# Beyond Vanilla Fine-Tuning: Leveraging Multistage, Multilingual, and Domain-Specific Methods for Low-Resource Machine Translation

SARUBI THILLAINATHAN, Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka and Department of Language Science and Technology, Saarland Informatics Campus, Saarland University, Germany SONGCHEN YUAN, Department of Computer Science, University of Toronto, Canada

EN-SHIUN ANNIE LEE, Department of Computer Science, University of Toronto, Canada and Faculty of Science, University of Ontario of Technology, Canada

SANATH JAYASENA, Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka SURANGIKA RANATHUNGA, Department of Computer Science and Engineering, University of Moratuwa, Sri Lanka and School of Mathematical and Computational Sciences, Massey University, New Zealand

Fine-tuning multilingual sequence-to-sequence large language models (msLLMs) has shown promise in developing neural machine translation (NMT) systems for low-resource languages (LRLs). However, conventional single-stage fine-tuning methods struggle in extremely low-resource NMT settings, where training data is very limited. This paper contributes to artificial intelligence by proposing two approaches for adapting msLLMs in these challenging scenarios: (1) continual pre-training (CPT), where the msLLM is further trained with domain-specific monolingual data to compensate for the under-representation of LRLs, and (2) intermediate task transfer learning (ITTL), a method that fine-tunes the msLLM with both in-domain and out-of-domain parallel data to enhance its translation capabilities across various domains and tasks. As an application in engineering, these methods are implemented in NMT systems for Sinhala, Tamil, and English (six language pairs) in domain-specific, extremely low-resource settings (datasets containing fewer than 100,000 samples). Our experiments reveal that these approaches enhance translation performance by an average of +1.47 bilingual evaluation understudy (BLEU) score compared to the standard single-stage fine-tuning baseline across all translation directions. Additionally, a multi-model ensemble further improves performance by an additional BLEU score.

Additional Key Words and Phrases: low-resource languages neural machine translation, large language models, fine-tuning, pre-training

#### **ACM Reference Format:**

Authors' addresses: Sarubi Thillainathan, Department of Computer Science and Engineering, University of Moratuwa, Katubedda, Sri Lanka and Department of Language Science and Technology, Saarland Informatics Campus, Saarland University, Saarbrücken, Germany, sarubi@coli.uni-saarland. de; Songchen Yuan, Department of Computer Science, University of Toronto, Toronto, Canada, ysc.yuan@mail.utoronto.ca; En-Shiun Annie Lee, Department of Computer Science, University of Toronto, Toronto, Canada and Faculty of Science, University of Ontario of Technology, Oshawa, Canada, annie.lee@cs.toronto.edu; Sanath Jayasena, Department of Computer Science and Engineering, University of Moratuwa, Katubedda, Sri Lanka, sanath@cse.mrt.ac.lk; Surangika Ranathunga, Department of Computer Science and Engineering, University of Moratuwa, Katubedda, Sri Lanka and School of Mathematical and Computational Sciences, Massey University, Auckland, New Zealand, s.ranathunga@massey.ac.nz.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

©~2025 Association for Computing Machinery.

Manuscript submitted to ACM

#### 1 INTRODUCTION

Neural Machine Translation (NMT) systems have achieved remarkable success in recent years, with significant improvements in translating both high-resource and low-resource languages. A promising approach that has emerged to improve the translation of these diverse language pairs is the fine-tuning (single-stage) Large Language Models, particularly multilingual sequence-to-sequence Large Language Models (msLLMs) such as mBART [28, 41] and mT5 [45]. This technique involves fine-tuning pre-trained msLLMs directly on the language pairs, leveraging their extensive pre-trained knowledge to deliver substantial performance gains compared to training an NMT system from scratch with randomly initialized weights. Studies [25, 28, 41, 42] have demonstrated that this method significantly boosts translation accuracy, making it a viable option for improving NMT outcomes.

However, while single-stage fine-tuning has shown impressive results, it often falls short for Low-Resource Languages (LRLs)<sup>1</sup>, particularly those that are underrepresented in the pre-trained msLLMs due to limited monolingual data. This problem is more notable in domain-specific translation. As demonstrated by [25], a straightforward, single-stage fine-tuning of msLLMs tends to yield suboptimal outcomes for these LRLs, requiring more advanced strategies to improve translation performance in such contexts.

Moreover, even generative LLMs, continue to struggle significantly with the translation [44]. The performance gap between these models and msLLMs remains particularly wide in LRL settings. This underscores the ongoing need for refined training approaches that better accommodate the nuances and scarcities associated with LRL.

In this study, we aim to address the existing limitations in the performance of msLLMs, particularly in domain-specific settings, by answering the following research questions:

- (1) How can the translation performance of msLLMs for LRLs be enhanced by addressing the underrepresentation of these languages in the training data?
- (2) Beyond conventional single-stage fine-tuning, what alternative techniques can be employed to leverage parallel corpora from different domains to enhance the adaptability and accuracy of msLLMs in translating LRLs?
- (3) Considering these advanced techniques, how can we develop a comprehensive strategy to significantly improve the performance of msLLMs in translating LRLs in domain-specific settings within NMT frameworks?

To answer the question (1) and (2), we investigate two techniques that utilize additional datasets to assist in the training process - Continual Pre-Training (CPT)<sup>2</sup> and Intermediate Task Transfer Learning (ITTL)<sup>3</sup>. For question (3), we combined the aforementioned CPT and ITTL two techniques to advance msLLMs for LRL-NMT in domain-specific contexts:

- (1) Continually Pre-Train msLLMs with domain-specific monolingual corpora, followed by
- (2) **Fine-Tuning these CPT-enhanced msLLMs** using in- and out-domain parallel corpora with ITTL. We propose two ITTL fine-tuning strategies
  - (a) a three-stage bilingual fine-tuning method, involving sequential fine-tuning with out-domain, mixed-domain, and in-domain parallel data
  - (b) a hybrid multilingual and bilingual fine-tuning approach, combining multilingual fine-tuning followed by bilingual fine-tuning.

<sup>&</sup>lt;sup>1</sup>To define the extent of resource scarcity, we adopt the threshold proposed by [36], classifying NMT's training data as low-resource or extremely low-resource when the available parallel corpus contains fewer than 0.5 million or 0.1 million sentences, respectively. Nonetheless, it is important to note that this threshold is not an absolute criterion.

<sup>&</sup>lt;sup>2</sup>CPT—which entails further pre-training of an LLM with monolingual data, refer to Section-6.3 for more details

<sup>&</sup>lt;sup>3</sup>ITTL-fine-tuning with the msLLMs with auxiliary parallel data before fine-tuning with the limited target domain data, refer to Section-6.2

(3) This methodology is further enhanced by the use of ensemble strategies to optimise results.

We selected Sinhala, Tamil, and English in extremely LRL domain-specific settings. Among these three languages, we experiment with the possible combination of six language pair translation directions. Our proposed CPT with ITTL strategies outperforms the baseline (direct fine-tuning) strongly for all the six directions where we obtained +2.6, +1.83, +1.41, +1.17, +1.03 and +0.8 BLEU score improvement respectively for  $Ta \rightarrow En$ ,  $Ta \rightarrow Si$ ,  $Si \rightarrow Ta$ ,  $Si \rightarrow En$ ,  $En \rightarrow Ta$  over strong direct fine-tuning baseline. Furthermore, our multi-model ensemble improves the performance even further by an average of (+2.13) BLEU score.

#### 2 METHODOLOGY

#### 2.1 Overview

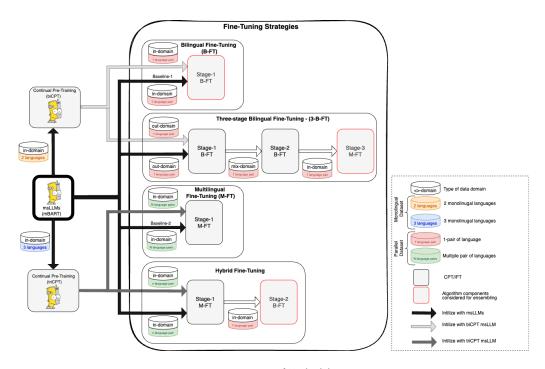


Fig. 1. Overview of Methodology

We use two single-stage fine-tuning baselines- Bilingual Fine-Tuning (B-FT) and Multilingual Fine-Tuning (M-FT) as described in [41]. In B-FT we fine-tune one language pair at a time with domain-specific parallel data. Meanwhile, in M-FT, we fine-tune more than one language pair simultaneously with domain-specific parallel data. M-FT is further categorized into three types:

- (1) Many-to-One Multilingual Fine-Tuning (M2O-FT):  $Many \rightarrow Language X$
- (2) One-to-Many Multilingual Fine-Tuning (O2M-FT): Language  $X \rightarrow Many$
- (3) Many-to-Many Multilingual Fine-Tuning (M2M-FT): Many → Many, with Language X as a pivot language.

In both B-FT and M-FT, training of the NMT model starts with the weights of the msLLM instead of random initialization. In other words, both the encoder and the decoder of the NMT model are initialized with the weights of the msLLM. The overview of this fine-tuning process is described in Fig. 1.

We investigate CPT and ITTL to obtain better results over single-stage fine-tuning of msLLMs for domain-specific LRL-NMT. These are the main three approaches:

- Continual Pre-Training (CPT) of the msLLM for domain adaptation.
- ITTL fine-tuning of msLLM with in- and out-domain parallel data for the NMT task.
- Ensembling different fine-tuned models.

Fig. 1 summarises how our techniques are implemented/related.

# 2.2 Continual Pre-training for Domain Adaptation

To mitigate the under-representation of LRLs within the msLLM, we extend the pre-training phase of the msLLM by integrating supplementary monolingual data. During this CPT process, the msLLM is exposed to a procedure wherein the monolingual data is deliberately 'noised', or altered, and then presented as the input (source). The model is then tasked with reconstructing or translating this noised input back into its original, unaltered form (target), a method known as self-supervised learning.

Our approach to CPT encompasses the utilization of various types of monolingual data, detailed as follows:

- Case A: Only with in-domain monolingual data.
- Case B: Only with out-domain monolingual data, which is larger in quantity.
- Case C: By combining both in-domain and out-domain monolingual data in two ways.
- Case C1: Mix in-domain CPT: Out-domain data together with in-domain data and create a mixed-domain dataset and then pre-train.

Case C2: Sequential CPT: Pre-training first with large out-domain and then in-domain data.

# 2.3 Intermediate Task Transfer Learning (ITTL)

We proposed two different ITTL fine-tuning strategies as follows: (exploration are visually represented and elaborated upon in Fig. 1)

- (1) Three-stage Bilingual Fine-tuning [3-B-FT]: We borrowed the idea of mix-fine-tuning from [10], where they first train a Transformer-based NMT model from scratch with out-domain parallel corpus and then fine-tune it on a parallel corpus that is a mix of the in-domain and out-domain corpora. However, they did not use any msLLMs. In our study, we propose the following on msLLMs:
  - (a) Initialize training with msLLM.
  - (b) B-FT with out-domain parallel data.
  - (c) B-FT with mixed domain parallel data where we up-sampled the in-domain data size to match with the larger out-domain corpus and created the mixed domain parallel data.
  - (d) Finally, B-FT with in-domain parallel data.
- (2) Multilingual Fine-tuning followed with Bilingual Fine-tuning [M-FT,B-FT]:
  - (a) Initialize training with the msLLM.
  - (b) Conduct M-FT using in-domain parallel data. This could be O2M-FT, M2O-FT, or M2M-FT, and select the best-performing model.

(c) Finally, tailer to one particular translation direction by doing B-FT with in-domain parallel data.

#### 2.4 Ensemble of Fine-tuned msLLMs

As we propose multiple CPT and ITTL techniques to improve msLLMs for domain-specific NMT, we utilize multiple similar models that share identical target vocabularies and the same decoding parameters; all are based on the same foundational msLLMs. To further enhance performance, we explore the potential of ensembling multiple fine-tuned msLLMs.

Ensembling is a machine learning technique that combines several base models to create an optimized predictive model [21]. In the context of NMT, ensembling involves translating the input sentence using multiple models, and then averaging the outputs from each to produce the final translation. We implement primarily two techniques for ensembling:

#### (1) Checkpoint Ensemble

This is the bare minimum way to conduct the ensemble from a single training process [7]. Previous studies applied checkpoint ensembling on RNN based NMT systems by combining the last N checkpoints of a single training [22, 38, 39]. In this study, we apply this checkpoint ensembling on msLLMs. We combine 3 different saved checkpoints from a particular single training. In other words, when we are fine-tuning the msLLM, we save the last ten checkpoints. We select 3 checkpoints that yield the best results in the validation test set among all the saved checkpoints. Then, we apply the checkpoint ensembling using these selected 3 checkpoints. In addition, we also experiment with the combination of 2 checkpoints. Finally, we report the best ensemble result among both of the above scenarios. Note that we can combine any number of models we want.

#### (2) Multi-model Ensemble

Besides our baseline (single-stage fine-tuning), we explore different techniques to improve the fine-tuning process of the msLLM (for example [3-B-FT] as shown in Fig. 1), resulting in multiple similar msLLMs that have been fine-tuned with in-domain parallel data at the final stage. We combine a maximum of 3 such msLLM fine-tuned models (due to our computational resource limitations) for the ensemble. We pick the top 3 best models out of these different approaches.

## 3 IMPLEMENTATION

#### 3.1 Dataset

For our experiments, we pick two languages that are underrepresented in the mBART model: Sinhala (Si) and Tamil (Ta), along with English (En).

# Domain-specific Parallel data

Our domain-specific parallel datasets are from official government documents of Sri Lanka [16], consisting of annual reports, crawled contents from government institutional websites, committee reports, procurement documents and acts. According to the statistics of our dataset given in Table 1, our dataset is smaller than 100k. Thus this creates an extremely low-resource domain-specific translation task.

## Out-domain Parallel data

We also gather publicly available out-domain parallel datasets from OPUS<sup>4</sup> and WMT MT tasks<sup>5</sup>. For Sinhala-English,

<sup>&</sup>lt;sup>4</sup>https://opus.nlpl.eu/

<sup>&</sup>lt;sup>5</sup>https://www.statmt.org/wmt20/translation-task.html

we use the FLoRes V1 training dataset<sup>6</sup> [20]. For Tamil-English, we use WMT20<sup>7</sup> MT news tasks' parallel datasets. For Sinhala-Tamil we use NLLB dataset from [11]. From NLLB, we pick the top 100,0000 samples after sentence similarity storing. Statistics of the dataset are given in Table 1.

#### Monolingual data for CPT

For monolingual data, we use domain-specific monolingual data obtained from [13] for our selected languages. For out-domain monolingual data, we use the online news site NewsFirst<sup>8</sup> [35] and FLoRes<sup>9</sup> [20]. The statics are given in Table 2.

Domain	Dataset	Language pair	No. of Sentence			
		Sinhala-Tamil (train)	66,348			
		Tamil-English (train)	66,348			
In-Domain	Government	Sinhala-Tamil (train)         66,348           Tamil-English (train)         66,348           Sinhala-English (train)         74,468           for all pairs (validation)         1,623           for all pairs (test)         1,603           Sinhala-English (train)         646,781           Tamil-English (train)         305,671				
		for all pairs (validation) 1,623				
		for all pairs (test)	1,603			
	FLoRes	Sinhala-English (train)	646,781			
Out-Domain	WMT20 News	Tamil-English (train)	305,671			
	NLLB	Sinhala-Tamil (train)	100,000			

Table 1. Statistics of the parallel dataset. In-domain dataset obtained from official government documents [16, 17]

Domain	Dataset	Language	No. of Sentences
		Sinhala	44,115
In-domain	Government Data	English	42,773
		Tamil	24,220
	News-first	Sinhala	650,000
Out-domain	News-IIIst	English	627,301
	FLoRes	Sinhala	646,781
	Lokes	English	646,781

Table 2. Statistics of the monolingual dataset.

# 3.2 msLLM selection

For our experiments, we select mBART50. It supports all the considered languages. In particular, mBART has shown promising results for supervised and unsupervised NMT [28, 41]. mBART is memory efficient and has shown relatively better results than mT5 [25]. We follow the same mBART model architecture — the standard sequence-to-sequence Transformer [43], with 12 layers of encoder-decoder with the model dimension of 1024 on 16 heads. For training, we use the FairSeq<sup>10</sup> tool.

<sup>6</sup>https://github.com/facebookresearch/flores/tree/main/floresv1

<sup>&</sup>lt;sup>7</sup>https://www.statmt.org/wmt20/translation-task.html

<sup>8</sup>https://www.newsfirst.lk/

<sup>9</sup>https://github.com/facebookresearch/flores/tree/main/floresv1

<sup>10</sup> https://github.com/pytorch/fairseq

#### 3.3 Baselines

As mentioned in Section 2.1, we consider Bilingual Fine-tuning (B-FT) and Multilingual Fine-Tuning (M-FT) as our baselines. For all directions, we fine-tune parallel data with 0.3 dropout, 0.2 label smoothing, 2500 warm-up steps, 3e-5 maximum learning rate as described by [28]. We use a maximum of up to 100k training updates for B-FT case. For the M-FT case, we applied the same setting as B-FT and used a maximum of 300k training updates. Additionally, we applied a temperature sampling rate of 1.5. Since our main goal is to conduct M-FT in a non-English centric manner, we chose Sinhala centric M-FT as well as Tamil centric M-FT. It is the most requested use case in Sri Lanka since the government documents are mostly produced in Sinhala and should be translated to Tamil and English [15]. Non-English translation, of course, has been identified as a common requirement in general for MT [14].

#### 3.4 Continual Pre-training for Domain Adaptation

We explore different data selections for the CPT strategies proposed in Section 2.2 as described below:

Case A: CPT with in-domain monolingual data.

We use available in-domain government monolingual data (see Table 2). However, the monolingual dataset is small - 44k, 42k, 24k for Sinhala, English and Tamil respectively.

Next, we utilize parallel data (as shown in Table 1) as monolingual data for each selected language pair and the available small monolingual government data to create a combined dataset. First, we use only the selected two languages for pre-training (biCPT). Then we move to CPT with three languages for M-FT, which we refer to as triCPT

Case B: CPT with out-domain monolingual data, which is larger in quantity.

We use the much larger out-domain monolingual data mainly obtained from news data crawled from Sri Lankan news websites [35] and FLoRes (see Table 2).

Case C: CPT by combining both in-domain and out-domain monolingual data.

Case C1: CPT with mixed domain (in-domain + out-domain) monolingual data. For this case, we took the FLoRes [20] out-domain monolingual data and mixed it with our in-domain government monolingual dataset mentioned in Case A.

Case C2: Multistage CPT on large out-domain and then in-domain data. Here we first pre-train with large out-domain monolingual data (case B) and then pre-train with in-domain monolingual data (case A)

# Pre-training (denoising) mBART50 msLLM

We further pre-train the mBART50 msLLM using the same objective function used for the initial pre-training of the model.

Our CPT data covers K languages:  $D = D_1, ..., D_K$ , where each  $D_i$  is monolingual data in a language i. Equation 1 is the noising function g that corrupts the text [28]. We train the model to predict the original text X given g(X). More formally, we aim to maximize  $L_\theta$ :

$$L_{\theta} = \sum_{D_i \in D} \sum_{X \in D_i} log P(x|g(x); \theta)$$
 (1)

- where X is an instance in language i
- the distribution *P* is defined by the msLLM.

We use the noising techniques used by the mBART model: Random Span masking and Order Permutation [28]. First, we select spans of text and replace it with a mask token. We mask 0.3% of words in each instance (with random masking 0.1) by randomly sampling a span length according to a Poisson distribution ( $\lambda = 3.5$ ). Also, we shuffle the order of sentences within each instance. The decoder input is the original text with one position offset. A language id symbol  $\langle LID \rangle$  is used as the initial token to predict the sentence as in [28].

## 3.5 ITTL Fine-tuning

We use two base models (mBART50 and best-performed CPT mBART50 which is an in-domain CPT model) to initialize the ITTL fine-tuning. We have two different ITTL fine-tuning techniques: three-stage fine-tuning and M-FT followed by B-FT as described in Section 2.3.

### 3.6 Preprocessing

The government corpus has been cleaned and verified manually with the help of professional translators [17]. Sentences containing only dates, special characters and numbers have been removed. Cleaning scripts of the Moses<sup>11</sup> [24] tool were used to remove misaligned sentences. English sentences were tokenized using Moses toolkit<sup>12</sup> [24], while a tokenizer developed by [15] was used for Sinhala and Tamil. We use the SentencePiece<sup>13</sup> model learned over monolingual Common Crawl (CC) data in the mBART [41] model, containing 250,000 sub-word tokens.

## 3.7 Addressing the Zero Width Joiner (ZWJ) issue

As mentioned earlier in Section 3.6, we use the SentencePiece<sup>14</sup> tokenizer. However, empty tokens where Unicode appears as space in output vocabulary have been removed while training the SentencePiece model. Due to that, the Zero Width Joiner character (U+200D) has been replaced by whitespace. Hence language scripts that require Zero Width Joiner get altered and result in a wrong output. This issue is there for languages like Sinhala, Kannada and Malayalam<sup>15</sup>. Ideally, we should not replace the Zero Width Joiner (U+200D) with whitespace since it indicates to join two chars without zero width (no whitespace). Also, the zero-width joiner must be present to decode the segmentation of decoder outputs to raw text successfully. These special characters should be kept as they are while learning the SentencePiece model. We resolved this issue and the fix was successfully merged with SentencePiece<sup>16</sup> to eliminate this ZWJ issue.

However, we cannot eliminate this issue when using the already trained (pre-trained) models as the models were trained without ZWJ over the large-scale monolingual data. Hence we propose a post-processing solution to add the ZWJ character in the possible occurring places for Sinhala languages<sup>17</sup>. Here we mainly cover frequently occurring "yansaya" and "rakāransaya" of the Sinhala script. Further details of the post-processing logic are provided in A.

#### 3.8 Evaluation setup

The final model is selected based on the validation likelihood. We use beam size 5 for decoding as used by [28]. The final results are calculated against the true-target tokenized data and reported in BLEU [33]<sup>18</sup>

<sup>11</sup> http://www.statmt.org/moses/

<sup>12</sup> http://www.statmt.org/moses/

<sup>13</sup> https://github.com/google/sentencepiece

<sup>14</sup> https://github.com/google/sentencepiece

<sup>&</sup>lt;sup>15</sup>https://en.wikipedia.org/wiki/Zero-width\_joiner

<sup>16</sup> https://github.com/google/sentencepiece/pull/630

<sup>17</sup> https://en.wikipedia.org/wiki/Sinhala\_script#Consonant\_conjuncts

<sup>&</sup>lt;sup>18</sup> script from [24]

For ensembling cases described in Section 2.4, we combine a maximum of 3 models and evaluate under the same decoding and BLEU scores calculation as mentioned above.

#### 4 RESULTS

# 4.1 Comparison Between Full Precision Fine-Tuning and Mixed-Precision Fine-Tuning

When training a neural network, usually we use Full Precision, a 32-bit floating-point (FP32) arithmetic calculation by default. Alternatively, we can use Mixed-Precision training, which combines single-precision (FP32) with half-precision (FP16) format. Mixed-Precision training has further performance benefits on NVIDIA GPUs. It requires shorter training time and lower memory requirements, enabling larger batch sizes, models, and inputs<sup>19</sup>.

Hence, for bilingual fine-tuning (baseline), we fine-tuned each pair of directions two times under the same hyper-parameters, one with the default setting Full Precision and the second one with Mixed-precision training. As shown in Table 3, there is not much BLEU score difference between Full Precision and Mixed-precision. However, we noticed that Mixed-precision training reduces the training time. Considering our resource limitation for training, we choose Mixed-precision training going forward.

Models	$Si \rightarrow Ta$	$Ta \rightarrow Si$	$Si \rightarrow En$	$En \rightarrow Si$	$Ta \rightarrow En$	$En \rightarrow Ta$
Bilingual full precision FT	29.82	37.87	37.84	35.89	33.63	25.96
Bilingual mixed precision FT	29.75	37.57	37.72	36.11	33.36	25.8

Table 3. Comparison between full precision training and mixed precision Fine-Tuning.

## 4.2 Continual Pre-training (CPT) for Domain Adaptation

Extensive evaluation results of the CPT specifically focusing on the  $Si \leftrightarrow En$  language pair are detailed in Table 4. Overall, CPT improves the results compared to both baselines. This clearly illustrates that under-represented languages require additional pre-training to learn a good representation in the msLLM. Among the several CPT techniques we proposed, we observe the highest improvement on Case A(ii), where we get +0.79 and +0.53 on  $Si \to En$  and  $En \to Si$ , respectively. Our experiments show that in-domain pre-training (Case A) plays a pivotal role compared to other pre-training strategies that utilize out-domain data. Even though we do not have adequate in-domain data, our experiments validate that having little in-domain data would have more impact than out-domain or mixed-domain data. Hence, we pick CPT on in-domain monolingual data in going forward.

When we look at biCPT+B-FT (raw 15) in Table 5, we can evidence that CPT helps the NMT model to perform better on extremely LRL pairs for all six directions, where we obtained a maximum of +0.97 BLEU improvement on  $Si \rightarrow Ta$ .

## 4.3 Multilingual Fine-tuning using the msLLM

The performance results of the M-FT baseline in comparison to the B-FT baseline are presented in the first ten rows of Table 5. Predominantly, the results indicate a marginal decrement in the performance of the M2M-FT relative to the B-FT baseline. This trend is postulated to stem from the insufficiency of parallel data required for training robust Multilingual NMT (MNMT) models.

Despite the general decline, overall an incremental enhancement in performance is observed for O2M-FT and M2O-FT configurations compared to M2M-FT (expect 3 scenarios). This increment aligns with previous English-centric MNMT

<sup>19</sup> https://pytorch.org/blog/accelerating-training-on-nvidia-gpus-with-pytorch-automatic-mixed-precision/

	Models	$Si \rightarrow En$	$En \rightarrow Si$
	baseline (B-FT)	37.72	36.11
Case A	(i) biCPT in-domain mono data then B-FT	38.21 (+0.49)	
Case A	(ii) biCPT combined in-domain parallel and mono data then B-FT	<b>38.51</b> (+0.79)	<b>36.64</b> (+0.53)
Case B	biCPT out-domain mono data then B-FT	38.09 (+0.37)	36.21 (+0.1)
	(i) mixed in-domain and out-domain mono biCPT then B-FT	38.3 (+0.58)	36.33(+0.22)
Case C	(ii) multistage biCPT then B-FT	38.38 (+0.66)	36 (-0.11)

Table 4. Bilingual fine-tuning on top of CPT models against bilingual fine-tuning on original msLLM model (mBART) for  $Si \leftrightarrow En$  pairs

results reported by others [2, 4, 23, 41]. In our experiments, Si - Ta acts as an LRL pair compared to the Si - En pair, as we have around 66k and 74k sentences, respectively. When we up-sample Si - Ta to match the size of Si - En, we were able to maximize the performance of O2M-FT and M2O-FT and beat the B-FT baseline of  $Si \rightarrow Ta$ ,  $Ta \rightarrow Si$  by +0.55 and +0.65, respectively. Similarly, we observe M2M-FT cases overall perform lower than [triCPT,O2M-FT], [triCPT,M2O-FT] cases. We obtained a maximum of +1.22 BLEU score improvement on triCPT M-FT models for  $Ta \rightarrow Si$ .

#### 4.4 ITTL Fine-tuning

Results of our two ITTL strategies [biCPT,3-B-FT] and [triCPT, M-FT,B-FT]  $^{20}$  are given in the last four raws of Table 5. Three-stage fine-tuning ITTL [biCPT,3-B-FT] significantly improves BLEU score compared to baseline B-FT for all the directions where we obtained a maximum of +1.81 score improvement for  $Si \rightarrow Ta$  and a minimum of +0.65 improvements on  $En \rightarrow Ta$ .

Our biCPT 3-stage fine-tuning results prove that utilizing the available in-domain monolingual and out-domain parallel data enhances fine-tuning performance in extremely low-resource domain-specific settings rather than only utilizing out-domain parallel data.

Our second ITTL approach, Fine-tuning from Multilingual Fine-tuned MNMT models [triCPT,M-FT,B-FT] shows quantifiable improvements except on  $Si \rightarrow En$ . For  $Si \rightarrow Ta$  and  $Ta \rightarrow Si$ , we obtain +1.41 and +1.83 BLEU score gains respectively, and also notably +1.03 on  $En \rightarrow Ta$  direction.

Overall, we prove that utilizing additional in-domain monolingual data for pre-training and in- and out-domain data fine-tuning improves the results on under-represented languages in the msLLM. As observed in our initial B-FT experiments (baseline) results, here also we can witness that, even though we fine-tune with large out-domain parallel data, we still require a reasonable amount of in-domain parallel corpus to learn a good language representation on a particular domain for morphologically rich languages.

#### 4.5 Ensemble

We select the top 3 well-performed models for each direction and rank them accordingly as shown in Table 6. Then we empirically evaluate all the possible combinations of multi-model ensembling and report the results as shown in Table 7. The checkpoints ensemble slightly outperforms the baseline single model in all directions. As we can see, the multi-model ensemble overall improves the performance in all the translation tasks except on  $Ta \to En$ , where a single model stands out. We observe that baseline B-FT also stands as a strong model on multi-model ensemble along with our best-ranked proposed models for the rest of the 5 directions. Finally, we obtain a maximum +2.21 BLEU point

<sup>&</sup>lt;sup>20</sup>biCPT where we pre-train using monolingual data of two languages and whereas triCPT with three languages. biCPT will follow with B-FT and triCPT follow with M-FT as we try to fine-tune using three languages

Models	Pivot Lang	$Si \rightarrow Ta$	$Ta \rightarrow Si$	$Si \rightarrow En$	$En \rightarrow Si$	$Ta \rightarrow En$	$En \rightarrow Ta$
	in M-FT						
B-FT (Baseline)	-	29.75	37.57	37.72	36.11	33.36	25.8
O2M-FT (Baseline)		30.3 (+0.55)	N/A	37.59 (-0.13)	N/A	N/A	N/A
M2O-FT (Baseline)	Si	N/A	37.62 (+0.05)	N/A	36.03 (-0.08)	N/A	N/A
M2M-FT (Baseline)		28.44 (-1.31)	36.38 (-1.19)	36.05 (-1.67)	34.93 (-1.18)	N/A	N/A
O2M-FT (Baseline)		N/A	38.22 (+0.65)	N/A	N/A	33.45 (+0.09)	N/A
M2O-FT (Baseline)	Ta	29.78 (+0.03)	N/A	N/A	N/A	N/A	26.27 (+0.47)
M2M-FT (Baseline)		28.85 (-0.9)	36.35 (-1.22)	N/A	N/A	32.64 (-0.72)	24.65 (-1.15)
O2M-FT (Baseline)		N/A	N/A	N/A	35.58 (-0.53)	N/A	26.24 (+0.44)
M2O-FT (Baseline)	En	N/A	N/A	37.75 (+0.03)	N/A	34.85 (+1.49)	N/A
M2M-FT (Baseline)		N/A	N/A	36.61 (-1.11)	34.48 (-1.63)	33.99 (0.63)	24.68 (-1.12)
M-FT(best),B-FT	Si	30.97 (+1.22)	37.54 (-0.03)	36.1 (-1.62)	35.88 (-0.23)	N/A	N/A
M-FT(best),B-FT	Ta	29.85 (+0.1)	39.17 (+1.6)	N/A	N/A	35.96 (+2.6)	26.22 (+0.42)
M-FT(best),B-FT	En	N/A	N/A	36.87 (-0.85)	36.42 (+0.31)	33.94 (+0.58)	26.12 (+0.32)
3-B-FT	-	30.44 (+0.69)	36.97 (-0.6)	38.27 (+0.55)	35.68 (-0.43)	33.54 (+0.18)	25.83 (+0.03)
biCPT,B-FT	-	30.72 (+0.97)	38.08 (+0.51)	38.51 (+0.79)	36.64 (+0.53)	34.29 (+0.93)	26.3 (+0.5)
triCPT,O2M-FT		30.5 (+0.75)	N/A	37.93 (+0.21)	N/A	N/A	N/A
triCPT,M2O-FT	Si	N/A	38.04 (+0.47)	N/A	35.87 (-0.24)	N/A	N/A
triCPT,M2M-FT		29.64 (-0.11)	37.15 (-0.42)	36.98 (-0.74)	34.51 (-1.6)	N/A	N/A
triCPT,O2M-FT		N/A	38.79 (+1.22)	N/A	N/A	33.95 (+0.59)	N/A
triCPT,M2O-FT	Ta	29.31 (-0.44)	N/A	N/A	N/A	N/A	26.43 (+0.63)
triCPT,M2M-FT		28.87 (-0.88)	36.46 (-1.11)	N/A	N/A	32.61 (-0.75)	24.25 (-1.55)
triCPT,O2M-FT		N/A	N/A	N/A	35.17 (-0.94)	N/A	26.63 (+0.83)
triCPT,M2O-FT	En	N/A	N/A	37.98 (+0.26)	N/A	35.23 (+1.87)	N/A
triCPT,M2M-FT		N/A	N/A	36.85 (-0.87)	35.74 (-0.37)	33.1 (-0.26)	25.37 (-0.43)
triCPT,M-FT(best),B-FT	Si	31.16 (+1.41)	38.9 (+1.33)	36.53 (-1.19)	36.17 (+0.06)	N/A	N/A
triCPT,M-FT(best),B-FT	Ta	30.19 (+0.44)	39.4 (+1.83)	N/A	N/A	35.15 (+1.79)	26.83 (+1.03)
triCPT,M-FT(best),B-FT	En	N/A	N/A	37.2 (-0.52)	36.68 (+0.57)	34.52 (+1.16)	26.75 (+0.95)
biCPT,3-B-FT	-	30.87 (+1.12)	39.38 (+1.81)	38.89 (+1.17)	36.91 (+0.8)	34.82 (+1.46)	26.45 (+0.65)

Table 5. CPT and ITTL against our strong single-stage fine-tuning baseline (B-FT, M-FT) - In M-FT, not applicable cases denotes with N/A. For example Si centric O2M-FT includes  $Si \rightarrow Ta$  and  $Si \rightarrow En$  directions and rest of the cases indicated with N/A.

Rank	$Si \rightarrow Ta$	$Ta \rightarrow Si$	$Si \rightarrow En$	$En \rightarrow Si$	$Ta \rightarrow En$	$En \rightarrow Ta$
1	triCPT+M-FT+B-FT (Si)	triCPT+M-FT+B-FT (Ta)	biCPT+3-B-FT	biCPT+3-B-FT	mbart50+M-FT+B-FT (Ta)	triCPT+M-FT+B-FT (Ta)
2	mbart50 + M-FT + B-FT (Si)	biCPT+3-B-FT	biCPT+B-FT	triCPT+M-FT+B-FT (En)	triCPT+M2O-M-FT (En)	triCPT+M-FT+B-FT (En)
3	biCPT+3-B-FT	mbart50+M-FT+B-FT (Ta)	mBART50+3-B-FT	biCPT+B-FT	triCPT+M-FT+B-FT (Ta)	triCPT+O2M-M-FT (En)

Table 6. Top 3 improved models from baseline B-FT

improvement on Si-En and a minimum of +1.78 BLEU score improvement on En-Si directions relative to our strong baseline single-stage fine-tuning (B-FT).

# 5 DISCUSSION

We explore the effectiveness of mixed precision training, the application of CPT and ITTL techniques, and the role of inand out-domain data in enhancing translation quality across multiple language pairs. The insights derived from these evaluations offer valuable directions for optimizing translation models, particularly in resource-constrained settings. Findings are: (1). Mixed precision training, while delivering similar BLEU scores as full precision training, significantly Manuscript submitted to ACM

Ensemble	Models	$Si \rightarrow Ta$	$Ta \rightarrow Si$	$Si \rightarrow En$	$En \rightarrow Si$	$Ta \rightarrow En$	$En \rightarrow Ta$
No	B-FT (Baseline)	29.75	37.57	37.72	36.11	33.36	25.80
Ensemble	Rank-1	31.16 (+1.41)	39.40 (+1.83)	38.89 (+1.17)	36.91 (+0.8)	35.96 (+2.6)	26.83 (+1.03)
Checkpoint	B-FT (Baseline)	30.07	37.86	37.77	36.21	33.40	25.82
Ensemble	Rank-1	31.26	39.30	39.58	36.85	34.54	26.76
	B-FT & Rank-1	31.41	39.48	38.97	37.42	35.19	26.77
	B-FT & Rank-2	30.93	38.75	38.68	37.17	34.48	27.12
	B-FT & Rank-3	31.16	38.80	38.26	37.54	34.65	26.57
	B-FT & Rank-1,2	31.89 (+2.14)	39.31	39.93 (+2.21)	37.69	35.55	27.82 (+2.02)
Multi-	B-FT & Rank-1,3	31.61	39.68 (+2.11)	39.70	37.89 (+1.78)	35.47	27.23
Model	B-FT & Rank-2,3	31.57	39.11	39.13	37.87	35.04	27.09
Ensemble	Rank-1,2	31.90	39.52	38.81	36.97	35.88 (+2.52)	27.73
	Rank-1,3	31.46	38.60	39.83	37.29	25.95	26.61
	Rank-1,2,3	31.77	39.24	39.68	37.57	35.75	27.52

Table 7. Ensemble results for all the six directions.

enhances training efficiency by reducing training time and memory requirements. This allows for larger batch sizes and increased model capacity without compromising translation accuracy. (2). CPT approach, especially for the Si - En language pair, shows notable improvements in translation quality over baseline models. The biCPT approach, utilizing in-domain monolingual data (Case A (ii)) followed by bilingual fine-tuning (B-FT), achieved the highest translation quality. In-domain pre-training was found to be more effective (even with limited in-domain monolingual data availability) than strategies relying on out-of-domain or mixed-domain monolingual data. (3). The M2M-FT setup typically underperformed compared to the B-FT baseline due to insufficient parallel data, essential for robust msLLMs. However, the O2M-FT and M2O-FT configurations generally surpassed M2M-FT in performance (except in three scenarios), aligning with previous results in English-centric MNMT[41]. Specifically, equalizing data volumes between Si - Ta and Si - En significantly improved results for Si - Ta and Ta - Si in the O2M-FT and M2O-FT setups. (4). The inclusion of in-domain monolingual data in pre-training, combined with the use of both in-domain and out-of-domain data in fine-tuning, is crucial for improving translation performance in underrepresented languages. Although large volumes of out-domain parallel data are beneficial, a substantial volume of in-domain parallel data is essential for capturing effective language representations, particularly in morphologically rich languages. (5). The baseline B-FT model remains robust, effectively enhancing the performance of multi-model ensembles in five out of six translation directions. Significant improvements were observed in translations between Si and En compared to the baseline model. The findings highlight the general superiority of multi-model ensembles in boosting translation performance across various tasks.

# 5.1 Manual Analysis of the Translated Output

An in-depth manual inspection of the translation outputs for each language direction reveals areas of model limitations. Examples of such translated sentences are included in Appendix B. A recurrent issue is the models' inability to accurately translate place names and roads, highlighting the challenge of domain-specific named entities with limited data coverage. Future enhancements may incorporate data augmentation strategies to integrate these entities.

A salient observation from our analysis is the diminished performance of our models when translating into Tamil. This reduced efficacy can be attributed to the intrinsic linguistic features of Tamil, such as its flexible word order [18] Manuscript submitted to ACM

and more complex inflectional system, compared to Sinhala. The free word-ordering nature of Tamil presents a unique challenge for translation models. Moreover, as indicated by [3], Tamil's extensive inflection further exacerbates these challenges. Our scrutiny of the Tamil translations revealed instances of free word order, compound words, and syntactic structures that closely mirror those of the references.

Additionally, the translated outputs often employed words that were synonymous with or similar to those in the reference sentences. This phenomenon underlines a fundamental principle in computational linguistics: as the linguistic complexity of a target language increases, so does the requisite volume of training data needed for a model to capture and reproduce the language's specific features accurately. Consequently, our findings suggest that even when two languages are provided with comparable amounts of monolingual data for pre-training, the complexity of the target language is a determining factor in the success of the fine-tuning process.

#### 6 RELATED WORK

First, we discuss msLLMs and their utility for NMT. Then we analyze CPT and ITTL with msLLMs for NMT.

# 6.1 Multilingual Sequence-to-sequence Large Language Models (msLLMs)

The BART [27] model is the first complete self-supervised sequence-to-sequence (encoder-decoder) model trained on a large scale (English) monolingual corpus. BART is implemented as a standard Transformer architecture [43] with a bidirectional auto-encoder and a left-to-right autoregressive decoder. Later BART was extended to incorporate more than one language, resulting in mBART25. mBART25 [28] is the first self-supervised multilingual sequence-to-sequence Large Language Model (msLLM) trained on a large-scale unlabeled monolingual corpus of 25 different languages. [41] extended this model upto 50 languages, which resulted in the mBART50 model. Another well-known sequence-to-sequence model is T5 [34], a Text-to-Text Transfer Transformer trained over a large English monolingual corpus similar to BART [27]. mT5 [45] is the multilingual version of T5, and includes 101 languages [45]. Subsequently introduced mT6 [9] has been additionally trained with parallel data on mT5. However, mBART is memory efficient and has shown relatively better results than mT5 for translation [25, 32].

# 6.2 Fine-tuning msLLMs for NMT Tasks

Since msLLMs are trained using a denoising objective, they cannot be directly used as an NMT model. However, when parallel data between a source and target language is available, an msLLM can be used to fully initialize the NMT (encoder-decoder) training, thus avoiding the need to train the NMT model from scratch. This process is referred to as *fine-tuning*. In other words, the parameters of an msLLM can be fine-tuned to incorporate the model in an NMT task in the context of a language pair. Studies show this leads to significant performance gains over NMT models based on the simple Transformer architecture due to the knowledge transfer from msLLMs [28, 41, 42].

In particular, the mBART model has shown promising results on supervised and unsupervised machine translation at both the sentence and document levels [28, 41]. Significant gains have been observed in LRL pairs such as  $English \leftrightarrow Vietnamese/Turkish$  as well [28, 41]. However, all the experiments conducted by the above research have been English-centric, i.e. English is either on the source or target side of the given language pair. [41] experimented with fine-tuning multiple language pairs simultaneously, called Multilingual Fine-Tuning (M-FT). In their experiments also, the tested cases are only English-centric. They experimented with M2O ( $Many \rightarrow English$ ), O2M ( $English \rightarrow Many$ ), and M2M ( $Many \leftrightarrow Many$ ) with English as a pivot language. Some studies considered non-English centric fine-tuning where only one language from the given pair is included in the mBART model [6, 31]. [42] demonstrated that fine-tuning

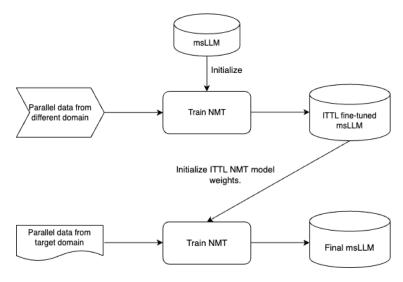


Fig. 2. Overview of ITTL

msLLMs for LRLs NMT, particularly in non-English pairs like Tamil to Sinhala, significantly outperforms standard Transformer-based models. Their study also highlighted the importance of the volume of monolingual data in training msLLMs for LRLs, suggesting a need for enhanced pre-training and fine-tuning approaches in LRL-NMT.

[1, 32] show that ITTL of msLLM models is extremely beneficial for domain-specific NMT, especially when target domain data is limited/unavailable in msLLMs. They analysed the possibility of using auxiliary parallel data to fine-tune the msLLMs before fine-tuning with the limited target domain data (as showed in Figure 2). Their observation improved the msLLM performance for domain-specific NMT of LRLs and strongly holds for languages missing or under-represented in the msLLM.

Previous studies mainly explored two variants: (1) fine-tune with data from an auxiliary domain (also referred to as intermediate task) and further fine-tune with the target domain [1, 32] and (2) fine-tune by combining data from different domains and further fine-tune with the target domain [1]. However, they did not further look into consecutive three-stage fine-tuning strategies and fine-tuning from M-FT models. Building upon the findings of [1, 32], which underscore the significance of ITTL in enhancing the performance of msLLMs, this study delves deeper into the exploration of various strategies for implementing ITTL.

# 6.3 Continual Pre-training of msLLMs (Extending the msLLMs)

CPT—which entails further pre-training of an LLM with monolingual data—has been found to improve the performance of encoder-based models such as BERT [12] and RoBERTa [29] in domain-specific text classification tasks [5, 19, 26, 46]. Further studies conducted CPT on parallel corpora [9, 37], demonstrate that integrating parallel data into msLLMs significantly enhances model performance across various language tasks, outperforming traditional methods that rely mainly on monolingual corpora. Although this continual pre-training strategy has the potential for domain adaptation of msLLMs in NMT, existing research has not adapted it to a particular domain.

CPT has been commonly employed to extend an msLLM to new languages [8, 30, 40, 41]. PARADISE [37] extended the conventional pre-training using denoising objective by (i) replacing words in the noised sequence according to a Manuscript submitted to ACM

multilingual dictionary, and (ii) predicting the reference translation according to a parallel corpus instead of recovering the original sequence. However, they have not focused on adapting to a particular domain via pre-training on msLLMs.

#### 7 CONCLUSION

This study highlights the effectiveness of refining CPT and ITTL on msLLMs with additional monolingual and parallel data, presenting it as a promising method for LRL-NMT. Previous research primarily English-focused settings, our work broadens these techniques to incorporate non-English-centric translation as well. We developed and assessed CPT and ITTL with various approaches using both in-domain and out-domain parallel data.

Our experimental results indicate that translation accuracy is significantly influenced by both the quantity of monolingual data and its domain relevance to the pre-training phase. Furthermore, we have determined that incorporating even a modest amount of in-domain monolingual data can yield substantial improvements in msLLM performance. Notably, the use of out-domain data in the ITTL fine-tuning process appears to be more beneficial for translations from Sinhala and Tamil into English than the reverse.

However, our proposed approaches are not specific to one particular msLLMs, they can be applied to generative LLMs as well, leaving that for feature work. Besides, the adaptation of M-FT models has not consistently surpassed the performance of B-FT models. Recognizing this, our future work will aim to refine the M-FT processes. We anticipate that such enhancements will facilitate the development of a singular, comprehensive MNMT model capable of managing all translation directions effectively.

## **ACKNOWLEDGMENTS**

This research was supported by the Accelerating Higher Education Expansion and Development (AHEAD) Operation of the Ministry of Education, Sri Lanka, funded by the World Bank.

# **REFERENCES**

- [1] David Adelani, Jesujoba Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Emezue, Colin Leong, Michael Beukman, Shamsuddeen Muhammad, Guyo Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ajibade, Tunde Ajayi, Yvonne Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaaba, Blessing Sibanda, Andiswa Bukula, and Sam Manthalu. 2022. A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Seattle, United States, 3053–3070. https://doi.org/10.18653/v1/2022.naacl-main.223
- [2] Roee Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively Multilingual Neural Machine Translation. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minnesota, 3874–3884. https://doi.org/10.18653/v1/N19-1388
- [3] M Anand Kumar, V Dhanalakshmi, KP Soman, and S Rajendran. 2010. A sequence labeling approach to morphological analyzer for tamil language. IJCSE) International Journal on Computer Science and Engineering 2, 06 (2010), 1944–195.
- [4] N. Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, Melvin Johnson, Maxim Krikun, Mia Xu Chen, Yuan Cao, George F. Foster, Colin Cherry, Wolfgang Macherey, Z. Chen, and Yonghui Wu. 2019. Massively Multilingual Neural Machine Translation in the Wild: Findings and Challenges. ArXiv abs/1907.05019 (2019).
- [5] Iz Beltagy, Kyle Lo, and Arman Cohan. 2019. SciBERT: A Pretrained Language Model for Scientific Text. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, 3615–3620. https://doi.org/10.18653/v1/D19-1371
- [6] Samuel Cahyawijaya, Genta Indra Winata, Bryan Wilie, Karissa Vincentio, Xiaohong Li, Adhiguna Kuncoro, Sebastian Ruder, Zhi Yuan Lim, Syafri Bahar, Masayu Khodra, Ayu Purwarianti, and Pascale Fung. 2021. IndoNLG: Benchmark and Resources for Evaluating Indonesian Natural Language

Generation. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 8875–8898. https://doi.org/10.18653/v1/2021.emnlp-main.699

- [7] Hugh Chen, Scott M. Lundberg, and Su-In Lee. 2017. Checkpoint Ensembles: Ensemble Methods from a Single Training Process. ArXiv abs/1710.03282
- [8] Peng-Jen Chen, Ann Lee, Changhan Wang, Naman Goyal, Angela Fan, Mary Williamson, and Jiatao Gu. 2020. Facebook AI's WMT20 News Translation Task Submission. In Proceedings of the Fifth Conference on Machine Translation. Association for Computational Linguistics, Online, 113–125. https://aclanthology.org/2020.wmt-1.8
- [9] Zewen Chi, Li Dong, Shuming Ma, Shaohan Huang, Saksham Singhal, Xian-Ling Mao, Heyan Huang, Xia Song, and Furu Wei. 2021. mT6: Multilingual Pretrained Text-to-Text Transformer with Translation Pairs. In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 1671–1683. https://doi.org/10.18653/v1/2021.emnlp-main.125
- [10] Chenhui Chu, Raj Dabre, and Sadao Kurohashi. 2017. An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Association for Computational Linguistics, Vancouver, Canada, 385–391. https://doi.org/10.18653/v1/P17-2061
- [11] Marta R Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, et al. 2022. No language left behind: Scaling human-centered machine translation. arXiv preprint arXiv:2207.04672 (2022).
- [12] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). Association for Computational Linguistics, Minneapolis, Minnesota, 4171–4186. https://doi.org/10.18653/v1/N19-1423
- [13] Koshiya Epaliyana, Surangika Ranathunga, and Sanath Jayasena. 2021. Improving Back-Translation with Iterative Filtering and Data Selection for Sinhala-English NMT. In 2021 Moratuwa Engineering Research Conference (MERCon). IEEE, 438–443.
- [14] Angela Fan, Shruti Bhosale, Holger Schwenk, Zhiyi Ma, Ahmed El-Kishky, Siddharth Goyal, Mandeep Baines, Onur Celebi, Guillaume Wenzek, Vishrav Chaudhary, Naman Goyal, Tom Birch, Vitaliy Liptchinsky, Sergey Edunov, Edouard Grave, Michael Auli, and Armand Joulin. 2022. Beyond English-Centric Multilingual Machine Translation. J. Mach. Learn. Res. 22, 1, Article 107 (jul 2022), 48 pages.
- [15] Fathima Farhath, Surangika Ranathunga, Sanath Jayasena, and Gihan Dias. 2018. Integration of bilingual lists for domain-specific statistical machine translation for sinhala-tamil. In 2018 Moratuwa Engineering Research Conference (MERCon). IEEE, 538–543.
- [16] Aloka Fernando, Surangika Ranathunga, and G. Dias. 2020. Data Augmentation and Terminology Integration for Domain-Specific Sinhala-English-Tamil Statistical Machine Translation. ArXiv abs/2011.02821 (2020).
- [17] Thilakshi Fonseka, Rashmini Naranpanawa, Ravinga Perera, and Uthayasanker Thayasivam. 2020. English to Sinhala Neural Machine Translation. In 2020 International Conference on Asian Language Processing (IALP). IEEE, 305–309.
- [18] Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Quantifying word order freedom in dependency corpora. In *Proceedings of the third international conference on dependency linguistics (Depling 2015)*. 91–100.
- [19] Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, 8342–8360. https://doi.org/10.18653/v1/2020.acl-main.740
- [20] Francisco Guzmán, Peng-Jen Chen, Myle Ott, Juan Pino, Guillaume Lample, Philipp Koehn, Vishrav Chaudhary, and Marc'Aurelio Ranzato. 2019. The FLORES Evaluation Datasets for Low-Resource Machine Translation: Nepali-English and Sinhala-English. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, 6098-6111. https://doi.org/10.18653/v1/D19-1632
- [21] Lars Kai Hansen and Peter Salamon. 1990. Neural network ensembles. IEEE transactions on pattern analysis and machine intelligence 12, 10 (1990), 993–1001.
- [22] Kenji Imamura and Eiichiro Sumita. 2017. Ensemble and Reranking: Using Multiple Models in the NICT-2 Neural Machine Translation System at WAT2017. In Proceedings of the 4th Workshop on Asian Translation (WAT2017). Asian Federation of Natural Language Processing, Taipei, Taiwan, 127–134. https://aclanthology.org/W17-5711
- [23] Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean. 2017. Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. Transactions of the Association for Computational Linguistics 5 (2017), 339–351. https://doi.org/10.1162/tacl\_a\_00065
- [24] Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondřej Bojar, Alexandra Constantin, and Evan Herbst. 2007. Moses: Open Source Toolkit for Statistical Machine Translation. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions. Association for Computational Linguistics, Prague, Czech Republic, 177–180. https://aclanthology.org/P07-2045
- [25] En-Shiun Lee, Sarubi Thillainathan, Shravan Nayak, Surangika Ranathunga, David Adelani, Ruisi Su, and Arya McCarthy. 2022. Pre-Trained Multilingual Sequence-to-Sequence Models: A Hope for Low-Resource Language Translation?. In Findings of the Association for Computational Linguistics: ACL 2022. Association for Computational Linguistics, Dublin, Ireland, 58–67. https://doi.org/10.18653/v1/2022.findings-acl.6

- [26] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 36, 4 (09 2019), 1234–1240. https://doi.org/10.1093/bioinformatics/btz682 arXiv:https://academic.oup.com/bioinformatics/article-pdf/36/4/1234/32527770/btz682.pdf
- [27] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, 7871–7880. https://doi.org/10.18653/v1/2020.acl-main.703
- [28] Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual Denoising Pre-training for Neural Machine Translation. Transactions of the Association for Computational Linguistics 8 (2020), 726-742. https://doi.org/10.1162/tacl\_a\_00343
- [29] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv abs/1907.11692 (2019).
- [30] Zihan Liu, Genta Indra Winata, and Pascale Fung. 2021. Continual Mixed-Language Pre-Training for Extremely Low-Resource Neural Machine Translation. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, 2706–2718. https://doi.org/10.18653/v1/2021.findings-acl.239
- [31] Lovish Madaan, Soumya Sharma, and Parag Singla. 2020. Transfer Learning for Related Languages: Submissions to the WMT20 Similar Language Translation Task. In Proceedings of the Fifth Conference on Machine Translation. Association for Computational Linguistics, Online, 402–408. https://aclanthology.org/2020.wmt-1.46
- [32] Shravan Nayak, Surangika Ranathunga, Sarubi Thillainathan, Rikki Hung, Anthony Rinaldi, Yining Wang, Jonah Mackey, Andrew Ho, and En-Shiun Annie Lee. 2023. Leveraging Auxiliary Domain Parallel Data in Intermediate Task Fine-tuning for Low-resource Translation. arXiv:2306.01382 [cs.CL]
- [33] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, 311–318. https://doi.org/10.3115/1073083.1073135
- [34] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Journal of Machine Learning Research 21, 140 (2020), 1–67. http://jmlr.org/papers/v21/20-074.html
- [35] MD Rajitha, LL Piyarathna, MMD Nayanajith, and S Surangika. 2020. Sinhala and english document alignment using statistical machine translation. In 2020 20th International Conference on Advances in ICT for Emerging Regions (ICTer). IEEE, 29–34.
- [36] Surangika Ranathunga, En-Shiun Annie Lee, Marjana Prifti Skenduli, Ravi Shekhar, Mehreen Alam, and Rishemjit Kaur. 2022. Neural Machine Translation for Low-Resource Languages: A Survey. ACM Comput. Surv. (nov 2022). https://doi.org/10.1145/3567592 Just Accepted.
- [37] Machel Reid and Mikel Artetxe. 2022. PARADISE: Exploiting Parallel Data for Multilingual Sequence-to-Sequence Pretraining. In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Seattle, United States, 800–810. https://doi.org/10.18653/v1/2022.naacl-main.58
- [38] Rico Sennrich, Alexandra Birch, Anna Currey, Ulrich Germann, Barry Haddow, Kenneth Heafield, Antonio Valerio Miceli Barone, and Philip Williams. 2017. The University of Edinburgh's Neural MT Systems for WMT17. In Proceedings of the Second Conference on Machine Translation. Association for Computational Linguistics, Copenhagen, Denmark, 389–399. https://doi.org/10.18653/v1/W17-4739
- [39] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Edinburgh neural machine translation systems for WMT 16. arXiv preprint arXiv:1606.02891
- [40] Raymond Hendy Susanto, Dongzhe Wang, Sunil Yadav, Mausam Jain, and Ohnmar Htun. 2021. Rakuten's Participation in WAT 2021: Examining the Effectiveness of Pre-trained Models for Multilingual and Multimodal Machine Translation. In Proceedings of the 8th Workshop on Asian Translation (WAT2021). Association for Computational Linguistics, Online, 96–105. https://doi.org/10.18653/v1/2021.wat-1.9
- [41] Yuqing Tang, Chau Tran, Xian Li, Peng-Jen Chen, Naman Goyal, Vishrav Chaudhary, Jiatao Gu, and Angela Fan. 2021. Multilingual Translation from Denoising Pre-Training. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. Association for Computational Linguistics, Online, 3450–3466. https://doi.org/10.18653/v1/2021.findings-acl.304
- [42] Sarubi Thillainathan, Surangika Ranathunga, and Sanath Jayasena. 2021. Fine-Tuning Self-Supervised Multilingual Sequence-To-Sequence Models for Extremely Low-Resource NMT. In 2021 Moratuwa Engineering Research Conference (MERCon). IEEE, 432–437.
- [43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (Long Beach, California, USA) (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 6000–6010.
- [44] Haoran Xu, Young Jin Kim, Amr Sharaf, and Hany Hassan Awadalla. 2024. A Paradigm Shift in Machine Translation: Boosting Translation Performance of Large Language Models. arXiv:2309.11674 [cs.CL] https://arxiv.org/abs/2309.11674
- [45] Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer. In Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Association for Computational Linguistics, Online, 483–498. https://doi.org/10.18653/v1/2021.naacl-main.41

[46] Rong Zhang, Revanth Gangi Reddy, Md Arafat Sultan, Vittorio Castelli, Anthony Ferritto, Radu Florian, Efsun Sarioglu Kayi, Salim Roukos, Avi Sil, and Todd Ward. 2020. Multi-Stage Pre-training for Low-Resource Domain Adaptation. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Online, 5461–5468. https://doi.org/10.18653/v1/2020.emnlp-main.440

# A ADDRESSING THE ZERO WIDTH JOINER (ZWJ) ISSUE

ZWJ post-processing fix is mainly required for  $En \to Si$  and  $Ta \to Si$  directions. i.e. when Sinhala is on the target side. We added the following two lines of code

```
sentence = sentence.replace("\u0DCA\u0dbb", "\u0DCA\u200D\u0dbb")
sentence = sentence.replace("\u0DCA\u0dba", "\u0DCA\u200D\u0dba")
just before the return statement of the following method in Fairseq:
    def post_process(sentence: str, symbol: str):
```

#### **B OUTPUT TRANSLATIONS**

Figure 3 highlights some output translations of the best-fine-tuned model against reference texts for all six translation directions.

	Sentence 1	Sentence 2	Sentence 3
Input (Si)	විමර්ශනයට ලක් කරන ලද ආගතන පම්බන්ධයෙන් පහත සඳහන් නිරීස්ෂණයන් පිදු කර ඇත (the following general observations are made in relation to the institutions examined .)	ආපියාලු පංචර්ධන බැංකු ආධාර යටතේ මාර්ත ජාල වැඩිදියුණු කිරීමේ වනාපෘතිය / ADB (Asian Development Bank Funded Road Network Improvement Project / Asian Development Bank)	ආයපනිස පහ පුනිපත්තිමය පහය : ඉදිනිර්ම් උපදේශන සේවය (institutional and Policy Support : construction Supervision Consultant)
Ref (Ta)	பரிசீலனை மேற்கொள்ளப்பட்ட நிறுவனங்கள் தொடர்பாகப் பின்வரும் அவதானிப்புக்கள் மேற்கொள்ளப்பட்டன .	ஆசிய அபிவிருத்தி வங்கி உதவியின் கீழ் வீதி வலையமைப்புகளை மேம்படுத்தும் செயற்றிட்டம் / ADB	நிறுவன ரீதியான மற்றும் கொள்கை ரீதியான உதவி : நிர்மாண ஆலோசனை சேவை
B-FT	விசாரணைக்குட்படுத்தப்பட்ட நிறுவனங்கள் தொடர்பாக பின்வரும் அவதானிப்புக்கள் மேற்கொள்ளப்பட்டன் .	ஆசிய அபிவிருத்தி வங்கி உதவியின் கீழ் வீதி வலையமைப்பு மேம்பாட்டுச் செயற்றிட்டம் / ADB	நிறுவன ரீதியான மற்றும் கொள்கை ரீதியான உதவி : நிர்மாண ஆலோசனைச் சேவை
Best Model	புலனாய்வு செய்யப்பட்ட நிறுவனங்கள் தொடர்பாக பின்வரும் அவதானிப்புக்கள் மேற்கொள்ளப்பட்டன .	ஆசிய அபிவிருத்தி வங்கியின் நிதி உதவியின் கீழ் வீதி வலையமைப்பு மேம்பாட்டுச் செயற்றிட்டம் / ADB	நிறுவன ரீதியான மற்றும் கொள்கை ரீதியான உதவி : நிர்மாண ஆலோசனைச் சேவை
Input (Ta)	பின்னதாவயில் இருந்து மாத்தறை வரை காலித் துறைமுகத்துக்கு நுழையுப் பாதை : 05 % (Pinnaduwa to Matara ; GPAR : 05 %)	கிழக்கு மாகாணத்தில் வறிய மக்களுக்கான உட்கட்டமைப்பு வசதிகளை மேற்படுத்தல் : JICA , உள்ளூர் நிதியம் (Pro-Poor Eastern Infrastructure Development Project : Japan International Coorporation Agency & Government of Sri Lanka)	ஆசிய அபிவிருந்தி வங்கியும் இவங்கை அரசும் (Asian Development Bank and Government of Sri Lanka)
Ref (Si)	පින්නදුව සිට මාතර දක්වා සහ ගාලු වරාය පිවිසුම : 05 %	නැගෙනහිර පළාතේ දුහි දුප්පත් ජනයා සඳහා වූ යථිනල පහසුකම් වැඩිදියුණු කිරීමේ වසාපතේය : JICA සහ දේශීය අරමුදල්	ආසියානු සංවර්ධන බැංකුව සහ ශූී ලංකා රජය
B-FT	පින්නදුව සිට මාතර දක්වා සහ ගානු වරාය පිවිසුම : 05 %	නැගෙනතිට පළාතේ දුහි දුප්පත් ජනයා සඳහා වූ යටිනල පහසුකම් වැඩිදියුණු කිරීමේ වසාපෘතිය : JICA සහ දේශීය අරමුදල්	ආසියානු පංවර්ධන බැංකුව සහ ශුී ලංකා රජය
Best Model	පින්නදුව සිට මාතර දක්වා සහ ගානු වරාය පිවිසුම : 05 %	නැගෙනහිර පළාතේ දුනී දුප්පත් ජනයා සඳහා වූ යටිනල පහසුසම් වැඩිදියුණු කිරීමේ වෘහපත්ය : JICA සහ දේශීය ආධාර	ආසියානු සංවර්ධන බැංකුව සහ ශී් ලංකා රජය
Input (Si)	සමාලෝටින වර්ෂය කුළු අයවැයගත හා තථන ආදායම් හා වියදුම් අතර පුමාණාත්මක විචලනයන් නිරීක්ෂණය වූ හෙයින් අයවැය ලේඛනය එලදායි පාලන කාරකයක් ලෙස භාවිත කර නොතිබුණු බව නිරීක්ෂණය විය .	අත්කර්වාර වාර්තාවට අනුව පරීක්ෂා කළ ආයතන 32 ද ඇතුලන්ය .	2014 වර්ෂයේ අයවැය ඇස්පෙමින්තුවෙන් තෝරානන් අමාතනාංශවල වැය ශීර්ෂයන් පිළිබඳ සාකච්ඡා කිරීම සඳහා වන පාර්ලිමේන්තු විශේෂ කාරක සභාව ,
Ref (En)	significant variances were observed between the budgeted and the actual income and expenditure for the year under review , thus indicating that the budget had not been made use of as an effective instrument of control .	includes 32 Institutions examined as per Interim Report .	Select Committee of Parliament to discuss the Heads of Expenditure of the Ministries selected from the Budget Estimates of the year 2014,
B-FT	Since considerable variances were observed between the budgeted and the actual income and expenditure during the year under review, it was observed that the budget had not been made use of as an effective instrument of control.	the total number of Institutions examined according to the Interim Report was 32.	Select Committee of Parliament to discuss the Heads of Expenditure of Ministries selected from the Budget Estimates of 2014 ,
Best Model	Significant variances were observed between the budgeted and the actual income and expenditure during the year under review, thus indicating that the budget had not been made use of as an effective control instrument.	this includes 32 institutions examined according to the interim report .	Select Committee to discuss the Heads of Expenditure of Ministries selected from the Budget Estimates of 2014 ,
Input (En)	Pro-Poor Eastern Infrastructure Development Project : Japan International Coorporation Agency & Government of Sri Lanka	Enhance technological knowledge need to conduct experiments on lesser known and underutilized plants .	Housing Project Operating Expenses
Ref (Si)	නැගෙනත්ර පළාතේ දුනි දුප්පත් ජනයා සඳහා වූ යටිනල පහසුකම් වැඩිදියුණු කිරීමේ වනපෘතිය : JICA සහ දේශීය අරමුදල්	දැනුවත් බවින් අඩු සහ භාවිතය අඩු ශාක ගවේෂණය කිරීමට අවශෘ තාක්ෂණික දැනුම දියුණු කිරීම .	නිවාස වනපෘතිය මෙහෙයුම් වියදම
B-FT	නැගෙනතිර පළාතේ දුයී දුප්පත් ජනයා සඳහා වූ යටිතල පහසුයාම් වැඩිදියුණු කිරීමේ වසාපෘතිය : ජපන් ජාතෘත්තර සහයෝගිතා නියෝජිතායනනය සහ ශුී ලංකා රජය	දැනුවත් බවින් අඩු සහ භාවිතය අඩු ශාක ගවේෂණය කිරීමට අවශෘ භාක්ෂණික දැනුම දියුණු කිරීම .	නිවාස වනපෘතිය මෙහෙයුම් වියදම
Best Model	ලැගෙනතිර පළාතේ දුම් දුප්පත් ජනයා සඳහා වූ යටිතල පහසුකම් වැඩිදියුණු කිරීමේ විනාපෘතිය : ජපන් ජාතෘත්තර සහයෝගිතා නියෝජිතායකනය හා ශුී ලංකා රජය	දැනුවත් බවින් අඩු සහ භාවිතය අඩු ශාක හවේණෙය කිරීමට අවශෘ තාක්ෂණික දැනුම වැඩි දියුණු කිරීම .	නිවාස වනාපසේය මෙහෙයුම් වියදම්
Input (Ta)	கீளாய்வாண்டில் பாதிடப்பட்ட மற்றும் உள்ளவாறான வருமானம் மற்றும் செலவினத்திற்கிடையேயும் முரண்கள் அவதானிக்கப்பட்டமையால் பாதீடு சிறந்த முகாமைத்துவக் கட்டுப்பாட்டுக் கருவியாக பயன்படுத்தப்பட்டிருக்காமை அவதானிக்கப்பட்டது	இடைக்கால அறிக்கையின்படி பரிசீலிக்கப்பட்ட 32 நிறுவனங்களை உள்ளடக்குகின்றது .	<ol> <li>ஒரு வருட காலப்பகுதிக்கும் மேலாக பிரதம நிறைவேற்று அதிகாரியின பதவி வெற்றிடமாயுள்ளது .</li> </ol>
Ref (En)	significant variances were observed between the budgeted and the actual income and expenditure for the year under review , thus indicating that the budget had not been made use of as an effective instrument of control .	includes 32 Institutions examined as per Interim Report .	the post of Chief Executive Officer had been vacant over a period of one year.
B-FT	It was observed that the conflicts were observed between the internal income and expenditure during the year under review and that the budget had not been made use of as an effective instrument of management control.	This includes 32 institutions examined as per the interim report .	the post of Chief Executive Officer had been vacant over a period of one year.
Best Model	It was observed that the differences were observed between the income and expenditure , which were affected during the year under review and that the budget had not been made use of as an effective management control instrument .	The Examined institutions include 32 institutions as per the & REAL report.	The post of Chief Executive Officer had been vacant for a more than one year period.
Input (En)	spiritual Leadership Training for Hindu Youths	issuance of Budget guidelines on the preparation of the Budget Estimates .	7.National Secretariat for Non-Governmental Organizations
Ref (Ta)	இந்து இளைஞர்களுக்கான ஆன்மீக தலைமைத்துவப் பயிற்சி	வரவு செலவுத்திட்ட மதிப்பீடுகளைத் தயாரித்தல் தொடர்பான வரவு செலவுத்திட்ட வழிகாட்டல்களை வழங்குதல்	7. அரச சார்பற்ற நிறுவனங்களுக்கான தேசிய செயலகம்
B-FT	இந்து இளைஞர்களுக்கான ஆன்மீக தலைமைத்துவப் பயிற்சி	வரவு செலவுத்திட்ட மதிப்பீடுகளைத் தயாரித்தல் தொடர்பான வரவு செலவுத்திட்ட வழிகாட்டல்களை வழங்குதல்	7. அரச சார்பற்ற நிறுவனங்களுக்கான தேசிய செயலகம்
Best Model	இந்து இளைஞர்களுக்கான ஆன்மீக தலைமைத்துவப் பயிற்சி	வரவு செலவுத் திட்ட மதிப்பீடுகளைத் தயாரிப்பதற்கான வரவு செலவுத்திட்ட வழிகாட்டல்களை வெளியிடுதல்	7. அரச சார்பற்ற நிறுவனங்களுக்கான தேசிய செயலகம்

Fig. 3. Output translations