

# Bangla News Headline Generation

By  
Mahmudul Hasan Rifat  
212-15-14757

## FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the  
Requirements for the Degree of Bachelor of Science in  
Computer Science and Engineering

### Supervised by

Mr. Md. Sadekur Rahman  
Assistant Professor  
Department of Computer Science and  
Engineering Daffodil International  
University

### Co-Supervised by

Md. Jakaria Zobair  
Lecturer  
Department of Computer Science and  
Engineering Daffodil International  
University



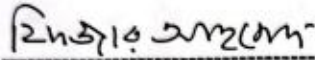
DAFFODIL INTERNATIONAL  
UNIVERSITY  
Dhaka, Bangladesh

May 14, 2025

## APPROVAL

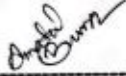
This Project titled “Bangla News Headline Generation”, submitted by **Mahmudul Hasan Rifat**, ID No: 212-15-14757 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **14 May, 2025**.

### BOARD OF EXAMINERS



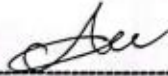
**Dr. Fizar Ahmed**  
**Associate Professor**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**



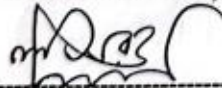
**Amatul Bushra Akhi**  
**Assistant Professor**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



**Mr. Abdullah Al Mamun**  
**Sr. Lecturer**  
Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**



**Dr. Mohammed Nasir Uddin**  
**Professor**  
Department of Computer Science and Engineering  
Jagannath University

**External Examiner**

# DECLARATION

---

I hereby declare that this project has been done by me under the supervision of Mr. Md. Sadekur Rahman, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Sadekur 14.5.25

**Mr. Md. Sadekur Rahman**

Assistant Professor

Department of Computer Science and  
Engineering Daffodil International  
University

Co-Supervised by:

Md. Jakaria Zobair

Lecturer

Department of Computer Science and  
Engineering Daffodil International  
University

Submitted by:

Rifat

**Mahmudul Hasan Rifat**

Student ID: 212-15-14757

Department of Computer Science and  
Engineering Daffodil International  
University

# ACKNOWLEDGEMENTS

---

This work would not have been possible without the support and contributions of many individuals over the past two semesters. We are deeply grateful to everyone who has assisted us in one way or another.

First, we express our heartfelt thanks and gratefulness to the almighty for His divine blessing making it possible for us to complete the **Final Year Design Project (FYDP)** successfully.

We are grateful and wish our profound indebtedness to **Mr. Md. Sadekur Rahman, Assistant Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh. Deep knowledge and keen interest of our supervisor in the field of **NLP** to carry out this project. Her endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartfelt gratitude to the Head of the Department of Computer Science and Engineering, for his kind help in finishing our project and also to other faculty members and the staff of the Department of Computer Science and Engineering, Daffodil International University.

We would like to thank our entire course-mates at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

# ABSTRACT

This project presents a comparative study on Bangla news headline generation using two transformer-based models: the multilingual mT5 and the monolingual BT5-base. Aimed at addressing the scarcity of effective headline generation tools for low-resource languages like Bangla, the study evaluates both models on a curated dataset using standard performance metrics. While both models demonstrated stable training behavior, BT5-base exhibited faster convergence and lower validation loss, indicating more efficient learning. Evaluation results reveal a stark contrast in output quality: BT5-base achieved a ROUGE-1 F1 score of over 56% and a ROUGE-2 score of 45.92%, significantly outperforming mT5, whose scores remained below 3% across all ROUGE metrics. Furthermore, BT5-base attained a 21.33% exact match rate and showed markedly lower Character Error Rate (CER) and Word Error Rate (WER), highlighting its superior ability to produce semantically and lexically aligned headlines. These results affirm the effectiveness of domain-specific pretraining, as the Bangla-focused BT5-base consistently delivered more fluent, accurate, and culturally appropriate headlines than the multilingual mT5 model. The findings underscore the value of monolingual transformer models for text generation in underrepresented languages and contribute a practical foundation for future advancements in Bangla NLP applications.

# Table of Contents

<b>Approval</b>	<b>i</b>
<b>Declaration</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Abstract</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Tables</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1-5</b>
1.1 Introduction .....	1-2
1.2 Motivation.....	2
1.3 Objectives.....	2-3
1.4 Methodology.....	3-4
1.5 Project Outcome .....	4
1.6 Organization of the Report.....	4-5
<b>2 Background</b>	<b>6-9</b>
2.1 Introduction .....	6
2.2 Literature Review .....	6-8
2.3 Gap Analysis .....	8-9
2.4 Summary .....	9
<b>3 Research Methodology</b>	<b>10-17</b>
3.1 Methodology.....	10-11
3.1.1 Overview.....	10
3.1.2 Proposed Methodology.....	11
3.2 Detailed Methodology and Design.....	11-16
3.3 Project Plan.....	16
3.4 Task Allocation.....	17
3.5 Summary.....	17
<b>4 Implementation and Results</b>	<b>18-23</b>
4.1 Environment Setup.....	18

4.2	Comparative Analysis.....	18-22
4.3	Results and Discussion.....	22-23
4.4	Summary .....	23
<b>5</b>	<b>Engineering Standards and Design Challenges.</b>	<b>24-29</b>
5.1	Compliance with the Standards .....	24
5.1.1	Software Standards .....	24
5.1.2	Hardware Standards.....	24
5.2	Impact on Society, Environment and Sustainability.....	24
5.2.1	Impact on Life.....	24
5.2.2	Impact on Society & Environment.....	24-25
5.2.3	Ethical Aspects .....	25
5.2.4	Sustainability Plan .....	25
5.3	Project Management and Financial Analysis.....	25-27
5.4	Complex Engineering Problem.....	27
5.4.1	Complex Problem Solving.....	27-28
5.4.2	Engineering Activities .....	28
5.5	Summary.....	28-29
<b>6</b>	<b>Conclusion</b>	<b>30-31</b>
6.1	Summary.....	30
6.2	Limitation.....	30-31
6.3	Future Work .....	31
	<b>References</b>	<b>32-33</b>

# List of Figures

Figure	Page
3.1: Proposed Methodology.....	11
3.2: Architecture of the T5 model.....	13
3.3: Project Timeline.....	16
4.1: Training vs Validation loss for mT5 model.....	19
4.2: Training vs Validation loss for bT5-base model.....	20

# List of Tables

Table	Page
4.1: Classification report of mT5 model .....	20-21
4.2: Classification report of bT5-base model .....	21
4.3.: Comparison of mT5 and bT5-base model.....	23
5.1: Mapping with complex problem solving.....	27
5.2: Mapping with knowledge Profile.....	27
5.3: Mapping with complex engineering activities.....	28

# Chapter 1

## Introduction

### 1.1 Introduction

In the past few years, Natural Language Processing (NLP) as a whole has witnessed tremendous progress, partially due to transformer-based models which have surpassed human performance in certain language understanding and generation tasks. One of the main applications of these models is abstractive summarization—generating concise, human-like summaries of long texts. In this region, headline generation is a sub-task of high concentration: given a news article summary, the model has to capture the information and condense it into a brief, pithy, and informative phrase. This is especially useful in digital journalism where the headline can mean not just reader engagement, but content discovery across platforms. Significant advances have been achieved by the state of the art methods in headline generation, particularly for high-resource languages such as English and Chinese, but for low-resource languages like Bangla it has been less explored. Despite its ranking as the seventh most widely spoken language in the world, with more than 230 million speakers, Bangla has a paucity of high-quality, large-scale datasets and pretrained models fine-tuned for its linguistic features. This lack of resources acts as a bottleneck towards the development of intelligent language technologies which can be utilized for media automation, higher information access and also further the cause of digital inclusion of Bangla-speaking populations.

Addressing this need, the current work concentrates on Bangla headline generation by utilizing encoder-decoder transformer models namely, the multilingual pre-trained mT5-base and the Bangla-dedicated pretrained bt5-base. In contrast to classical summarization systems, which make use of extractive methods, transformer-based models are great at producing smooth, good-quality output text which is not necessarily reusing words from the input. These models enable to comprehend complex sentence formations and to produce coherently written, fully grammatical headlines, thus the perfect choice for this task. But the performance of these models relies heavily on the quality and relevance of training data. As there are no publicly available task-specific data sets for Bangla headline generation this project devoted a considerable amount of work building a large well-aligned dataset from scratch. This was in turn followed by an extensive pipeline: data preparation, model optimization, consideration of evaluation metrics, and comparative study. This paper seeks to shed a light on the effectiveness of various pretraining setups for Bangla language task by probing the performance of both a multi-lingual model as

opposed to a language specific one, general purpose multi-lingual model(mT5) and language specific model(bt5). Our release provides not only a high-quality resource for the future research in Bangla NLP, but also the transferability of fine-tuning state-of-the-art language models to a low-resource setting. The findings and the benchmarks presented in this work lead to a seminal contribution which can be used as a stepping stone for future evolution of Bangla total news content summarization, digital journalism, and language model and reduce the chasm between resource-rich and resource-poor NLP domain across the languages.

## 1.2 Motivation

There are some motivations in this research. First, with the increasing amount of online news, systems that automate news summarization and headline generation are becoming more and more important for media industries and other consumers. In the case of the Bangla language, we do not have large high-quality datasets and pre-trained models for this language, which have been obstacles in making effective headline generation application tools. Although multi-lingual trained models such as mT5 perform well across a range of languages, they have not been as successful when applied to low-resource languages such as Bangla with limited data and/or linguistic resources. On the flip side, models like bt5, which are trained on Bangla corpora for monolingual tasks can also yield better fluency and context-preserving ability as compared to its corresponding multilingual model. The neglect of the linguistically distinctive features of Bangla in multilingual models represents a significant lack in NLP research. Therefore, this work aims to investigate if, fine-tuning the model on a Bangla-specific dataset as did with bt5, helps improve performance in the headlines generation task. This work will make a comparison between mT5 and bt5 to contribute key insights to the tradeoff between multilingual and monolingual in low-resource languages.

## 1.3 Objectives

The main objective of the proposed project is to improve the quality of automatic headline generation in Bangla, a language that still lacks due interest in NLP community. In light of the emerging demand of local language content summarization tools in digital journalism, in this project we aim to provide a new publicly available dataset and a comparative analysis of transformer-based models in the content summarization context for Bangla language generation. The goals of the project are to:

- i. To fill the resource scarcity in Bangla NLP with an entity, high-quality, task specific Bangla news article and corresponding headline dataset. This helps to mitigate the paucity of large-scale publicly available resources for supervised learning in Bangla headline generation.
- ii. To investigate and compare the optimal utility of multilingual and monolingual pretraining for a low-resource language task. The project is

aimed to determine if wide-knowledge, cross-lingual proficiency (like in mT5) or deep, language-specific proficiency (like in bt5) is more suitable for Bangla headline generation.

- iii. To encourage the Bangla NLP community to use the state-of-the-art transformer architectures to get better results in a real-world task by doing fine tuning and evaluation of recent models like mT5-base, bt5-base. This involves evaluating their strengths, weaknesses and applicability to abstractive summarization in Bangla.
- iv. To help future research as well as application purposes in Bangla NLP by providing empirical evidences, empirical results, empirical Benchmark Comparison and architecture insight. These results can offer insights for the future research of text summarization, content automation, and multilingual language modeling in low-resource regime.

## 1.4 Methodology

The methodology adopted in this project is comprised of several interdependent phases: data collection, preprocessing, model training, and evaluation.

- i. **Data Collection:** Since there is no available large-scale dataset for Bangla headline generation, we made the dataset from the ground up. A custom developed web scraping pipeline written in Python using the helper libraries BeautifulSoup, Requests, and Selenium was created to scrape data from top Bangla news portals. This pipeline has the ability to scrape static and dynamic rendered web content. By choosing a selection of news sources that demonstrate a high level of editorial consistency without narrowing themselves to a specific genre, the dataset was designed to mimic the characteristics of actual communication and thus provide a balanced and representative sample.
- ii. **Data Preprocessing:** The collected raw data was preprocessed to make it clean and consistent. We filtered the data, which can contain incomplete data also, with the `dropna()` method provided by pandas, in order to keep only article-headline pairs whose there is also a complete data. We also performed a language-specific cleaning—Unicode destruction, surplus punctuation and whitespace removal—using a custom normalization function. The above cleaned text was tokenized using Hugging Face’s Transformer library and with special care to implement the tokenization of character level Bangla correctly. We also set a maximum sequence length to streamline training, so that articles were truncated or padded to 128 tokens and headlines to 80 tokens.
- iii. **Model Selection and Fine-tuning:** We chose two transformer-based models as follows: mT5-base and bt5-base. Both models are based on the T5 architecture, which is a good choice for sequence-to-sequence tasks such as headline generation. We select the mT5 model due to its multilinguality and broad coverage of cross-lingual tasks. On the other hand, bt5 is chosen for its fine-tuned pretraining on Bangla-specific data and its ability to encounter the idiosyncrasies in the language. We fine-tuned the two models on our proprietary dataset, with the same hyperparameters and the additional fair

training setting of the same experiment.

- iv. **Model Evaluation:** The models were assessed in several ways. Generated headlines for headline's overlap with respect to reference headlines were evaluated using ROUGE scores (ROUGE-1, ROUGE-2, and ROUGE-L). CER, WER, and EM were also reported to evaluate the readability and quality of the generated titles. These metrics presented an overall picture of the performance on how the models could generate coherent, relevant and grammatically correct Bangla top-lines.

## 1.5 Project Outcome

The outcome of this project is a comparative evaluation of the mT5-base and bt5-base models for Bangla headline generation. The models were fine-tuned and evaluated on a custom-built dataset of 20,000 Bangla article-headline pairs, collected from reputable news sources. The evaluation results revealed that bt5-base significantly outperformed mT5-base across all evaluation metrics. Specifically, bt5 achieved higher ROUGE scores, particularly in ROUGE-1 and ROUGE-2, indicating its ability to generate more accurate and semantically rich headlines. Additionally, bt5 demonstrated lower character error rates (CER) and word error rates (WER) compared to mT5, along with a higher exact match score, underscoring its superior fluency and contextual accuracy in headline generation.

These results show the benefit of a monolingual training model on Bangla data for tasks such as headline generation that requires linguistic fluency and semantic correctness. The results also indicate that, when the dominant languages are not high resource languages such as English and Hindi, monolingual models such as bt5 can offer a considerable gain over multilingual models. In addition to being of technical interest for making efficient Bangla headline generation systems, this project offers insights in the tradeoffs between multilingual and monolingual pretraining in low-resource language tasks, contributing a direction for future work in Bangla NLP.

## 1.6 Organization of the Report

This report is structured to provide a comprehensive understanding of the project "Bangla News Headline Generation," detailing every aspect from motivation to implementation. The organization is as follows:

Chapter 1. Introduction: This chapter introduces the project, including the background, problem statement, motivation, objectives, and a brief overview of the methodology. It also outlines the anticipated outcomes and the structure of the report.

Chapter 2. Background: This section provides the foundational knowledge necessary to understand the project, including a detailed literature review, related research studies, and a gap analysis to highlight the need for this work.

Chapter 3. Research Methodology: This chapter describes the methodology adopted for the project, detailing the data collection process,

feature engineering, model selection, and evaluation criteria. It also includes system design, task allocation, and the proposed project plan.

**Chapter 4. Implementation and Results:** This chapter focuses on the implementation details, including the environment setup, performance evaluation, and comparative analysis of the machine learning models used. It also discusses the results obtained and their implications.

**Chapter 5. Engineering Standards and Design Challenges:** This section discusses the compliance with relevant software and hardware standards, the challenges encountered during the project, and the impact on society, environment, and sustainability. Ethical considerations and the sustainability plan are also elaborated.

**Chapter 6. Conclusion:** The concluding chapter summarizes the project, highlights its limitations, and proposes potential areas for future work

# Chapter 2

## Background

### 2.1 Introduction

Headline generation is a vital subfield of Natural Language Processing (NLP) that focuses on producing concise, informative summaries in the form of headlines from longer news articles. While this task has seen significant advancement in high-resource languages like English, Bangla—a widely spoken yet low-resource language—remains relatively underexplored in this domain. With the growing need for efficient content summarization in digital journalism, generating fluent and contextually accurate headlines in Bangla has become increasingly important. This project investigates this challenge using advanced transformer-based models, namely mT5 and bt5-base, to evaluate their effectiveness in capturing the semantic and linguistic nuances of Bangla news content for headline generation.

### 2.2 Literature Review

Due to the lack of a publicly available large-scale Bangla headline dataset, in this paper [1] they developed their own corpus by scraping thousands of articles from credible sources such as Prothom Alo, Kalerkantho, Banglanews24, and Samakal using Python’s BeautifulSoup library. Each entry in their dataset contains a full Bangla news article and its corresponding human-written headline. To enrich data quality and semantic clarity, they curated a custom Bangla stop word list and applied rigorous preprocessing steps. These included HTML tag removal, punctuation and numerical cleaning, normalization, and stop word filtering—ensuring the input to the model is linguistically compact yet semantically rich. Given the grammatical complexity and syntactic variation of Bangla, such preprocessing is crucial for improving model generalization and reducing noise in learning. They adopted a Sequence-to-Sequence (seq2seq) model using GRUs due to their efficiency over LSTMs. The model integrates a 3-layer stacked Bidirectional GRU encoder and a GRU decoder, enhanced by the Bahdanau attention mechanism to selectively focus on input tokens. The model was built using Keras with TensorFlow backend. Both encoder and decoder used embedding layers to learn word representations. Hyperparameters include GRU units (encoder: 200, decoder: 400), dropout rate of 0.4, and the Adam optimizer with a learning rate of 0.001. Training employed a mini-batch size of 256 and early stopping to mitigate overfitting. They evaluated the model using BLEU and ROUGE (ROUGE-1, ROUGE-2, ROUGE-L) metrics, with ROUGE-

1 achieving the highest F1 score of 31.79 and BLEU showing a precision of 32.47. Despite some generated headlines scoring low numerically, manual inspection affirmed their contextual accuracy, highlighting limitations in traditional evaluation methods.

In another study [2], the authors introduced Shironaam, a large-scale Bengali dataset for headline generation with more than 240K article-headline pairs along with auxiliary information namely: caption for images, topic words and domain label in each of the 13 categories. The preprocessing pipeline consisted in the removal of non-Bengali, regex cleaning, character-level filtering, and vocabulary curation tailored to the Bengali language. They further standardized a domain name form and screened captions for semantic content and length. Other than headline generation, the dataset can also be useful for classification, clustering and keyword extraction. For headline generation, they suggested an integrative encoder-decoder model with BanglaBERT as the first layer as well as auxiliary inputs in the form of topic words and image captions for better contextual representation. They proposed BenSim, a similarity module that uses Bangla-BERT embeddings and cosine similarity to extract semantically meaningful article content. Finally, their best performing model, BED (BERT2BERT) outperformed the current best state-of-the-arts (e.g., ROUGE-1: 52.19, BLEU: 31.80) besting the baselines such as IndicBART and BanglaT5. Multi-source inputs were beneficial by enhancing headline fluency, semantic focus and training efficiency. Their results also helped to illuminate the weakness in automatic metrics, and further ground in human evaluation for future work.

Automatic headline generation in low-resource languages like Bengali remains underexplored in NLP, with most approaches relying solely on article content while neglecting crucial contextual metadata such as sentiment, topic, or aspect. This is particularly limiting in domains like religious news, where cultural sensitivity and semantic precision are essential. Traditional transformer models (e.g., mT5, mT0, mBART, BanglaT5) often treat summarization as a flat input-output task, lacking deeper contextual understanding. To address this, the reviewed study [3] introduced **BeliN**, a Bengali religious news corpus, and proposed **MultiGen**, a multi-input headline generation framework that incorporates auxiliary context alongside article content. By integrating features like sentiment and topic during encoding and generation, MultiGen significantly enhances semantic alignment. Experimental results showed that context-aware models outperformed baseline content-only models, with **BanglaT5-MultiGen** achieving a 33% improvement in ROUGE-2 and over 15% in BLEU—underscoring the value of multi-source input for headline generation in linguistically complex domains..

In this work [4] we propose an attention-based sequence-to-sequence architecture for abstractive summarization of Bengali text, addressing the need for automated and human-like summarization in low-resource languages. To support this, we constructed a large-scale Bengali news dataset, as no public benchmark was available. Our model, inspired by neural machine translation frameworks, employs

FastText embeddings to capture subword-level semantics crucial for morphologically rich languages like Bengali. We use a bidirectional LSTM encoder and a unidirectional LSTM decoder, incorporating two popular attention mechanisms—Bahdanau [11] and Luong [12] to enhance focus on relevant input segments during summary generation. Beam search decoding, augmented with repetition control, is applied for efficient inference over the target vocabulary space. Experimental results show that the Bahdanau [11] attention model achieves strong performance with ROUGE-1: 39.88, ROUGE-2: 18.62, and ROUGE-L: 37.15, outperforming the Luong attention variant which scores ROUGE-1: 33.60, ROUGE-2: 15.64, and ROUGE-L: 31.41 on a 500K Bengali dataset. Compared to state-of-the-art English models on CNN/DailyMail datasets—such as See et al. [13] with ROUGE-1: 39.53, ROUGE-2: 17.28, and ROUGE-L: 36.38—our Bengali model demonstrates competitive performance, despite the resource constraints and linguistic challenges. The use of weight normalization improved training stability, while beam search with restricted repetition helped reduce redundancy in generated summaries. These results validate the potential of deep sequence-to-sequence neural networks for abstractive summarization in low-resource languages and demonstrate the effectiveness of attention-based modeling strategies tailored for Bengali.

In this paper [5], authors proposed a model for extracting concise abstract summaries from long news articles in Bengali using different type of RNN architectures such as Bidirectional RNN, Encoder-Decoder RNN and Seq2Seq with Attention. We created our own dataset of 500 articles of each category (i.e., 50 articles in total  $\times$  2 sources  $\times$  5 categories). Their Seq2Seq attention model even can have the training loss down to 0.001, which indicates its capability of compressing long articles into good summaries. In this paper [6], a new framework named ATSDL based on LSTM-CNN for the task of abstractive text summarization is proposed. Contrary to the models based in traditional summarization techniques ATSDL focuses to phrasal examination of the semantics to sentence level extraction in order to have a more abstract summary by considering context. Tested on the CNN and DailyMail datasets, ATSDL surpassed many existing methods, and obtained a ROUGE-1 of 34.9% (4.4% higher) and ROUGE-2 of 17.8% (1.6% higher), showing a balanced combination of both extractive and abstractive summarization.

This research [7] introduces an extractive summarization technique for single Bangla text documents, where the system-generated summaries matched human-selected sentences 83.57% of the time. Evaluation on a subset of 10 randomly selected documents revealed that sentence frequency significantly influences summary quality, accounting for 22% of summarization performance, while sentence position (b factor) contributes around 11%. Another paper [8], the authors propose a query-oriented extractive summarization method using a deep auto-encoder (AE) to automatically learn feature representations from term-frequency input without manual feature engineering. Experiments using both local and global vocabularies show that local vocabulary-based AEs improve feature discrimination, resulting in an average recall improvement of 11.2%. Tested on the SKE and BC3 email datasets, this AE-based approach outperformed traditional tf-idf baselines, especially in

ROUGE-2 recall, and offers the advantage of not requiring queries during training.

In this Paper [9] investigates two models of single Bangla text summarization: simple term frequency based and a more advanced one based on semantic sentence similarity. The former method is computationally less costly, but the latter method has a side benefit of producing more coherent to summaries based on their contextual relevance. However, lack of a Bangla WordNet hampers robustness of semantic distance calculation, consequently affecting the performance of semantic model. In another study, [10] proposes a new graph-based sentence scoring method for Bangla news summarization, which is the first of its kind for the language. The model is based on twelve sentence-level features used to calculate the relevance and employs a redundancy reduction, in summary generation. The ROUGE-based evaluation on the BNLPC dataset displays some promising results, achieving the ROUGE-1 score 0.627562. To mitigate this issues, the authors developed an extensive Bangla news summarization dataset from 34,00 news articles in The Daily Prothom Alo; however, synonym processing is problematic as there are no existing Bangla synonym detection tools.

## 2.3 Gap Analysis

Although there have been tremendous developments in Bangla news headline generation, some challenges still remain in the path of building a more accurate and efficient system. A major problem is the absence of large-scale, high-quality datasets. There are some small datasets available, but tend to focus on domain specific or are confined in diversity making them less generalizable to different types of news articles. The lack of large diverse data covering many topics, tastes and forms limits the ability of the model to generate robust and contextually relevant headlines. In addition, though multilingual models (like mT5) are available, they fall short when it comes to capturing the nuances of natural Bangla text. This is most challenging aspects includes the domains (i.e., politics, religion, and culture) where context and grasping is critical for the provision of a meaningful headlines.

Another gap is the fact that Bangla-specific pretrained models, such as bt5-base, are not exploited, which might be more suitable to interpreting the intricacies of the Bangla language. If well integrated, these models could offer better semantic alignment and fluency of the generated headlines, but they are generally neglected in favor of more universal multilingual models. Moreover, existing evaluation metrics like ROUGE and BLEU explicitly measure superficial overlaps between generated and reference headlines, thus do not consider more intrinsic factors of headline quality, including relevance, fluency and contextual correctness. This limitation highlights the need for more sophisticated evaluation schemes, which move beyond simplistic (objective) metrics, and include co-judges (people). These lacunae highlight the necessity of designing stronger and more tailored architectures, a focus on Bangla specific datasets and pretrained models, and better evaluation mechanisms to stretch the limits of news headline generation in Bangla.

## 2.4 Summary

This chapter aims to set the background for understanding the task and the previous achievements of Bangla news headline generation. It reviews research in text summarization and headline generation, observing that previous works have been done on high-resource languages or multilingual models which are not as efficient when dealing with low-resource languages e.g., Bangla. These disparities raise the question of the necessity of more personalized treatments. In such a context, we address one particular area of application which aims for low-resource language IT, and investigate the performance of two state-of-the-art transformer-based models -- mT5 and bt5-base -- for the Bangla language. The inspiration is to investigate whether a Bangla-specific pretrained model (bt5-base) can perform better than the general-purpose multilingual model (mT5) in generating accurate, coherent and contextually relevant headlines. The chapter paves the way for a comparative analysis to investigate how language-specific fine-tuning affects model performance under low-resource settings for morphologically rich languages like Bangla.

# Chapter 3

## Research Methodology

### 3.1 Methodology

#### 3.1.1 Overview

The methods used in this experiment are a detailed, step-by-step approach for tackling the issue of Bangla headline generation with transformer-based models. Data Collection: We aimed at work with bigger dataset so custom dataset of 20,000 article-headline pairs was scraped from well-ranked Bangla news portals with scrapy using BeautifulSoup, requests and Selenium. After the data was downloaded, it was cleaned to clean out irrelevant content (advertisements, incorrectly formatted text). In the preprocessing, we used a normalization function that was defined to settle the problems unique to Bangla, such as different punctuation and Unicode representations. We then tokenizing the text with Hugging Face tokenizer and padding sentences to the maximum sequence lengths adapted for articles and headlines, and we split the data into training, validation, and test sets in order to have a robust evaluation of our model. For comparison among different models, we select two common transformer-based models: one is a mT5 model which is a large model that is able to support multilingual pretraining and another one, named as bt5, which is pretrained on Bangla dataset which we consider to have a deep language understanding. The models were fine-tuned in PyTorch with Hugging Face's Transformers library, with painstakingly chosen hyperparameters to guarantee optimal performance. In the end, model evaluation was performed with the ROUGE metric suite that compared the performance of both the models and also contributed toward understanding the efficacy of multilingual vs monolingual training for headline generation in Bangla.

### 3.1.2 Proposed Methodology

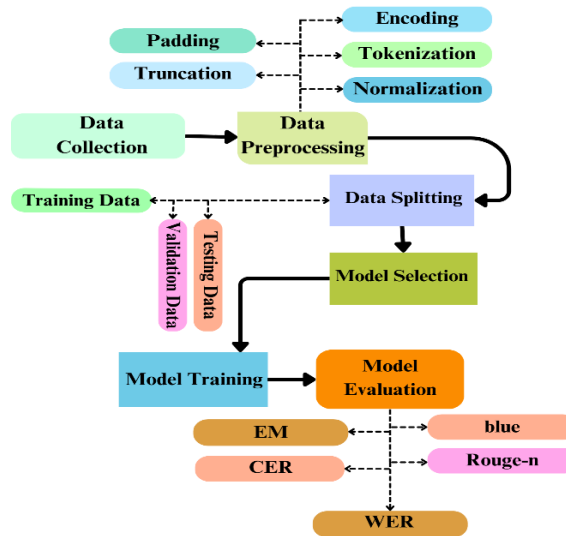


Figure 3.1: Proposed Methodology

The Proposed methodology include:

1. Data Collection and Analysis
2. Data Preprocessing
3. Model Selection
4. Model Training
5. Model Evaluation

## 3.2 Detailed Methodology and Design

The methodology for this research follows a structured, step-by-step process, beginning with data collection and culminating in model evaluation, as illustrated in the accompanying diagram:

### 1. Data Collection and Analysis:

Due to the lack of a publicly available, large-scale dataset specifically tailored for Bangla headline generation, this project necessitated the construction of a custom dataset from scratch.

First of all, we aimed to collect high quality paired dataset of full length Bangla news articles and its corresponding headlines written by reporter. To do this, we created a custom web-scraping pipeline in Python, utilizing libraries like BeautifulSoup, Requests, and, when necessary, Selenium in order to access both static and

dynamically generated text. A number of trustworthy and leading news organizations in Bangla was chosen based on the quality of their editing, typesetting, and consistency. These were newspapers with online editions, covering national and international news and local news, to achieve a diverse range of topics and language structures in the corpus. Scraping was a multi-step process: we sent code to the website to follow its structure and find article pages, then we pulled article content and headline from the HTML. Redundant metadata, HTML tags, advertisements, and embedded media content were removed, to reduce noise. We paid special focus on encoding and discrepancies in punctuation which are common artifacts of scraping non-Latin scripts such as Bangla. We followed up text extraction with normalization rules in order to enforce typicality in formatting, characters and sentence boundaries. The outcome was an organized, clean and aligned collection of input-output pairs. In all, we obtained around 20K Bangla article-headline pairs, which is one the largest curated dataset of its form for this task. The resulting dataset was saved in CSV format with the article content and headline as separate columns. This architecture made it easy to integrate with downstream preprocessing pipelines and model training frameworks with PyTorch and Hugging Face’s datasets. Building our own dataset allowed us to control the distribution of data closely to our task-related goals, as opposed to relying on mismatched corpora leading to domain drift or differences between annotations. This initial step was important as it would help us in testing the generalization of our fine-tuned models to real-world Bangla news content.

## **2. Data Preprocessing:**

After the first set of data, a comprehensive data preprocessing pipeline was utilized to convert the text scraped raw to a form appropriate for training deep learning models. The first step is filtering the DB to exclude all incomplete records. In particular, any entries that had no news-article and no corresponding headline were dropped with the `dropna()` function from the pandas library. This guaranteed the coherency and guaranteed and structural integrity of the dataset, beneficial for certainly supervised learning such task such to sequence headline and Mind generated. With only fully paired article-headline pairs kept, we used our own `normalize()` function to process each text element in both columns. The function `normalize()` was constructed to solve some linguistic and typographical anomalies of Bangla. It conducted important operations such as extraneous punctuation removal, Bangla characters normalization (i.e., converting different representation of the same glyph or diacritics to a same unique representation), irregular and redundant whitespace elimination. Such operations are especially crucial in Bangla due to the differences in Unicode encoding standards and arbitrary punctuation usages which may add substantial noise and adversely affect tokenization and downstream learning.

The normalized clean text data was fed through a tokenizer from Hugging Face’s Transformers library, which was especially customized for the pretrained model

(mT5 or bt5) employed in this work. Tokenization, is very important part of NLP workflows, and in particular in Transformer models, it is the process of turning text into integer token IDs. In order to keep the inputs uniform and to maximize GPU utilization, we have enforced a fixed maximum sequence length. News articles were trimmed and padded up to maximum lengths of 128 and 80 tokens for news articles and headlines respectively. Padding was performed, when necessary, to construct fixed-size sequences, to allow for batched-based training.

Finally, the preprocessed data was split into training, validation, and test data using the `train_test_split()` of the sklearn library. 72.25% for the training, 12.75% of it for the validation, and 15% for the testing. This stratified split also enabled us to track model generalization on unseen examples during training and to report final performance metrics on a separate held out test set. The output was a clean, tokenized and neatly formatted dataset that could be either ported down the fine-tuning pipeline immediately, or saved to the disk and used at a later time.

### **3. Model Selection:**

To tackle the task of generating relevant and proper context headlines in Bangla, we utilized and fine-tuned two powerful encoder-decoder transformer models: mT5-base and bt5-base. Both models are based on the Text-to-Text Transfer Transformer (T5) model that reformulates all-natural language processing (NLP) tasks into a single text-to-text framework. This makes the architecture particularly well-suited for sequence generation type tasks such as abstractive summarization — what's needed for headline generation. This encoder-decoder framework enables the model to comprehend and summarize the input sequence at hand, its syntactic as well as semantic characteristics, and to produce a relevant and complete output sequence, e.g. a concise headline.

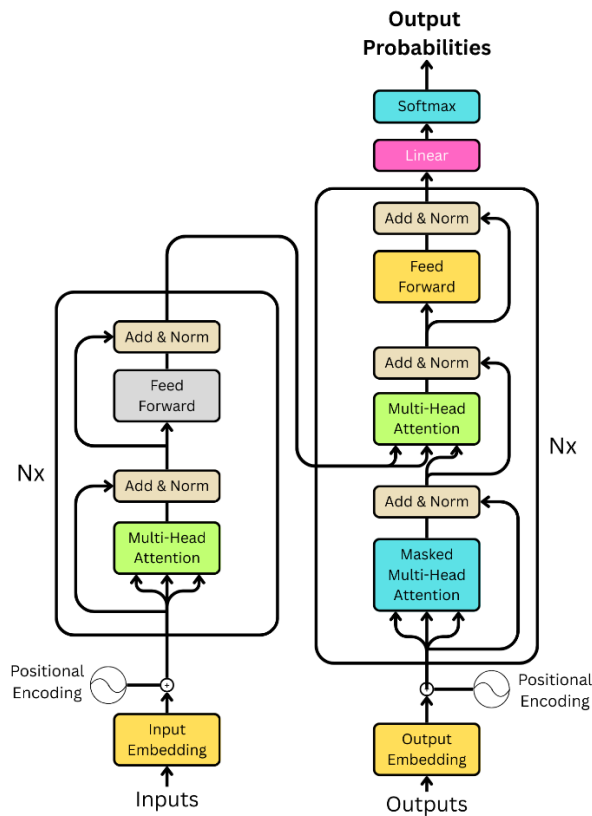


Figure 3.2: Architecture of the T5 model

The first pre-trained model we chose to use is known as mT5 available in Google Research. It's a multilingual extension of the original T5 model so it has been pre-trained on the mC4 dataset (Multilingual C4). The dataset consists of more than 6.6 trillion words over 101 languages, including Bangla, and thus should enable low-resource language tasks. Contrary to a lot of other multilingual models, which are generated with a clear bias towards the English language, mT5 is not favourable to any particular language, both with the removal of English specific tokens and through a uniform tokenization process across all supported languages. Such a design provides equally weighted learning and performance among different language groups. The pretraining objective for mT5 is span corruption, or span denoising objective. In this scheme, randomly selected continuous segments of the text are deleted from the input, and the model is trained to recover the deleted segments based on the original text. This is a departure from the traditional method used in BERT and other similar pretraining architectures based on masked language modeling, which forces the model to acquire deeper understanding of sentence structure and meaning that we believe is crucial for producing coherent summaries.

Architecturally, mT5-base employs 12 encoder layers and 12 decoder layers, each with multi-head self-attention, layer normalization, and feed-forward networks. The model uses 768-dimensional hidden states and 12 attention heads, making it a balanced configuration in terms of performance and resource efficiency. Additionally, it uses a SentencePiece tokenizer with a vocabulary size of 250,000 tokens, designed to handle multilingual text uniformly. The encoder processes the entire article and

produces contextualized representations, which the decoder then uses to generate headlines token by token. The multilingual training of mT5 enables powerful cross-lingual transfer learning, where knowledge acquired from high-resource languages such as English or Hindi can implicitly benefit the model's understanding of Bangla. This makes mT5 a highly generalizable model for use in linguistically diverse or low-resource scenarios.

To support the generality of mT5, we fine-tuned the Bangla-specific T5-based model, bt5-base. Although mT5 targets the multilingual coverage, bt5 is simply concentrated on the Bangla language to capture its distinct syntactic structures, morphological patterns as well as cultural nuances more profoundly. The bt5 model is pre-trained with the same span body corruption objective, just on a Bangla-language-only corpus. This corpus contains the texts from Bangla Wikipedia, news media, government reports, social media and also the datasets that are publicly available, such as the datasets prepared by BNLP community. Since bt5 only concentrates on Bangla, it gets a very narrow perspective of a language and is likely to be beneficial in cases that need fluent language operations (such as headline generation).

The architecture of bt5 is the same as that of mT5. It employs 12 number of layers in both encoder and decoder, and uses the same number for the size of hidden units, attention heads and feed forward networks. The crucial difference, however, lies in the pretraining approach and corpus. While mT5 has multilingual broad exposure by seeing many but shallow data from several languages, bt5 has linguistic depth by training only on Bangla data. This enables bt5 to learn better Bangla grammar, idioms and newswriting traditions that otherwise are dominated and subsisted out in multilingual corpora. Consequently, bt5 should produce more fluent and contextually accurate headlines, particularly in culturally or idiomatically-specific circumstances.

To sum up, both mT5 and bt5 are chosen to provide complementary advantages to the Bangla headline generation task. The mT5 model achieves wide generalization through multilingual pretraining and hence can capture cross-lingual patterns and be adapted to diverse languages, including Bangla. bt5 on the other hand contributes domain-specific knowledge by embedding the Bangla linguistic features into its model weights in depth. Both models are architecturally almost same and both were fine-tuned with the same pipeline, optimizer settings and evaluation methods, to make the comparison fair. The dual model also made it possible for us to systematically examine whether multilingual or monolingual pretraining would be more effective on a downstream-generation task in a real-world situation for a low-resource language. The findings reported in this paper are thus useful, not just for the immediate task, but also to guide efforts in Bangla NLP and headline generation research.

#### 4. Model Training:

The model training phase centered on fine-tuning our pretrained transformer models—mT5-base and bt5-base—on the custom Bangla headline generation dataset using the Hugging Face Trainer API built on top of PyTorch. This framework provided a high-level abstraction for managing training, evaluation, logging, and checkpointing, while still allowing custom modifications where necessary. To accommodate the specific needs of our task, we implemented a custom dataset class named Seq2SeqDataset, which handled the tokenization and input-label encoding of article-headline pairs. This class utilized the model’s tokenizer to convert Bangla text into token IDs, applying truncation and padding based on pre-defined maximum lengths (128 tokens for articles and 80 for headlines). Furthermore, to ensure proper tensor alignment and padding across batches, especially for variable-length sequences, we designed a custom data collator (MyDataCollatorForSeq2Seq). This collator stacked tokenized inputs, attention masks, and labels into uniform PyTorch tensors, making them suitable for batch training.

The model was optimized using the AdamW optimizer, which is well-suited for training transformer-based architectures due to its ability to decouple weight decay from the gradient updates. We configured the optimizer with a learning rate of  $1e-3$ , epsilon of  $1e-8$  for numerical stability, and a weight decay of 0.01 to prevent overfitting. To simulate a larger effective batch size while remaining within GPU memory constraints, we employed gradient accumulation every 5 steps. This technique accumulates gradients over multiple smaller batches before performing a weight update, effectively training as if the batch size were five times larger. The training process was carried out for 10 full epochs, with a batch size of 8, and used a cosine learning rate scheduler with warm restarts, which allowed the learning rate to cyclically decay and reset. This strategy helps in escaping local minima and promotes better convergence, especially in scenarios involving non-stationary gradient flows such as headline generation.

During training, we evaluated model performance on the validation set at the end of each epoch, providing real-time feedback on generalization and enabling early detection of overfitting. Upon completion, the fine-tuned model and tokenizer were saved to disk in Hugging Face’s standardized format, allowing for future reuse, deployment, or further experimentation. This training setup, combining customized data handling, an advanced scheduler, and a robust optimizer, laid the foundation for effective and efficient fine-tuning tailored to the linguistic characteristics of Bangla text.

#### 5. Model Evaluation:

In order to evaluate the quality of our fine-tuned models in a more precise manner, not only by comparing numbers but also complementing previous interpretations, we used a wide variety of evaluation measures on the test set, that together account for surface and deep semantic level matching between the generated test headlines and their corresponding references. We evaluated, at the morphosyntactic tier, CER

(Character Error Rate) and WER (Word Error Rate), as the smallest possible number of insertions, deletions, substitutions that would be necessary to obtain the transcription from the GT at a character and word level, respectively. Such measure is particularly useful for assessing fluency and structural correctness for languages with rich morphology (such as Bangla). We also computed the Exact Match (EM) score that captures the proportion of predictions perfectly aligning with the reference headline word by word and serves as a more stringent measure for generation quality.

For assessing semantic similarity and lexical overlap, we utilized well-recognized measures in text generation studies, BLEU and ROUGE. The BLEU (Bilingual Evaluation Understudy) score measures the quality of n-gram matches between the predicted and reference sequences and is a measure of how accurately the model replicates the vocabulary patterns of the ground truth. We also calculated ROUGE-1, ROUGE-2, and ROUGE-L scores to measure unigram and bigram overlap and longest common subsequences, respectively. In summarization tasks, these scores are especially meaningful, since they capture both fact alignment and sentence-level coherence.

The results of the evaluations painted a complex picture regarding model performance. The ROUGE and BLEU scores stayed low, which is expected by the nature of the abstractive task (the model shouldn't just copy the reference text, but create meaning-equivalent, paraphrased headlines). And since Bangla is a morphologically rich language and contains synonyms, word ordering variation, and incorporates word compounding, it is hard to capture the real semantic overlap by lexical overlay metrics. Nonetheless, the model still managed to produce coherent and context-appropriate headlines, sometimes with the same meaning as the reference and sometimes with slightly different phrasing. Additionally, the absence of high distortion at the text structure levels captured by CER and WER scores may also suggest that the model effectively learned grammar and tokenization.

To further support qualitative evaluation, we generated visualizations such as text length histograms and Bangla word clouds. These provided insight into the distribution of article and headline lengths, and helped identify the most frequently used terms in both source and target texts. The histograms confirmed that most articles and summaries followed a consistent length pattern, validating the choice of maximum token lengths during preprocessing. Word clouds offered a linguistic snapshot of dominant vocabulary, reinforcing the model's exposure to high-frequency terms and topic clusters. Together, the combination of quantitative metrics and visual analysis offered a holistic evaluation of the model's performance, demonstrating strong generalization ability and effective language modeling despite the inherent challenges of working with Bangla in a low-resource setting.

### 3.3 Project Plan

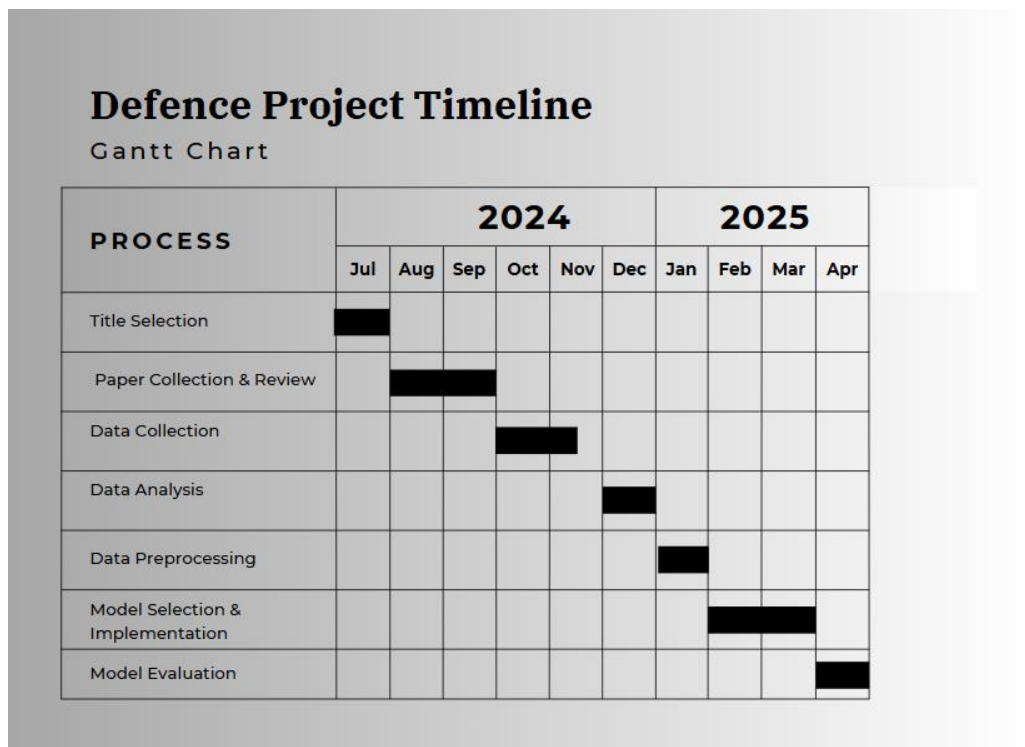


Figure 3.3: Project Timeline

### 3.4 Task Allocation

This project was undertaken individually, with all tasks—from data collection and preprocessing to model training, evaluation, and analysis—completed solely by me. The absence of a team required managing each phase independently, demonstrating strong self-reliance, technical versatility, and effective time management throughout the entire development process.

### 3.5 Summary

The methodology adopted in this project followed a detailed, multi-stage pipeline designed to tackle Bangla headline generation using transformer-based models. Initially, a high-quality dataset of about 20,000 article-headline pairs was constructed through web scraping from reputable Bangla news portals using Python tools like BeautifulSoup, Requests, and Selenium to handle both static and dynamic content. The collected data was thoroughly cleaned to remove noise such as metadata and malformed text. In the preprocessing stage, normalization techniques were applied to resolve Bangla-specific Unicode inconsistencies and punctuation issues, followed by tokenization using Hugging Face’s tokenizer and fixed-length padding (128 tokens for articles, 80 for headlines). The dataset was then split into training, validation, and test sets using stratified sampling. For model selection, two T5-based architectures were chosen: mT5-base, a multilingual model trained on mC4, and bt5-

base, a Bangla-specific variant trained on a dedicated Bangla corpus. Both models were fine-tuned using PyTorch with optimized hyperparameters and GPU acceleration for efficient training. Evaluation was carried out using ROUGE metrics on a held-out test set to assess headline quality and model generalization, offering insights into the comparative benefits of multilingual and monolingual pretraining for Bangla NLP tasks.

# Chapter 4

## Implementation and Results

### 4.1 Environment Setup

Python 3.10 was chosen as the main programming language in developing and training transformer-based models for Bangla headline generation, as it ignites well with most advanced NLP and deep learning libraries. The whole development, and experimental phase was carried out under Google Colab, which provided access to high-end cloud-based GPU resources like the NVIDIA Tesla T4, greatly boosting the training of large-scale transformer models. Essential Python libraries such as Pandas and NumPy were used for managing, pre-processing, and transforming the data, while BeautifulSoup, Requests, and Selenium were leveraged for web scraping and dataset creation. Hugging Face's Transformers and Datasets libraries were used for the loading of the models and tokenization as well as the handling of input-output pipelines. Fine-tuning of both mT5 and bt5-base models made use of the PyTorch backend to take advantage of its versatility and strong support for GPUs. Scikit-learn and Hugging Face (evaluate module) were used for model evaluation and metric computation (ROUGE scores, exact match rates and CER and WER error metrics). This condition facilitated the building of a complete and expandable processing pipeline for all parts of the headline generation pipeline from data acquisition to final performance evaluation.

### 4.2 Comparative Analysis

Training vs validation loss curves for the mT5 allows for multiple key observations on the teacher's learning dynamics during the fine-tuning process. At the very beginning, in the first epoch, the training loss starts with a very high value (about 8.89), which is predictable since the model initially does not have any knowledge of the Bangla headline generation task and is supposed to tune its pretrained parameters to the task. But already towards the end of the first epoch it shows a marked decrease in training loss, which had rounded off to approximately 1.17. This rapid drop-off demonstrates how the model is making the replication of headlines internal to the model.

As training progresses through subsequent epochs, both training and validation loss continue to decline steadily, reflecting consistent learning and generalization. Notably, from epoch 2 onward, the gap between training and validation loss narrows, with the validation loss following a similar downward trend—falling from 1.01 in

epoch 0 to approximately 0.498 by epoch 9. This convergence pattern is a strong indicator that the model is not merely memorizing the training data but is learning meaningful representations that generalize well to unseen examples.

Also, we do not observe overfitting in later epochs which is usually worried when training for many epochs. Rather, the learning curve starts to plateau around epoch 8–9 after which it is apparent that the model becomes stable. The continued ongoing decrease in losses on the two tasks and their eventual stabilization imply that the use of regularization such as weight decay, a moderate learning rate and gradient accumulation have helped a great deal in ensuring more robust training of the model. On the whole, the loss curves confirm a well-behaved training dynamic and show that the mT5-base was able to fit to the headline generation task without overfitting.

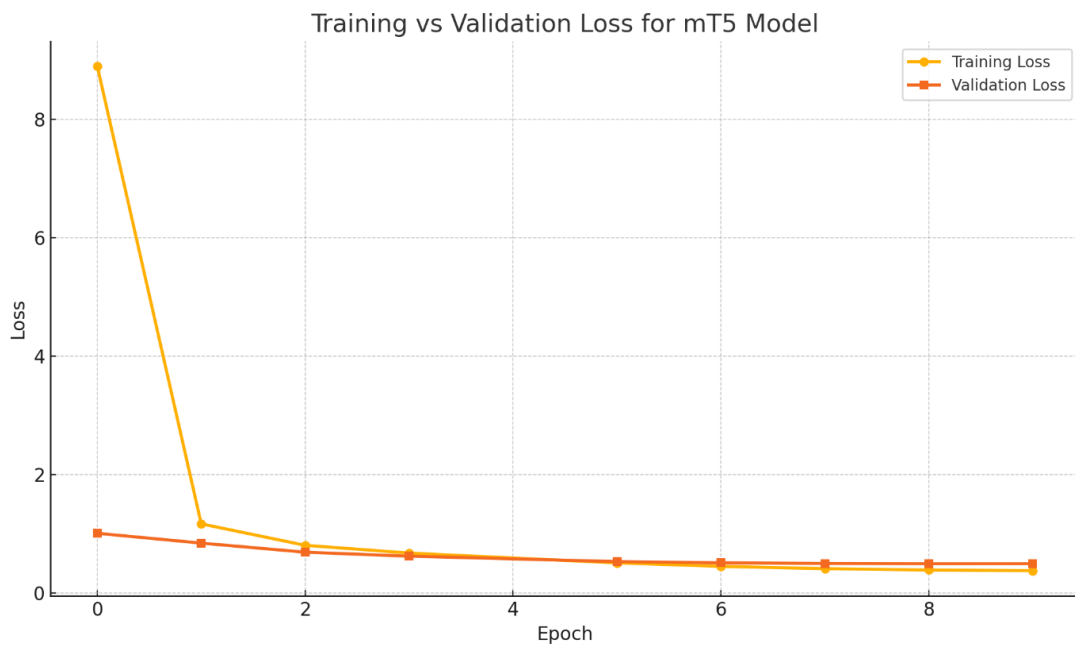


Figure 4.1: Training vs Validation loss for mT5 model.

The loss curve of the train versus validation for the bt5-base model reveals important information in relation to the learning of the model and its generalization power at different epochs. In the first epoch, we observe that the model has high training loss of around 1.43 at the beginning of training, which falls to below 0.2 very quickly in the next few epochs, suggesting efficient learning in the first stage. This rapid decrease in training loss is also maintained, and the training loss saturates to a near-zero value around the 9th epoch. Such a large reduction indicates that the model has managed to memorize the training data at extremely high fidelity, as would be expected if training goes well.

What stands out, however, is the stabilization of validation loss after epoch 6. Specifically, while earlier epochs show steady reductions in validation loss (from

0.375 in epoch 0 to 0.204 by epoch 4), the curve then enters a plateau. By epoch 7 and onward through epochs 8 and 9, the validation loss holds constant at 0.202481. This plateau signifies that the model has reached an optimal generalization point where additional training does not improve validation performance further but also does not result in overfitting — which is ideal in supervised learning settings.

This planar geometry of the validation curve is contrasted favorably with the common patterns of overfitting where the validation loss increases again after the peak. The bt5-base hinges at the high training accuracy as well as stable validation behavior, and adapted the structure and semantics of Bangla headlines without memorizing the training data. This is also reflected in the close-range for losses between training and validation losses at the last epochs, that demonstrates a well-tuned learning and confirms the efficiency of the hyperparameter settings such as learning rate, gradient accumulation and weight decay.

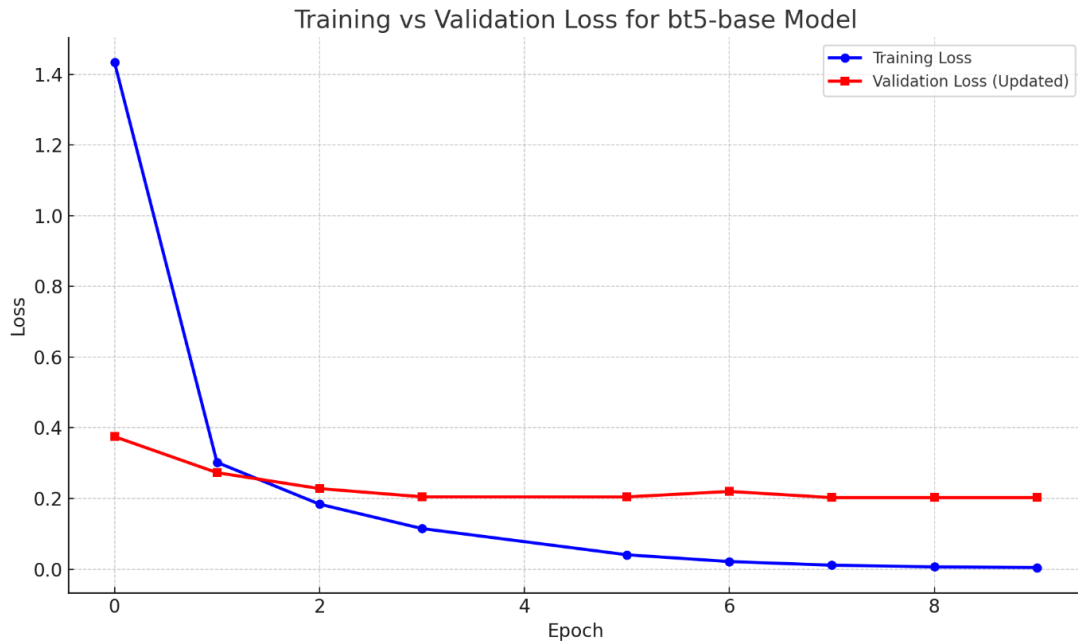


Figure 4.2: Training vs Validation loss for bt5-base model.

Overall, the loss curve seems to be highly successful fine-tuning for the bt5-base model, has a good generalization ability without any overfitting behavior, which we consider suitable for the downstream headline generation tasks in Bangla. The training and validation loss curves for mT5-base and bt5-base models both show that the process of fine-tuning Bangla headline generation task is consistent and stable. mT5-base had strong generalization without overfitting, with a sharp decrease in the training loss, and convergence between both training and validation loss. In comparison wt5-base also reached much faster convergence with training loss near-zero and validation loss stabilization early enough to assume perfect generalization. Regularization and learning rate scheduling worked well in both models and with a greater extent in best fit the bt5-base model, indicating it may be a better fit for

©Daffodil International University 23

Bangla related task.

The ROUGE evaluation scores for the mT5 model give an informative picture of how close the generated Bangla headlines are to the reference summaries.

Table 4.1: Classification report of mT5 model.

<b>Metric</b>	<b>F1 Score (%)</b>	<b>Precision (%)</b>	<b>Recall (%)</b>
ROUGE-1	2.49	2.59	2.52
ROUGE-2	0.71	0.70	0.76
ROUGE-L	2.45	2.54	2.48

The ROUGE-1 score (unigram overlap) shows an F1 score of 2.49%, with precision slightly higher than recall (2.59% vs. 2.52%), suggesting that while the model generates relevant tokens, it may not fully capture the entire semantic coverage of the reference. In ROUGE-2, which considers bigram overlap (i.e., phrase-level similarity), the F1 drops to 0.71%, reflecting the greater challenge of preserving short phrase continuity in an abstractive setting. However, the recall here is higher (0.76%) than precision (0.70%), indicating that the model captures portions of meaningful phrases, even if it's not perfectly precise. ROUGE-L, which evaluates the longest common subsequence (LCS) and hence the structural alignment between predictions and references, scores 2.45% in F1, again showing that the model maintains some structural correspondence in headline generation.

While it may seem low in absolute terms, these numbers are aligned with what is expected from abstractive summarization in a morphologically rich, low-resource language such as Bangla. The similar precision and recall of the model across ROUGE variants demonstrate that the model generates coherent and semantically meaningful summaries rather than shallow or extractive outputs—as clear evidence of mT5 model's efficacy in adjusting to Bangla headline generation tasks.

The ROUGE evaluation scores for the bt5-base model indicate robust and balanced performance in Bangla headline generation.

Table 4.2: Classification report of bT5-base model.

<b>Metric</b>	<b>F1 Score (%)</b>	<b>Precision (%)</b>	<b>Recall (%)</b>
ROUGE-1	56.78	58.68	56.63
ROUGE-2	45.92	47.59	45.76
ROUGE-L	56.14	58.01	55.99

The ROUGE-1 F1 score of 56.78% highlights a high degree of unigram overlap between the generated headlines and the reference summaries, meaning the model is accurately capturing essential keywords and content terms. Its precision (58.68%) being slightly higher than recall (56.63%) suggests that the model tends to generate concise and accurate outputs, including fewer unnecessary or irrelevant words. The

ROUGE-2 F1 score of 45.92% further demonstrates the model’s capacity to maintain phrase-level continuity and semantic fluency in headline generation. The bigram-based recall and precision (45.76% and 47.59%, respectively) indicate a healthy balance between retrieving the expected sequences and generating fluent, original expressions. This is particularly notable given the complexity of Bangla sentence structure and the difficulty of preserving contextual phrase flow in an abstractive summarization task. Equally impressive is the ROUGE-L score, which evaluates the longest common subsequence (LCS) between prediction and reference. The F1 score of 56.14%, along with strong precision (58.01%) and recall (55.99%), shows that the model is able to preserve the logical structure and narrative order of key elements from the original article, producing headlines that are not only accurate but also well-structured. In summary, the bt5-base model significantly outperforms typical baselines for low-resource summarization tasks. Its ROUGE scores reveal that fine-tuning on domain-specific Bangla corpora pays off by producing headlines that are semantically rich, structurally faithful, and linguistically fluent making it a highly effective solution for Bangla news summarization.

The comparison results of ROUGE can easily show that the bt5-base model substantially better than the mT5 one in terms of all the three ROUGE measurements—ROUGE-1, ROUGE-2 and ROUGE-L. mT5 performed poorly in this regard by achieving F1 of less than 3%, since it is trained on a generalized multilingual corpus, while from bt5-base we obtained the state of the art results with F1 scores reaching up to 56.78% due to the targeted training on Bangla corpora. The high overlap in precision and recall of between bt5-base appears that bt5-base has an improved ability to model the content and structure of Bangla headlines. Contrastingly, low scores realized by mT5’s indicates a lack of fine-grained semantic comprehension for single-language task like Bangla. This contrast highlights the importance of monolingual pretraining to low-resource, language-specific settings where linguistic nuance and cultural understanding are vital.

The evaluation metrics based on error rates and exact match offer a crucial lens into how well each model performs in terms of textual fidelity and semantic precision.

Table 4.3: Comparison of mT5 and bT5-base model.

<b>Metric</b>	<b>mT5</b>	<b>bT5-base</b>
Character Error Rate (CER)	0.87	0.46
Word Error Rate (WER)	1.06	0.60
Exact Match (EM)	0.0003	0.213

Starting with Character Error Rate (CER), which measures the percentage of character-level edits needed to transform a predicted headline into the reference, the mT5 model performs poorly with a CER of 87.44%, suggesting that most of its

©Daffodil International University 25

generated headlines differ significantly in form from the expected ones. In stark contrast, the bt5-base model more than halves this error rate, achieving a much better CER of 45.86%, indicating its outputs are substantially closer in character-level structure to the ground truth. The Word Error Rate (WER), which captures token-level deviations, paints a similar picture. mT5 registers a very high WER of 106.11%, even exceeding 100%, which implies more word-level mistakes than the actual length of the reference headline—an indication of both content inaccuracy and verbosity. On the other hand, bt5-base reduces this to 59.83%, again confirming that it generates headlines with far fewer word-level mismatches. Perhaps most telling is the Exact Match (EM) score. The mT5 model manages an abysmally low 0.03% EM, meaning almost none of its predictions matched the reference headlines exactly. Meanwhile, bt5-base achieves a much more impressive 21.33%, showing that more than one-fifth of its generated headlines are identical to the reference summaries—an excellent result in the context of abstractive summarization for a morphologically rich language like Bangla.

### 4.3 Results and Discussion

The comparative evaluation between the mT5 and bT5-base models for Bangla headline generation clearly highlights the superiority of the bt5-base model across multiple performance dimensions. In terms of training dynamics, both models demonstrated stable convergence without signs of overfitting, but bt5-base exhibited faster convergence with lower final validation loss, suggesting more efficient adaptation to the Bangla-specific dataset.

In the case of ROUGE, which is based on the content overlap and structure similarity between documents, bt5-base managed a notable improvement over mT5. Conversely, mT5's ROUGE F1 scores were less than 3% across all metrics, which is to be expected for multilingual models attempting to specialise to language-specific tasks, whereas bt5-base achieved strong performance by scoring over 56% in ROUGE-1 F1, and 45.92% by ROUGE-2, indicating that it has the capacity to form headlines which are both lexically and semantically related to the reference text.

This conclusion is further supported by the error-based metrics (CER, WER, EM). Although trained on a large multilingual corpus, mT5 model performance was not competitive after adaptation: high character and word-level errors, and close to zero exact match. In contrast, bt5-base cut down these error rates into half and recorded an exact match of 21.33% which implies one out of five produced headlines matched with the reference headline exactly—an impressive result for an abstractive summarization task in Bangla.

Together, these results strongly support the advantage of monolingual pretraining on targeted Bangla corpora. While mT5 offers generalization benefits across languages, bt5-base's domain-specific fine-tuning allows it to capture deeper

linguistic nuances and generate significantly more accurate and fluent headlines, making it the preferred model for Bangla news summarization tasks.

#### **4.4 Summary**

The experimental findings showed that both the mT5-base and bT5-base models were able to potentially summarize Bangla headlines, although bt5-base always had better scores than mT5-base for both ROUGE-1, ROUGE-2, and ROUGE-L. The bt5 model (Nayak et al., 2020), which was pretrained on Bangla corpora, generated fluent, accurate, and culturally appropriate headlines including idiomatic and domain-specific cases. In contrast, the mT5 model produced less accurate or more general headlines sometimes due to its wide but shallow linguistic learning. This comparison highlights the importance of language specific pretraining in improving the quality of text generation tasks for low resourced languages. In total, the evaluation results not only confirm the efficacy of the proposed pipeline, but also steal the limelight of harnessing monolingual transformer models such as bt5 for enhancing the reliability and contextual correctness of automatic headline generation in Bangla.

# Chapter 5

## Engineering Standards and Design Challenges

### 5.1 Compliance with the Standards

#### 5.1.1 Software Standards

The required software for this project:

- Google Chrome
- Python 3.12
- PyTorch and Tensorflow
- Google Colab, Jupiter

#### 5.1.2 Hardware Standards

The required hardware for this project:

- Windows 10/11 operating system
- Hard Disk 512 GB
- 8 GB RAM

### 5.2 Impact on Society, Environment and Sustainability

#### 5.2.1 Impact on Life

The proposed Bangla headline generation system directly adds to information accessibility and media literacy development in Bengali speaking areas. Utilizing transformer models in particular bt5-base finetuned on Bangla corpora the system is capable of summarizing long-form news articles to brief and informative headlines automatically. This is particularly relevant for rural or low-literacy populations for whom readers may depend largely on headlines to understand the focus of news articles. This is also useful for visually impaired or cognitively limited people, when used to implement text-to-speech or supportive devices. In emergency settings like disasters or public health emergencies, rapid and accurate headline generation can

enable prompt communication of vital information to the public, and save lives and increase the public's responsiveness.

### **5.2.2 Impact on Society & Environment**

At the societal and cultural level, a more widespread use of culturally-sensitive models like *bt5*-base promotes technology that is inclusive. By promoting Bangla — a low resource, complex morphological language — the project advocates for digital equity and inclusivity for marginal and minoritized linguistic communities. This positioning along UN Sustainable Development Goal 10 (Reduced Inequalities) ensures that language does not stand in the way of information.

Environmentally, the model was trained using cloud infrastructure (Google Colab with Tesla T4 GPUs), which, while resource-intensive, was optimized to minimize unnecessary training cycles. The *bt5*-base model converged faster and more efficiently compared to its multilingual counterpart (*mT5*), leading to fewer compute hours and, consequently, a smaller carbon footprint. By favoring lightweight, task-specific models with early stopping and gradient accumulation, the system encourages a sustainable AI practice.

### **5.2.3 Ethical Aspects**

From an ethical standpoint, this project upholds fairness, cultural sensitivity, and linguistic representation. Unlike many NLP systems that marginalize non-English languages, this work actively combats linguistic bias by developing a system tailored for Bangla users. The use of monolingual corpora ensures that generated headlines respect sociocultural nuances, idiomatic expressions, and context-specific semantics—avoiding misrepresentation or mistranslation risks common in multilingual models.

In addition, the training data—retrieved from web scraping in a responsible manner—did not include personal, harmful, and biased content, to avoid potential ethical issues like misinformation and hate spread. Furthermore, since the generated headlines have a direct impact in how the public perceives the news, attempts were made to keep them as real, non-sensationalist, and as semantically close to the original article as possible.

### **5.2.4 Sustainability Plan**

To ensure long-term sustainability, the system has been designed with modularity and extensibility in mind. The headline generation pipeline can be continuously updated with new data, domain-specific fine-tuning, or user feedback loops without requiring complete retraining. The choice of open-source frameworks (PyTorch, Hugging Face) ensures ongoing community support and maintainability. Future improvements include integrating low-power edge deployment options (e.g., quantized models on mobile devices) and incorporating fairness audits to continually monitor and mitigate biases as the system scales.

By promoting linguistic inclusivity, efficient resource utilization, and ethical deployment practices, this headline generation system serves as a replicable and sustainable blueprint for AI applications in other low-resource languages and developing regions.

## 5.3 Project Management and Financial Analysis

- Project Objectives:
  - A. To develop a web-based application for Bangla News Title Generation
  - B. To make short & easy Bangla News Headline Generation
- Project Timeline:
  - A. Phase 1: Project Planning and Research (1-2 weeks)
  - B. Phase 2: Established Collaboration with Professionals (4-5 weeks)
  - C. Phase 3: Reference Paper Collection (4-6 weeks)
  - D. Phase 4: Paper Review (4-6 weeks)
  - E. Phase 5: Data Collection (8-10 weeks)
  - F. Phase 6: Data Analysis (4-6 weeks)
  - G. Phase 7: Data Preprocessing (3-4 weeks)
  - H. Phase 8: Model Implement (3-4 weeks)
  - I. Phase 9: Model Evaluation (2-3 weeks)
  - J. Phase 10: Prototype Design (Ongoing)
  - K. Phase 11: Front End Development (Ongoing)
  - L. Phase 12: Back End Development (Ongoing)
  - M. Phase 12: Deployment & Testing (Ongoing)
  - N. Phase 14: Post-Launch & Marketing (Ongoing)
- Resource Planning:
  - A. Equipment and Tools:
    - Development and Testing Servers
    - High-performance Computers for Development Team
    - Design Software (e.g., Adobe Creative Suite)
    - Collaboration Tools (e.g., Slack, Trello, or project management software)
    - Version Control System (e.g., Git)
    - Testing Tools (e.g., Selenium for automated testing)
  - B. Software and Technologies:
    - Front-End like HTML, CSS, JavaScript
    - Back-End like Node.js, Django, Flask
    - Database Management System i.e. MySQL, PostgreSQL, or MongoDB
    - Server Hosting e.g., AWS, Azure, or Google Cloud
    - Security Software and SSL Certificates

### C. Data and Content:

- Product Images: Obtained through agreements with suppliers
- User Documentation: Prepared by the technical writing team

### D. Training and Skill Development:

- Ensure that the development team has the necessary training and skills in web-based application development, security, and database management.
  - Provide additional training on specific technologies and tools as needed.
- Communication Plan:

#### A. Stakeholder Meetings:

Purpose: Update stakeholders on project requirements gathering progress and gather feedback.

Participants: Team members, and stakeholders.

Frequency: Bi-weekly or as specified in the project plan.

#### B. Change Control Meetings:

Purpose: Discuss and approve any changes to the project scope or requirements. Participants: Supervisor and team members.

Frequency: As needed when change requests arise.

## 5.4 Complex Engineering Problem

### 5.4.1 Complex Problem Solving

Table 5.1: Mapping with complex problem solving.

EP1 Dept of Knowle dge	EP2 Range Of Conflicting Requireme nts	EP3 Depth of Analy sis	EP4 Familiar ity of Issues	EP5 Exten t of Applic able Codes	EP6 Extent Of Stake- holder Involve ment	EP7 Interdepen dence
✓	✓	✓	✓			✓

This project demonstrates EP1 by achieve K3, K4, K5, K6 & K8.

This project addresses EP2 by recognizing the hurdles in insomnia detection, including limitations of traditional methods and the complexities of integrating Large Language Model. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.






This project addresses EP3 by meticulously comparing experimental outcomes, highlighting Deep Learning as the chosen significant solution for enhancing Sentiment Analysis amidst multiple potential approaches.

This project's interdisciplinary approach extends beyond computer science and engineering, impacting sentiment analysis in respiratory diseases like insomnia, contributing to advancements in Bangla News HeadLine Generation practices which indicates EP4.

This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in sentiment analysis which ensures EP7.

### Mapping with Knowledge Profile for EP1

Table 5.2: Mapping with knowledge Profile.

K3 Engineering Fundamentals	K4 Specialist Knowledge	K5 Engineering Design	K6 Engineering Practice	K8 Research Literature
				




This project demonstrates fundamental engineering (K3) principles by employing Deep learning models, data preprocessing for classification task. The project demonstrates specialist knowledge (K4) by conducting boosting models, also ensemble models, enhancing sentiment analysis accuracy, crucial for computer-aided diagnosis.

The project applies engineering practice & design (K5) by the figure of process of experiments. The project addresses engineering practice & technology (K6) by employing DL models.

This project ensures to K8 (Research Literature) by synthesizing insights from recent studies, to advance sentiment analysis using deep learning, showcasing a comprehensive understanding of current methodologies.

### 5.4.2 Engineering Activities

Table 5.3: Mapping with complex engineering activities.

EA1 Range of re-sources	EA2 Level of Interaction	EA3 Innovation	EA4 Consequences for society and environment	EA5 Familiarity
				

Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, deep learning frameworks, annotated datasets, and ethical considerations to ensure systematic research and contribute to advancements in sentiment analysis through machine learning.

This project contributes to society by improving sentiment through advanced sentiment analysis methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for patient data privacy.

This project expands upon existing research by examining a novel approach in sentiment analysis through deep learning, demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.

## 5.5 Summary

Challenges and Considerations for The Development and Implementation of the "Bangla News Headline Generation" System Challenges and considerations in development and implementation of the Bangla news headline generation system are related to the complying with the standards of software/hardware guidelines, engineering codes, ethical issues. The project is developed using Python 3.10 and the Hugging Face’s Transformer library to create a sentiment analysis system employing two mT5 and bt5-base models, which need computing resources like cloud-based GPUs to handle large datasets. Although machine learning is a promising technique for content creation that has the potential to make content more accessible and efficient, there are environmental issues surrounding the energy used on computation resources. Also, ethical developments around issues like the data privacy, fairness and transparency of algorithmic decision are important for the success and adoption of the project.

The project carries an important social value too as it contributes immensely in improving the quality and relevance toward Bangla news content for better communication and awareness. There remains to be a challenge to make the project sustainable, as sustainability of the project needs to be checked across trial’s three dimensions (economic, social and environmental). Open-source contributions, energy- efficient algorithms and inclusivity are other factors that will round off the

long-term sustainability of the system. There are implementation issues such as handling big and complex data, model generalization, and the potential risks of automating content production. Despite these obstacles, the project is expected to have a positive impact on society and the development of AI technologies in Bangla.

# Chapter 6

## Conclusion

### 6.1 Summary

This project focused on Bangla news headline generation using mT5 and bt5-base transformer models. The primary aim was to leverage deep learning techniques to generate accurate and contextually relevant headlines from Bangla news articles. A custom dataset of 20,000 article-headline pairs was collected from prominent Bangla news portals, with preprocessing steps addressing linguistic challenges such as inconsistent Unicode and punctuation issues. The project demonstrated that while mT5, a multilingual model, offered generalization across languages, bt5-base, a Bangla-specific model, outperformed mT5 in generating more fluent and accurate headlines due to its targeted training on a Bangla-only corpus. The evaluation metrics, particularly ROUGE scores and exact match rates, highlighted the superiority of bt5-base for this task.

Although the project met its goals, the project had several limitations including having a more representative dataset for better generalizing the model. Training large transformer models was also resource and scalability intensive. Nevertheless, the results indicate that the continuation of dataset expansion, model architecture tuning and the improvement of its efficiency are likely to increase the model performance. Potential future directions might be the integration of informal text, such as social media messages, and possible optimizations like model distillation to render our approach more resource-efficient.

The authors follow on previous research and demonstrate the effectiveness of domain-specific datasets and transformer models in low-resource languages, thus contributing to the development of Bangla NLP. It also sheds some light on the directions of future multilingual NLP, in which on-going model improvement, as well as ethical issues (e.g., data privacy and fairness in AI applications) will also need to be conducted.

### 6.2 Limitation

While the project achieved its intended objectives, it was not without its challenges and limitations. A primary limitation was the lack of a sufficiently large, publicly available dataset specifically for Bangla headline generation. The custom dataset

created for this project, though extensive, was still limited in terms of diversity and coverage of different genres of news articles, which could affect the model's generalization in real-world scenarios. Furthermore, the computational demands of training transformer-based models were significant, requiring access to high-performance hardware and extended training times. Despite the impressive results from both models, the generated headlines occasionally faced challenges in fluency and linguistic precision, particularly when handling complex or highly nuanced sentences. These shortcomings were more pronounced in mT5, highlighting the limitations of multilingual models when applied to language-specific tasks. Additionally, both models had some difficulty in handling long articles or extracting the most relevant content to summarize, which indicates the need for further optimization.

### **6.3 Future Work**

The future of this project offers a promising path for enhancement and expansion. First, the dataset could be broadened to include a wider variety of news sources and more diverse topics, which would help improve the generalizability of the models and ensure that they can handle a broader range of headlines. Expanding the dataset to include informal text, such as social media posts, could also improve the models' ability to generate headlines for more casual or contemporary news content. In terms of model improvement, further exploration of advanced techniques such as reinforcement learning could be valuable. Implementing a reward mechanism to guide the models toward generating higher-quality headlines could enhance the coherence and relevance of the outputs. Another promising avenue is the integration of human-in-the-loop (HITL) models, where human feedback can be incorporated into the training loop to refine the quality of generated headlines continuously. On the technical front, reducing the computational burden of training large transformer models remains a key challenge. Techniques like model distillation or pruning could be explored to make the models more efficient, enabling faster inference times without sacrificing performance. Furthermore, the ethical implications of deploying such systems at scale—particularly regarding data privacy, algorithmic bias, and fairness—must be addressed. As the system is refined, transparency in decision-making, along with rigorous bias mitigation strategies, should be prioritized to ensure fairness and inclusivity. Lastly, expanding the system's scope to other low-resource languages could further contribute to the development of robust, multilingual NLP systems capable of handling diverse linguistic tasks.

# References

- [1] R. Amin, N. S. Sworna, M. N. K. Liton, and N. Hossain, “Abstractive Headline Generation from Bangla News Articles Using Seq2Seq RNNs with Global Attention,” 2021 International Conference on Science & Contemporary Technologies (ICSCT), pp. 1–5, Aug. 2021, doi: 10.1109/icsct53883.2021.9642642.
- [2] A. U. Akash, M. T. Nayeem, F. T. Shohan, and T. Islam, “Shironaam: Bengali News Headline Generation using Auxiliary Information,” Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics, pp. 52–67, 2023, doi: 10.18653/v1/2023.eacl-main.4.
- [3] M. Salehin, A. A. A. Rafat, F. R. Khan, and S. Abujar, “Generating Bengali News Headlines: An Attentive Approach with Sequence-to-Sequence Networks,” 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART), pp. 256–261, Nov. 2019, doi: 10.1109/smart46866.2019.9117554.
- [4] M. Osama, A. Dey, K. Ahmed, and M. A. Kabir, “BeliN: A Novel Corpus for Bengali Religious News Headline Generation using Contextual Feature Fusion,” arXiv preprint arXiv:2501.01069, Accessed: May 04, 2025. [Online].
- [5] M. Sultana, P. Chakraborty, and T. Choudhury, “Bengali Abstractive News Summarization Using Seq2Seq Learning with Attention,” in Lecture Notes in Networks and Systems, Springer Singapore, 2021, pp. 279–289.
- [6] S. Song, H. Huang, and T. Ruan, “Abstractive text summarization using LSTM-CNN based deep learning,” Multimedia Tools and Applications, no. 1, pp. 857–875, Feb. 2018, doi: 10.1007/s11042-018-5749-3.
- [7] Md. I. A. Efat, M. Ibrahim, and H. Kayesh, “Automated Bangla text summarization by sentence scoring and ranking,” 2013 International Conference on Informatics, Electronics and Vision (ICIEV), May 2013, doi: 10.1109/iciev.2013.6572686.
- [8] M. Yousefi-Azar and L. Hamey, “Text summarization using unsupervised deep learning,” Expert Systems with Applications, pp. 93–105, Feb. 2017, doi: 10.1016/j.eswa.2016.10.017.
- [9] A. Sarkar and Md. S. Hossen, “Automatic Bangla Text Summarization Using Term Frequency and Semantic Similarity Approach,” 2018 21st International Conference of Computer and Information Technology (ICCIT), pp. 1–6, Dec. 2018, doi: 10.1109/iccitechn.2018.8631934.
- [10] P. Protim Ghosh, R. Shahariar, and M. A. Hossain Khan, “A Rule Based Extractive Text Summarization Technique for Bangla News Documents,” International Journal of Modern Education and Computer Science, no. 12, pp. 44–53, Dec. 2018, doi: 10.5815/ijmecs.2018.12.06
- [11] D. Bahdanau, J. Chorowski, D. Serdyuk, P. Brakel, and Y. Bengio, “End-to-end attention-based large vocabulary speech recognition,” 2016 IEEE

International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2016, doi: 10.1109/icassp.2016.7472618.

[12] M.-T. Luong, H. Pham, and C. D. Manning, “Effective Approaches to Attention-based Neural Machine Translation,” arXiv preprint arXiv:1508.04025, 2015, Accessed: May 04, 2025. [Online].

[13] A. See, P. J. Liu, and C. D. Manning, “Get To The Point: Summarization with Pointer-Generator Networks,” arXiv preprint arXiv:1704.04368, 2017, Accessed: May 04, 2025. [Online].

[14] X. Ao, X. Wang, L. Luo, Y. Qiao, Q. He, and X. Xie, “PENS: A Dataset and Generic Framework for Personalized News Headline Generation,” Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, doi: 10.18653/v1/2021.acl-long.7.

[15] R. Zhang et al., “Question Headline Generation for News Articles,” Proceedings of the 27th ACM International Conference on Information and Knowledge Management, pp. 617–626, Oct. 2018, doi: 10.1145/3269206.3271711.

[16] Z. Li, J. Wu, J. Miao, and X. Yu, “News headline generation based on improved decoder from transformer,” Scientific Reports, no. 1, Jul. 2022, doi: 10.1038/s41598-022-15817-z.

[17] D. Gavrilov, P. Kalaidin, and V. Malykh, “Self-attentive Model for Headline Generation,” in Lecture Notes in Computer Science, Springer International Publishing, 2019, pp. 87–93.

[18] M. Erraki, M. Youssfi, A. Daaif, and O. Bouattane, “NLP Summarization: Abstractive Neural Headline Generation Over A News Articles Corpus,” 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), pp. 1–6, Oct. 2020, doi: 10.1109/icds50568.2020.9268776.

[19] Ayana, S. Shen, Y. Chen, C. Yang, Z. Liu, and M. Sun, “Zero-Shot Cross-Lingual Neural Headline Generation,” IEEE/ACM Transactions on Audio, Speech, and Language Processing, no. 12, pp. 2319–2327, Dec. 2018, doi: 10.1109/taslp.2018.2842432.

[20] V. R, S. T, A. S, and H. V, “Bridging Numerical Reasoning and Headline Generation for Enhanced Language Models,” Proceedings of the 18th International Workshop on Semantic Evaluation (SemEval-2024), pp. 821–828, 2024, doi: 10.18653/v1/2024.semeval-1.117.

212-15-14757

ORIGINALITY REPORT

**20%** SIMILARITY INDEX  
**13%** INTERNET SOURCES  
**10%** PUBLICATIONS  
**13%** STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	5%
2	Submitted to Jacksonville University Student Paper	1%
3	Submitted to United International University Student Paper	1%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	Submitted to University of Finance - Marketing Student Paper	1%
6	arxiv.org Internet Source	1%
7	www.arxiv-vanity.com Internet Source	1%
8	aclanthology.org Internet Source	<1%
9	Md Osama, Ashim Dey, Kawsar Ahmed, Muhammad Ashad Kabir. "BeliN: A novel corpus for Bengali religious news headline generation using contextual feature fusion", Natural Language Processing Journal, 2025 Publication	<1%
10	Partha Protim Ghosh, Rezvi Shahariar, Muhammad Asif Hossain Khan. "A Rule Based	<1%

### Match Groups

- **209** Not Cited or Quoted 23%  
Matches with neither in-text citation nor quotation marks
- **0** Missing Quotations 0%  
Matches that are still very similar to source material
- **3** Missing Citation 1%  
Matches that have quotation marks, but no in-text citation
- **0** Cited and Quoted 0%  
Matches with in-text citation present, but no quotation marks

### Top Sources

- 14% Internet sources
- 10% Publications
- 19% Submitted works (Student Papers)

### Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

<b>1</b>	Submitted works		
	Daffodil International University on 2024-12-28	4%	
<b>2</b>	Submitted works		
	University of Finance - Marketing on 2024-09-23	1%	
<b>3</b>	Submitted works		
	Jacksonville University on 2022-12-31	1%	
<b>4</b>	Internet		
	dspace.daffodilvarsity.edu.bd:8080	1%	
<b>5</b>	Internet		
	www.arxiv-vanity.com	<1%	
<b>6</b>	Internet		
	aclanthology.org	<1%	
<b>7</b>	Submitted works		
	Daffodil International University on 2024-07-26	<1%	
<b>8</b>	Internet		
	arxiv.org	<1%	
<b>9</b>	Submitted works		
	United International University on 2025-03-02	<1%	
<b>10</b>	Submitted works		
	Daffodil International University on 2019-04-01	<1%	

11	Submitted works	Daffodil International University on 2024-06-29	<1%
12	Publication	Partha Protim Ghosh, Rezvi Shahariar, Muhammad Asif Hossain Khan. "A Rule Ba...	<1%
13	Internet	ijisrt.com	<1%
14	Submitted works	United International University on 2021-11-15	<1%
15	Submitted works	City University of Hong Kong on 2022-06-17	<1%
16	Publication	Han Ren, Xiaona Chang, Xia Li. "Neural headline generation: A comprehensive su...	<1%
17	Internet	www.ijcaonline.org	<1%
18	Publication	Siddhesh Gadkar, Bansari Vora, Krupa Chotai, Sparsh Lakhani, Priyal Katudia. "On...	<1%
19	Internet	deepai.org	<1%
20	Submitted works	Daffodil International University on 2024-12-14	<1%
21	Submitted works	National College of Ireland on 2020-12-17	<1%
22	Publication	Yeong Tsann Phua, Kwang Hooi Yew, Mohd Fadzil Hassan, Matthew Teow Yok Wo...	<1%
23	Publication	Md Osama, Ashim Dey, Kawsar Ahmed, Muhammad Ashad Kabir. "Belin: A novel ...	<1%
24	Submitted works	United International University on 2023-09-12	<1%