslc	0.0078 | 0.1654	0.0084 | 0.1578	0.0084 | 0.1489	0.0084 | 0.1492	0.0079 | 0.1484
billsum	cnn	0.2689 | 0.404	0.27 | 0.3246	0.2701 | 0.3335	0.2699 | 0.2724	0.2694 | 0.5687
billsum	multi_news	0.3853 | 0.4522	0.385 | 0.402	0.3849 | 0.3589	0.385 | 0.3191	0.3856 | 0.6068
billsum	billsum	0.4523 | 0.5368	0.4523 | 0.4743	0.4522 | 0.4189	0.453 | 0.3692	0.4531 | 0.4643
billsum	pubmed	0.5437 | 0.467	0.544 | 0.4262	0.5442 | 0.432	0.5441 | 0.4203	0.544 | 0.6119
billsum	reddit_tifu	0.1286 | 0.3956	0.1303 | 0.3122	0.13 | 0.2878	0.1301 | 0.2438	0.1289 | 0.545
billsum	aeslc	0.0084 | 0.2412	0.0079 | 0.1592	0.0079 | 0.1413	0.0084 | 0.1375	0.0079 | 0.1597
pubmed	cnn	0.2698 | 0.6919	0.2697 | 0.627	0.2697 | 0.6031	0.2698 | 0.5814	0.2701 | 0.7033
pubmed	multi_news	0.3849 | 0.7748	0.3853 | 0.6904	0.3853 | 0.6616	0.3849 | 0.5699	0.3854 | 0.8761
pubmed	billsum	0.452 | 0.8053	0.4528 | 0.7863	0.4531 | 0.7965	0.4528 | 0.7716	0.4522 | 0.8804
pubmed	pubmed	0.5444 | 0.6132	0.5444 | 0.5591	0.5446 | 0.5469	0.5437 | 0.5263	0.5444 | 0.5675
pubmed	reddit_tifu	0.1297 | 0.5976	0.13 | 0.4888	0.1299 | 0.4892	0.1296 | 0.4252	0.1301 | 0.601
pubmed	aeslc	0.0079 | 0.3295	0.0079 | 0.2509	0.0079 | 0.2571	0.0079 | 0.259	0.0084 | 0.258
reddit_tifu	cnn	0.2699 | 0.1744	0.2688 | 0.1574	0.2693 | 0.155	0.2696 | 0.0754	0.269 | 0.4088
reddit_tifu	multi_news	0.3849 | 0.566	0.385 | 0.2095	0.3852 | 0.2432	0.3852 | 0.0339	N/A
reddit_tifu	billsum	0.453 | 0.7803	0.4544 | 0.6456	0.4538 | 0.7495	0.4535 | 0.1928	0.4538 | 0.9722
reddit_tifu	pubmed	0.5438 | 0.3746	0.5445 | 0.4049	0.5436 | 0.3765	0.5441 | 0.2422	0.544 | 0.8442
reddit_tifu	reddit_tifu	0.1307 | 0.1676	0.1296 | 0.1111	0.1298 | 0.0913	0.1295 | 0.084	0.1291 | 0.2595
reddit_tifu	aeslc	0.0078 | 0.1191	0.0084 | 0.068	0.0079 | 0.0763	0.0079 | 0.0138	0.0079 | 0.1941
aeslc	cnn	0.2702 | 0.0377	0.2699 | 0.0448	0.2694 | 0.0664	0.2693 | 0.0417	0.2695 | 0.1154
aeslc	multi_news	0.3851 | 0.0178	0.3856 | 0.0182	0.3852 | 0.0171	0.3844 | 0.0062	0.3856 | 0.0938
aeslc	billsum	0.4534 | 0.0255	0.4535 | 0.0055	0.4539 | 0.01	0.4531 | 0.0129	0.4537 | 0.0399
aeslc	pubmed	0.5433 | 0.0231	0.5436 | 0.0321	0.5443 | 0.0108	0.5445 | 0.01	0.5443 | 0.2039
aeslc	reddit_tifu	0.1296 | 0.011	0.1305 | 0.03	0.13 | 0.03	0.13 | 0.02	0.1289 | 0.0617
aeslc	aeslc	0.0079 | 0.0086	0.0084 | 0.0061	0.0079 | 0.0076	0.0078 | 0.0061	0.0079 | 0.0081

Figure 9: BLUE: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5	beamsearch_5_α_0.5	beamsearch_10_α_0.5	beamsearch_10_α_0.1	beamsearch_10_α_0.9
base	cnn	7.434	8.1764	8.3794	8.4174	8.215
base	multi_news	3.0505	3.8242	3.7505	2.84	4.1654
base	billsum	2.1748	2.7755	2.7239	2.64	2.9896
base	pubmed	2.8936	3.2114	3.076	1.9833	4.2745
base	reddit_tifu	2.2867	2.3226	2.3191	2.332	2.3029
base	aeslc	2.4254	2.2997	2.3694	2.2275	2.3575
cnn	cnn	14.425	13.5281	12.6422	10.9263	14.1885
cnn	multi_news	1.4331	0.8877	0.8849	0.2527	2.8845
cnn	billsum	4.807	1.7625	1.6234	1.0079	5.0149
cnn	pubmed	3.0847	0.8838	0.7238	0.4212	3.0419
cnn	reddit_tifu	2.5296	3.0526	2.9692	3.5678	1.844
cnn	aeslc	2.1683	2.4368	2.3011	3.0219	2.3166
multi_news	cnn	4.8318	5.5219	4.9069	4.9295	4.8171
multi_news	multi_news	13.1772	13.3808	12.9807	12.7873	13.9081
multi_news	billsum	2.4333	3.4515	4.6368	4.5674	4.7414
multi_news	pubmed	2.1454	2.9982	3.2318	3.2428	3.187
multi_news	reddit_tifu	1.2127	1.1168	1.1832	1.1702	1.1706
multi_news	aeslc	0.8117	0.8761	1.1026	1.1053	1.1108
billsum	cnn	4.4674	4.3165	4.2582	3.905	5.0225
billsum	multi_news	1.3953	0.9606	0.4871	0.4204	1.7915
billsum	billsum	30.7906	31.2473	27.8634	23.9812	31.5621
billsum	pubmed	1.3346	1.0063	0.8648	0.7406	1.4382
billsum	reddit_tifu	1.7808	2.0054	2.1638	2.122	1.6097
billsum	aeslc	2.4126	2.7003	2.9178	3.0232	2.718
pubmed	cnn	1.6289	2.0017	2.2508	2.344	2.0577
pubmed	multi_news	1.0348	0.5395	0.4521	0.3129	1.0146
pubmed	billsum	1.5974	1.6522	1.4588	1.446	1.5245
pubmed	pubmed	15.7253	15.6704	15.2355	14.5062	15.6786
pubmed	reddit_tifu	1.6866	2.1389	2.2229	2.5322	1.583
pubmed	aeslc	1.0905	1.1474	1.1467	1.1387	1.1413
reddit_tifu	cnn	4.4702	4.6138	5.5171	3.2821	6.5069
reddit_tifu	multi_news	0.124	0.0486	0.0495	0.0111	N/A
reddit_tifu	billsum	1.0198	0.6335	0.5655	0.3296	0.7137
reddit_tifu	pubmed	1.0674	1.2786	1.4398	0.651	5.9314
reddit_tifu	reddit_tifu	6.4737	6.1387	5.8654	5.76	4.9569
reddit_tifu	aeslc	1.8399	1.7689	2.0162	1.6595	2.7744
aeslc	cnn	0.5478	0.9124	0.7178	0.2832	1.2466
aeslc	multi_news	0.0015	0.0009	0.0016	0	0.0888
aeslc	billsum	0.2093	0.1697	0.172	0.1698	0.2223
aeslc	pubmed	0.0039	0.005	0	0	0.1576
aeslc	reddit_tifu	0.7613	0.5248	0.5542	0.431	0.7563
aeslc	aeslc	12.6212	13.1251	12.9268	11.9428	13.4071

Figure 10: Length Scorer char: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (target|predicted|relative)	beamsearch_5_α_0.5 (target|predicted|relative)	beamsearch_10_α_0.5 (target|predicted|relative)	beamsearch_10_α_0.1 (target|predicted|relative)	beamsear (target|p
base	cnn	295.555 | 449.585 | 1.6352	295.24 | 418.715 | 1.536	295.86 | 427.075 | 1.5645	295.34 | 395.86 | 1.4702	295.715 |
base	multi_news	1103.785 | 793.375 | 0.7473	1102.92 | 610.815 | 0.5818	1104.81 | 586.295 | 0.5584	1105.165 | 444.455 | 0.4351	1104.575 |
base	billsum	861.44 | 976.365 | 1.8193	858.32 | 696.625 | 1.3442	858.82 | 629.72 | 1.2006	861.18 | 294.635 | 0.5403	859.485 |
base	pubmed	950.18 | 757.93 | 0.9048	951.06 | 541.52 | 0.6339	948.105 | 501.9 | 0.607	949.225 | 322.54 | 0.4056	946.81 | 7
base	reddit_tifu	104.62 | 347.565 | 4.6295	105.01 | 333.555 | 4.4367	105.075 | 331.225 | 4.4315	104.865 | 308.86 | 4.1892	105.455 |
base	aeslc	28.81 | 101.115 | 4.6556	28.855 | 91.74 | 4.2326	28.725 | 87.08 | 4.0506	28.87 | 86.885 | 4.0215	28.865 | 8
cnn	cnn	295.595 | 328.835 | 1.1691	295.855 | 236.395 | 0.8324	295.685 | 248.775 | 0.8887	294.78 | 175.595 | 0.6369	295.31 | 3
cnn	multi_news	1105.83 | 320.315 | 0.3021	1103.68 | 266.015 | 0.2537	1105.07 | 268.28 | 0.2529	1104.775 | 177.68 | 0.1673	1104.345 |
cnn	billsum	859.575 | 350.985 | 0.6823	860.91 | 142.23 | 0.2639	858.835 | 131.82 | 0.247	859.23 | 90.77 | 0.1739	860.925 |
cnn	pubmed	948.835 | 346.185 | 0.4215	949.49 | 179.325 | 0.2298	949.075 | 173.25 | 0.2261	949.465 | 110.945 | 0.139	948.755 |
cnn	reddit_tifu	105.725 | 340.165 | 5.1581	105.125 | 185.385 | 2.3576	105.16 | 173.41 | 4.1783	105.06 | 106.5 | 1.4133	104.9 | 43
cnn	aeslc	28.84 | 139.41 | 6.3889	28.75 | 102.965 | 4.9182	28.86 | 103.89 | 4.916	28.76 | 81.79 | 3.8222	28.78 | 11
multi_news	cnn	295.73 | 580.51 | 2.1585	296.05 | 550.47 | 2.045	295.44 | 548.13 | 2.0406	296.245 | 544.815 | 2.0317	296.025 |
multi_news	multi_news	1105.28 | 1006.0 | 0.9272	1106.88 | 943.09 | 0.8679	1104.04 | 915.685 | 0.8481	1104.0 | 871.23 | 0.8061	1104.985 |
multi_news	billsum	860.485 | 1108.885 | 2.1269	862.56 | 1046.375 | 1.9791	862.775 | 1038.295 | 1.9883	860.285 | 1032.1 | 1.995	861.21 | 1
multi_news	pubmed	949.995 | 1155.68 | 1.4479	950.355 | 1138.005 | 1.4186	949.515 | 1126.56 | 1.3838	949.44 | 1124.695 | 1.3807	948.795 |
multi_news	reddit_tifu	105.03 | 553.89 | 9.1795	105.245 | 525.135 | 8.5623	105.34 | 509.015 | 8.4026	104.655 | 506.505 | 8.3179	104.945 |
multi_news	aeslc	28.875 | 152.44 | 7.2757	28.715 | 147.49 | 7.1059	28.81 | 148.53 | 7.1244	28.86 | 148.465 | 7.1158	28.78 | 14
billsum	cnn	296.26 | 352.98 | 1.2784	294.98 | 246.165 | 0.9017	295.22 | 223.31 | 0.845	294.595 | 196.055 | 0.7382	295.16 | 3
billsum	multi_news	1103.605 | 544.725 | 0.5261	1105.195 | 317.94 | 0.3044	1105.07 | 267.055 | 0.2638	1104.555 | 240.115 | 0.2384	1104.88 |
billsum	billsum	859.16 | 1010.94 | 1.5288	860.565 | 769.09 | 1.1661	860.485 | 633.47 | 1.0198	858.42 | 536.39 | 0.8623	858.49 | 7
billsum	pubmed	949.195 | 503.44 | 0.6307	949.245 | 393.535 | 0.4932	949.925 | 393.635 | 0.5071	950.43 | 362.24 | 0.4712	948.945 |
billsum	reddit_tifu	105.065 | 274.845 | 3.7209	105.085 | 162.09 | 3.3732	105.32 | 159.455 | 3.1568	105.105 | 138.95 | 2.9417	105.03 | 3
billsum	aeslc	28.755 | 133.98 | 6.3855	28.825 | 109.775 | 5.0169	28.82 | 105.67 | 4.9534	28.825 | 100.77 | 4.5427	28.875 | 1
pubmed	cnn	295.545 | 507.925 | 1.9221	296.51 | 359.09 | 1.3665	295.35 | 356.185 | 1.3424	295.405 | 315.115 | 1.1872	295.835 |
pubmed	multi_news	1105.075 | 801.1 | 0.7657	1106.695 | 554.49 | 0.5293	1104.79 | 517.86 | 0.4774	1104.24 | 365.75 | 0.3431	1103.585 |
pubmed	billsum	859.75 | 916.765 | 1.7427	860.54 | 715.73 | 1.2768	860.165 | 689.635 | 1.1791	859.185 | 607.655 | 1.0613	858.02 | 9
pubmed	pubmed	948.69 | 957.46 | 1.1056	948.685 | 882.435 | 0.994	949.29 | 859.935 | 0.9838	950.415 | 807.48 | 0.9144	950.82 | 9
pubmed	reddit_tifu	105.05 | 396.105 | 5.6207	104.885 | 287.97 | 6.4116	104.635 | 285.1 | 6.3034	104.89 | 222.03 | 5.4298	104.87 | 4
pubmed	aeslc	28.88 | 146.325 | 6.9415	28.76 | 144.03 | 6.7625	28.79 | 138.07 | 6.5237	28.84 | 137.94 | 6.5288	28.77 | 13
reddit_tifu	cnn	295.605 | 127.21 | 0.4732	295.41 | 139.45 | 0.5457	296.185 | 154.375 | 0.5906	295.605 | 82.67 | 0.3173	295.795 |
reddit_tifu	multi_news	1103.79 | 219.02 | 0.2102	1104.92 | 111.74 | 0.1007	1105.215 | 134.145 | 0.1221	1104.79 | 18.69 | 0.017	N/A
reddit_tifu	billsum	856.465 | 397.245 | 0.6656	861.33 | 212.055 | 0.4114	856.74 | 240.76 | 0.4112	859.235 | 33.35 | 0.0629	855.93 | 4
reddit_tifu	pubmed	949.075 | 271.825 | 0.3691	948.17 | 274.195 | 0.3585	948.825 | 295.405 | 0.3805	947.43 | 141.11 | 0.1741	949.185 |
reddit_tifu	reddit_tifu	104.225 | 94.61 | 1.6539	104.845 | 77.17 | 0.9609	104.995 | 68.66 | 0.8664	105.44 | 63.715 | 0.8097	104.875 |
reddit_tifu	aeslc	28.815 | 57.42 | 2.5882	28.78 | 34.215 | 1.5835	28.78 | 36.54 | 1.6689	28.89 | 19.04 | 0.9766	28.735 | 6
aeslc	cnn	295.435 | 58.64 | 0.2273	296.48 | 53.075 | 0.2021	295.875 | 65.04 | 0.245	295.64 | 23.24 | 0.0889	295.175 |
aeslc	multi_news	1104.93 | 39.17 | 0.0381	1105.405 | 57.045 | 0.049	1104.365 | 54.665 | 0.047	1105.875 | 18.97 | 0.0187	1104.65 |
aeslc	billsum	863.955 | 39.45 | 0.0675	857.145 | 20.97 | 0.0429	859.66 | 20.895 | 0.0459	857.14 | 18.065 | 0.043	861.205 |
aeslc	pubmed	948.85 | 37.925 | 0.0469	948.1 | 43.7 | 0.045	948.98 | 20.43 | 0.024	950.67 | 18.7 | 0.0222	950.46 | 3
aeslc	reddit_tifu	104.85 | 15.34 | 0.2492	105.12 | 25.29 | 0.4298	104.95 | 25.98 | 0.4353	104.805 | 12.47 | 0.1968	105.055 |
aeslc	aeslc	28.8 | 27.87 | 1.1766	28.885 | 24.555 | 0.9756	28.845 | 24.965 | 0.9918	28.85 | 22.93 | 0.8969	28.835 | 2

Figure 11: Length Scorer word: Comparison of validation results on cross-domain trained models

Model	Testset	beamsearch_1_α_0.5 (target|predicted|relative)	beamsearch_5_α_0.5 (target|predicted|relative)	beamsearch_10_α_0.5 (target|predicted|relative)	beamsearch_10_α_0.1 (target|predicted|relative)	beamsea (target|p
base	cnn	51.68 | 75.71 | 1.5779	51.65 | 70.41 | 1.4732	51.73 | 71.875 | 1.5057	51.575 | 66.405 | 1.4167	51.685 | 7
base	multi_news	184.315 | 141.605 | 0.7974	184.385 | 100.87 | 0.5803	184.38 | 98.11 | 0.5635	184.34 | 73.435 | 0.4357	184.3 | 12
base	billsum	129.54 | 162.23 | 2.048	129.02 | 112.755 | 1.4882	129.11 | 103.115 | 1.3328	129.465 | 48.45 | 0.5943	129.64 | 1
base	pubmed	156.285 | 136.065 | 0.9896	156.065 | 93.63 | 0.6622	156.38 | 86.13 | 0.6279	155.835 | 55.41 | 0.4218	156.41 | 1
base	reddit_tifu	20.18 | 67.92 | 4.4569	20.115 | 65.295 | 4.3109	20.14 | 64.565 | 4.2617	20.15 | 60.585 | 4.039	20.25 | 67
base	aeslc	4.29 | 17.47 | 5.4749	4.28 | 15.74 | 4.9701	4.29 | 15.01 | 4.7066	4.3 | 14.885 | 4.7156	4.29 | 14.
cnn	cnn	51.81 | 58.45 | 1.1851	51.63 | 41.4 | 0.8322	51.585 | 43.495 | 0.8912	51.74 | 30.86 | 0.641	51.715 | 6
cnn	multi_news	184.325 | 55.695 | 0.3172	183.995 | 45.81 | 0.2635	184.065 | 46.28 | 0.263	184.27 | 30.845 | 0.1746	184.35 | 7
cnn	billsum	129.515 | 59.415 | 0.7663	129.49 | 24.255 | 0.2898	129.35 | 22.79 | 0.2759	129.355 | 15.72 | 0.1923	129.25 | 6
cnn	pubmed	156.315 | 58.515 | 0.4346	156.15 | 29.745 | 0.2352	155.79 | 29.15 | 0.2396	156.17 | 18.18 | 0.1395	156.59 | 5
cnn	reddit_tifu	20.2 | 68.63 | 4.8161	20.115 | 37.305 | 2.3986	20.21 | 35.115 | 2.8613	20.13 | 21.465 | 1.3861	20.1 | 86.
cnn	aeslc	4.28 | 24.81 | 7.8608	4.28 | 18.175 | 6.0586	4.28 | 18.33 | 6.1235	4.27 | 14.53 | 4.7118	4.29 | 20.
multi_news	cnn	51.77 | 101.16 | 2.1482	51.7 | 94.895 | 2.016	51.635 | 94.24 | 2.0117	51.705 | 93.735 | 2.0052	51.71 | 96
multi_news	multi_news	184.255 | 172.275 | 0.955	184.38 | 160.235 | 0.8862	184.23 | 155.475 | 0.8648	184.16 | 148.875 | 0.826	184.15 | 1
multi_news	billsum	129.34 | 192.695 | 2.4537	129.46 | 179.625 | 2.2877	129.215 | 176.35 | 2.3269	128.845 | 175.525 | 2.3322	129.21 | 1
multi_news	pubmed	156.22 | 199.435 | 1.5187	156.265 | 192.39 | 1.4654	156.175 | 191.465 | 1.4208	156.12 | 190.805 | 1.4232	155.97 | 1
multi_news	reddit_tifu	20.21 | 103.79 | 7.5298	20.195 | 98.68 | 7.145	20.15 | 95.97 | 6.9665	20.22 | 95.39 | 6.8185	20.135 | 9
multi_news	aeslc	4.28 | 27.65 | 8.9317	4.3 | 26.47 | 8.5218	4.28 | 26.54 | 8.6099	4.27 | 26.55 | 8.6052	4.29 | 26.
billsum	cnn	51.715 | 60.375 | 1.2563	51.565 | 42.155 | 0.8838	51.66 | 39.06 | 0.8375	51.64 | 34.555 | 0.739	51.62 | 60
billsum	multi_news	184.2 | 89.83 | 0.5225	184.21 | 52.765 | 0.3027	184.225 | 45.11 | 0.2674	184.205 | 40.59 | 0.2405	184.2 | 11
billsum	billsum	129.55 | 153.89 | 1.5579	129.64 | 117.315 | 1.1861	128.885 | 96.285 | 1.0258	129.84 | 82.135 | 0.8735	129.365 |
billsum	pubmed	156.47 | 78.11 | 0.5887	155.76 | 60.38 | 0.457	156.235 | 60.94 | 0.4754	156.245 | 56.01 | 0.4383	156.075 |
billsum	reddit_tifu	20.17 | 48.33 | 3.1382	20.1 | 29.105 | 2.2168	20.15 | 27.99 | 2.0119	20.15 | 24.595 | 1.835	20.145 | 6
billsum	aeslc	4.285 | 21.965 | 7.125	4.29 | 18.39 | 5.8343	4.28 | 17.385 | 5.771	4.28 | 16.655 | 5.3529	4.29 | 18.
pubmed	cnn	51.615 | 93.515 | 2.0163	51.58 | 66.615 | 1.4428	51.7 | 65.625 | 1.4121	51.61 | 58.445 | 1.2561	51.6 | 84.
pubmed	multi_news	184.245 | 150.705 | 0.8749	184.145 | 104.16 | 0.6058	184.135 | 95.99 | 0.5431	184.34 | 68.495 | 0.3917	184.355 |
pubmed	billsum	128.98 | 163.085 | 2.0936	128.845 | 129.165 | 1.6207	129.5 | 127.97 | 1.5182	129.405 | 113.175 | 1.3577	129.515 |
pubmed	pubmed	156.22 | 161.93 | 1.1322	156.105 | 149.15 | 1.0197	156.01 | 145.12 | 1.0034	156.275 | 136.4 | 0.9307	156.155 |
pubmed	reddit_tifu	20.115 | 80.64 | 5.5781	20.175 | 58.42 | 4.2213	20.09 | 57.95 | 4.2779	20.22 | 45.485 | 3.22	20.135 | 8
pubmed	aeslc	4.3 | 25.4 | 8.2452	4.29 | 25.27 | 8.0682	4.29 | 24.49 | 7.8305	4.29 | 24.47 | 7.8687	4.29 | 24.
reddit_tifu	cnn	51.705 | 22.61 | 0.4747	51.71 | 24.015 | 0.5402	51.645 | 25.955 | 0.5695	51.685 | 13.96 | 0.3049	51.61 | 54
reddit_tifu	multi_news	184.435 | 73.355 | 0.4374	184.17 | 28.18 | 0.1567	184.485 | 33.565 | 0.1826	184.455 | 3.95 | 0.0219	N/A
reddit_tifu	billsum	129.63 | 97.175 | 1.1085	128.87 | 71.065 | 0.9121	129.01 | 88.085 | 0.9953	129.515 | 6.26 | 0.0762	129.515 |
reddit_tifu	pubmed	155.76 | 50.34 | 0.4128	156.405 | 54.32 | 0.4106	156.36 | 52.53 | 0.4018	156.195 | 24.23 | 0.1818	156.085 |
reddit_tifu	reddit_tifu	20.11 | 19.03 | 1.3798	20.205 | 15.34 | 0.9847	20.2 | 13.53 | 0.8757	20.175 | 12.705 | 0.8337	20.18 | 32
reddit_tifu	aeslc	4.32 | 11.17 | 3.553	4.29 | 6.59 | 2.0353	4.29 | 7.11 | 2.2098	4.29 | 4.0 | 1.304	4.3 | 12.7
aeslc	cnn	51.675 | 9.48 | 0.2088	51.67 | 7.935 | 0.1711	51.63 | 8.555 | 0.1848	51.655 | 3.65 | 0.0792	51.7 | 15.
aeslc	multi_news	184.3 | 6.49 | 0.0376	184.26 | 6.455 | 0.035	184.14 | 6.13 | 0.033	184.12 | 3.09 | 0.0184	184.285 |
aeslc	billsum	129.695 | 5.69 | 0.0656	129.52 | 3.16 | 0.0412	129.28 | 3.07 | 0.0428	129.48 | 2.635 | 0.0385	129.31 | 8
aeslc	pubmed	156.375 | 5.82 | 0.0433	156.08 | 6.21 | 0.0394	156.225 | 3.27 | 0.0239	156.13 | 2.94 | 0.0218	156.395 |
aeslc	reddit_tifu	20.21 | 2.8 | 0.207	20.13 | 3.74 | 0.3201	20.1 | 3.86 | 0.3264	20.24 | 2.06 | 0.1437	20.19 | 8.
aeslc	aeslc	4.295 | 4.12 | 1.1896	4.28 | 3.6 | 1.0031	4.295 | 3.645 | 1.0093	4.28 | 3.29 | 0.9056	4.29 | 3.8