Pokhara Engineering College Journal (PECJ)

Applied Sciences and Engineering Insights

ISSN: 3021-9795 (Print) / 3059-9628 (Online)
(Volume-3, Issue- I)

Abstractive Summarization of News Articles Using BART

Mohan Bikram KC[1]*, Ayush Sharma Kaundinya[2], Pramish Adhikari[3], Smita Adhikari[4]

[1,2,3] *Department of Computer Engineering, Gandaki College of Engineering and Science,Nepal*

[4] *Department of Electronics and Computer Engineering, Pashchimanchal Campus, Pokhara, Nepal*

*Corresponding email: kcmohan64@gmail.com

Received:22- January-2026; Revised: 15- February-2026; Accepted: 20- March-2026
DOI: https://doi.org/10.3126/pecj.v3i1.93538

Abstract

Abstractive text summarization generates concise summaries that preserve key concepts through newly constructed sentences rather than extraction from the source. As online news content grows rapidly, effective summarization systems are essential for efficient information consumption. This study investigates abstractive summarization using BART-Large, a transformer-based encoder-decoder model, with a novel training approach: fine-tuning on a combined dataset merging XSum and CNN/DailyMail. While XSum emphasizes highly abstractive single-sentence summaries and CNN/DailyMail targets longer multi-sentence outputs, combining these datasets during training aims to produce a more versatile model capable of handling diverse summarization demands. Performance is evaluated using ROUGE metrics, which measure n-gram overlap between generated and reference summaries. The fine-tuned model achieves a ROUGE-2 score of 22.86, demonstrating competitive performance against existing approaches. Qualitative analysis reveals that the model produces fluent, coherent summaries while maintaining factual consistency with source documents. These results indicate that exposing models to varied summarization demands during fine-tuning can improve flexibility without sacrificing quality. This work offers a practical direction for building more generalizable summarization systems and highlights the potential of multi-dataset training strategies for domain adaptation in natural language generation tasks.

Keywords: Abstractive Text Summarization, BART-Large, News Articles, ROUGE Evaluation, Natural Language Processing

1. Introduction

With expansion of online media, individuals are exposed to an excessive volume of news content. Every moment, numerous articles appear across digital platforms which makes it challenging for readers to remain informed on factual information efficiently. This flooding of

data has created a need for automated methods that can produce concise and informative summaries of lengthy news articles. In the field of natural language processing, text summarization has emerged as a practical approach to address this challenge by enabling readers to learn the core ideas of a text in a shorter time.

Text summarization uses two methods i.e., extractive and abstractive. The extractive method identifies the most relevant sentences or phrases in the source and arranges them to form a summary. An abstractive method models the meaning of the source and produces new sentences that deliver the same information with clarity. Abstractive summarization is more complex because it depends on deeper knowledge of linguistic structure, grammar and context. In recent years advances in deep learning and progress of models based on transformers have led to substantial gains in the quality of abstractive summarization.

This study implements the Large variant of the Bidirectional and Auto-Regressive Transformer for the abstractive summarization of news articles. The model is fine-tuned with two widely recognized datasets, XSum, which contains short single sentence summaries, and CNN/DailyMail, which includes longer multiple sentence summaries. These datasets enable evaluation of the ability of the model to adapt to different lengths for summary and writing styles. Performance is evaluated using the Recall Oriented Understudy for Gisting Evaluation (ROUGE) metrics, which quantify the degree of similarity between generated summaries and human summaries.

1.1 Problem Statement

The rapid growth of online news has reduced the ability of readers to keep pace with the continuous production of information. A very large number of articles are published continuously across many platforms, which overwhelms readers and makes it difficult to determine what is most important. Although automatic summarization systems are available, many continue to rely on extractive methods that select segments of the source text. These summaries can be produced quickly, but they often read as unnatural and do not accurately represent the meaning or intent of the article.

1.2 Objective

The main objective of the study is to fine-tune the BART-Large model on two benchmark datasets: XSum and CNN/DailyMail for producing both short and long summaries while maintaining factual accuracy.

2. Literature Review

Lewis et al. (2019) introduced BART, a denoising autoencoder for sequence pretraining that uses a bidirectional encoder similar to BERT and a left-to-right decoder similar to GPT. The model is trained by corrupting input text, for example by shuffling sentences and using a span-masking scheme that replaces continuous segments with a single mask token, and then learning to reconstruct the original sequence. BART performs well on both text generation and text understanding tasks: it reaches performance comparable to RoBERTa on GLUE and SQuAD, and improves results for abstractive dialogue, summarization, and question answering by as much as 6 ROUGE points. In machine translation, target-side pretraining with BART yields a BLEU gain of 1.1. Ablation studies show which pretraining objectives and design choices matter most for downstream performance, and indicate that BART can be applied effectively to a wide range of natural language processing tasks.

Zhang et al. (2019) created PEGASUS, a Transformer-based model. It's designed specifically for abstractive summarization. Instead of randomly masking words like most models, it skips entire unimportant sentences. Then it learns to turn the best parts of the text into coherent summaries. They tested it on 12 different dataset news, research papers, stories, even law documents. PEGASUS achieved strong results across all datasets. Even when there's little data, it outperforms older models, and its summaries often read as natural as a human wrote them.

Raffel et al. (2019) introduced T5, a Transformer model that treats every natural language processing task as a text-to-text problem. They ran a large set of experiments comparing different transfer learning methods, pretraining objectives, model architectures, and training datasets across many language tasks. Using large-scale pretraining on the Colossal Clean Crawled Corpus (C4), T5 reached state-of-the-art results in tasks such as summarization, question answering, and text classification. The authors also released pretrained models and the C4 dataset, making it easier to study and apply transfer learning in natural language processing.

Text summarization reduces long texts to concise summaries, and large language models use context to produce human-like summaries. Basyal and Sanghvi (2023) compared MPT-7B-Instruct, Falcon-7B-Instruct, and OpenAI text-davinci-003 on CNN/DailyMail and Extreme Summarization datasets. Text-davinci-003 gave the most accurate summaries by Bilingual Evaluation Understudy, Recall Oriented Understudy for Gisting Evaluation, and Bidirectional Encoder Representations from Transformers score. The study shows that large language models can improve automatic summarization and guide future natural language processing.

Shakil et al. (2024) present a comprehensive survey on abstractive text summarization, categorizing state-of-the-art techniques into five groups: traditional sequence-to-sequence models, pre-trained large language models, reinforcement learning approaches, hierarchical methods, and multi-modal summarization. The authors provide comparative analysis tables offering insights into model complexity, scalability, and appropriate applications. The paper highlights key challenges including inadequate meaning representation, factual consistency, controllable summarization, cross-lingual summarization, and evaluation metrics, while proposing solutions leveraging knowledge incorporation and other innovative strategies. The survey also explores emerging research directions such as domain-specific summarization, long-document summarization, and handling noisy data, providing researchers with a structured overview to advance abstractive summarization research.

Venkataramana, Srividya, and Cristin (2022) proposed using Bidirectional and Auto Regressive Transformers, a deep learning encoder decoder model, for efficient abstractive summarization. The model uses attention to identify key features and predict missing tokens, improving summary quality. Compared with Bidirectional Encoder Representations from Transformers, Text To Text Transfer Transformer, and Robustly Optimized Bidirectional Encoder Representations from Transformers Pretraining Approach, the model performed well on large scale text summarization. This approach shows that deep learning can automate and speed processing of extensive datasets.

Chari et al. (2024) proposed a system using pretrained Bidirectional and Auto Regressive Transformers fine-tuned on CNN/DailyMail to generate abstractive summaries from text files and Uniform Resource Locators. It incorporates the encoder from BERT into Generative Pre-trained Transformer. This model enhances coherence and factual consistency. The repetition, faithfulness and control of length are considered as challenges in this study. In this way, it

indicates how a deep learning model can summarize large-scale and diverse documents efficiently.

Automatic text summarization condenses long sentences into brief summaries. Adhik, Lakshmi, and Muralidharan (2024) developed a high-quality summarizer based on a bidirectional encoder and autoregressive decoder model. Their proposed system received a quality score of about 78%, as ROUGE metrics were adopted for assessing the accuracy of generated summaries. It removes unnecessary information and emphasizes only the important ones, which proves that transformer-based models work well in text summarization.

Deep learning models are normally used to summarize large volumes of text data efficiently. Dharrao et al. (2024) compared BART, T5, and PEGASUS on the BBC business article dataset. Model performance was evaluated using ROUGE and METEOR scores. T5 attained the highest accuracy of ROUGE-1: 0.354 and METEOR: 0.35. The results showed that a transformer-based model can effectively get relevant information from business news.

Abstractive summarization produces concise summaries for both brief and lengthy texts. Wilman, Atara, and Suhartono (2024) proposed a method that uses a pre-trained Bidirectional and Auto-Regressive Transformers model with chunking to process long English documents. The text is split into chunks; each chunk is encoded and summarized with the model, and then the summaries are combined into a final version. The results obtained by the CNN/DailyMail pre-trained model were better than those of the Extreme Summarization version. This approach improves summary quality for lengthy texts such as news and research articles.

Ghanem et al. (2025) proposed a hybrid deep learning system that combines Bidirectional Encoder Representations from Transformers, known as BERT, a Transformer encoder and decoder, and long short term memory networks to produce concise summaries that fit the context. The system uses attention to focus on key information and limit repetition. On the Hagupit, SHShoot, and Hyderabad Blast datasets it achieved higher ROUGE scores than earlier methods. The results show practical ways to automate short text summarization on social media.

To find an effective model for summarizing news, Krishna et al. (2025) compared several deep learning approaches. They tested Text-to-Text Transfer Transformer-base, Text-to-Text Transfer Transformer-large, Bidirectional and Auto-Regressive Transformers Cable News

Network, and Pre-training with Extracted Gap-sentences for Abstractive Summarization. The researchers initially prepared the data and conducted exploratory analysis, before assessing the models using Recall-Oriented Understudy for Gisting Evaluation and Bilingual Evaluation Understudy scores. All the results indicate very clear differences in summary quality. This work demonstrates both the promise and the difficulty of using deep learning to generate accurate summaries from long documents.

Nnadi and Bertini (2024) present a survey on abstractive text summarization focusing on datasets, models, and evaluation metrics. The survey covers transformer-based models for summarization including BART, PEGASUS, Longformer, and LongT5 for single documents, and PRIMERA, CENTRUM, and REFLECT for multi-document tasks. A key challenge the authors highlight is that standard transformers struggle with longer sequences (beyond 512-1024 tokens) because self-attention gets computationally expensive as length increases. They evaluate models using ROUGE, METEOR, BertScore, and factual consistency checks, finding that training across multiple datasets and domains improves performance. Interestingly, factual inconsistency rates have dropped compared to the roughly 30% reported in earlier work, though generating reliably accurate summaries remains a hurdle for practical use.

Chen and Song (2021) proposed a hybrid text summarization method combining TextRank with BART to address topic deviation problems in abstractive summarization. Their approach first uses TextRank for extractive summarization and BART for abstractive summarization separately, then splices the results to create a new text that increases the weight of key sentences. This combined text is then fed into the BART model again to generate the final summary. Experimental results demonstrate improvements over a single BART model, with average recall scores increasing by 1.5%, 0.5%, and 1.3% for ROUGE-1, ROUGE-2, and ROUGE-L respectively.

Dhanda and Gupta (2024) compared extractive and abstractive summarization techniques. Algorithms discussed included TextRank, Sequence to Sequence, and Bidirectional and Auto-Regressive Transformers. Results indicated that though TextRank works well with short texts, and Sequence to Sequence models generate summaries of high quality, the Bidirectional and Auto-Regressive Transformers model beats them when applied to benchmark datasets. Hence, based on Recall-Oriented Understudy for Gisting Evaluation scores, transformer-based models generate summaries that are more accurate as well as coherent. This study is an eye-opener regarding the strengths and shortcomings of various summarization algorithms.

3. Methodology

3.1 Datasets

Two popular benchmark datasets were used for the purpose of model training and evaluation with regards to the abstractive summarization task, the XSum dataset, along with the CNN/DailyMail dataset version 3.0.0. The XSum dataset consists of 204,045 train samples, 11,332 validation samples, and 11,334 test samples, which feature a collection of BBC articles summarized in a single sentence with a highly abstractive nature. The CNN/DailyMail dataset consists of 287,113 train samples, 13,368 validation samples, and 11,490 test samples, with a feature set that provides more extractive two-sentence abstracts. These two datasets were merged together in order to provide a well-rounded corpus consisting of a combined dataset of 491,158 train samples, 24,700 validation samples, and 22,824 test samples.

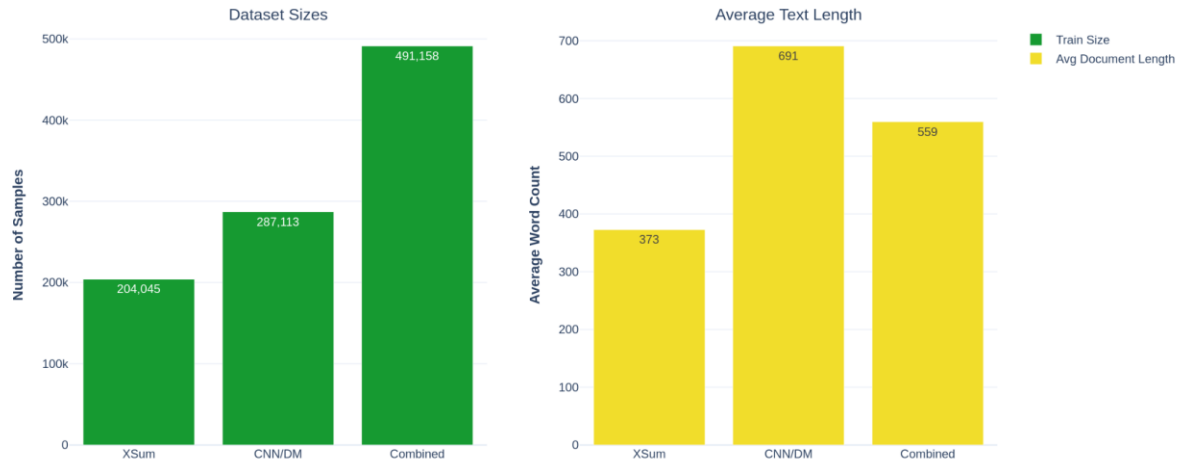

Figure 1: Sample Distribution and Average Document Length Across XSum, CNN/DailyMail, and Combined Datasets

3.2 Data Preprocessing

During the text processing stage, the BART tokenizer with a vocabulary size of 50,265 tokens was used to convert text into model-compatible representations. The input documents were tokenized and truncated to a maximum length of 1,024 tokens to preserve essential content while maintaining computational efficiency. Target summaries were similarly processed with a maximum length of 128 tokens. Dynamic padding was implemented at the batch level, with sequences padded to the longest sequence within each batch. DataCollatorForSeq2Seq was used for the conversion of summary tokens into label format, with padding tokens replaced with -100 to exclude them from loss computation during training.

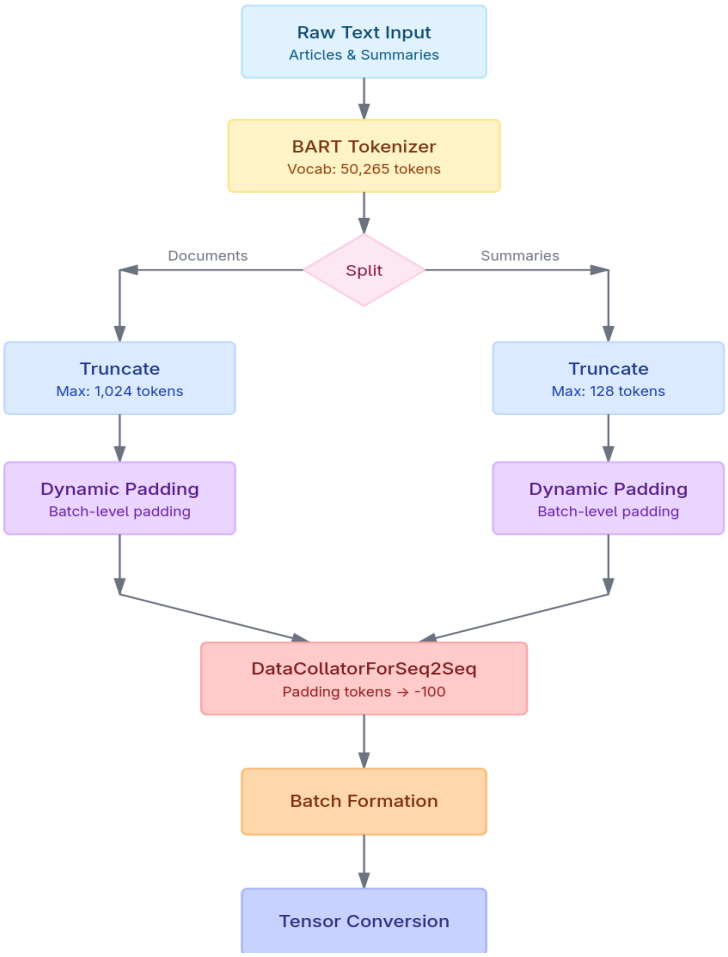

Figure 2: Data Preprocessing Pipeline for BART-Large Training

3.3 Model Architecture

BART-Large with the pre-trained checkpoint 'facebook/bart-large-xsum' was employed as the foundation for abstractive summarization. The model consists of a transformer-based encoder-decoder architecture with 12 encoder layers and 12 decoder layers. Each layer consists of 16 attention heads, with a hidden dimension of 1,024. This is a model that consists of roughly 406 million parameters, with a bidirectional encoder that encodes the entire input text document, and an autoregressive decoder that generates its output sequentially, on a token basis. This model takes advantage of the self-attention mechanism through cross-attention layers in both its encoder and decoder modules. This feed forward network utilizes a dimension of 4,096, along with GELU activation functions that facilitate complex transformations. In efforts to reduce memory usage during the training process, the application of gradient-checkpointing was considered.

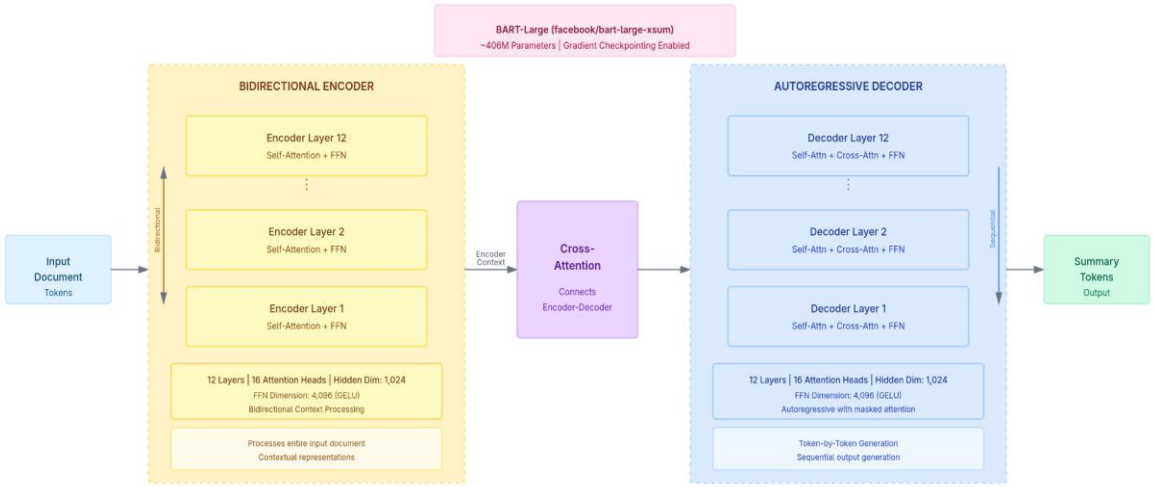

Figure 3: BART-Large Encoder-Decoder Architecture for Abstractive News Summarization

3.4 Training and Inference Configuration

Fine-tuning was performed using the AdamW optimizer, with hyperparameters being tuned through a two-stage procedure. In the first stage, hyperparameters that relate to training optimization were considered, such as learning rates of (1×10^{-5}, 2×10^{-5}, 3×10^{-5}), the number of training epochs of (3, 4), and the warmup ratios of (0, 0.05, 0.1) with respect to the total number of training steps. In this case, the most effective configuration consisted of a learning rate of 1×10^{-5} with weight decay of 0.01, along with 3 training epochs with a batch size of 8. This effective batch size was achieved through the use of a per-device batch size of 2 along with gradient accumulation over 4 steps. Mixed precision training with FP16 was employed to accelerate computation and reduce memory usage while maintaining numerical stability. During this, the learning rate was set to linearly decay after the completion of the warmup period. Also, this procedure saves a model checkpoint every epoch, with the best model being selected based on validation loss.

In the second stage of hyperparameter tuning, the generation quality hyperparameters were further tuned during inference. Beam search settings were evaluated with beam sizes of (2, 4, 6), length penalty coefficients of (1.5, 2.0, 2.5), and maximum output sequence lengths of (96, 128) tokens. Although other hyperparameter combinations may work well, the best hyperparameter configuration involved beam search decoding with a beam size of 4 and a length penalty of 1.5. Max length was fixed as 96, with min length fixed as 20 tokens, thus allowing for the generation of summaries that were single sentence, as in the case of the news summary task in XSum, and multiple sentence summaries, as seen in CNN/DailyMail. In addition, n-gram blocking was applied to prevent the repetition of 3-grams within generated

summaries for improved coherence. Finally, the early stopping hyperparameter was enabled, such that the beam search decoding would terminate once all beams produced the end-of-sequence token.

3.5 Evaluation Metrics

Model performance was evaluated with the ROUGE metrics, computing ROUGE-1, ROUGE-2, ROUGE-L, and ROUGE-Lsum metrics. These metrics calculate n-gram overlap between generated summaries and human-written references. ROUGE-1 captures unigram overlap to account for content selection, while ROUGE-2 looks at bigram overlap to evaluate fluency, and finally, ROUGE-L calculates with the use of the longest common subsequence. During hyperparameter search, the focus was on maximizing the value of ROUGE-2 and that of ROUGE-L.

4. Results and Discussion

4.1 Hyperparameter Optimization

The hyperparameter optimization was conducted in two stages to identify the optimal configuration for our BART-based news summarization model on 30% of the total dataset. Stage 1 focused on training parameters while Stage 2 optimized generation quality metrics.

4.1.1 Training Parameter Optimization

Table 1: Training Hyperparameter Optimization Results-Top 5 Configurations by ROUGE-2 Score

Learning Rate	Epochs	Warmup Ratio	ROUGE-1	ROUGE-2	ROUGE-L
1×10^{-5}	3	0.05	46.52	23.21	38.74
1×10^{-5}	3	0.00	46.34	23.10	38.61
1×10^{-5}	3	0.10	46.35	23.16	38.43
1×10^{-5}	4	0.05	45.85	22.92	38.27
3×10^{-5}	3	0.05	45.90	22.63	38.11

From the optimization of the training process, the ideal value of the learning rate was determined to be 1×10^{-5}, together with a number of training epochs set to 3, along with a warmup ratio of 0.05, thus providing the best results, as shown by the highest ROUGE-2 score

of 23.21. Lower learning rates, such as 1×10^{-5}, performed better compared to higher rates like 2×10^{-5} and 3×10^{-5}, showing reduced ROUGE-2 scores with increased rates. These results confirm that slower updates of the parameters play a crucial role in the fine-tuning process on the aggregated heterogeneous dataset. Integration of a 5% warmup period led to minor but consistent improvements across all learning rates, stabilizing the training process during the initial adaptation phase.

Figure 4: ROUGE-2 Scores Across Learning Rate, Epochs, and Warmup Ratio Variations

4.1.2 Generation Quality Optimization

Table 2: Inference Generation Parameter Optimization Results: Top 5 Configurations by ROUGE-2

Beam Width	Length Penalty	Max Length	ROUGE-1	ROUGE-2	ROUGE-L
4	1.5	96	47.56	24.06	39.64
4	2.0	96	47.45	23.94	39.60
4	2.5	96	47.39	23.86	39.49
6	1.5	96	47.19	23.77	39.45

2	1.5	96	47.17	23.62	39.38

Experiments on generation quality show that beam search with the beam size set to 4 and the length penalty set to 1.5 produced the best-quality summaries, with a ROUGE-2 score of 24.06. It is important to note that varying the maximum length parameter between 96 or 128 tokens did not impact the result, as the same scores were generated in both cases. It is presumed that 96 tokens is sufficient to capture the essential information in both single-sentence as well as multi-sentence summaries. Also, further increase in the beam size beyond 4 resulted in diminishing returns, with slight performance degradation at a beam size of 6, indicating that excessive exploration during decoding may introduce noise rather than improve summary quality.

Figure 5: Generation Parameter Impact Analysis: ROUGE Scores Across Beam Width, Length Penalty, and Maximum Length Variations

4.2 Model Performance

Following hyperparameter optimization, the model was trained on the complete combined dataset using the identified optimal configuration.

Table 3: Final Model Performance: ROUGE Scores on Validation and Test Sets

Metric	Validation Set	Test Set
ROUGE-1	46.82	46.43
ROUGE-2	22.91	22.86

ROUGE-L	38.86	38.67
ROUGE-Lsum	39.13	38.93

The final model achieved a ROUGE-2 score of 22.86 on the test set, demonstrating strong generalization capability with minimal overfitting (validation-test gap of only 0.09 points). The average generated summary length of 29.6 tokens indicates the model learned to produce concise summaries appropriate for news articles.

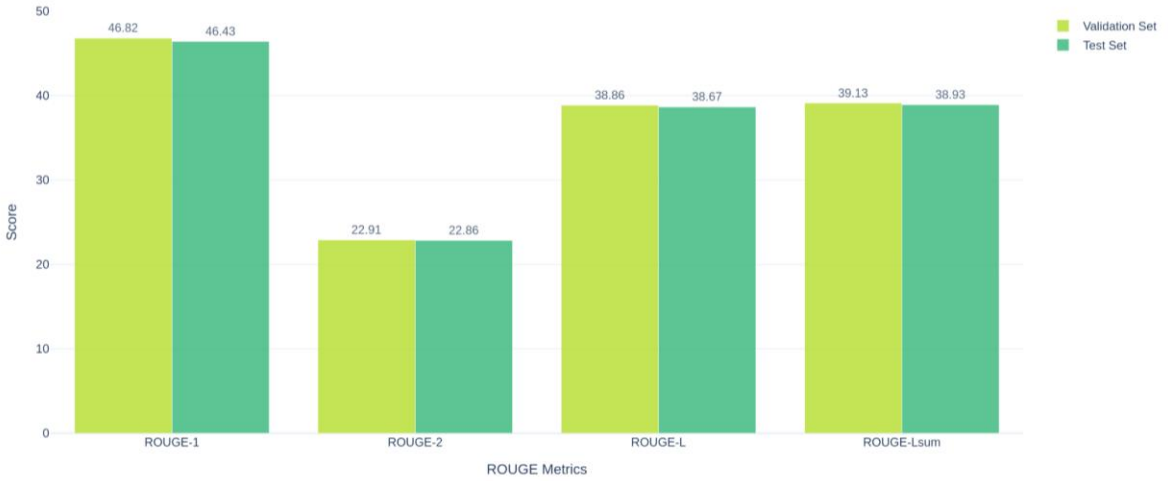

Figure 6: Comparison of ROUGE Metrics Between Validation and Test Sets

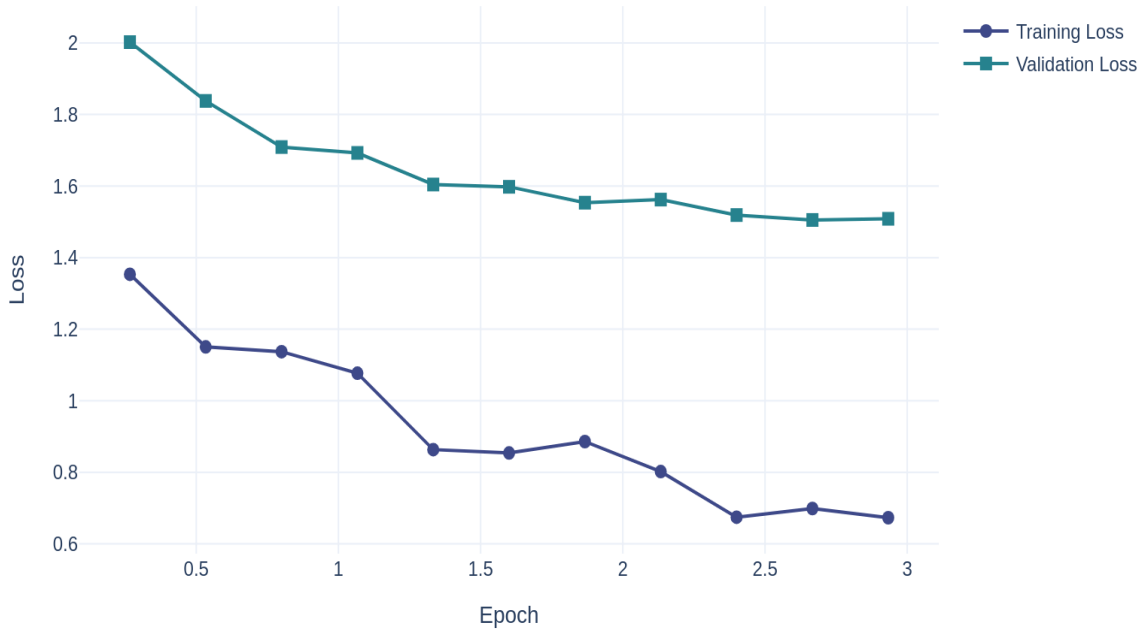

Figure 7: Cross-Entropy Loss Progression Across Training Epochs for Training and Validation Sets

4.3 Discussion

Our hyperparameter tuning produced several findings for training Transformer models on heterogeneous summarization datasets. The optimal learning rate of 1×10^{-5} accords with best practices in the fine-tuning of pretrained models. Conservative learning rates reduce catastrophic forgetting while allowing effective adaptation. Given the diversity of the combined dataset, the positive effect of warmup set to five percent of training steps was notable. It enables gradual adaptation to the distinct summarization styles in the Extreme Summarization dataset which uses highly abstractive single sentences and in the CNN/DailyMail dataset which contains more extractive multiple sentence summaries.

Generation parameters showed consistent patterns. A beam width of four provided a balance between exploration and exploitation during decoding. A length penalty of 1.5 achieved a balance between brevity and completeness and prevented both short summaries that omit key information and verbose outputs that lack focus. The model was relatively insensitive to maximum length settings of 96 and 128 tokens and this indicates that it learned appropriate length control from the training data itself rather than from hard constraints.

Training dynamics showed efficient convergence. Training loss decreased by 45 percent from 1.353 to 0.6731 over three epochs while maintaining a gap between training and validation losses that indicates good generalization without overfitting.

4.3.1 Comparison with Existing Approaches

Table 4: Performance Comparison of fine-tuned BART-Large with other Summarization Models

Model	Dataset	ROUGE-1	ROUGE-2	ROUGE-L	Parameters
Our Model	XSum + CNN/DailyMail	46.43	22.86	38.67	406M
BART-Large (Lewis et al., 2019)	XSum only	45.14	22.27	37.25	406M
BART-Large (Lewis et al., 2019)	CNN/DailyMail only	44.16	21.28	40.90	406M

PEGASUS (Zhang et al., 2019)	XSum	47.21	24.56	39.25	568M
PEGASUS (Zhang et al., 2019)	CNN/DailyMail only	44.17	21.47	41.1	568M
T5-Large (Raffel et al., 2019)	CNN/DailyMail	43.52	21.55	40.69	770M
ProphetNet (Qi et al., 2020)	CNN/DailyMail	44.20	21.17	41.30	393M

5. Conclusion

This study demonstrates the effectiveness of fine-tuning the Large Bidirectional and Autoregressive Transformer for abstractive news summarization across datasets with diverse characteristics. A systematic two stage hyperparameter optimization yielded coherent and concise summaries that balance the highly abstractive style of the Extreme Summarization dataset with the more extractive style of the CNN/DailyMail articles. The results confirm that Transformer based encoder decoder architectures adapt well to varied summarization requirements when configured appropriately. These findings advance automated news summarization systems by enabling summaries that allow readers to process large volumes of online content efficiently.

6. Recommendation for future research

This work provides an initial basis for abstractive summarization across multiple domains and several directions remain for further research. Future research should develop resource efficient fine-tuning methods to increase reproducibility across computing environments. Adaptive sampling can address dataset imbalance and reduce the observed bias toward extractive styles. Expanding the training corpus with scientific legal and medical documents will improve generalization beyond news. Evaluation protocols need to extend beyond the ROUGE metrics. Semantic similarity metrics such as the Bidirectional Encoder Representations from Transformers score known as BERTScore and factuality checking software such as the Fact Consistency Classifier known as FactCC together with systematic human reviews will provide a more comprehensive measure of summary quality and factuality. There are two principal

technical limitations in this approach. The first is factual inconsistency in which the model generates information that is not present in the source text and this may be mitigated with factual verification modules, retrieval augmented generation, or reinforcement learning with human feedback. The second is a fixed input limit of 1024 tokens which prevents processing of long or multiple topic documents. Transformer architectures that support long context, sliding window encoding, or hierarchical chunking can address this constraint.

References

- Lewis, M. et al. (2019) BART: Denoising Sequence-to-Sequence Pre-training for natural language generation, Translation, and Comprehension. https://arxiv.org/abs/1910.13461v1.
- Zhang, J. et al. (2019) PEGASUS: Pre-training with Extracted Gap-sentences for Abstractive Summarization. https://arxiv.org/abs/1912.08777.
- Raffel, C. et al. (2019) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. https://arxiv.org/abs/1910.10683.
- Basyal, L. and Sanghvi, M. (2023) Text summarization using large language models: a comparative study of MPT-7B-Instruct, Falcon-7b-Instruct, and OpenAI Chat-GPT models. https://arxiv.org/abs/2310.10449.
- Shakil, H., Farooq, A. and Kalita, J. (2024) 'Abstractive text summarization: State of the art, challenges, and improvements,' Neurocomputing, 603, p. 128255. https://doi.org/10.1016/j.neucom.2024.128255.
- Venkataramana, A., Srividya, K. and Cristin, R. (2022) 'Abstractive text summarization using BART,' 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), pp. 1–6. https://doi.org/10.1109/mysurucon55714.2022.9972639.
- Parabrahma Chari, S. et al. (2024) AUTOMATIC TEXT SUMMARIZATION USING BART, International Journal of Creative Research Thoughts (IJCRT). journal-article, pp. c411–c413. https://www.ijcrt.org/papers/IJCRT2402277.pdf.
- Adhik, C., Lakshmi, S.S. and Muralidharan, C. (2024) 'Text summarization using BART,' AIP Conference Proceedings, 3075, p. 020038. https://doi.org/10.1063/5.0217004.
- Dharrao, D. et al. (2024) 'Summarizing business news: Evaluating BART, T5, and PEGASUS for effective information extraction,' Revue D Intelligence Artificielle, 38(3), pp. 847–855. https://doi.org/10.18280/ria.380311.

- Wilman, P., Atara, T. and Suhartono, D. (2024) 'Abstractive English Document Summarization Using BART Model with Chunk Method,' Procedia Computer Science, 245, pp. 1010–1019. https://doi.org/10.1016/j.procs.2024.10.329.
- Ghanem, F.A. et al. (2025) 'Deep Learning-Based Short Text Summarization: an integrated BERT and transformer Encoder–Decoder approach,' Computation, 13(4), p. 96. https://doi.org/10.3390/computation13040096.
- Krishna, K.M.R. et al. (2025) 'Deep learning for text summarization using NLP for automated news digest,' Scientific Reports, 15(1), p. 36343. https://doi.org/10.1038/s41598-025-20224-1.
- Nnadi, G.O. and Bertini, F. (2024) 'Survey on Abstractive Text Summarization: Dataset, models, and Metrics,' arXiv (Cornell University) [Preprint]. https://doi.org/10.48550/arxiv.2412.17165.
- Chen, Y. and Song, Q. (2021) News Text Summarization Method based on BART-TextRank Model, 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). https://doi.org/10.1109/iaeac50856.2021.9390683.
- Dhanda, N. and Gupta, K.K. (2024) 'A novel approach to text summarization using machine learning,' Asian Journal of Research in Computer Science, 17(4), pp. 95–104. https://doi.org/10.9734/ajrcos/2024/v17i4432.
- Narayan, S., Cohen, S.B. and Lapata, M. (2018) 'Don’t give me the details, just the summary! Topic-Aware convolutional neural networks for extreme summarization,' arXiv (Cornell University) [Preprint]. https://doi.org/10.5281/zenodo.2399761.
- Hermann, K.M. et al. (2015) 'Teaching machines to read and comprehend,' arXiv (Cornell University) [Preprint]. https://doi.org/10.48550/arxiv.1506.03340.
- Qi, W. et al. (2020) 'ProphetNET: Predicting future N-Gram for Sequence-to-Sequence pre-training,' arXiv (Cornell University) [Preprint]. https://doi.org/10.48550/arxiv.2001.04063.