

An Attention-Driven Multilingual Ethnic Language Translator (MELT) for the Marma Language: A Comparative Analysis of LLaMa-3 & NLLB-200

by

Md. Mahedul Hasan Mahe (221014011)
Amena Akter (223014065)
Sumaia Afroz Sayka (223014034)
Meherun Nesa Ritu (223014089)

*Capstone project report (CSE 4098C) submitted in partial fulfillment of the
requirements for the degree of*

Bachelor of Science in Computer Science and Engineering

Under the supervision of

Md. Kaviul Hossain



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
UNIVERSITY OF LIBERAL ARTS BANGLADESH

SPRING 2026

© *Md. Mahedul Hasan Mahe, Amena Akter,
Sumaia Afroz Sayka, and Meherun Nesa Ritu*

All rights reserved

DECLARATION

Project Title An Attention-Driven Multilingual Ethnic Language Translator (MELT) for the Marma Language: A Comparative Analysis of LLaMa-3 & NLLB-200
Authors Md. Mahedul Hasan Mahe, Amena Akter, Sumaia Afroz Sayka, and Meherun Nesa Ritu
Student IDs 221014011, 223014065, 223014034, and 223014089
Supervisor Md. Kaviul Hossain

We declare that this capstone project report entitled *An Attention-Driven Multilingual Ethnic Language Translator (MELT) for the Marma Language: A Comparative Analysis of LLaMa-3 & NLLB-200* is the result of our own work except as cited in the references. The capstone project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Md. Mahedul Hasan Mahe	Amena Akter
221014011	223014065
Department of Computer Science and Engineering	Department of Computer Science and Engineering
University of Liberal Arts Bangladesh	University of Liberal Arts Bangladesh

Sumaia Afroz Sayka	Meherun Nesa Ritu
223014034	223014089
Department of Computer Science and Engineering	Department of Computer Science and Engineering
University of Liberal Arts Bangladesh	University of Liberal Arts Bangladesh

Date: May 03, 2026



Department of Computer Science and Engineering
University of Liberal Arts Bangladesh
688 Beribadh Road, Mohammadpur
Dhaka - 1207

CERTIFICATE

This is to certify that capstone project report entitled **An Attention-Driven Multilingual Ethnic Language Translator (MELT) for the Marma Language: A Comparative Analysis of LLaMa-3 & NLLB-200**, submitted by **Md. Mahedul Hasan Mahe** (221014011), **Amena Akter** (223014065), **Sumaia Afroz Sayka** (223014034) and **Meherun Nesa Ritu** (223014089) are undergraduate students of the **Department of Computer Science and Engineering** has been examined. Upon recommendation by the examination committee, we hereby accord our approval of it as the presented work and submitted report fulfill the requirements for its acceptance in partial fulfillment for the degree of *Bachelor of Science in Computer Science and Engineering*.

Place: Dhaka

Date: May 03, 2026

Md. Kaviul Hossain
Lecturer

Rezwan Al Islam Khan, PhD
Associate Professor

Dewan Ahmed Muhtasim
Lecturer

ACKNOWLEDGEMENTS

We would like to express our sincere gratitude to our supervisor, *Md. Kaviul Hossain* for his invaluable guidance, constructive feedback and constant encouragement throughout this study. We are also thankful to Professor Mahmud Hasan Khan, PhD, Department of English and Humanities at the University of Liberal Arts Bangladesh, for his crucial linguistic support in the preparation and validation of the dataset. We also acknowledge the contributions of Marma native speakers - Domiching Marma, Usain Marma and Wai Wai Swe Marma whose trilingual skills greatly enhanced the process of primary data collection. We are especially grateful to Toai Ma Ching Marma who is the respected Marma Headman and the community reviewers of Bandarban for their support in the multi-layered validation process which ensures the accuracy and cultural integrity of the final dataset.

Lastly, we acknowledge the utilization of the publicly available data set in the IEEE Data Port that played a significant role as the starting point of the current research.

Md. Mahedul Hasan Mahe, Amena Akter, Sumaia Afroz Sayka, and Meherun Nesa Ritu
University of Liberal Arts Bangladesh

Date: May 03, 2026

Dedicated to
my loving Mother

– *Md. Mahedul Hasan Mahe*

To my beloved parents
the source of my inspiration

– *Amena Akter*

To my mother
a good soul

– *Sumaia Afroz Sayka*

To my dear mother

– *Meherun Nesa Ritu*

ABSTRACT

Bangladesh is a country of many languages and cultures where ethnic groups reside collaboratively in harmony and tranquility. But there remains a language barrier among these groups and the locals which hinders ethnographic research and development. This study addresses such digital divide affecting the Marma language, an indigenous language spoken by approximately 200,000 people in Bangladesh, but lacks standardized digital resources and translation tools. In this research, we developed an Attention-Driven Multilingual Ethnic Language Translator (MELT) that facilitates bidirectional translation between English, Bangla, and Marma. Using field data and secondary sources, a new ethnographically validated parallel corpus of over 12,000 sentences was created. To reduce data sparsity, transliteration and curriculum learning were used. The NLLB-200 multilingual model, and an improved LLaMA-3 (8B) model were the two architectures that were benchmarked. In low-resource settings, our experimental results show that the refined LLaMA-3 significantly outperforms multilingual NMT models and conventional neural baselines, especially when it comes to managing Marma script variability and semantic consistency, supporting linguistic inclusion and digital accessibility. Word Accuracy, Character Accuracy, Word and Character Error Rates, and BLEU-4 scores have been used to evaluate the performance of these models. According to our evaluation, LLaMa-3 outperformed the other models with a BLEU-4 score of 71.04, Word Accuracy Score of 64.3%, WER of 0.22, and a CER of 0.18.

Keywords: *Marma language, Neural Machine Translation (NMT), Large Language Models (LLM), LLaMa-3, NLLB-200, Attention mechanisms, BLEU-4.*

Contents

1	Introduction	1
1.1	Problem statement	4
1.2	Aims, Objectives & Motivation	5
1.2.1	Aims	5
1.2.2	Objectives	5
1.2.3	Motivation	6
1.3	Project Specification	6
2	Literature Review and Analysis	10
2.1	Literature Review	10
2.2	Literature Analysis	12
2.2.1	Model Analysis	12
2.2.2	Dataset Analysis	17
3	Methodology	20
3.1	Data Collection, Modification and Validation	20
3.1.1	Data Collection	20
3.1.2	Secondary Data Acquisition:	20
3.1.3	Primary Data Expansion (Fieldwork):	21
3.1.4	Data Modification and Column Expansion	21
3.1.5	Data Validation	22
3.1.6	Final Dataset Description	23
3.2	Dataset Preprocessing	23
3.3	Model Training	26
3.4	Model Evaluation	26
3.5	Comparative Analysis	27

3.6	App Development and Deployment	27
4	Design and Development	29
4.1	LLaMA-3 Model Architecture	29
4.2	NLLB-200 Model Architecture	33
5	Plans and Deliverables	39
5.1	Timeline (Gantt Chart)	39
5.2	Required Budget	39
6	Result and Analysis	42
6.1	Result Analysis of LLaMA-3 Model	42
6.2	Result Analysis of NLLB-200 Model	43
6.3	Comparative Analysis	50
6.4	Web Application	54
7	Conclusion	61
7.1	Social, Legal, Ethical, and Environmental Issues	61
7.1.1	The Social Impact and Responsibility	62
7.1.2	Safety Design and Ethical Consideration	62
7.1.3	Law and Regulatory Obligations	63
7.1.4	Environmental Considerations	63
7.2	Brief Summary	64
7.3	Future Works	65
	References	66
A	Dataset Validation Evidence	69

List of Figures

3.1	Flow Chart of this Project	28
4.1	LLaMA Structure	30
4.2	LLAMA Architecture Model	36
4.3	NLLB-200 Base Architecture	37
4.4	NLLB-200 Architecture Model	38
5.1	Gantt Chart of this Project	41
6.1	Assertive Sentence Training and Performance Analysis for LLaMA-3	43
6.2	Model Capability and Performance Analysis	44
6.3	Interrogative Sentence Training and Performance Analysis	44
6.4	Imperative Sentence Training and Performance Analysis	45
6.5	Performance Analysis of Assertive Sentence (NLLB-200)	46
6.6	Performance Analysis Chart of Assertive Sentence (NLLB-200)	47
6.7	Performance Analysis of Interrogative Sentence (NLLB-200)	48
6.8	Performance Analysis Chart of Interrogative Sentence (NLLB-200) . .	48
6.9	Performance Analysis of Imperative Sentence (NLLB-200)	49
6.10	Performance Analysis Chart of Imperative Sentence (NLLB-200) . . .	49
6.11	Comparison Table	50
6.12	Comparison Chart	51
6.13	User Interface of MELT Application	55
6.14	Bangla To English	55
6.15	Bangla To Marma (English Transliterate)	56
6.16	Bangla To Marma (Bangla Transliterate)	56
6.17	English To Bangla	57
6.18	English To (Marma English Transliterate)	57
6.19	English To (Marma Bangla Transliterate)	58

6.20	Marma (English Transliterate) To Bangla	58
6.21	Marma (English Transliterate) To English	59
6.22	Marma (Bangla Transliterate) To Bangla	59
6.23	Marma (Bangla Transliterate) To English	60
A.1	Data Validation Evidence	70
A.2	Turnitin Similarity Check	71

List of Tables

2.1	Literature analysis	13
3.1	Expanded Column Description	22
5.1	Budget for this project	40
6.1	Performance Comparison (BLEU-4 Scores)	51
6.2	Comparative Analysis (MELT vs Kokborok)	52

Chapter 1

Introduction

Language translation is the process of converting text from one language into another particular language while maintaining its semantic, grammatical, and contextual meaning. As the world becomes more globalized and digital communication standards are established, multilingual translation becomes an important domain of natural language processing (NLP) [Siddique et al. \(2020\)](#); [Laskar et al. \(2022\)](#); [Sharma & Sharma \(2021\)](#). Widely spoken languages, including English, Bengali and Hindi, already have well-established digital support and translation tools [Geetha & Priyakumari \(2024\)](#); [Ansary \(2024\)](#); [Sen et al. \(2022\)](#). But thousands of ethnic languages, such as more than 30 in Bangladesh and most of them do not include standardized linguistic resources, annotated corpus, and computational translation systems [Podder et al. \(2023\)](#); [Chakma et al. \(2024\)](#); [Mahi et al. \(2025\)](#).

Among them, the Marma language is used by a large native population in the Chittagong Hill Tracts of Bangladesh and other parts of India and spoken by approximately 0.2 million people [Mahi et al. \(2025\)](#). Even though this language is mostly written online in Bengali script, its context is still unknown to normal Bengali readers [Chakma et al. \(2024\)](#); [Mahi et al. \(2025\)](#). For this, it has not yet been covered by machine translation systems, language processing models, and available linguistic technologies [Podder et al. \(2023\)](#); [Chakma et al. \(2024\)](#); [Mahi et al. \(2025\)](#). This digital divide reduces communication between the Marma community and the broader population and limits access to vital services like education, health, information on travel and administrative services.

Marma is also linguistically Burmese origin, and phonologically and syntactically different structure compared to Indo-Aryan languages, like Bengali [Podder et al. \(2023\)](#); [Chakma et al. \(2024\)](#); [Mahi et al. \(2025\)](#). These unique scripts and

linguistic format of this language also make this translation even more difficult. The absence of standard orthography, parallel corpora, and pre-trained embeddings is a major problem for traditional machine learning and neural machine translation (NMT) methods [Rahman et al. \(2020\)](#); [Mumin et al. \(2019\)](#); [Khan & Sarfaraz \(2019\)](#); [Islam et al. \(2023\)](#).

Transliteration is an excellent pre-processing technique to surmount script-diversity and data scarcity by mapping words phonetically into a common script, thereby improving normalization and model training [Khan & Sarfaraz \(2019\)](#). For Marma, transliteration into a phonetic script like Bengali and English simplify orthography and allows smoother normalization of data [Laskar et al. \(2022\)](#); [Kowsher et al. \(2022\)](#); [Mondal et al. \(2024\)](#); [Ansari et al. \(2022\)](#). In this work, it is used as an intermediate normalization which enables two-way bilingual representation between the ethnic and the mainstream languages [Mahi et al. \(2025\)](#); [Laskar et al. \(2022\)](#); [Mondal et al. \(2024\)](#).

Sequence-to-Sequence models such as RNN, LSTM, and GRU with attention mechanisms have been widely studied in prior research on machine translation for the Bengali and others languages [Siddique et al. \(2020\)](#); [Khan & Sarfaraz \(2019\)](#); [Islam et al. \(2023\)](#); [Tiwari et al. \(2020\)](#). The transformer-based architectures were more useful to improving the quality of the translations, as they are capable of capturing long-range interaction and contextual relations [Vaswani et al. \(2017\)](#); [Israr et al. \(2023\)](#); [Dhar et al. \(2021\)](#). The existing literature is characterized by high-resource or moderately-resourced language pairs. So, highly low-resource languages like Marma have not been examined in recent research on translation.

Recent advances in LLM models, LLaMA and NMT models, NLLB-200 offer new opportunities to improve translation quality based on transfer learning, multilingual pretraining, and attention-based representations [Touvron et al. \(2023\)](#); [NLLB Team et al. \(2022\)](#). These models are effective when using cross-linguistic understanding and a low-resource translation case.

This research proposes an attention-based multilingual ethnic language translator for the Marma language by evaluating traditional NMT architectures and modern LLM models such as LLaMA 3, NLLB-200 and a modified Transformer model, which is used for low-resource conditions. The model performance is measured by using BLEU-4 score, word accuracy, character accuracy, word error rate (WER) and character error rate (CER) to ensure linguistic accuracy.

With this method of research, the work will help preserve the Marma language, as well as make it more accessible and computationally rich, making it possible to introduce it into the environment of modern digital ecosystems and help it further develop and enhance social communication.

1.1 Problem statement

Indigenous languages are seriously under-represented in the field of computational linguistics and the Marma language is no different. While Marma is spoken by around 200,000 people in Bangladesh and its neighboring countries, there are limited resources, such as large-scale parallel corpora, standardized linguistic resources, pretrained language representations, and robust computational translation systems for this language [Mahi et al. \(2025\)](#); [Podder et al. \(2023\)](#). The scarcity of digital language resources for Marma significantly hinders access to technological innovations that are common in other languages, such as English, Bengali and Hindi, e.g., healthcare platforms, travel services, administrative communication, and educational materials [Mahi et al. \(2025\)](#); [Podder et al. \(2023\)](#). In recent years, some initiatives in this field have explored the possibilities of computational language processing, such as the Multilingual Ethnic Language Dataset (MELD) and early trials of neural machine translation for indigenous languages; yet, the available resources and models are still limited in scope, size, and translation quality, making them unsuitable for practical applications [Chakma et al. \(2024\)](#); [Mahi et al. \(2025\)](#). In addition, the language itself has some difficulties because of its script (based on Burmese language), unique phonological features, mixed orthography, and lack of annotated data that hinders current multilingual and attention-based translation methods. In view of the above concerns, this study introduces the concept of creating an attention-based multilingual translation system for Marma, Bengali and English. In the study, the researchers will systematically assess and benchmark neural machine translation architectures, as well as the state-of-the-art multilingual language models such as LLaMA-3, NLLB-200 and a custom-built attention-based Transformer model, and compare them in one consistent setup. Limited amounts of parallel corpora and annotated data, the use of script and orthographic variations, computational limitations due to training and fine-tuning of advanced models on limited data, and lack of previous studies for benchmarking the translation performance are the limitations of the research. The outcomes are the creation of a working multilingual translation system for the Marma language, the development of a thorough comparative analysis of NMT models and LLM-based models, and a valuable step towards the digital preservation, accessibility and computational processing of indigenous languages in the future.

1.2 Aims, Objectives & Motivation

1.2.1 Aims

This research aims to develop an attention-based multilingual ethnic language translator of the Marma language, as well as Bengali and English languages to facilitate linguistics research, comparative model testing, and digital access of a low-resource indigenous language. It will consider script diversity and orthographic variation and attempt to perform an empirical comparison of LLaMA-3, NLLB-200, and a customized attention-based Transformer model by using pre-processing methods such as transliteration, normalization, and data augmentation. In the end, the work aims at the contribution of low-resource NLP research, aiding the digital preservation of the Marma language, and giving a basis to the future translation systems.

1.2.2 Objectives

The objectives of this study are -

- To build and test a trilingual parallel corpus (English–Bangla–Marma) with more than 12,000 sentences with cultural and linguistic authenticity.
- To fine-tune and optimize advanced multilingual translation models such as LLaMA-3 (8B) and NLLB-200 for Marma language translation.
- To adopt a curriculum learning framework that gradually builds up models from basic linguistic tasks to complex sentence structures.
- To assess and compare the performance of different translation systems based on well-established measures like BLEU-4, Word Accuracy, WER and CER.
- To design a practical multilingual (Marma-Bangla-English) translation platform on the web which will enable two way translation for practical use in accessing and effective communication.

1.2.3 Motivation

This study is highly inspired by the ongoing digital and communicative segregation of the Marma community in Bangladesh. As our country boasts of its multi-lingual nature, Marma speakers are simply digitally erased since there is not a single standardized translation tool or language resources. Through our fieldwork and our talks with professionals such as Professor Mahmud Hasan Khan, we realized that lack of a credible translator is a huge obstacle to researchers, medical practitioners and government officials who are attempting to interact with the community.

At present, an outsider in the community, who would like to know or speak Marma, has a mountain to climb. No standard keyboard, no digital corpora, no automated systems exist to fill the gap. The result of this inaccessibility is usually disastrous delays in vital services such as education and healthcare. We aspire to make sure that the voice of the Marma people is no longer mangled in translation by creating the Multilingual Ethnic Language Translator (MELT), which will allow preserving the language and making it a digitally empowered language.

1.3 Project Specification

This project focused on the design and development of a multilingual ethnic language translator of the Marma language with Bengali and English that enables to operate under extremely low-resource conditions. The design requirements established with measurable performance criteria that deployed to determine the success of the final system in addressing the identified requirements and limitations.

1. Specification of translation accuracy

- The system obtained BLEU-4 score almost 18 to 44 in the translation tasks of Marma (English and Bangla transliteration) to Bengali, English and vice versa when trained and fine-tuned with low resource conditions.
- Character Error Rate (CER) was approximate 0.18 and Word Error Rate (WER) was 0.22 which guarantees acceptable intelligibility to be used practically.
- These measures directly indicate translation and semantic accuracy.

2. Specification of Multilingual Coverage

- This system provide the support of bilingual translation of Marma (English and Bangla transliteration), Bengali and English.
- There are ten pairs of translators (Marma-Bengali, Marma-English) working under one framework.

3. Specification of Model Comparison

The system implemented and evaluated with three translation architectures:

- LLaMA-3 based translation model
- NLLB-200 multilingual model and
- Customized Attention-based Transformer model
- The relative analysis was done on the same data sets and measurements to promote fairness and reproducibility.

4. Specification of Low-Resource Data Handling

- The system effectively worked based on a parallel corpus of approximate 12k sentence pairs and 800 word pairs.
- To minimize the script inconsistency and orthographic inconsistency, 100 percent of Marma input data subjected for transliteration and normalization.

5. Specification of Pre-processing and Linguistic Robustness

The pre-processing pipeline:

- Transliteration
- Script normalization
- POS tagging and tokenization
- Create a dictionary of word data
- three types of Sentence Segmentation

6. Specification of Computational Feasibility

- Training and fine-tuning on the model was trainable with a small amount of computational resources such as a single GPU or comparable functionalities in the clouds.
- The customized Transformer model was not require more than 24 hours of training time to have a full experimental run.

7. Specification Usability and Deployment Preparedness

- The system provide a web-based friendly interface.
- The automatic response time of short sentences when translation is involved was less than 3 seconds.

8. Specification Evaluation and Validation

The system evaluated by using -

- BLEU-4 score
- Word Accuracy and Character Accuracy
- WER and CER
- Evaluation results was recorded and replicable to facilitate further studies.

9. Requirement Evolution and Validation

The problem requirement specifications are viewed as dynamic and can be refined through an iterative process by the researchers and stakeholders. Any change in the requirements baseline will be considered to guarantee:

- Congruence with the purpose of the original research.
- Transparency of changes.
- Adoption of revised requirements by the stakeholders.

Success Criteria Summary

The project was considered as successful when -

- The specified quantitative performance indicators are met by the system.
- Shows good performance in the low-resource environment,
- Allows a comparative analysis between NMT and LLM-based models in a systematic way, and grants a useful translators platform that helps to preserve and digitally include the Marma language.

Chapter 2

Literature Review and Analysis

2.1 Literature Review

[Podder et al. \(2023\)](#) create a CNN-based model to recognize Chakma handwritten characters, achieving high accuracy with a small amount of data and showing that convolutional networks can distinguish between script-specific features. The work focuses on curated datasets to preserve and educate, but is restricted to character recognition and generalization is not possible because of small data. [Chakma et al. \(2023\)](#) take this step further and suggest an early NMT prototype that works like small parallel corpora with pre-processing, normalization, and domain adaptation that can produce stable yet primitive translation results.

To alleviate the problem of data scarcity, the multilingual dataset in [Mahi et al. \(2025\)](#) offers aligned Chakma, Garo, Marma, Bengali and English sentences and can be benchmarked and used in cross-lingual experiments. Although it lowered the barriers to entry, the domain imbalance and orthographic inconsistencies are still an issue.

[Siddique et al. \(2020\)](#) show that LSTM models are able to capture long-range dependencies, whereas tokenization and subword modeling can enhance the performance of morphologically rich languages. On the same note, [Mumin et al. \(2019\)](#) demonstrates that subword segmentation and regularization can be used to improve the translation of English into Bangali with limited data. Nevertheless, both of them use comparatively bigger datasets, making them applicable only to the conditions of micro-resources.

Strategies in optimization are important in low-resource NMT. Label smoothing, dropout, and learning-rate scheduling help to increase stability and reduce overfit-

ting as demonstrated in [Rahman et al. \(2020\)](#), whereas [Israr et al. \(2023\)](#) proposes attention-specific dropout as a way to achieve better generalization. These methods are efficient, but sensitive to dataset size and are highly tuned.

Multimodal systems [Laskar et al. \(2022\)](#); [Israr et al. \(2023\)](#) also use image context to enhance lexical disambiguation and fluency and have been observed to make measurable gains in grounded translation. Their use of aligned image-text data however restricts their use in indigenous language environments. Wider systems like [Geetha & Priyakumari \(2024\)](#) offer advice on preprocessing, feature engineering, and assessment of multilingual systems, but do not empirically validate the highly diverse low-resource scripts.

Comparisons of architecture also guide design decisions. As demonstrated in [Khan & Sarfaraz \(2019\)](#); [Tiwari et al. \(2020\)](#), RNN, LSTM, GRU, and convolutional models are compared, with trade-offs in effectiveness and contextual modeling. GRU-based frameworks, such as those in [Shalu & M. \(2021\)](#), have quicker training and competitive accuracy and thus are well-suited to smaller environments, where their evaluations are often performed on larger datasets.

In high-resource research like [A. \(2022\)](#) there is a better alignment and fluency, and [Dhar et al. \(2021\)](#) has verified such advantages in Bangla-English translation, especially with long and morphologically complex sentences. These gains might, however, not necessarily be directly applied to very low-resource settings.

Transformer architecture as shown in [Vaswani et al. \(2017\)](#), is a significant step forward, substituting recurrence with self-attention to allow parallelization and better quality of translation. Pretrained models like BanglaBERT [Kowsher et al. \(2022\)](#) also improve the performance using transfer learning. However, Transformers are so data and computationally intensive that indigenous languages need to be adapted.

Recurrent models are pertinent baselines. LSTM models [Oruh et al. \(2022\)](#) are also able to deal with vanishing gradient problems and sequence modeling and [Islam et al. \(2023\)](#) points to the typical problems (OOV words and long-range dependencies) and the usefulness of subword segmentation in enhancing stability.

Back-translation, described in [Mondal et al. \(2024\)](#), creates pseudo-parallel information and can greatly enhance the quality of translation, when filtered appropriately. However, it's success relies on the quality of reverse models. In [Ansari et al. \(2022\)](#), cross-script issues are tackled with GRU-based transliteration, which

enhances OOV handling and alignment, but performance is adversely impacted by ambiguous orthography.

The linguistic issues of English Bengali translation, as presented in [Ansary \(2024\)](#) and [Akan et al. \(2020\)](#) are morphology, compounding, and idiomatic expressions, both of which require culturally sensitive metrics of evaluation rather than superficial ones. Nevertheless, these methods tend to be not standardized.

Surveys like [Sen et al. \(2022\)](#) pinpoint more general restrictions in Bangla NLP, like the lack of annotated data and inconsistent evaluation methods, but do not completely cover the problem of the indigenous language. Transference methods of cross-lingual transfer in [Das et al. \(2023\)](#) note the significance of shared embedding spaces, which can be used in multilingual MT, and [Zargar \(2021\)](#) offers feasible advice on model selection and optimization in spite of the lack of empirical profundity.

2.2 Literature Analysis

2.2.1 Model Analysis

Transformer-based models (Self-attention, BERT, BanglaBERT, etc.)

Transformers and pretrained Transformer encoders (self-attention, multi-head attention, positional encodings and language specific BERT variants) have recently become the most popular technique due to the ability to capture long-range information without recurrence as well as the potential of extreme parallelism in training. This architectural strength transfers into steady improvements in BLEU scores and translation fluency in cases where a large amount of parallel data or effective ways to pretrained models are available [Vaswani et al. \(2017\)](#); [Tiwari et al. \(2020\)](#); [Mondal et al. \(2024\)](#). Language-specific pretrained models like BanglaBERT are highly popular to pre-train encoders or use to provide contextualized embeddings, and fine-tuning them results in improved lexical choice, adequacy, and sample efficiency relative to starting with Transformers trained without pretraining [Kowsher et al. \(2022\)](#).

In low-resource scenarios such as Chakma-Bangla or small Bangla subsets, researchers note that Transformers require transfer learning and further pretraining on similar high-resource corpora. Back-translation and data augmentation are

Table 2.1: Literature analysis

Ref.	Study (Year)	Language Pair	Setting	Architecture	Tr.	LLM	ML-NMT	Trl.	TL	BT	CL	Metrics	Best BLEU
Podder et al. (2023)	Self-ChakmaNet (2023)	Chakma (OCR)	Indig./LR	CNN (SeIONN)	×	×	×	×	×	×	×	Acc.	N/A
Chakma et al. (2024)	ChakmaNMT (2024)	Chakma ↔ Bn	Indig./LR	RNN/Transf./BanglaT5	✓	~	×	✓	✓	✓	×	BLEU, chrF	17.8
Mahit et al. (2025)	MELD (2025)	Ch/Garo/Marma → Bn/En	Indig./LR	Classical ML	×	×	×	×	×	×	×	Acc., F1	N/A
Siddique et al. (2020)	Siddique (2020)	En → Bn	LR	RNN (GRU/LSTM)	×	×	×	×	×	×	×	X-entropy	N/A
Rahman et al. (2020)	Rahman (2020)	En → Bn	LR	Multi-enc. + attn.	×	×	×	×	✓	×	×	BLEU	N/A
Laskar et al. (2022)	Laskar (2022)	En → Bn (multimodal)	Mod./LR	Attn. + CNN vision	~	×	×	×	×	×	×	BLEU	N/A
Geetha & Priyakumari (2024)	Geetha (2024)	Multilingual	Mixed	Hybrid pipeline	×	×	✓	×	×	×	×	N/A	N/A
Mumin et al. (2019)	Mumin (2019)	En → Bn	LR	Attn. Enc-Dec + subword	×	×	×	×	×	×	×	BLEU	N/A
A. (2022)	A.M. (2022)	De → En	High	Attn. Enc-Dec	~	×	×	×	×	×	×	BLEU	N/A
Vaswani et al. (2017)	Vaswani (2017)	En-De / En-Fr	High	Transformer	✓	×	×	×	×	×	×	BLEU	28.4 / 41.8
Khan & Sarfaraz (2019)	Khan (2019)	En-Hi	LR	RNN/LSTM/GRU	×	×	×	×	×	×	×	Acc.	N/A
Dhar et al. (2021)	Dhar (2021)	Bn ↔ En	Mod.	Transformer (multi-head)	✓	×	×	×	×	×	×	BLEU-4	21.4 / 25.4
Islam et al. (2023)	Islam (2023)	Bn → En	Mod.	Seq2Seq RNN	×	×	×	×	×	×	×	BLEU	N/A
Tiwari et al. (2020)	Tiwari (2020)	En-Hi	Mod./High	LSTM-S2S + ConvS2S	×	×	×	×	×	×	×	BLEU	N/A
Laskar et al. (2019)	Laskar (2019)	En → Hi (multimodal)	Mod.	Attn. + image feat.	~	×	×	×	×	×	×	BLEU	N/A
Mondal et al. (2024)	Mondal (2024)	En-Bn	LR	Transformer + BT	✓	×	×	×	~	×	×	BLEU	N/A
Ansari et al. (2022)	Ansari (2022)	Hi → En (translit.)	Cross-scr.	Multilayer GRU	×	×	×	×	×	×	×	Acc.	N/A
Shalu & M. (2021)	Shalu (2021)	En → Hi	Mod./High	GRU Enc-Dec	×	×	×	×	×	×	×	BLEU	N/A
Touvron et al. (2023)	LLaMA (2023)	En-centric (LM)	High	Decoder-only Transf.	✓	✓	×	×	×	×	×	QA acc.	N/A
NLLB Team et al. (2022)	NLLB-200 (2022)	200 languages	High+LR	MoE Transformer	✓	×	×	×	×	×	×	spBLEU	+44%
Ours	MELT (2026)	Marma ↔ Bn ↔ En	Indig./LR	LLaMA-3 / NLLB / Transf.	✓	✓	✓	✓	✓	×	×	BLEU-4, WA, CA, WER, CER	35.2 (max 71.0)

Note: Tr. = Transformer, LLM = Large Language Model, ML-NMT = Multilingual Neural Machine Translation, Trl. = Transliteration, TL = Transfer Learning, BT = Back-Translation, and CL = Curriculum Learning.

popular additions alongside Transformers in order to overcome data sparsity, but these mechanisms do not completely address the issue-parallel data is still limiting translation quality [Chakma et al. \(2024\)](#); [Mondal et al. \(2024\)](#). One more consistent message is the relevance of word segmentation and tokenization which splits schemes such as SentencePiece/BPE or language-specific tokenization style are deemed essential in a morphologically rich but unstructured language like Bangla or Hindi, as the effectiveness of pretrained encoders suffers drastically when poor subword handling occurs [Kowsher et al. \(2022\)](#); [Mondal et al. \(2024\)](#); [Chakma et al. \(2024\)](#).

In spite of their good performance, Transformers have some practical challenges. Some papers mention several times that they have a large computation and memory expense during training and inference, which limits their availability to resource-limited environments [Vaswani et al. \(2017\)](#); [Kowsher et al. \(2022\)](#). To solve these problems, some articles use hybrid or parameter-roe capability designs, including pretrained transformer encoders and lightweight decoders, conv or transformer with recurrent, or multimodal hybrids that combine text and images with transformers [Laskar et al. \(2022\)](#); [Vaswani et al. \(2017\)](#); [Laskar et al. \(2019\)](#). There is a general agreement among the papers that Transformer and pretraining produces the best results where compute and data are not limiting, but again, transfer learning, back-translation, and consideration of domain and tokenization are necessary to achieve effectiveness with Bangla, Hindi, and Chakma translation tasks [Kowsher et al. \(2022\)](#); [Mondal et al. \(2024\)](#); [Chakma et al. \(2024\)](#).

In table. 2.1, it is found that many techniques and models were previously used, like transformer, LLM, multilingual NMT, and the strategies they used, like curriculum learning, transliteration, translation, and the metrics used in the papers to evaluate the scores, etc., for the translation purpose. It is shown that no paper has worked on the Marma language previously; the papers that work on low-resource languages like Chakma have only made a small dataset, not a full translation project. Our project on the Marma Language has done all of these.

RNN Architectures (RNN, LSTM, GRU) with Attention Mechanism

Recurrent Neural Network (RNN) architectures which include vanilla RNNs, LSTMs, and GRUs remain popular used baselines for sequence to sequence tasks because they explicitly handle temporal dependencies and are relatively simple

to implement. Though Vanilla RNNs have issues with long-range dependencies and suffer from the vanishing-gradient problem [Khan & Sarfaraz \(2019\)](#); [Oruh et al. \(2022\)](#); [Zargar \(2021\)](#). LSTMs overcome these constraints by utilizing gated memory operations, and more recently have been effective in neural machine translation decoding Bangla-English, Chakma-Bangla, and Hindi-English and demonstrating better ability in dealing with long sentences and improving translation quality over traditional statistical models [Mondal et al. \(2024\)](#); [Oruh et al. \(2022\)](#); [Mondal et al. \(2024\)](#). A simplified version, called GRUs, has reset and update gates and performs comparatively to LSTMs with fewer parameters and faster convergence which makes them especially useful in low-resource language pairs like English-Hindi or English-Bangla [Tiwari et al. \(2020\)](#); [Ansari et al. \(2022\)](#); [Shalu & M. \(2021\)](#) To address the drawbacks of the fixed-length context in recurrent encoders, the integration of attention mechanisms such as global soft attention and additive attention. Attention enables dynamic weighting of encoder states during decoding, which increases BLEU scores, alignment quality and translation fluency [Vaswani et al. \(2017\)](#); [Israr et al. \(2023\)](#); [Dhar et al. \(2021\)](#); [Islam et al. \(2023\)](#). Combinations of LSTM and GRU models with attention and mitigation strategies such as back-translation, subword tokenization, and corpus cleaning have proven to be an effective way of addressing data sparsity in low-resource settings such as Chakma-Bangla and Bangla-English [Chakma et al. \(2024\)](#); [Mumin et al. \(2019\)](#); [Mondal et al. \(2024\)](#). Despite these advantages, attention mechanisms may not entirely solve the problem of small or noisy datasets which can still give rise to unstable alignments, inappropriate lexical decisions, and orthographic sensitivity [Siddique et al. \(2020\)](#); [Dhar et al. \(2021\)](#); [Mondal et al. \(2024\)](#). Comparative analysis shows that although RNN with attention allows producing solid benchmarks, they are slower than Transformers computationally due to the sequential processing of tokens and cannot be scaled to large data sets [Vaswani et al. \(2017\)](#); [Kowsher et al. \(2022\)](#); [Oruh et al. \(2022\)](#).

Multimodal Neural Machine Translation (Text with Images)

Multimodal papers integrate elements of computer vision (such as CNN or Transformer vision backbone) into an attention-based NMT to mitigate ambiguity such as low-context sentences and support low-resource settings in English-Bangla or Hindi directions of the task. Two filaments emerge English-Bengali multimodal

NMT (WAT 2022): the multimodal task provides provable improvement on the text-only MET on both challenge sets when images are present [Laskar et al. \(2022\)](#).

English Enriched Vision Zeroth-level Feature Transfer (EVZFT) (WAT 2019): adds image features via concatenation or attention on top of standard attention-based encoder-decoder also shows improved performance when the alignment quality improves and image and text domains but gains are confined to a narrower range than larger models [Laskar et al. \(2019\)](#).

General recommendations are (a) visual noise or lack of a tight image to text alignment may adversely affect BLEU (b) visual information-based representations are useful when working with concrete nouns and ambiguous verbs (c) in Bangla-Hindi pipelines, pre-processing steps (tokenization, transliteration of proper nouns) may be significant [Laskar et al. \(2022,0\)](#).

Classical Machine Learning and Hybrid Pipelines (with NMT)

Some papers demonstrate classical MT elements (SMT phrase-based baselines, TF-IDF or LMs) or cloud or ASR-TTS chains and hybridize them with neural models. Reviews and operational systems outline statistical or mathematical MT and neural (seq2seq or Transformer) and in both cases SMT can serve as a useful base or as data (such as phrase tables to augment data). A multilingual ML system integrates cloud translation APIs, classical ASR or TTS and lightweight neural components to make end-to-end multilingual speech translation, although not a pure NMT research model, it does demonstrate pipeline hybrids, which are common in production-type settings [Geetha & Priyakumari \(2024\)](#).

Back-Translation (BT) and Synthetic Data Strategies

As present on [Mondal et al. \(2024\)](#), back-translation is highly prioritized as the most effective low-resource data augmentation alongside training a reverse model (T-S), the synthetic target should be translated back into the synthetic source language and synthetic target data should be combined with real parallel data to retrain the model. This study, briefly describes the BT workflow and suggests holding out monolingual data to address overfitting during the final stages, they consider the synthetic pairs and the target-side monolingual as pseudo-parallel corpora yielding significantly higher generalization on low-resource NMT. According to

Rahman et al. (2020); Mondal et al. (2024), BT is typically combined with subword tokenization (BPE or SentencePiece), domain cleaning and occasionally iterative BT or noisy channel varieties. The finding that BT effects are greater in cases where it is implemented with domain match and quality filters on synthetic text also features prominently in low-resource English-Bangla studies Mondal et al. (2024); Rahman et al. (2020). The main caveats are (a) bad reverse models would create noisy synthetic data should require existence of quality filtering, (b) no need to drown real parallel in synthetic (maintain the usual ratio inverted, (c) need to re-tune dropout and label-smoothing once switch to mixed data Mondal et al. (2024); Rahman et al. (2020).

2.2.2 Dataset Analysis

As described on Podder et al. (2023), Chakma Handwritten Character Dataset is the first large scale corpus of Chakma script recognition. It has 47,000 RGB images (224 224) of 47 classes (vowels, consonants, numerals) and 1,000 samples per class were gathered through the EthnicData mobile app. The data is divided into 70/10/20; training, validation and testing. Cleaning, labeling and validation are considered as part of pre-processing, and on-the-fly augmentation (rotation, resizing, perspective transformation) enhances generalization. This dataset helped the Self-ChakmaNet model to reach 99.84 percent accuracy in recognizing characters.

The earliest bilingual ChakmaBangali parallel corpus to NMT is the ChakmaNMT dataset, which is 29,849 aligned pairs of sentences translated into the Bible. Pre-processing involves token-based refinement and filtering of sentences to minimize noise. It is divided into training, validation (1, 250 pairs) and testing (1, 250 pairs), which has formed a baseline on low-resource MT even though the domain is not diverse Chakma et al. (2024).

The MELD (Multilingual Ethnic Language Dataset) presents a translated corpus of transliterated multilingual data of the Chakma, Garo, and Marma languages, in alignment with the Bengali and English languages. It has 3,046 sentences gathered through the native speakers and manually transliterated and meaning-based translated and multi-stage checked. Stop-word removal, punctuation filtering and frequency analysis are part of pre-processing. The maximum accuracy (93.44

percent) was obtained by baseline classification models (Naive Bayes, Logistic Regression, Random Forest, Decision Tree) proving the suitability of the model to multilingual NLP tasks [Mahi et al. \(2025\)](#).

An English-Bangali parallel data has been analyzed with smaller 4,000 sentence pairs, which were manually aligned using the bilingual texts. It contains 19,606 English (2,839 unique) and 19,000 Bangla words (3,527 unique). The pre-processing contains normalization, punctuation, TensorFlow tokenization, and padding (7 tokens English, 8 Bangla). The data is divided into 80 per cent training and 20 per cent testing, which facilitates the development of RNN-based MT [Siddique et al. \(2020\)](#).

A longer parallel corpus that is built on the TDIL includes back-translation and transliteration, 81,547 sentence pairs, and 68,905 more pairs of transliteration. Normalization, tokenization, and padding are involved in pre-processing, and after the 95/5 split and external validation. GRU-based translation-transliteration models are supported by the dataset and enhance the handling of named entities and the accuracy of translation [Rahman et al. \(2020\)](#).

In the case of English-Hindi translation, 1,000 pairs of sentences were pre-processed by standardizing the data, removing stop-words, mapping contractions, and tokenizing. Semantic representation was done using word embeddings (25-dimensional Word2Vec). The dataset was divided into 80/20 to train and test, which facilitated comparative analysis of RNN, LSTM and GRU models [Khan & Sarfaraz \(2019\)](#).

A bigger Urdu-English corpus consists of 200, 000 sentence pairs, which were gathered out of news, books, and the Internet. Pre-processing involves normalization, tokenization, and length filtering, and then the division into training (180k), validation (10k), and testing (10k). Regardless of its size, issues like domain imbalance, morphological complexity and alignment errors remain [Israr et al. \(2023\)](#).

The other parallel corpus is in Bangali-English which comprises 21 000 pairs of sentences taken online. The dataset is then broken down into training (18k), validation (1.5k), and testing (1.5k) after cleaning, normalization and filtering. It is a valuable source of NMT system creation and testing [Dhar et al. \(2021\)](#).

Lastly, the EnglishHindi CLARIN corpus contains 124,318 and 97,662 sentences in English and Hindi, respectively. Preprocessing includes the normalization, removal of stop-words, tokenizing and filtering of rare words. Word2Vec (CBOW)

and GloVe are used to create word embeddings. The data is divided into 80 training and 20 test, which helps to develop comparative NMT models under low-resource conditions [Shalu & M. \(2021\)](#).

Chapter 3

Methodology

This study is based on a systematic and modular research process to implement an Attention-based Multilingual Ethnic Language Translator of the Marma Language. As shown in the methodology flowchart in Figure 3.1, the general workflow starts with data collection to model development, evaluation, comparative analysis and final deployment.

3.1 Data Collection, Modification and Validation

This research is based on the principle of creating a trustworthy and linguistically consistent dataset that can be used in low-resource language translation.

3.1.1 Data Collection

In order to facilitate the creation of a strong multilingual translation framework of the low-resource indigenous languages, the data that was used in this study was gathered in a hybrid approach that involves the use of secondary archived materials and massive fieldwork.

3.1.2 Secondary Data Acquisition:

First, the publicly accessible dataset has been acquired at IEEE DataPort and it contains about 8,000 pairs of sentences. This dataset was initially comprised of three columns, which are English, Chakma and Marma (English Transliterated). Even though it was beneficial as a baseline, the dataset did not contain standard

Bangla translations and the Marma writing in Bangla phonetics, which are needed to provide the cross-lingual alignment of our attention-driven model.

3.1.3 Primary Data Expansion (Fieldwork):

It is based on our first effort during the Capstone Phase A in which we had developed a baseline of 2,022 sentences that we realized that we required more contextual variety. In response to the feedback on the size of datasets, we made this primary collection larger by direct cooperation with the Marma indigenous population.

The given stage was enabled by the professional support of Mahmud Hasan Khan, PhD, Professor, Department of English & Humanities (DEH) ULAB, who was our linguistic professional. We were also able to interact with the native speakers under his guidance to overcome the linguistic barrier. The primary data were collected with the help of Domiching Marma, Usain Marma, and Wai Wai Swe Marma, who have the trilingual knowledge of the Bangla, English, and Marma languages. On this joint endeavor, we were able to:

- Increased the sentence-level dataset consisting of the first 2,022 rows to around 3,000 sentences.
- Acquired a separate word-level corpus of more than 800 entries to increase vocabulary coverage.

3.1.4 Data Modification and Column Expansion

Massive data cleaning and structural normalization were conducted to provide uniformity across the different data sets and provide an opportunity to train the models effectively. The IEEE DataPort dataset was extended past its three-column format to the standard four-column format to perfectly match our newly gathered primary data. The resulting consolidated data contains the additional and extended columns as depicted in Table 3.1.

In the case of the IEEE dataset, as well as the newly obtained sentence dataset, missing columns were created by manual translation and transliteration, and then underwent reviews, which were repeated. Particular consideration was put on semantic and phonetic accuracy in all entries.

Table 3.1: Expanded Column Description

Column Name	Description
English	Standard English sentences serving as the source or pivot language.
Bangla	Bangla translations of the corresponding English sentences (manually added)
Marma (Bangla Transliteration)	Marma sentences written using the Bangla script.
Marma (English Transliteration)	Marma sentences transliterated using the English (Roman) script.

3.1.5 Data Validation

The process of data validation was done in a multi-phased, community-based process so that it was possible to deal with linguistic autonomy and cultural validity.

Preliminary Validation

Primary validation was done in the initial phase with the help of our linguistic collaborator, Domiching Marma. Since she is good in reading and writing in Bangla, English and Marma, his task was to fix any lexical mistakes, misunderstandings in a sentence, and the discrepancies between the forms of transliterated and native languages in the enlarged data.

Expert and Community Validation

Following constructive inquiries from the defense panel regarding the credibility of dataset validation, we sought further academic advice regarding indigenous language verification. It was clarified by Professor Mahmud Hasan Khan that for indigenous languages lacking formal certification boards, validation by proficient native speakers and community leaders is academically recognized as the standard for ground truth.

According to this principle, direct consultation with Toai Ma Ching Marma, who is a respected Marma Headman, was carried out to complete the final validation. Most importantly, it was not an isolated review but the Headman organized a joint validation exercise, which consisted of a focus group of 3-4 experienced community members in Bandarban. Through this, a group of native experts with this panel reviewed key parts of the data. Their reading and writing abilities in all the three target languages (Marma, Bangla and English) and the cultural control of the Headman guaranteed the strict and multi-dimensional confirmation of the final corpus. As shown in Fig. A.1

3.1.6 Final Dataset Description

The completed data set comprised around 12,000 parallel entries and it took several months of strict collection, modification and multiple-level validation to complete the data set. The breakdown is as follows:

- Less than 8,000 sentences taken through IEEE DataPort.
- Almost 3,000 sentences of the primary collection (the initial 2,022 rows were extended).
- Below 800 lexical/word-level entries.

All entries of the corpus have the same four-column format (English, Bangla, Marma-Bangla, Marma-English). This corpus is uniquely suited to the training and evaluation of "An Attention-Driven Multilingual Ethnic Language Translator of Marma Language: Comparative Analysis of NMT and LLM-Based Translation Models" through the combination of secondary public data and field-validated indigenous language data.

This dataset consists of ongoing community interaction and professional scrutiny by a Marma Headman, which implies that it is a dependable, ethically based, and linguistically sound source of data on low-resource translation studies

3.2 Dataset Preprocessing

A preprocessing pipeline based on systematic data was set up in order to prepare high-quality input data to be used in fine-tuning Large Language Models (LLMs) and training the custom attention-based translation model. As the dataset is based on low-resource multilingual data, preprocessing has been aimed at data cleaning, normalization, linguistic enrichment, and sentence-level organization.

Script Normalization

Data cleaning and normalization of the dataset were used to transform textual representations to a standardized form and eliminate orthographic variation. Script normalization was done -

- By removing all types of symbols and extra whitespaces.
- By using the consistent scripts in all language columns.
- Duplicate rows and entries containing missing values were removed to minimize noise
- Every character was coded into UTF-8 to achieve standardization of encoding.

The clean data was finally saved in the CSV format to be processed and fine-tuned further.

Tokenization and Padding

The process of tokenization was carried out with model-specific tokenizers to guarantee the ability to make the process compatible across architectures -

- During the fine-tuning, LLaMA-3 used a tokenizer based on a Byte-Pair Encoding (BPE).
- In NLLB-200, SentencePiece tokenizer was used.
- The Attention-based customized transformer model used WordLevel tokenizer to ensure token representation.

In batch preparation, padding was used to normalize the sequence lengths whether automatically by the tokenizer (LLMs) or manually in the preprocessing code (in the case of the custom model) to guarantee efficient and stable training.

Dictionary Creation

Based on the cleaned dataset, a dictionary-based dataset was created to facilitate linguistic matching. The dictionary includes 1, 726 entries at a word level in Marma, Bengali, and English. This data was originally in CSV format, but it was converted to JSON format to represent the data in a structured form.

Part-Of-Speech(POS) Tagging

The dictionary-based dataset was created using Manual Part-of-Speech (POS) tagging. The words were marked with their grammatical category to have explicit syntactic information. The POS tagging enhanced the understanding of the model in the comprehension of word roles and sentence structure especially to the low-resource Marma language. The dictionary data with POS-tagging was translated into the JSON format and was augmented and analyzed.

5. Sentence Segmentation

The cleaned data was sorted into various types of sentences such as assertive, imperative, interrogative, optative and exclamatory. The final dataset did not include optative and exclamatory sentences since there was not enough data samples. The other categories of sentences allowed more organized learning and helped to achieve better translation performance.

Evaluation Metrics

To measure the quality of translation within the system objectively, the following measures are identified:

1. BLEU Score (Bilingual Evaluation Understudy) The metric used as a standard of machine translating, where n-gram overlap between candidate and reference is calculated.

$$BLEU = BP \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (3.1)$$

2. Accuracy Determines the proportion of correct predictions of tokens.

$$Accuracy = \frac{\text{Correct Predictions}}{\text{Total Predictions}} \times 100 \quad (3.2)$$

3. Training Loss Plots the convergence of the model with respect to epochs. When the loss is smaller, the predicted probability distribution will be closer to the ground truth.

3.3 Model Training

We implemented and benchmarked two distinct architectures to evaluate the trade-off between model size and translation quality. Training was conducted on Google Colab Pro (A100/T4 GPUs). In order to make sure that the models developed a rich, hierarchical knowledge of the Marma language, we used a sequential fine-tuning approach (curriculum learning) to the Llama-3 and NLLB-200 models. In this case, we took the manual Parts of Speech (POS) tagging in the separate word data and made the data augmentation to achieve the optimal outcome of the grammatical knowledge. Although the training methodology was the same, for LLaMA-3, we used a strict 90/10 train-test split with 1 epoch per curriculum stage to prevent overfitting on the small dataset. For NLLB-200, we utilized a 70/30 split with 3 epochs to ensure robust convergence.

3.4 Model Evaluation

To strictly evaluate the translation performance of the two architectures, we used a multi-metric evaluation scheme that comprised of BLEU scores, Word Error Rate (WER), Character Error Rate (CER), Word Accuracy (WA) and Character Accuracy (CA). For Llama-3 model evaluation, we did the fine-tuning process where made the trainer show the loss after 50 steps every time to show whether overfitting is happening or not. On the other hand, in the NLLB-200 model, we used a larger test set (30%) of the data to make final prediction, which means that the sequence-to-sequence prediction would be stronger in the larger hidden part of the data. The whole process of the methodology is shown in Fig. 3.1

3.5 Comparative Analysis

Comparison Between LLaMA and NLLB in MELT

As the results were compared to reveal the best and the worst outcomes, it is evident that the fine-tuned Llama-3 model has a better overall average of BLEU since it generates an average of 35.22 contrasting with 29.33 of NLLB. Llama-3 has a decisive benefit in high resource translation, especially in the English-to-Bangla direction where it outperforms the baseline by more than 13 points (60.70 vs. 46.97) and it is the best generative model of the low-resource language of Marma, improving on the performance of the baseline by over twice when translating either the English or Bangla to Marma. Although NLLB-200 is more competitive (37.23 vs. 30.63) in Romanized Marma script interpreting English, Llama-3 has stronger features in creating coherent text in Marma and its even performance in the two-way tasks, which makes it the most effective solution to be used in this research project.

3.6 App Development and Deployment

In its final form, a web-based translation system is developed to demonstrate the feasible usability. The system offers a user-friendly interface for multilingual translation. The model is available as REST APIs using FastAPI, allowing input/output in Marma, Bengali and English.

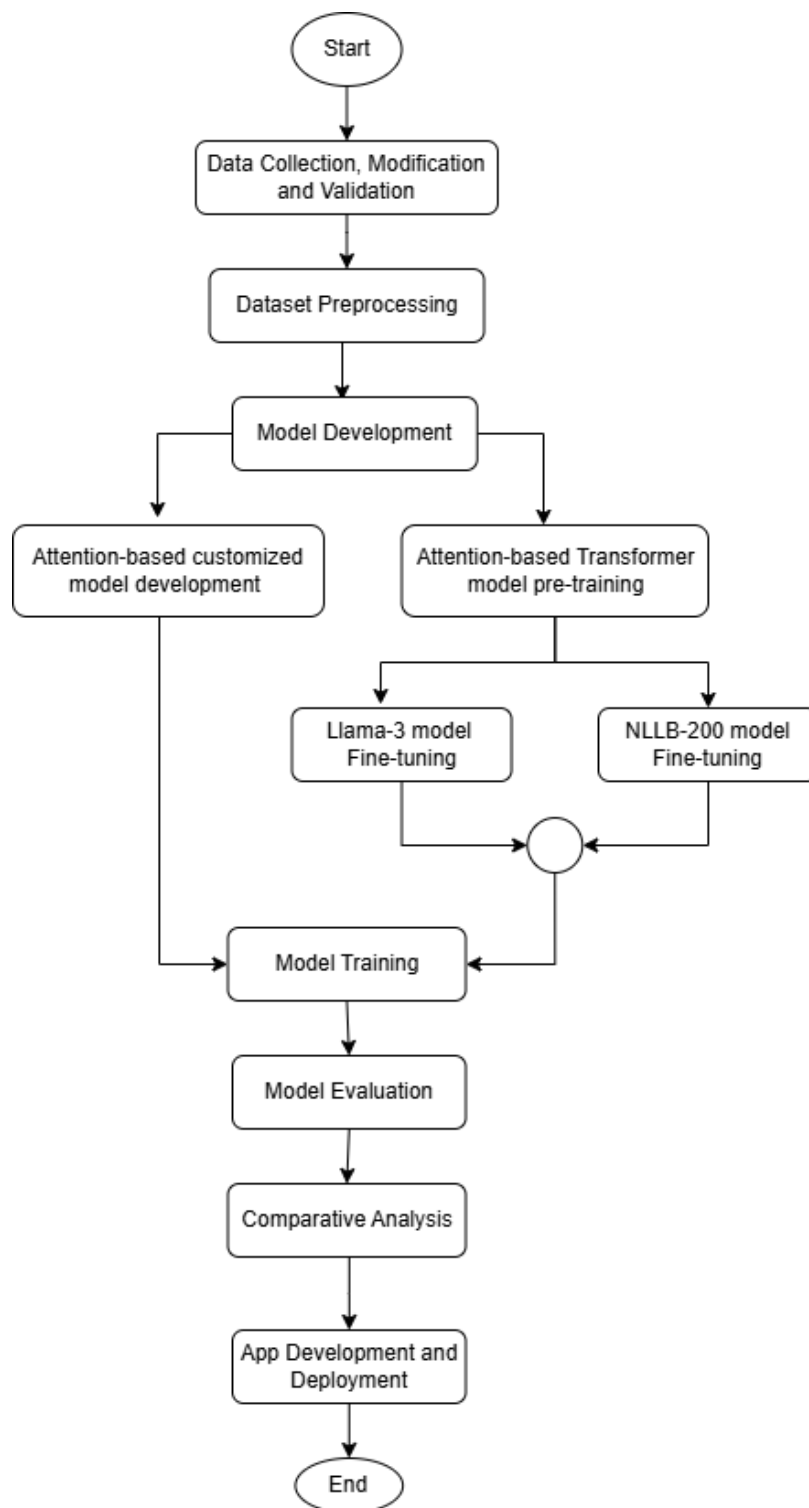


Figure 3.1: Flow Chart of this Project

Chapter 4

Design and Development

4.1 LLaMA-3 Model Architecture

The LLaMA models are a family of foundation language models designed to be useful both in downstream tasks and inference but with specific parameters 6.7B, 13B, 33B, and 65B. They are decoder-only transformers which focus more on smaller models which are trained longer with the best compute budgets. They had been trained (on 1T to 1.4T publicly available tokens) with next-token prediction. It is possible to do full open-sourcing when LLaMA uses only open data sources, unlike proprietary models such as GPT-3 and PaLM [Touvron et al. \(2023\)](#).

The preparation stage is the input stage, the first stage. Every translation starts with raw text which must be converted into a mathematical representation by the model.

1. Input (Bangla / English / Marma / POS):

This is the raw sentence or prompt that is presented to the model.

- **Tokens:** Text is broken into tokens (sub-word units). The size of Llama-3 vocabulary is enormous to enable the model to accommodate complex scripts.
- **Embeddings:** The tokens are converted to high-dimensional representations. The initial sense of the words in the vacuum of any context.

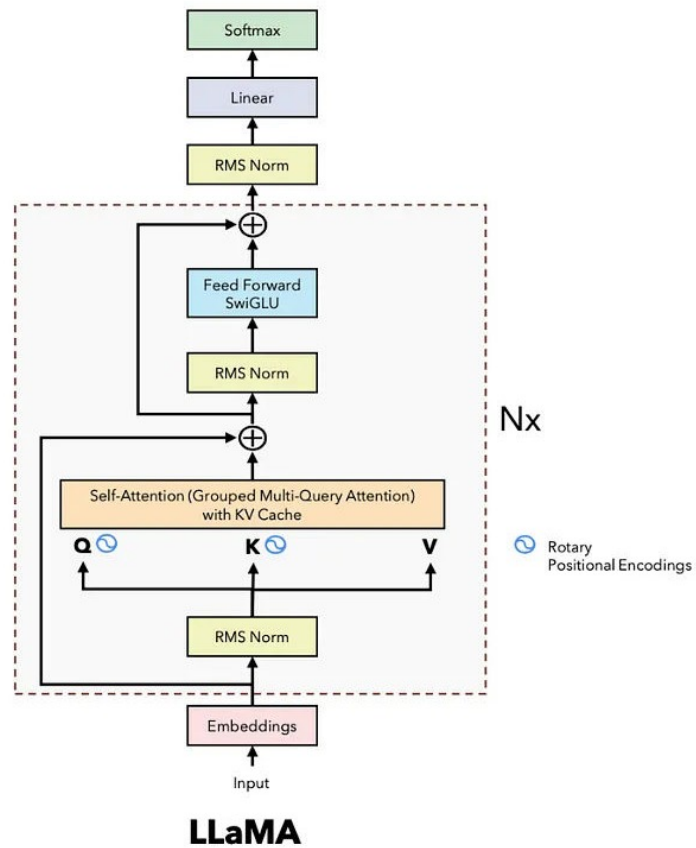


Figure 4.1: LLaMA Structure

2. The Heart: Decoder Blocks.

From Fig. 4.1 and Fig. 4.2, it is shown that Llama-3-8B consists of 32 Decoder Blocks. The diagram shows Layer 1 to 31 as a single block and then it magnifies Layer 32 to show what is going on in it.

Root Mean Square Layer Normalization (RMS Norm)

Data is normalized prior to the insertion of data into a layer to ensure that the mathematical values are kept constant. The equation generally is:

This prevents exploding gradients and enables quicker learning in the model by ensuring that the inputs of every layer are on an equal scale.

Grouped Multi-Query Attention (GQA) + LoRA

Here, the model familiarizes himself with context. GQA is an optimization which causes Llama-3 to be faster through sharing of "key" and value heads with multiple query heads.

- **LoRA Adapters (Q, V, O):** We do not change the large original weight matrices, but add small (Rank=16, Alpha=16) matrices known as adapters.
- **Function:** The original model is fixed and these adapters are trained to the patterns of our data of translation (as in Marma grammar).

Residual

The dotted lines of "Residuals" is a sort of bypass. They recapitulate the output of a layer to its input. This allows the gradients to traverse the network more efficiently when training, without the model forgetting information that it has previously learned.

FeedForward Network (SwiGLU) + LoRA

After attentions are provided, data is manipulated by a FeedForward Network (FFN) in the model. Llama-3 uses SwiGLU activation which is more efficient in deep learning than the standard ReLU.

Adapters (Gate, Up, Down):

In this case, like with attention layer, LoRA adapters are added to make the reasoning of the model more aligned with our specific task.

3. Output:

Once the data has been passed through 32 layers, it must then be decoded into a word.

- **Linear Head:** The resultant vector is appended to the entire vocabulary (e.g. 128,000+ possible tokens).
- **Softmax:** The raw scores are converted to probability by this

- **Output Token:** The one with the highest probability is selected as the next word in translation.

4. Training Loop (Training Only)

The feedback loop in the center of the diagram only occurs in the fine-tuning stage.

- **Loss Calculation (CrossEntropy):** The model uses the word in your Marma as a reference to the token predicted by the model. The discrepancy is the "Loss".
- **AdamW 8-bit Optimizer:** This is the brain of the training process. It makes use of the Loss to determine how to update the LoRA adapters. AdamW (Adam with decoupled weight decay) is used with $\beta_1 = 0.9$, $\beta_2 = 0.95$, $\epsilon = 1e - 8$, and weight decay $\lambda = 0.1$. Peak LR= 3×10^{-4} for 8B, with cosine decay and 6k-step linear warmup (2% of steps).

Update rule:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}$$

$$\theta_t = \theta_{t-1} - \eta \left(\frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} + \lambda \theta_{t-1} \right)$$

where g_t is the gradient, η is LR, and weight decay is applied separately from adaptive terms. By updating the weights of the model, AdamW optimizer teaches it after each training step. It keeps track of two things: m_t = average of past gradients, v_t = proportion of change of each weight. Then, $\hat{m}_t$ and $\hat{v}_t$ are used for correcting biasness for the m_t and v_t values, as when the training starts, they remain very small. Ultimately, these corrected values are used by the main update rule so that it can adjust the weight θ_t by the use of learning rate η , here if the weight changes more, then it gets a smaller update and vice versa, meaning if it changes less, then it gets a bigger update, making the training quick and faster. Then $\lambda \theta_{t-1}$ = weight decay, it is used to prevent the model from overfitting as it shrinks the weights in every step.

- **Sequential Fine-Tuning:** It optimizes the adapters sequentially.

- **8-bit Optimization:** 8-bit accuracy, rather than 32-bit, makes it substantially less memory-intensive.

4.2 NLLB-200 Model Architecture

NLLB-200 Model Description

NLLB-200 (No Language Left Behind) is a highly multilingual neural machine translation (NMT) model developed in 2022 by Meta AI. It supports translation across 200 languages including both high-resource languages (e.g., English, French) and low-resource languages (such as Central Aymara, Igbo).

The flagship model is a sparsely gated Mixture of Experts (MoE) Transformer with 54.5 billion parameters. Distilled dense versions are also available with 3.3B, 1.3B, and 600M parameters.

The model was trained on over 18 billion sentence pairs, incorporating bitext mining, back translation, and self-supervised pretraining to enhance performance on low-resource languages. It achieves a 44% improvement in BLEU scores on low-resource languages (particularly African and Indian) compared to prior baselines. The model and training code are open-sourced via Fairseq [NLLB Team et al. \(2022\)](#).

The base architecture is shown in Fig. 4.3 and in Fig. 4.4, it shows the architecture of NLLB-200 what we implemented in our project. Key innovations include:

- Bitext mining using LASER3 embeddings,
- Curriculum learning,
- Knowledge distillation.

1. The Input Stage

This starts from the source input text after inserting it in the pipeline. This text is tokenized and is ready to be processed numerically because what the transformer model works on is high-dimensional vectors, not raw characters.

2. The Encoder (Understanding Section)

The encoder works on the whole sentence of the source in parallel to generate a mathematical model of the meaning.

Input Embeddings:

Transforms discrete tokens into continuous vectors, with semantically similar words being closer to each other.

Positional Encoding:

Includes additional values to embed to preserve word order information as transformers do not naturally understand order.

12 Transformer Layers:

A combination of Self-Attention and Feed-Forward Networks (FFN).

- Self-attention enables the model to give emphasis on the relevance of various words in a sentence as compared to each other.
- The FFN additionally enhances these characteristics by non-linear transformations.
- The 12-layered model enables more and more complicated linguistic hierarchies to be represented.

3. The Context Bridge

The complete output of the encoder is passed on to the decoder. This context bridge gives the decoder the background knowledge that he needs to initiate a translation.

4. Generating Section (Decoder)

The generative part of the architecture is the decoder. While fine-tuning, to update weights, this part gets the main importance.

Start-of-Sentence Token: It is a signal that initiates the generation process.

Output Embeddings: Takes all past generated tokens to forecast the next word in the sequence.

Cross-Attention and Masked Self-Attention:

Masked self-attention makes sure that the model only sees previous tokens when being trained.

Cross-attention enables the decoder to attend back to the encoders' representations to make sure the translation is faithful to the source meaning.

12 Transformer Layers: Like the encoder, these layers operate on features block by block (12 consecutive blocks) to select the most precise word selection.

Language Modeling Head: It is the final linear layer that calculates the probability of the tokens in the vocabulary by selecting a candidate that is most similar to the next word.

5. The Training Mechanism

The training loop is a process that is repeated and modifies model parameters according to performance.

CrossEntropy Loss: This is a mathematical operation that quantifies the distance between the predicted token and the actual ground truth token in the dataset by the model.

AdamW Optimizer: An optimization algorithm that computes the changes needed in the weights of the model.

Gradients: The errors are computed, and the errors are propagated backwards (the dashed line in the diagram) to update the decoder trainable parameters.

6. The Output

The final product is a fully translated sentence in the target language, based on the smoothed out probabilities of the decoder language modeling head.

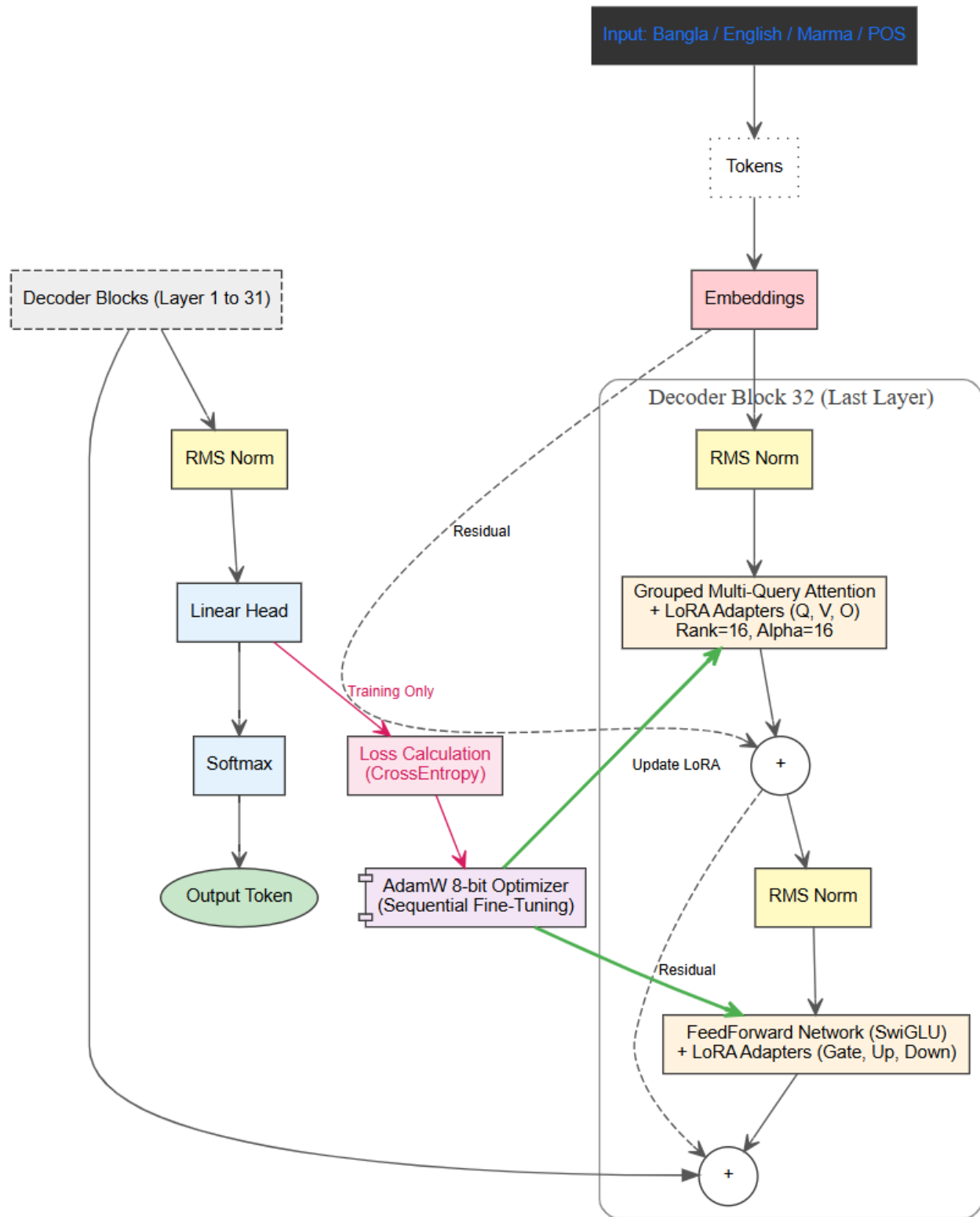


Figure 4.2: LLAMA Architecture Model

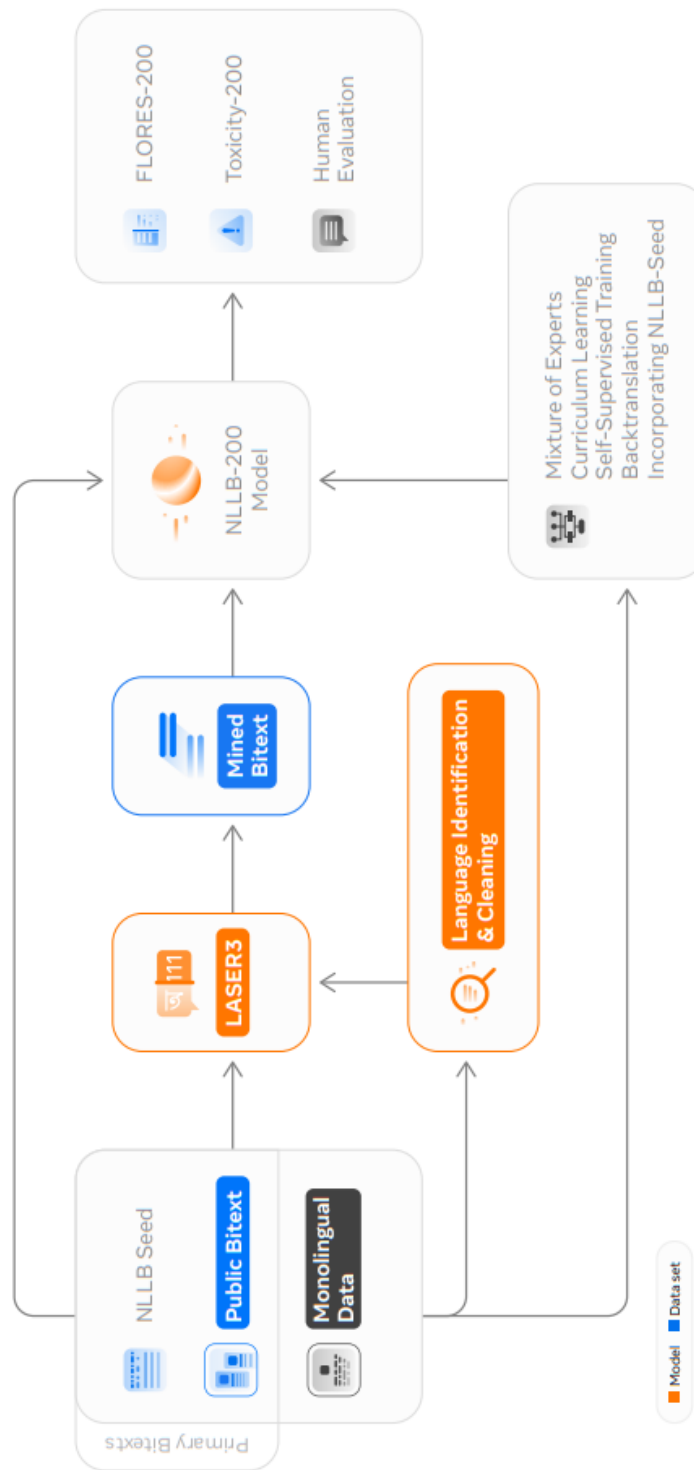


Figure 4.3: NLLB-200 Base Architecture

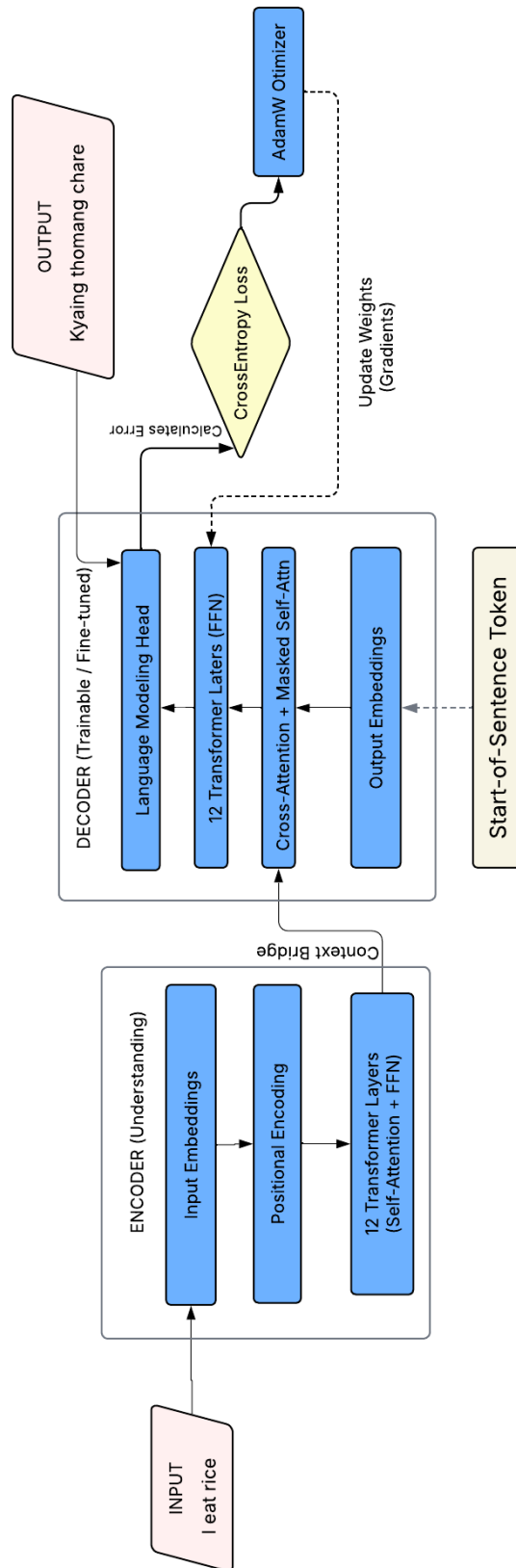


Figure 4.4: NLLB-200 Architecture Model

Chapter 5

Plans and Deliverables

5.1 Timeline (Gantt Chart)

The Gantt chart as shown in Figure 5.1, presents the entire project schedule in three semesters since the start of data collection up to the final deployment. It enumerates steps such as data collection and validation, pre-processing, model development, training, evaluation, comparative analysis and writing up. The timeline guarantees a systematic development, appropriate overlap among tasks and all the project milestones are completed in the given project time frame.

5.2 Required Budget

The approximate project budget as present on Table 5.1, is BDT 108,000 which will include hardware, cloud services, field collection, data verification and miscellaneous. The budget helps to finance model training, data collection of indigenous language, expert validation and system deployment preparation to assure that the proposed multilingual translation system is feasible and reliable.

Table 5.1: Budget for this project

Category	Description	Estimated Cost (BDT)
Hardware	Laptop/PC (mid-to-high spec for model training and testing)	60,000
Cloud Services	Google Colab Pro / Kaggle / AWS credits for GPU/TPU usage; Stable internet connection for dataset download, training, and testing	20,000
Field Data Collection	Travel, food, and logistics for collecting Chakma, Marma, and Garo datasets. Organizing small workshops/discussions with indigenous community members	10,000
Dataset Validation Support	Honorarium/compensation for community validators and language experts	8,000
Translation Tools	Subscription or licensing of translation-related APIs/tools (if required)	7,000
Miscellaneous / Contingency	Report printing, dataset documentation, validation acknowledgement letters, unexpected technical or operational expenses	3,000
Total		108,000

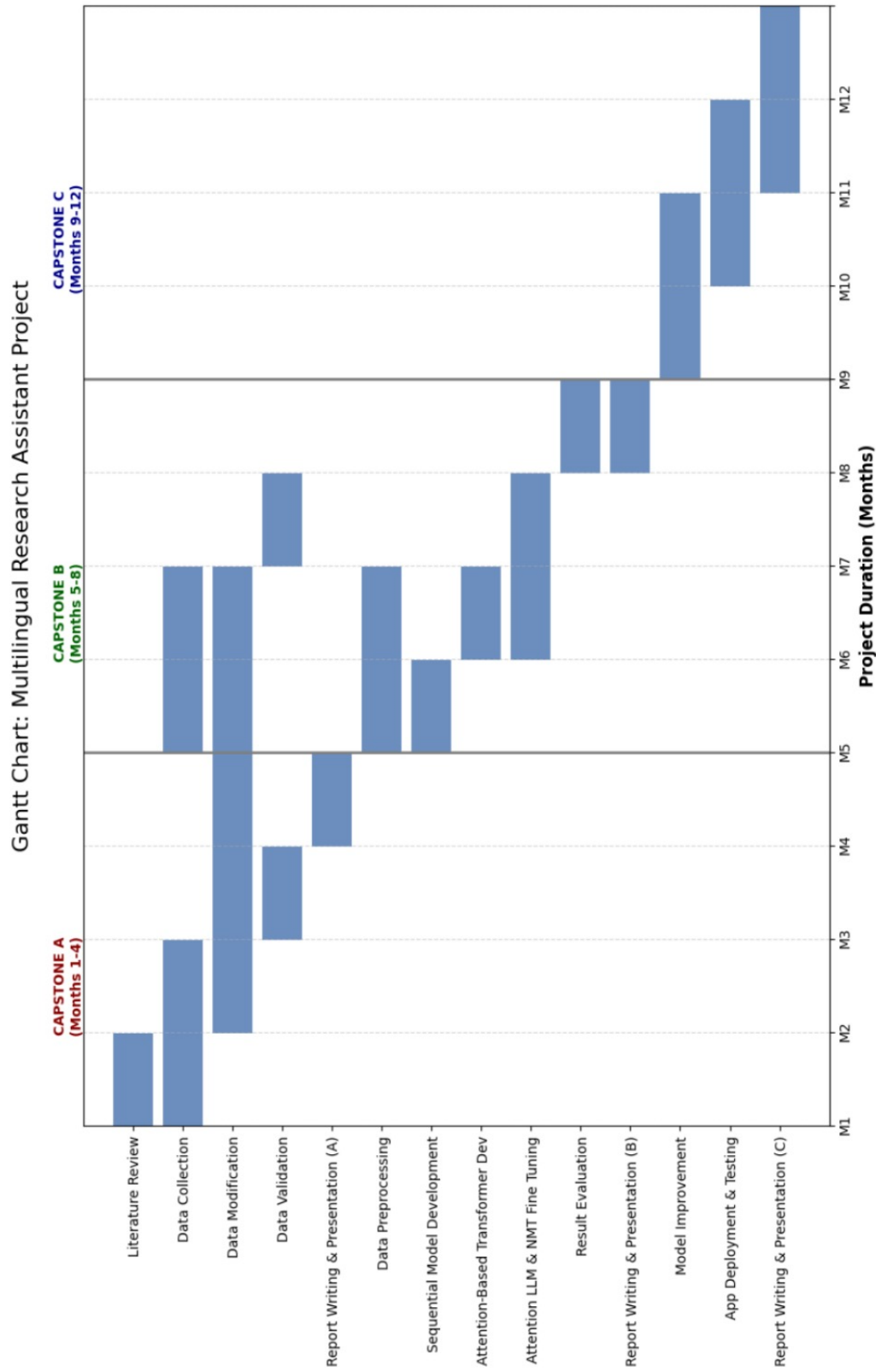


Figure 5.1: Gantt Chart of this Project

Chapter 6

Result and Analysis

6.1 Result Analysis of LLaMA-3 Model

The findings prove the effectiveness of the fine-tuned LLaMA-3 (8B) model in low-resource translation of the Marma. The model was enhanced gradually over time using a curriculum learning strategy and progressive data cleaning to achieve an accurate sentence-level translation at word-level grammatical understanding. Figure 6.2 illustrates that LLaMA-3 achieved 63.0% accuracy on word-level POS tagging and in Table 6.1, sentence-level translation in Marma-to-English was approximate 35-40 scores, which is a significant improvement over traditional NMT and RNN-based models. As shown in Table 6.1, language pairs with high resources acquired a BLEU-4 score of over 60 and the highest score of 71.04 which highlights the models strong performance in well-resourced scenarios. For Marma, Marma to English achieves a BLEU-4 score of 30.63 and also English to Marma reaches 44.88, which reflects cross-lingual generalization effectively. Moreover, the consistent improvement observed in Bangla to Marma directions also indicate that LLaMa-3 is able to leverage multilingual knowledge transfer. Particularly, it is observed that Marma (Bangla Script) to English BLEU-4 improved across a range of 5.11 to 22.34 due to fine-tuning on different formats of the dataset. The process of improvement is also confirmed by the incremental fine-tuning process, where the model was gradually trained on word-level POS-tagged dataset, after assertive, followed by interrogative and finally imperative sentence-level datasets and the final result is provided in Figure 6.4. In Fig. 6.1, it shows the result after training the model on assertive sentences. In Fig. 6.3, shows the result after training the model on the interrogative sentences and in Fig. 6.4, shows the result after training the model

on the imperative sentences. On the whole, the findings indicate that LLaMA-3 is well-generalized and does not overfit and provides a strong digitally inclusive translation model of the Marma language.

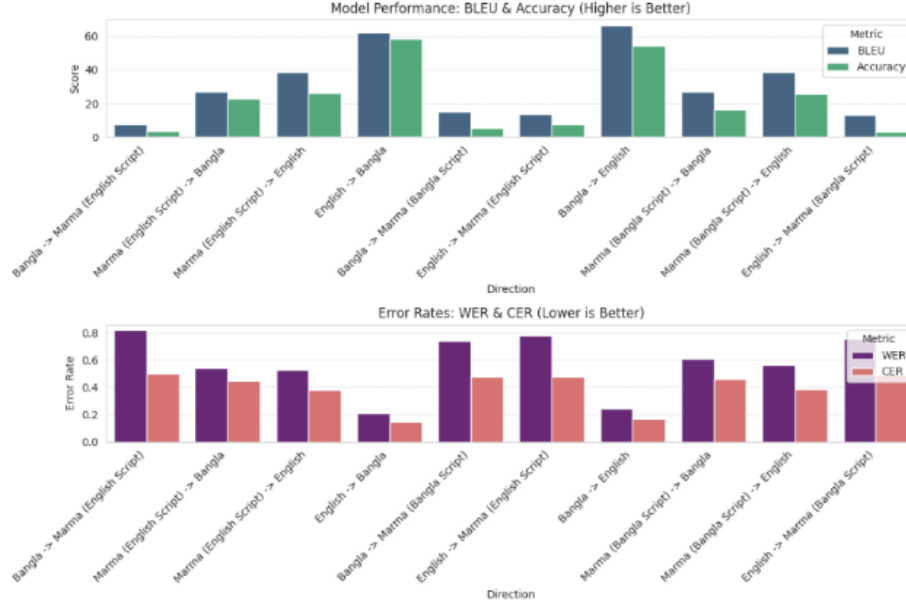
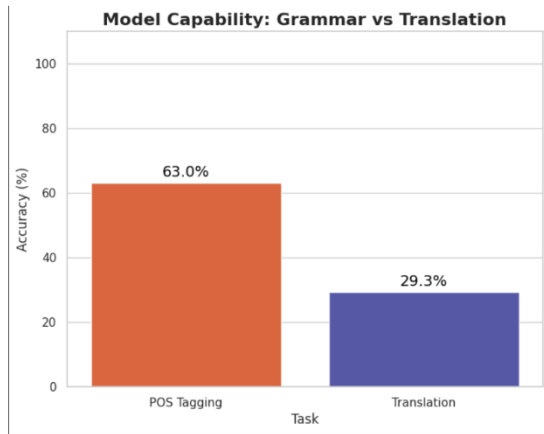


Figure 6.1: Assertive Sentence Training and Performance Analysis for LLaMA-3

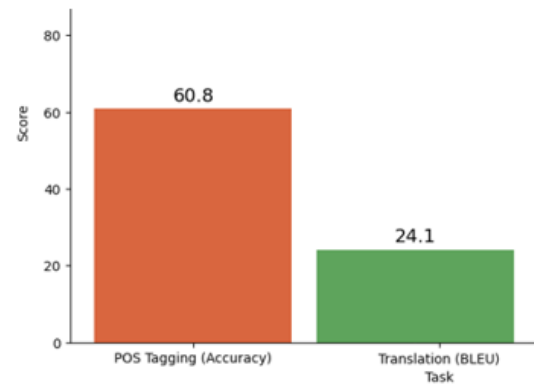
6.2 Result Analysis of NLLB-200 Model

We compared the fine-tuned LLaMA-3 (8B) to the baseline of NLLB-200 (600M) on 10 translation directions. Table 6.1 represents the final performance of the two models after completing the entire training pipeline, which comprised refinement of the noisy and the cleaned dataset, word-level POS tagging, and the last step of fine-tuning at the sentence level using the separated assertive, interrogative, and imperative datasets. As shown in Fig. 6.2b, it indicates that NLLB-200 model achieved 60.8% accuracy on the word-level POS tagging task, which demonstrates word-level linguistics learning effectively.

After training the model on assertive sentences we got scores for each direction shown in Fig. 6.5 and Fig. 6.6. Then after this training we used the saved model on the interrogative sentences and got the result which is shown in Fig. 6.7 and in Fig. 6.8 lastly, we did the training on the imperative sentences where the results



(a) LLaMA-3: Word-level Performance



(b) NLLB-200: Word-level Performance

Figure 6.2: Model Capability and Performance Analysis

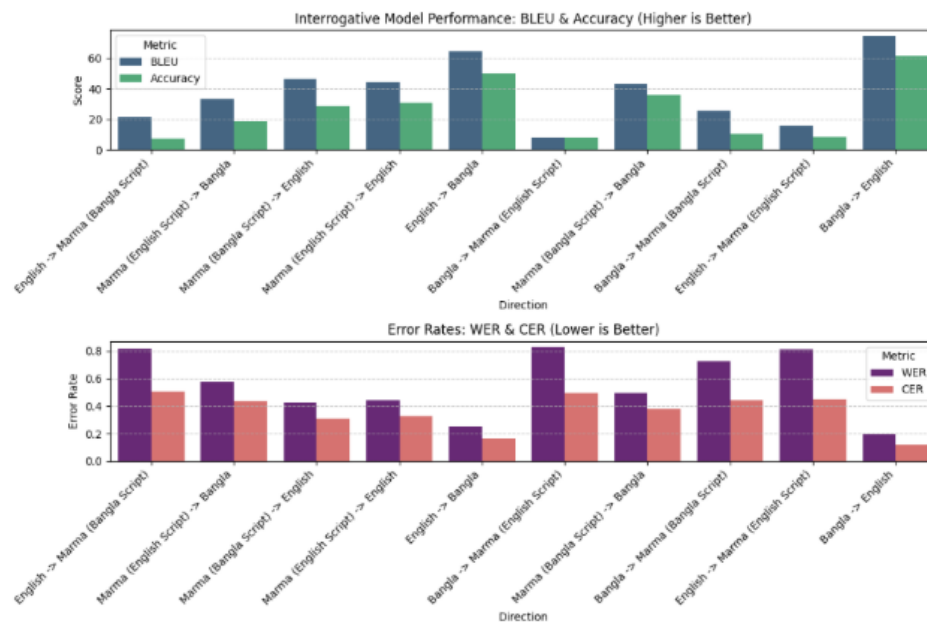


Figure 6.3: Interrogative Sentence Training and Performance Analysis

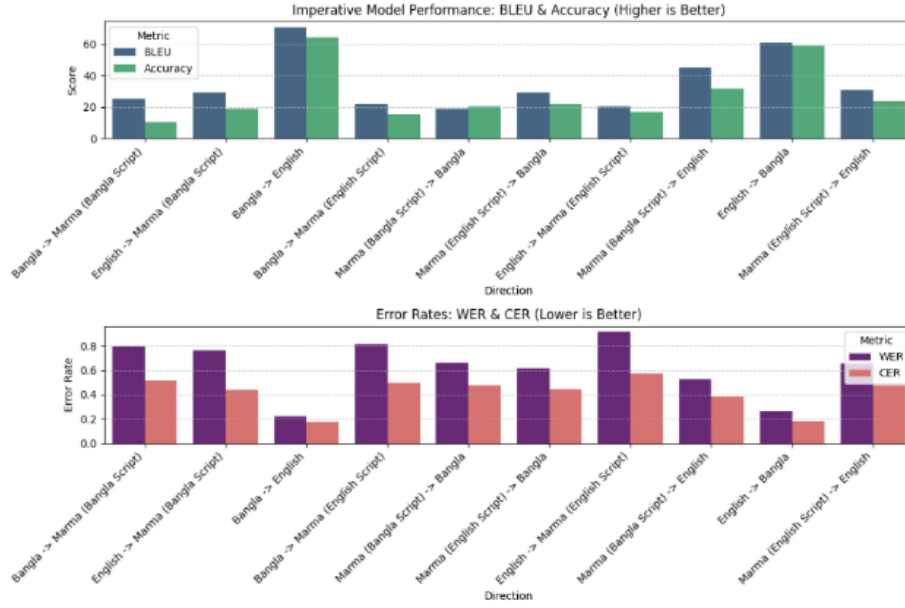


Figure 6.4: Imperative Sentence Training and Performance Analysis

are shown in Fig. 6.9 and Fig. 6.10. The findings are also presented in Figure 6.9 and 6.10 that the NLLB-200 model works effectively in managing Marma script transliteration and syntactic sentence construction when the resources are low. Bangla to English has the highest resource translation pairs with BLEU of 68.48, then the English - Bengali with a BLEU score of 46.97 which shows the reliability of quality of translation with high resources. In Marma-related translation to English, average results are obtained, with a BLEU score of 36.36 and 37.23 of Marma (Bangla) to English and Marma (English) to English, respectively. The model scored a BLEU-4 score of above 65 for high-resource language pairs and for Marma-related translation into English, moderate performance is observed while maintaining improved word accuracy above 50% and strong character accuracy levels of internal script mappings of Marma across assertive sentences, interrogative sentences, and imperative sentences. The results of BLEU scores 24.50 and 25.96 in the translation of Marma-to-Bangla are a sign of consistent syntactic map at the cost of few linguistic resources. On the other hand, Bangla and English translations to Marma also have much lower scores in BLEU with a range of 12.86 to 13.92, but Character Accuracy is still one of the highest, indicating an accurate preservation of the script and minimal character-level errors. These findings reveal that NLLB-

200 can serve as a reliable multilingual baseline and a solid base for the digital preservation of the Marma language.

📊 Detailed Scores per Direction:

Direction	BLEU	WER	CER	WA (%)	CA (%)	Samples
:-----:	-----:	-----:	-----:	-----:	-----:	-----:
ben_Beng-eng_Latn	68.9989	0.241055	0.162996	75.8945	83.7004	161
ben_Beng-mrm_Beng	10.1999	0.78424	0.532446	21.576	46.7554	161
ben_Beng-mrm_Latn	8.7207	0.697897	0.438614	30.2103	56.1386	161
eng_Latn-ben_Beng	62.5718	0.204586	0.159262	79.5414	84.0738	172
eng_Latn-mrm_Beng	12.4043	0.761649	0.508401	23.8351	49.1599	172
eng_Latn-mrm_Latn	10.6791	0.700544	0.453643	29.9456	54.6357	172
mrm_Beng-ben_Beng	15.3497	0.651489	0.510476	34.8511	48.9524	170
mrm_Beng-eng_Latn	18.1765	0.723708	0.530065	27.6292	46.9935	170
mrm_Beng-mrm_Latn	76.064	0.143389	0.0389447	85.6611	96.1055	170
mrm_Latn-ben_Beng	14.0186	0.660617	0.519244	33.9383	48.0756	164
mrm_Latn-eng_Latn	18.9904	0.716946	0.518167	28.3054	48.1833	164
mrm_Latn-mrm_Beng	66.1743	0.210526	0.055805	78.9474	94.4195	163

Figure 6.5: Performance Analysis of Assertive Sentence (NLLB-200)

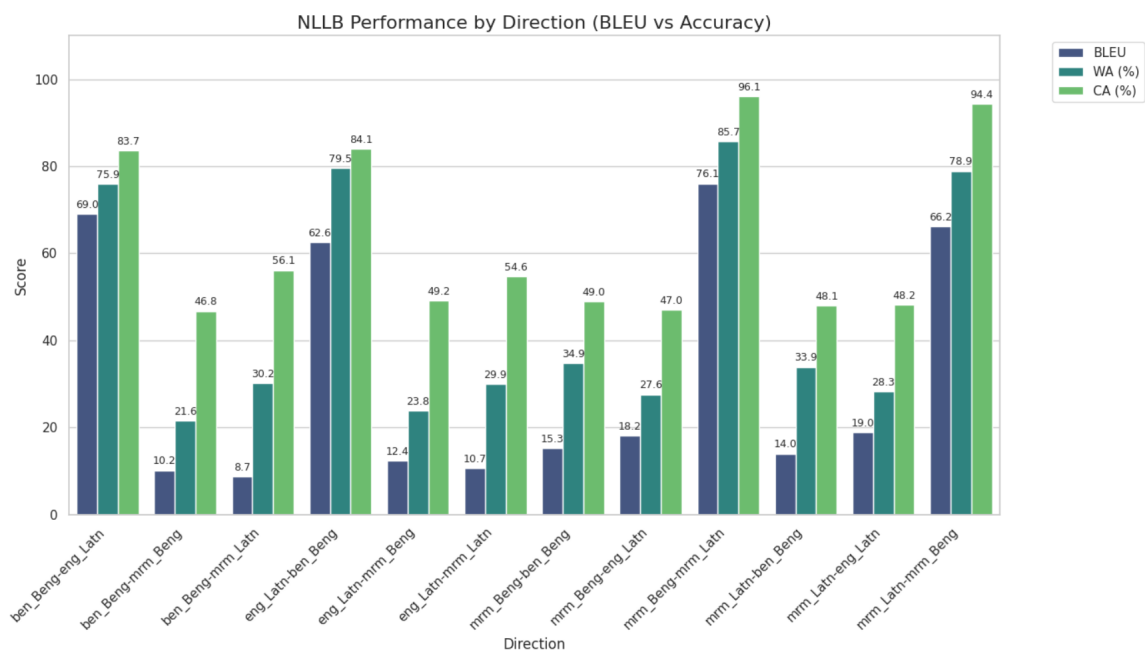


Figure 6.6: Performance Analysis Chart of Assertive Sentence (NLLB-200)

📊 Detailed Scores per Direction:

Direction	BLEU	WER	CER	WA (%)	CA (%)	Samples
ben_Beng-eng_Latn	74.3358	0.170819	0.118692	82.9181	88.1308	156
ben_Beng-mrm_Beng	15.8549	0.744565	0.465731	25.5435	53.4269	156
ben_Beng-mrm_Latn	10.8169	0.745455	0.442353	25.4545	55.7647	156
eng_Latn-ben_Beng	55.0166	0.288162	0.176395	71.1838	82.3605	184
eng_Latn-mrm_Beng	13.2229	0.767477	0.482711	23.2523	51.7289	184
eng_Latn-mrm_Latn	12.6561	0.763359	0.443822	23.6641	55.6178	184
mrm_Beng-ben_Beng	17.3871	0.629032	0.46151	37.0968	53.849	166
mrm_Beng-eng_Latn	28.3929	0.5625	0.427807	43.75	57.2193	166
mrm_Beng-mrm_Latn	73.6135	0.16179	0.0377778	83.821	96.2222	165
mrm_Latn-ben_Beng	22.3202	0.601113	0.460678	39.8887	53.9322	161
mrm_Latn-eng_Latn	25.526	0.610915	0.435145	38.9085	56.4855	161
mrm_Latn-mrm_Beng	60.6861	0.275439	0.0877193	72.4561	91.2281	161

Figure 6.7: Performance Analysis of Interrogative Sentence (NLLB-200)

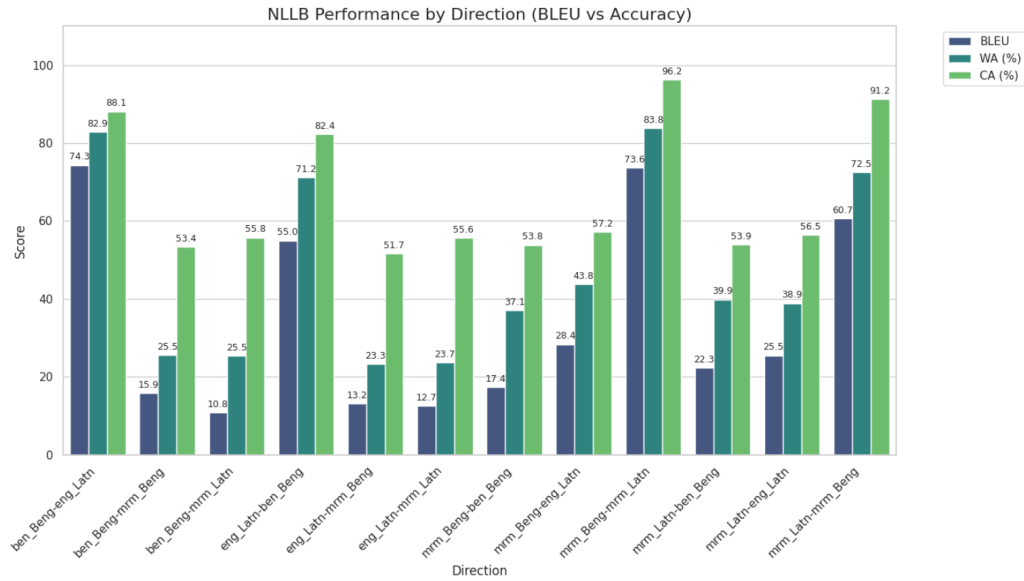


Figure 6.8: Performance Analysis Chart of Interrogative Sentence (NLLB-200)

📊 Detailed Scores per Direction:

Direction	BLEU	WER	CER	WA (%)	CA (%)	Samples
ben_Beng-eng_Latn	68.4762	0.246835	0.171035	75.3165	82.8965	157
ben_Beng-mrm_Beng	12.8584	0.881356	0.530908	11.8644	46.9092	157
ben_Beng-mrm_Latn	13.9215	0.829424	0.493303	17.0576	50.6697	157
eng_Latn-ben_Beng	46.9661	0.376289	0.223149	62.3711	77.6851	188
eng_Latn-mrm_Beng	13.8907	0.848592	0.508025	15.1408	49.1975	188
eng_Latn-mrm_Latn	13.1441	0.819643	0.484623	18.0357	51.5377	188
mrm_Beng-ben_Beng	24.4973	0.626667	0.462333	37.3333	53.7667	167
mrm_Beng-eng_Latn	36.3604	0.572874	0.437158	42.7126	56.2842	167
mrm_Beng-mrm_Latn	65.9092	0.198807	0.0571327	80.1193	94.2867	167
mrm_Latn-ben_Beng	25.9619	0.608879	0.444303	39.1121	55.5697	155
mrm_Latn-eng_Latn	37.229	0.5625	0.422283	43.75	57.7717	155
mrm_Latn-mrm_Beng	57.151	0.297727	0.0729123	70.2273	92.7088	154

Figure 6.9: Performance Analysis of Imperative Sentence (NLLB-200)

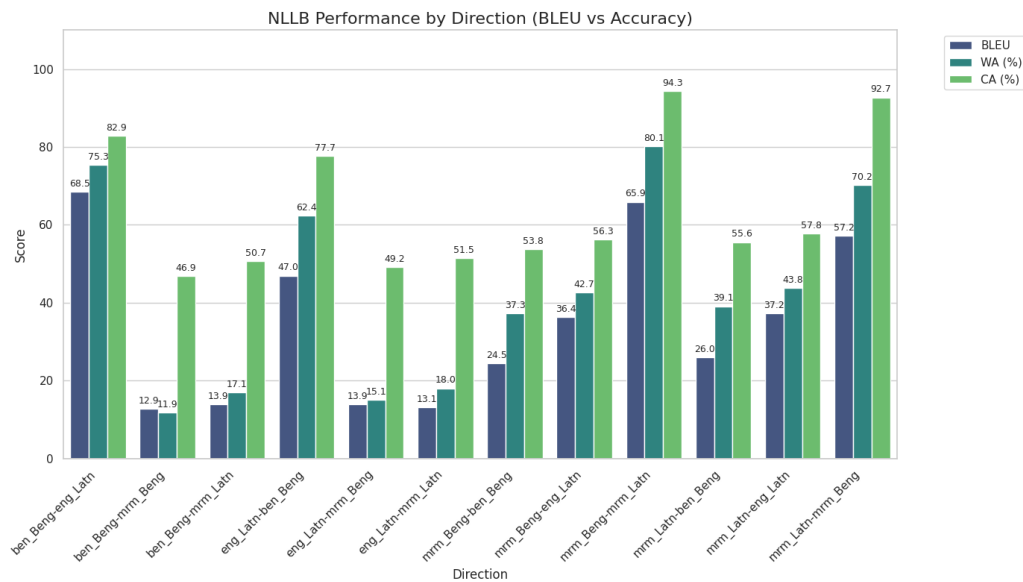


Figure 6.10: Performance Analysis Chart of Imperative Sentence (NLLB-200)

6.3 Comparative Analysis

As the results were compared to reveal the best and the worst outcomes, it is evident that the fine-tuned Llama-3 model has a better overall average of BLEU since it generates an average of 35.22, contrasting with 29.33 of NLLB, as shown in Table 6.12. Llama-3 has a decisive benefit in high-resource translation, especially in the English-to-Bangla direction, where it outperforms the baseline by more than 13 points (60.70 vs. 46.97) and it is the best generative model of the low-resource language of Marma, improving on the performance of the baseline by over twice when translating either English or Bangla to Marma. Although NLLB-200 is more competitive (37.23 vs. 30.63) in Romanized Marma script interpreting English, Llama-3 has stronger features in creating coherent text in Marma and its even performance in the two-way tasks, which makes it the most effective solution to be used in this research project, as shown in Fig. 6.11 and Fig. 6.1

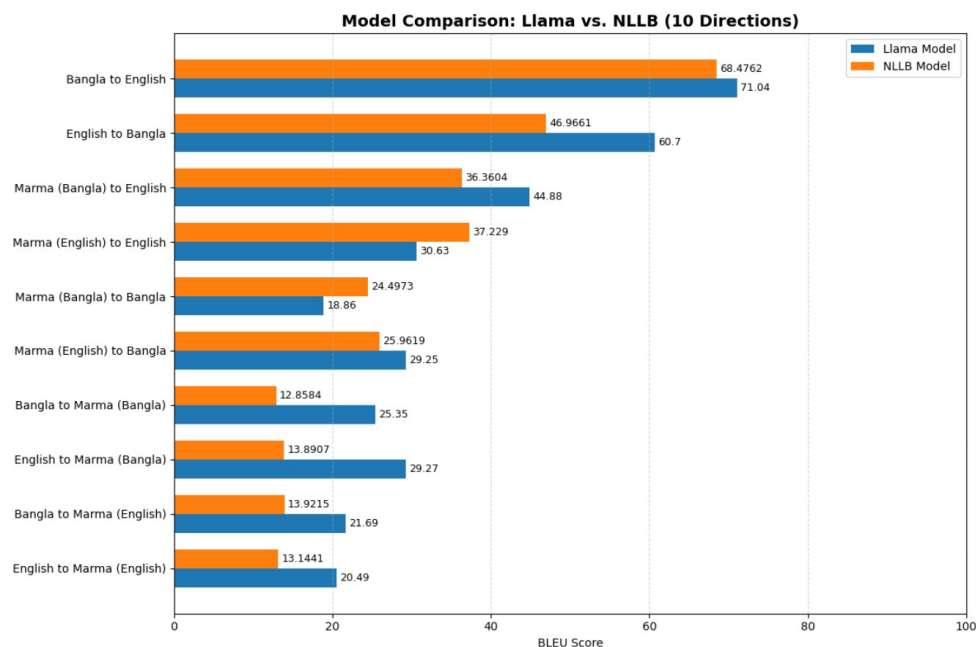


Figure 6.11: Comparison Table

Comparison between MELT and Kokborok In 6.2, the paper kokborok language is a sino-tibetan language just like Marma, both are member of sino-tibetan family and both of them are low-resource language, that's why, we can compare between

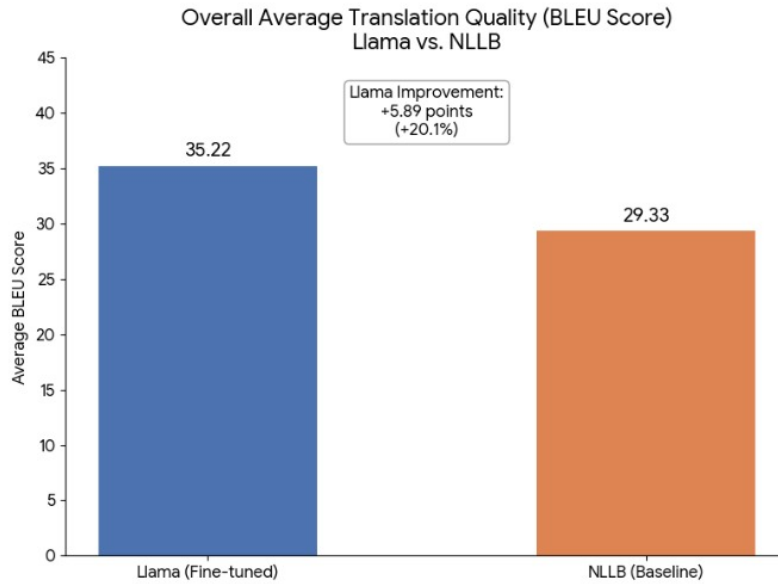


Figure 6.12: Comparison Chart

Table 6.1: Performance Comparison (BLEU-4 Scores)

Translation Direction	NLLB-200	LLaMA-3
Bangla → English	68.48	71.04
English → Bangla	46.97	60.70
Marma (Bangla) → English	36.36	44.88
Marma (English) → English	37.23	30.63
Marma (Bangla) → Bangla	24.50	18.86
Marma (English) → Bangla	25.96	29.25
Bangla → Marma (Bangla)	12.86	25.35
English → Marma (Bangla)	13.89	29.27
Bangla → Marma (English)	13.92	21.69
English → Marma (English)	13.14	20.49
Average Score	29.33	35.22

Table 6.2: Comparative Analysis (MELT vs Kokborok)

Feature	MELT (Marma)	KokborokNyalang & Debbarma (2026)
Language	Marma	Kokborok
Data Strategy	Ethnographic fieldwork (~12 thousand sentence pairs)	Synthetic back-translation (36 thousand sentence pairs)
Learning Strategy	Curriculum Learning (10 directions)	Joint bidirectional fine-tuning (2 directions)
Model	LLaMA-3 (8B), NLLB-200	NLLB-200-distilled-600M
Result	English→ Marma (English Transliteration): 20.49 BLEU, Marma (English Transliteration)→English: 30.63 BLEU	EN→Kokborok: 15.25 BLEU, Kokborok→EN: 38.56 BLEU Nyalang & Debbarma (2026)
Script Support	Roman and Bangla	Roman (trp_Latn)
Key Finding	LLM outperforms NMT in semantic consistency	LaBSE filtering unreliable for unseen languages
Deployment	FastAPI web application	Public release of models & data

them, others low-resource languages like Chakma, Garo, these are not from Sino-Tibetan family, that's why their structures will not be the same as Marma and they will not be appropriate to compare with Marma Language. The Kokborok cite paper worked on almost 36 thousand sentence pairs; we could manage only 12 thousand sentence pairs. Furthermore, we translated in all ten directions, including the four languages, two transliterated versions of the Marma language, Bangla, and English. Where our project translates from each language to the rest of the others. Our project performed better on LLaMA-3 model where their project did better on NLLB-200 distilled-600M, when coming to the BLUE score, our translation direction from English to Marma (English): 20.49 which is better than their project's translation direction, English to Kokborok: 15.25, then, in other direction Kokborok to English: 38.56 where we got from Marma (English Transliteration) to English: 30.63 as BLEU score, so, here Kokborok paper got a little bit higher score than us. Ultimately, it can be understood that our model did well on translation tasks as it could work on very small amounts of data, and it had no previous knowledge of the Marma Language. It got a very good BLEU score close to the Kokborok paper and somewhat better in one direction too. Also, it worked on 10 directions and got a good score in each direction, which is absent in the Kokborok paper.

6.4 Web Application

Created a translator website where it can translate in 10 directions including the four languages-Marma (English Transliterated), Marma (Bangla Transliterated), Bangla, and English. This application can be used as a way of communication with the ethnic group. In this website, the user needs to write the sentence in the source section and then can get the output in the output section. In both the source and output sections, the user can choose their preferred language. The same language cannot be selected in both the source and output sections simultaneously. Otherwise, this translator website will let the user translate from each language to others in all 10 directions total. Marma is written in two scripts—one using the Bangla script and another in English script - to make communication easier.

1. Used Stacks:

Backend: The backend framework was created using FastAPI, used as a server for the translation models through REST APIs that ensured quick asynchronous processing.

Model Deployment: The model was deployed in a scalable environment using the Hugging Face ecosystem to manage model weights and inference.

Frontend: Built using Python with a user-friendly interface adaptable to different devices.

2. Used Model:

Fine-tuned LLaMA-3 (8B) model was used to develop the whole system.

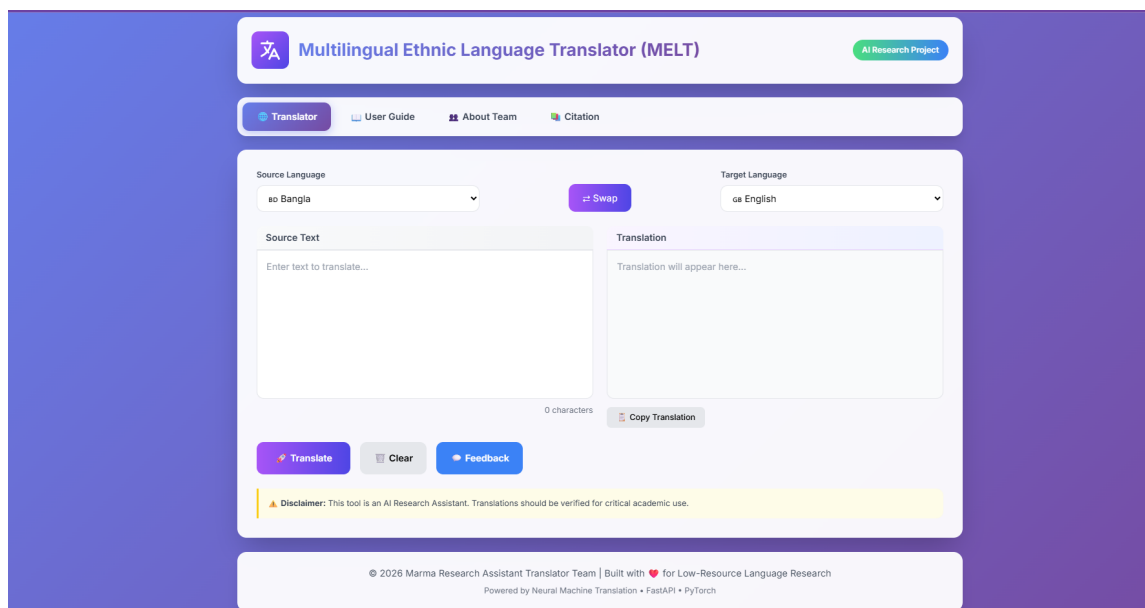


Figure 6.13: User Interface of MELT Application

In this Fig. 6.13, it shows the whole user interface of the project is shown, where the user can choose the source and the target language. Then, after clicking on the swap button, it will swap the source and target languages easily. To get the translation, the user needs to click on the translate button; he can also copy the translated text as well. He can clear the input by using the clear option, and then can give his feedback too.

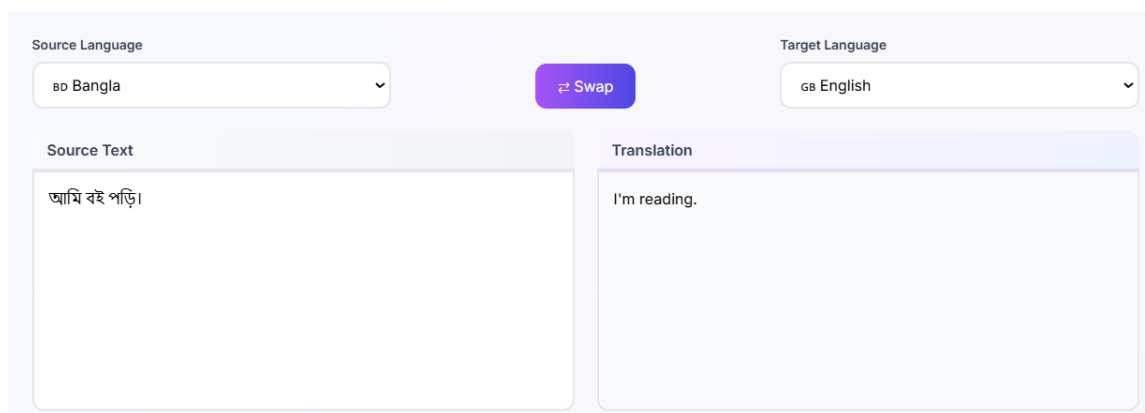


Figure 6.14: Bangla To English

In Fig. 6.14, it shows the translation from Bangla to English.

The interface features a light purple header with two dropdown menus: 'Source Language' set to 'বাংলা' and 'Target Language' set to 'Marma (English Script)'. A purple 'Swap' button is positioned between them. Below the header, the 'Source Text' box contains the Bangla text 'আমি বই পড়ি।'. The 'Translation' box on the right displays the English transliteration 'nga boi phya.'.

Figure 6.15: Bangla To Marma (English Transliterate)

In Fig. 6.15, it shows the translation from Bangla to Marma (English).

The interface is identical to Figure 6.15, but the 'Target Language' dropdown is set to 'Marma (Bangla Script)'. The 'Source Text' box still contains 'আমি বই পড়ি।'. The 'Translation' box on the right now displays the text in Bangla script: 'ঙা বোই রো ফাই।'.

Figure 6.16: Bangla To Marma (Bangla Transliterate)

In Fig. 6.16, it shows the translation from Bangla to Marma (Bangla).

The interface features a light purple header bar. On the left, 'Source Language' is set to 'en English' in a dropdown menu. In the center, a purple button with a double-headed arrow and the text 'Swap' is visible. On the right, 'Target Language' is set to 'bn Bangla' in a dropdown menu. Below the header, there are two main text areas. The left area, labeled 'Source Text', contains the text 'I eat rice'. The right area, labeled 'Translation', contains the Bengali text 'আমি ভাত খাই।'.

Figure 6.17: English To Bangla

In Fig. 6.17, it shows the translation from English to Bangla.

The interface is similar to Figure 6.17. The 'Source Language' is 'en English'. The 'Target Language' dropdown now shows 'Marma (English Script)' with a small icon to its left. The 'Source Text' area contains 'I read books'. The 'Translation' area contains the transliterated text 'nga boi ruwai.'.

Figure 6.18: English To (Marma English Transliterate)

In Fig. 6.18, it shows the translation from English to Marma (English).

The interface features a light purple header with two dropdown menus: 'Source Language' set to 'en English' and 'Target Language' set to 'Marma (Bangla Script)'. A purple 'Swap' button is positioned between them. Below the 'Source Language' dropdown is a text box labeled 'Source Text' containing the text 'I read books'. To the right, a text box labeled 'Translation' contains the Marma text 'ঙা বোই রো ফাই।'.

Figure 6.19: English To (Marma Bangla Transliterate)

In Fig. 6.19, it shows the translation from English to Marma (Bangla).

The interface features a light purple header with two dropdown menus: 'Source Language' set to 'Marma (English Script)' and 'Target Language' set to 'bn Bangla'. A purple 'Swap' button is positioned between them. Below the 'Source Language' dropdown is a text box labeled 'Source Text' containing the text 'nga boi faite'. To the right, a text box labeled 'Translation' contains the Bangla text 'আমার বইটা পড়ে।'.

Figure 6.20: Marma (English Transliterate) To Bangla

In Fig. 6.20, it shows the translation from Marma (English) to Bangla.

The interface has a light purple background. At the top, there are two dropdown menus: 'Source Language' on the left and 'Target Language' on the right. The 'Source Language' dropdown is set to 'Marma (English Script)' with a small icon of an open book. The 'Target Language' dropdown is set to 'English' with a small icon of a book. Between these two dropdowns is a purple button with a double-headed arrow and the word 'Swap'. Below the 'Source Language' dropdown is a text box labeled 'Source Text' containing the text 'nga boi faite'. Below the 'Target Language' dropdown is a text box labeled 'Translation' containing the text 'Read my book.'

Figure 6.21: Marma (English Transliterate) To English

In Fig. 6.21, it shows the translation from Marma (English) to English.

The interface has a light purple background. At the top, there are two dropdown menus: 'Source Language' on the left and 'Target Language' on the right. The 'Source Language' dropdown is set to 'Marma (Bangla Script)' with a small icon of an open book. The 'Target Language' dropdown is set to 'Bangla' with a small icon of a book. Between these two dropdowns is a purple button with a double-headed arrow and the word 'Swap'. Below the 'Source Language' dropdown is a text box labeled 'Source Text' containing the text 'ঙা বই ফাইতে'. Below the 'Target Language' dropdown is a text box labeled 'Translation' containing the text 'আমি বইটা পড়ছি।'.

Figure 6.22: Marma (Bangla Transliterate) To Bangla

In Fig. 6.22, it shows the translation from Marma (Bangla) to Bangla.

The image shows a web-based translation interface. At the top, there are two dropdown menus for language selection. The 'Source Language' dropdown is set to 'Marma (Bangla Script)' and the 'Target Language' dropdown is set to 'English'. Between these dropdowns is a purple button with a double-headed arrow and the text 'Swap'. Below the language selection, there are two main text areas. The left area, labeled 'Source Text', contains the Marma text 'ঙা বই ফাইতে'. The right area, labeled 'Translation', contains the English text 'I'm studying.'.

Figure 6.23: Marma (Bangla Transliterate) To English

In Fig. 6.23, it shows the translation from Marma (Bangla) to English.

Chapter 7

Conclusion

This research presented a robust, Attention-driven Multilingual Ethnic Language Translator (MELT) system designed to bridge the digital divide for the Marma language. By developing the first ethnographically validated parallel corpus and implementing a novel curriculum learning strategy, the study successfully demonstrated that Large Language Models can be effectively adapted for low-resource indigenous languages. The experimental results indicate that a fine-tuned LLaMA-3 model significantly outperforms traditional NMT baselines, achieving an average BLEU score of 35.22 and demonstrating superior capability in handling script diversity, transliteration complexity, and syntactic sentence construction despite limited linguistic resources. Beyond technical metrics, this proposed framework adheres to strict ethical guidelines, contributes toward reducing the digital language gap by ensuring reliable cross-lingual communication, and supports the digital preservation of the Marma language. There's still room to grow though. Expanding the corpus and adding real-time user feedback will take us even further.

7.1 Social, Legal, Ethical, and Environmental Issues

The development of an attention-based multilingual ethnic language translator in the Marma language has important social, ethical, legal, and environmental responsibilities. As engineers and researchers, it is our ethical duty to design a system that is safe, inclusive, legal, and sustainable and to serve the interests of the people and to protect vulnerable communities.

7.1.1 The Social Impact and Responsibility

The primary social aim of the proposed system is to reduce linguistic barriers experienced by the Marma community and ensure equal access to education, health information, government and non-governmental services and digital communication. The lack of effective translation resources for Marma has been associated with social exclusion and restricted involvement in mainstream services in the past. Special attention has been paid to make sure that the system does not violate cultural identity and linguistic integrity. The translation models are developed as a support system and not to substitute the human interpreters, and there is minimum possibility of cultural misrepresentation. The translator helps in communication and not as a legal or medical decision maker, thus decreasing the chances of social harm through misinterpretation or over-dependence on automated translations.

7.1.2 Safety Design and Ethical Consideration

The ethical responsibility in this project is concerned with accuracy, transparency, fairness, and misuse prevention. Since the Marma language is low-resource, the risk of translation mistakes can be expected. For this reason, the Design for Safety (DfS) approach was also taken during the development process in order to make sure that the system will not harm users or communities. A Hazard and Failure Analysis (HFA) was performed to find possible risks, and they included:

- Wrong or inaccurate translations in sensitive domains like medical or legal communication because of the lack of training data.
- Cultural misrepresentation or semantic distortion
- Pretrained multilingual bias.
- Completely depending on automated translation, which is not checked by humans.

To prevent such risks, several safety mechanisms were included in the design:

- It is stated that it is a support tool and not a certified legal or medical translator.
- Multiple evaluation metrics of uncertainty and error rates (BLEU, WER, CER, and accuracy) were used.

- Comparative study between NMT and LLM-based systems in terms of detecting systematic failures.
- Clear communication of model restrictions and performance thresholds.

The system clearly does not claim that it is always accurate and it records its limitations, which is also in agreement with the professional engineering standards of ethics that demand honesty and responsibility.

7.1.3 Law and Regulatory Obligations

The proposed system will meet local and international laws regarding software development, data usage, AI and research ethics. The dataset to be used in the project is publicly available, and ethically collected with no copyright or intellectual property laws violated. The system does not store or process any personal or sensitive user data, which aligns with the data protection principles. Also, the system honors the linguistic and cultural rights of indigenous populations, which promotes international standards like UNESCO guidelines on the maintenance of endangered languages.

7.1.4 Environmental Considerations

The system design also took into consideration environmental sustainability. Large-scale language model training can be computationally expensive and potentially energy-intensive. To deal with this, the project focuses on:

- Resource-efficient and low-resource model training, without unnecessary large-scale retraining.
- The system uses fine-tuning and transfer learning to train large-scale models using pretrained multilingual models (including NLLB-200 and LLaMA-3). This saves a lot of computational cost, training time and energy.
- The attention-based Transformer model was customized to fit one GPU or equivalent cloud resource and operate under a limited time, which renders the solution environmentally friendly and can be trained at institutions with limited resources. This strategy is both in line with green computing and responsible AI implementation.

The project is sustainable by minimizing energy use by maximizing computing power, and minimizing its environmental impact which is compatible with sustainable engineering principles. So, the suggested multilingual translation system is socially responsible, ethically safe, law-abiding, and environmentally friendly.

7.2 Brief Summary

This study bridges the gap between the languages and technologies utilized by the Marma community in Bangladesh by developing a Multilingual Translation Model, named MELT. Due to the scarcity of language resources, the team created a corpus of about 12,000 sentences in English, Bangla, Marma (Bangla Transliterated), and Marma (English Transliterated). This data set was curated to ensure that it aligns well with the language and cultural aspects.

Two highly advanced machine learning models, LLaMA-3 (8B) and NLLB-200 were used to evaluate which is better suited for translating. The experiments revealed that the LLaMA-3 model outperformed NLLB-200 as it understood varied writing styles and preserved the intended meanings more efficiently. As a result, LLaMA-3 produced a mean BLEU-4 score of 35.22 and the best word accuracy of 64.3%.

Lastly, the translation system was converted into a user-friendly web app using the FastAPI framework. This software can contribute to preserving the Marma language in digital form, furthering studies on the Marma people, and developing technologies for native peoples.

7.3 Future Works

This system shows good performance but it faces some limitations. The parallel corpus ($\sim 12k$ pairs) is still small and does not have enough optative, exclamatory, and positive/negative sentence segmentation, which limits the capacity of the model to represent expressive, sentiment-driven forms. The POS-tagged subset on the word-level is relatively very small, limiting grammatical learning and syntactic regularity, and creating noise of tokenization. There were also computational limitations that allowed less parameter tuning and experimentation with larger models.

The future directions of the work will involve the expansion of the multilingual Marma corpus by adding more diverse types of sentences, the development of larger linguistic resources, the combination of advanced training methods and multimodal features, and the implementation of the system as a web/mobile application to be tested on native speakers.

References

- A., M. (2022). A neural machine language translation system from German to English. *International Research Journal of Engineering and Technology (IRJET)*, 9(4), 3094–3103.
- Abbaszade, M., et al. (2021). Application of quantum natural language processing for language translation. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2021.3108768>
- Akan, M. F., Islam, A. B. S., & Ahmad, J. (2020). An insight into the intricacies of English to Bangla translation. *International Journal of Comparative Literature and Translation Studies*, 8(1), 10–20. <https://doi.org/10.7575/aiac.ijelts.v8n.1p.10>
- Ansari, M. Z., et al. (2022). Hindi to English transliteration using multilayer gated recurrent units. *Indonesian Journal of Electrical Engineering and Computer Science*, 27(2), 1083–1090. <https://doi.org/10.11591/ijeecs.v27.i2.pp1083-1090>
- Ansary, M. A. A. (2024). Neural machine translation between English and Bangla: A philosophical survey observing architectures and performances. *IOSR Journal of Computer Engineering*, 26(5), 17–25. <https://doi.org/10.9790/0661-2605011725>
- Aurelius, E., Sikoko, A. H., Rakun, E., & Azizah, K. (2025). Transfer learning and custom loss applied to transformer-based text translation for sign language animated subtitles. *IEEE Access*. <https://doi.org/10.1109/ACCESS.2025.3542785>
- Chakma, A., Chakma, A., Khisa, S., Tripura, C., Hasan, M., & Shahriyar, R. (2024). ChakmaNMT: A low-resource machine translation on Chakma language. *arXiv:2410.10219*.
- Das, R. K., et al. (2023). Sentiment analysis in multilingual context: Comparative analysis of machine learning and hybrid deep learning models. *Heliyon*, 9(9), e20281. <https://doi.org/10.1016/j.heliyon.2023.e20281>
- Dhar, A. C., et al. (2021). Bangla English machine translation using attention-based multi-headed transformer model. *Journal of Computer Science*, 17(10), 1000–1010. <https://doi.org/10.3844/ajbbsp.2021.1000.1010>

- Geetha, K., & Priyakumari, M. (2024). Machine learning based on multilingual translation. *International Journal of New Innovations in Engineering and Technology*, 24(1), 1421–1426.
- Islam, R., et al. (2023). Bangla to English translation using sequence to sequence learning model based recurrent neural networks. In *Lecture Notes of the Institute for Computer Sciences*. https://doi.org/10.1007/978-3-031-34619-4_3
- Israr, H., et al. (2023). Neural machine translation models with attention-based dropout layer. *Computers, Materials & Continua*, 75(2), 2982–3007. <https://doi.org/10.32604/cmc.2023.035814>
- Khan, A., & Sarfaraz, A. (2019). RNN-LSTM-GRU based language transformation. *Soft Computing*. <https://doi.org/10.1007/s00500-019-04281-z>
- Kowsher, M., et al. (2022). Bangla-BERT: Transformer-based efficient model for transfer learning and language understanding. *IEEE Access*, 10, 91855–91870. <https://doi.org/10.1109/ACCESS.2022.3197662>
- Laskar, S. R., Dadure, P., Manna, R., Pakray, P., & Bandyopadhyay, S. (2022). English to Bengali multimodal neural machine translation using transliteration-based phrase pairs augmentation. In *Proceedings of the 9th Workshop on Asian Translation (WAT2022)* (pp. 111–116).
- Laskar, S. R., et al. (2019). English to Hindi multi-modal neural machine translation and Hindi image captioning. In *Proceedings of the 6th Workshop on Asian Translation* (pp. 62–67).
- Mahi, M. H., Khan, A. R., Anik, M. H., Noori, S. R. H., Mahmud, A., & Mojumdar, M. U. (2025). MELD: A multilingual ethnic dataset of Chakma, Garo, and Marma in Bengali script with English and standard Bengali translation. *Data in Brief*, 61, 111745. <https://doi.org/10.1016/j.dib.2025.111745>
- Mondal, S. K., et al. (2024). Enhancement of English-Bengali machine translation leveraging back-translation. *Applied Sciences*, 14(15), 6848. <https://doi.org/10.3390/app14156848>
- Mumin, M. A. A., Seddiqui, M. H., Iqbal, M. Z., & Islam, M. J. (2019). Neural machine translation for low-resource English-Bangla. *Journal of Computer Science*, 15(11), 1627–1637. <https://doi.org/10.3844/jcssp.2019.1627.1637>
- NLLB Team, et al. (2022). No language left behind: Scaling human-centered machine translation. arXiv:2207.04672.
- Nyalang, B., & Debbarma, B. (2026). Towards high-quality machine translation for Kokborok: A low-resource Tibeto-Burman language of Northeast India. arXiv:2604.19778.

- Oruh, J., Viriri, S., & Adegun, A. (2022). Long short-term memory recurrent neural network for automatic speech recognition. *IEEE Access*, 10, 30069–30079. <https://doi.org/10.1109/ACCESS.2022.3159339>
- Podder, K. K., et al. (2023). Self-ChakmaNet: A deep learning framework for indigenous language learning using handwritten characters. *Egyptian Informatics Journal*, 24(4), 100413.
- Rahman, M. S., Mridha, M. F., & Akhand, M. A. H. (2020). Improving neural machine translation for low-resource languages using multiple encoders and parallel attention. In *Proceedings of the 2020 IEEE Region 10 Symposium (TENSYP)* (pp. 1591–1594). <https://doi.org/10.1109/TENSYP50017.2020.9230727>
- Sen, O., et al. (2022). Bangla natural language processing: A comprehensive analysis of classical, machine learning, and deep learning based methods. *IEEE Access*.
- Shalu, P., & M., M. (2021). Neural machine translation for English to Hindi using GRU. <https://ssrn.com/abstract=3851323>
- Sharma, A., & Sharma, V. (2021). Language translation using machine learning. *International Research Journal of Modernization in Engineering Technology and Science*, 3(6), 1245–1250.
- Siddique, S., Ahmed, T., Talukder, M. R. A., & Uddin, M. M. (2020). English to Bangla machine translation using recurrent neural network. *International Journal of Future Computer and Communication*, 9(2), 46–51. <https://doi.org/10.18178/ijfcc.2020.9.2.564>
- Tiwari, G., et al. (2020). English-Hindi neural machine translation- LSTM Seq2Seq and ConvS2S. In *Proceedings of the International Conference on Communication and Signal Processing* (pp. 871–875).
- Touvron, H., et al. (2023). LLaMA: Open and efficient foundation language models. [arXiv:2302.13971](https://arxiv.org/abs/2302.13971).
- Vaswani, A., et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*. Long Beach, CA.
- Zargar, S. A. (2021). Introduction to sequence learning models: RNN, LSTM, GRU. <https://doi.org/10.13140/RG.2.2.37099.522>

Appendix A

Dataset Validation Evidence

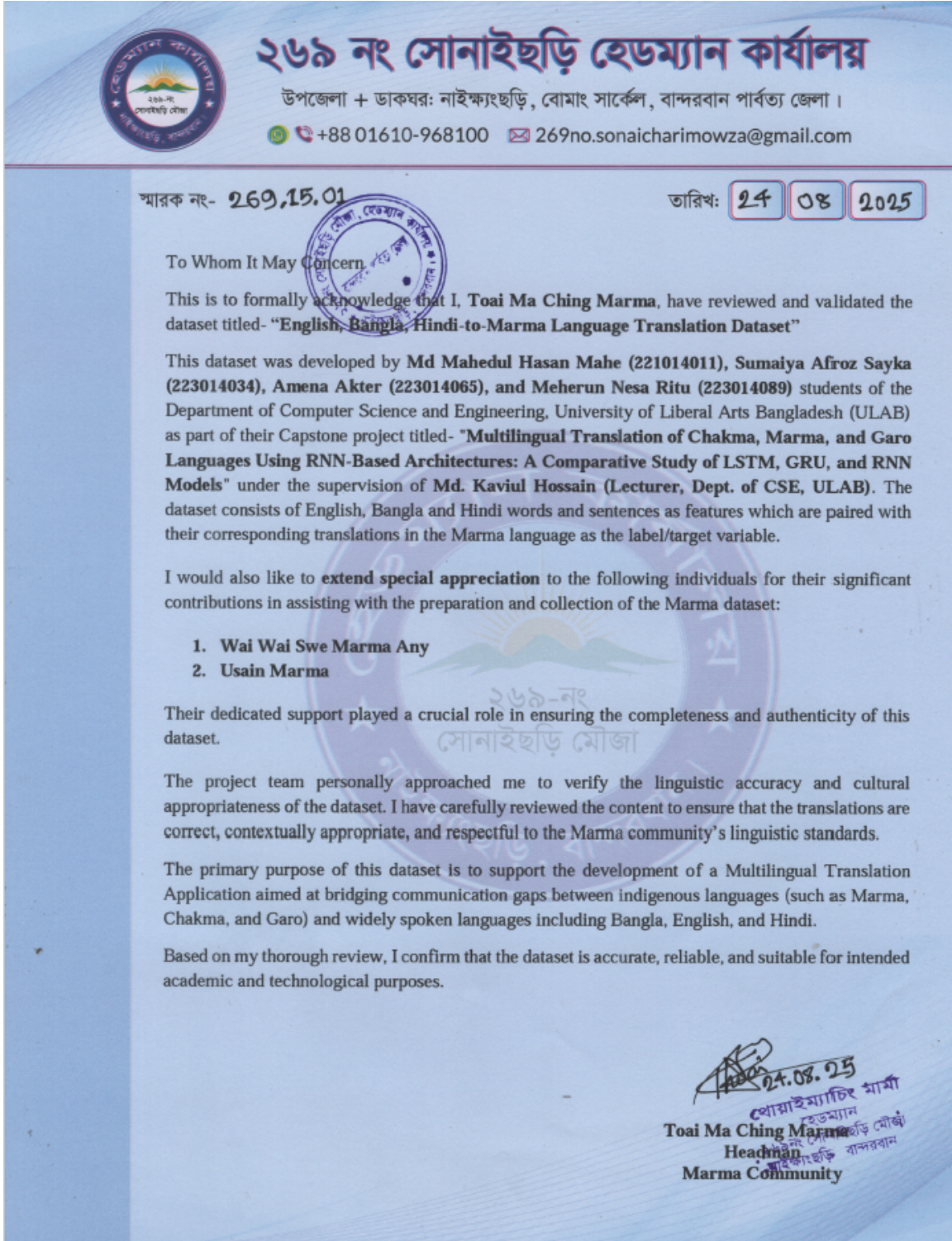


Figure A.1: Data Validation Evidence

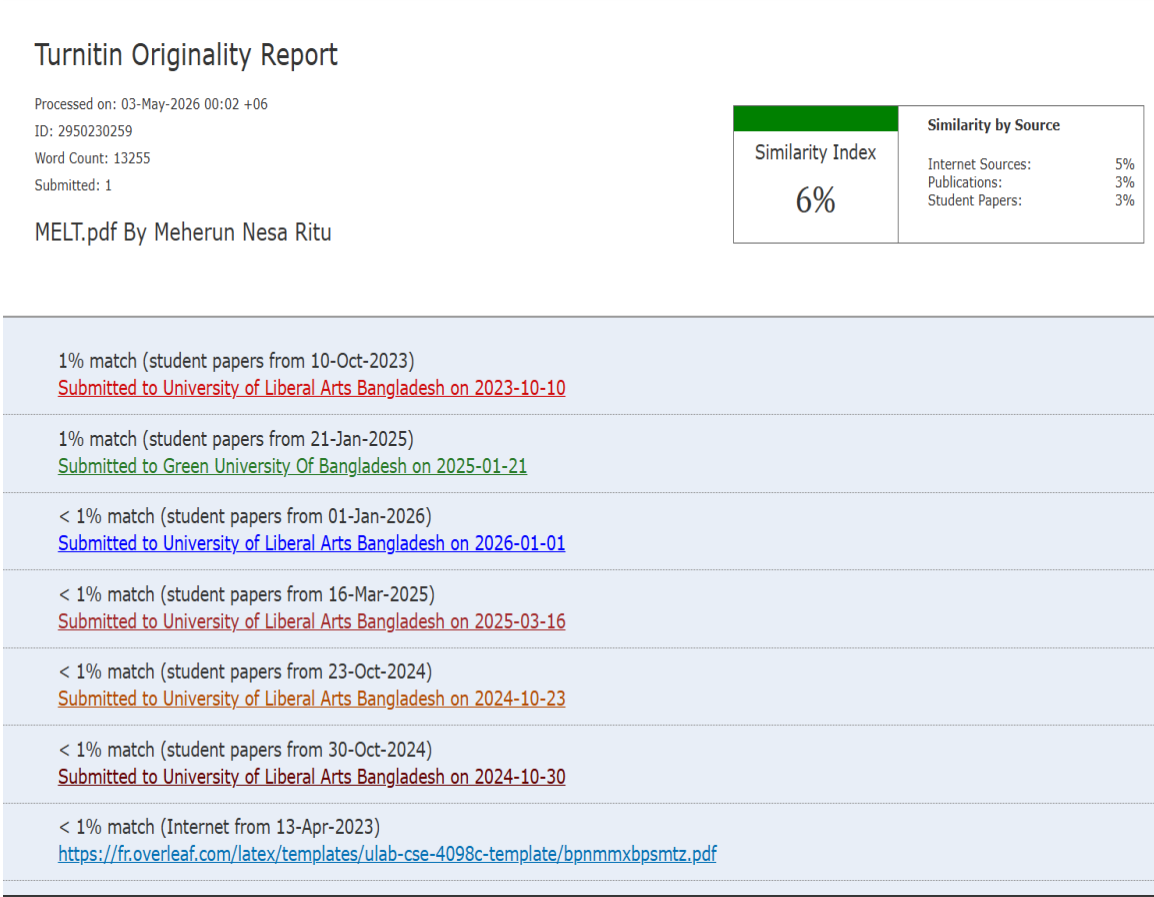


Figure A.2: Turnitin Similarity Check