

LTRC_Plural: Better Data and Better Rewards for Low-Resource Indic MT

Vishnuraj Yuvrajsinh Bodana Heli Hingrajiya

Priyanka Dasari Parameswari Krishnamurthy

Language Technologies Research Centre, IIIT Hyderabad

{vishnuraj.arjunasamy, yuvrajsinh.bodana, heli.hingrajiya,
dasari.priyanka}@research.iiit.ac.in
param.krishna@iiit.ac.in

Abstract

We present LTRC_Plural, our submission to the WMT 2026 Shared Task on Low-Resource Indic Language Translation, covering four language pairs English $\leftrightarrow$ Assamese, Mizo, Khasi, and Tagin spanning a wide range of resource levels, from moderately-resourced Assamese down to Tagin, which barely appears in existing multilingual MT systems at all. We compare three models (NLLB-200-1.3B, Sarvam-Translate, and TranslateGemma) across three setups: fine-tuning on the official shared-task data alone, fine-tuning with added external parallel data, and a reinforcement-learning stage (RLOO) that optimizes a combined BLEU/chrF++ reward on top of the fine-tuned model. Adding external parallel data helps most for Khasi, improving English-to-Khasi by 18 BLEU, and for Mizo, where our system places first among contrastive submissions in both directions; the same augmentation leaves Assamese essentially unchanged despite contributing twice as many sentence pairs, suggesting that data provenance matters more than volume at these scales. The RLOO system degrades performance on Assamese and Mizo, where supervised fine-tuning had already produced competitive systems, but improves over the primary system on both Khasi directions and on every metric for English-to-Tagin, where it obtains the highest METEOR score of any contrastive submission. Reward-based training therefore appears most useful precisely where supervised fine-tuning has least to work with.

1 Introduction

India is home to hundreds of languages beyond the 22 with constitutional “Scheduled” status, many of which are low-resource not because they lack speakers but because they lack digitized parallel text, standardized orthography, and sustained investment in language technology. This gap is especially pronounced in North-East India, one of the most linguistically diverse regions in the world.

This paper presents our submission to the WMT 2026 Shared Task on Low-Resource Indic Language Translation. We focus on four pairs:

- **English $\leftrightarrow$ Assamese (en-as):** Assamese is an Indo-Aryan language and the official language of Assam, spoken by roughly 15 million people as a first language, according to the 2011 census.
- **English $\leftrightarrow$ Mizo (en-lus):** Mizo (also called Lushai) is a Sino-Tibetan (Kuki-Chin) language, the official language of Mizoram, with on the order of 1.1 million speakers.
- **English $\leftrightarrow$ Khasi (en-kha):** Khasi is an Austroasiatic (Mon-Khmer) language spoken mainly in Meghalaya by roughly 1.4 million people, notable for using Latin rather than a Brahmic script.
- **English $\leftrightarrow$ Tagin (en-tgj):** Tagin is a Sino-Tibetan (Tani) language spoken in Arunachal Pradesh by a population of roughly 63,000 according to 2011 census, and represents one of the most severely under-resourced pairs in this shared task, with essentially no presence in standard multilingual MT systems.

The four pairs span a wide resource range, from moderate-resource Assamese to extremely low-resource Tagin, with Mizo and Khasi in between. Our experiments were conducted using three models: one encoder-decoder variant, NLLB-200¹ (Team et al., 2022; Koishekenov et al., 2023), alongside two decoder-only systems, Sarvam-translate² and TranslateGemma³. Native support for our target languages varies across these three models: NLLB-200 handles Assamese and Mizo,

¹<https://huggingface.co/facebook/nllb-200-1.3B>

²<https://huggingface.co/sarvamai/sarvam-translate>

³<https://huggingface.co/google/translategemma-4b-it>

Sarvam-Translate handle Assamese, and TranslateGemma handles Assamese, Mizo, and Khasi. None of them support Tagin.

Reliable MT for these languages has direct practical value, supporting government communication, education, language documentation, and digital access for otherwise excluded communities. Beyond this applied motivation, they serve as a useful testbed for studying how far cross-lingual transfer, multilingual fine-tuning, and monolingual-data augmentation can compensate for the absence of large bilingual corpora which is the core challenge this shared task addresses.

Our contributions are:

1. A systematic comparison of fine-tuning with shared-task data alone versus shared-task data augmented with external parallel sources.
2. We apply a method rooted in reinforcement learning, where the reward function is determined by a weighted combination of BLEU and chrF++ scores. The model is rewarded for translations that score highly on both BLEU and chrF++.

2 Related Works

Machine translation for Indic languages initially developed within the statistical machine translation (SMT) framework, but unlike European languages, which benefited from resources such as the Europarl corpus (Koehn, 2005), Indian languages had very limited publicly available parallel data, making data scarcity a major challenge for early Indic MT systems (Kakum et al., 2023). With the emergence of neural machine translation (NMT), the field gradually moved away from SMT. Attention-based encoder-decoder models (Bahdanau et al., 2014), followed by the Transformer architecture (Vaswani et al., 2017), became the dominant approaches. Subsequent work in WMT shared tasks also demonstrated the benefits of multilingual training, particularly for low-resource language pairs (Pal et al., 2023). More recent studies have explored multilingual training specifically within Indic languages, showing that limiting training to Indic-to-Indic translation pairs can further improve performance when parallel resources are scarce (Pakray et al., 2024). As these multilingual models grew larger, training all of their parameters for every new language or task became increasingly expensive, which is where parameter-efficient fine-

tuning (PEFT) comes in. LoRA (Hu et al., 2022) introduced low-rank adaptation as a lightweight alternative to full fine-tuning, and QLoRA (Dettmers et al., 2023) pushed this further by combining low-rank adaptation with quantization to cut memory requirements. Building on this, recent multilingual MT systems such as ALMA (Xu et al., 2024) and Tower (Alves et al., 2024) have shown that LoRA-based adaptation can reach strong translation performance at a fraction of the usual computational cost. Alongside these adaptation techniques, cross-lingual transfer has become a central strategy for extending translation capability to low-resource languages by leveraging related high-resource languages. Early work established that multilingual NMT models could achieve reasonable performance on unseen language pairs purely through parameter sharing (Ha et al., 2016; Arivazhagan et al., 2019), though the underlying mechanisms of successful transfer remained poorly understood at the time. Later studies pointed to language family membership as one of the strongest predictors of multilingual model success which outperformed surface-level similarity measures (Tan et al., 2019). (Dabre et al., 2020) found that multilingual Indic models benefit from careful language grouping based on linguistic relationships. Beyond data and adaptation strategies, reinforcement learning offers a different way to push translation quality further. It has driven recent gains in LLM reasoning by optimizing directly against a task reward rather than token-level likelihood (Ouyang et al., 2022; Guo et al., 2025), but applying this to MT is harder because translation quality cannot be checked with a simple rule-based verifier the way math or code can. Leave-one-out policy-gradient methods such as RLOO (Kool et al., 2019; Ahmadian et al., 2024) offer a lower-variance alternative to PPO-style RLHF without requiring a learned reward model, making them practical for optimizing directly against automatic metrics like BLEU and chrF++, as we do in our RLOO-based contrastive system.

3 Data

3.1 Official shared task data

We use the parallel training data released by the WMT 2026 Low-Resource Indic Language Translation shared task organizers (Pakray et al., 2026) for four language pairs. The corpora differ by an order of magnitude in size. Assamese and Mizo each

Pair	WMT Data			External Data		
	Train	Dev	Test	Train	Dev	Test
English–Assamese	52,923	500	500	50,000	500	500
English–Mizo	48,995	500	500	49,496	464	500
English–Khasi	25,000	500	500	22,358	482	500
English–Tagin	4,204	500	500	–	–	–

Table 1: Data statistics for the official WMT shared-task data and the external parallel data used for augmentation.

provide close to 50,000 training pairs and Khasi 25,000, while Tagin provides only 4,204, under a tenth of the largest pair. Tagin also lost the largest share of data during preprocessing, at 2.6% against well under 1% for the other pairs, while the Khasi corpus required no removals at all.

3.2 Typological differences among the language pairs

The four language pairs differ typologically as well as in resource level, and in our experiments these two factors are confounded. Assamese is written in the Bengali-Assamese abugida and shares substantial vocabulary and script conventions with Bengali and Hindi, both heavily represented in the pre-training data of NLLB-200, Sarvam-Translate, and TranslateGemma; this proximity to high-resource Indic languages plausibly aids cross-lingual transfer independent of shared-task data size. Mizo is tonal, with four lexical tones that are not consistently marked in its standard Latin orthography, so a written sentence can be phonologically ambiguous in ways no amount of parallel text resolves; its verb morphology also involves stem alternation and prefixal subject agreement not present in English. Khasi is largely isolating with verb-initial word order, diverging structurally from both English SVO and the SOV order of the Indic languages in our data, and encodes grammatical gender through particles rather than inflection. Tagin has no widely adopted standardized script and only a small number of published grammars and dictionaries, meaning its scarcity extends beyond the shared-task corpus to the language’s documentation more broadly.

Because typology and resource level vary together across these four pairs, our experiments cannot fully separate the effect of one from the other; we discuss this confound and what would be needed to isolate it in Section 8.

3.3 External parallel data

For our experiments we supplement the official data with external parallel corpora. The Assamese data was randomly sampled with a seed value of 42 from the AI4Bharat Samanantar corpus (Ramesh et al., 2022), a large mined collection covering Indian languages. The Mizo and Khasi data were collected from (Ghosh et al., 2025). The two sources differ in kind as well as in origin: Samanantar is automatically mined from comparable text, whereas the Mizo and Khasi corpora were manually developed, and this difference is relevant when interpreting the differing effect of augmentation across pairs.

No external parallel data was available for Tagin. Our Tagin systems therefore rely on the official data alone in every experiment. Table 1 summarizes the sizes of the official WMT training corpora and the additional external parallel data used for each language pair after preprocessing.

3.4 Preprocessing

We applied simple preprocessing to all data. Exact duplicate sentence pairs were removed. Stray whitespace and control characters, which would otherwise be treated as part of the surrounding tokens, were stripped. All text was normalized to Unicode NFC, which matters for the Indic scripts in our data, where the same character can be encoded as more than one sequence of codepoints and would otherwise be treated as distinct by the tokenizer. Finally, punctuation and quotation marks were normalized to consistent forms and repeated whitespace was collapsed.

3.5 Data splits

For each language pair in the WMT data, we held out 500 sentence pairs as an internal development set for checkpoint selection and a further 500 as an internal test set, sampled with a fixed seed of 42. For the external Assamese data we sampled 500 development and 500 test pairs using the same

seed, while for Mizo and Khasi we used the splits provided with those corpora. The two sets do not overlap and the remaining sentences were used for training.

All results reported in this paper are on the official test set as scored by the organizers. The internal splits were used only for checkpoint selection during development.

4 Preliminaries

Supervised fine-tuning maximises the likelihood of one reference translation token by token, penalising a fluent paraphrase as heavily as a genuine error and giving the model no feedback on the translations it actually produces. Reinforcement learning relaxes both constraints: the model generates its own translations, each is scored against a quality measure, and the model is updated toward higher-scoring output. We use RLOO (REINFORCE Leave-One-Out), an online policy-gradient method, which can be broken down into four steps: generating completions, computing the reward, computing the advantage, and computing the loss.

Generating completions: At each training step we take a batch of source sentences and, for each one, let the current model generate G different translations.

Computing the reward: Each generated translation receives a score from a reward function. A KL divergence penalty is subtracted, measuring how far the model has moved from its starting point, called the reference policy.

$$D_{KL}[\pi_\theta \parallel \pi_{\text{ref}}] = \sum_{t=1}^T \log \frac{\pi_\theta(o_{i,t} \mid q, o_{i,<t})}{\pi_{\text{ref}}(o_{i,t} \mid q, o_{i,<t})} \quad (1)$$

The final reward for a translation o_i is then:

$$r_i = R(o_i, q) - \beta D_{KL}[\pi_\theta \parallel \pi_{\text{ref}}] \quad (2)$$

where $\beta > 0$ controls how strongly the model is held to the reference policy.

Computing the advantage: A raw reward is hard to learn from, because it does not tell the model whether the score is relatively good or bad for that particular sentence. RLOO addresses this by comparing each translation against the other translations of the same sentence. For each completion i , the baseline is the average reward of its siblings:

Setting	NLLB-1.3B	Sarvam / T-Gemma
Adapters	per pair	shared
LoRA r / α	32 / 64	32 / 64
LoRA dropout	0.05	0.05
Learning rate	2×10^{-4}	5×10^{-4}
Epochs	5	5
Batch size	32	32
Max seq. length	256	256

Table 2: Training configuration for Experiments 1 and 2.

$$b_i = \frac{1}{G-1} \sum_{j \neq i} r_j \quad (3)$$

and the advantage is the difference between its own reward and this baseline:

$$A_i = r_i - b_i \quad (4)$$

Because the baseline never includes the completion’s own reward, it gives an unbiased estimate. This is the “leave-one-out” the method is named for, and it is what allows RLOO to work without the separate value network that PPO requires, making it considerably cheaper to train.

Computing the loss: The model is then updated to make high-advantage translations more likely and low-advantage ones less likely:

$$\mathcal{L}_{\text{RLOO}}(\theta) = -\frac{1}{G} \sum_{i=1}^G \hat{A}_i \log \pi_\theta(o_i \mid q) \quad (5)$$

5 Methodology

We conducted three experiments. Experiment 1 fine-tunes each model on the official shared task data alone. Experiment 2 adds external parallel data while keeping all hyperparameters fixed. Experiment 3 continues training the checkpoints from Experiment 1 with RLOO (Section 4). Each experiment changes one factor, so the differences between them are easy to interpret.

Data splits and preprocessing are described in Section 3. All models were trained with QLoRA, which keeps the base weights frozen in 4-bit NF4 precision and trains only a small set of adapter weights, allowing large models to be fine-tuned without discarding pretrained knowledge. All experiments ran on 2×L40S GPUs with 48GB VRAM each.

Language	Direction	Primary		Contrastive-1		Contrastive-2	
		BLEU↑	CHRF++↑	BLEU↑	CHRF++↑	BLEU↑	CHRF++↑
Assamese	en→lang	14.09	47.75	14.01	46.63	9.41	41.22
	lang→en	22.09	53.11	23.27	53.48	18.99	50.8
Mizo	en→lang	19.02	48.07	22.85	51.02	16.47	44.96
	lang→en	28.07	58.49	33.45	62.17	20.35	53.32
Khasi	en→lang	14.47	36.28	32.46	54.77	20.8	43.35
	lang→en	7.94	31.03	23.72	49.83	11.82	36.81
Tagin	en→lang	2.25	19.17	1.65	21.76	3.88	26.14
	lang→en	2.62	16.91	1.3	16.7	2.28	17.64

Table 3: BLEU and chrF++ scores for all three submitted systems, grouped by language pair and direction. Bold indicates the best score for each row, so that comparisons are read within a language pair and direction across systems. Full results including METEOR, TER, BERTScore, and COMET for each system are given in Appendix A.

5.1 Experiment 1: Fine-tuning on shared task data

We fine-tuned NLLB-200-1.3B, Sarvam-Translate, and TranslateGemma on the official WMT training data alone (Section 3). This forms our primary submission. Training settings are given in Table 2.

For NLLB we trained a separate adapter per language pair. NLLB has no language tag for Khasi or Tagin, so we added new tags (kha_Latn, tgj_Latn) and kept the encoder and decoder embedding weights tied so that each tag shares one representation on both sides. NLLB selects the output language by reading this tag, and training several new pairs jointly forces the model to learn the tags while it learns the translations, which is unstable with limited data; separate adapters avoid this.

For Sarvam-Translate and TranslateGemma we trained a single adapter shared across all pairs. These models take the target language as text in the prompt rather than as a tag, so there is no language identification (LID) interference and no vocabulary change is required. We used rank-stabilised LoRA (rsLoRA) (Kalajdzievski, 2023), which scales the adapter update so that training stays stable at higher learning rates. Sarvam-Translate is built on a multi-modal backbone, but the image encoder remained frozen as our task is text-only.

5.2 Experiment 2: Adding external parallel data

Models, adapter settings, and all hyperparameters are unchanged from Experiment 1 (Table 2), and only the training data differs, so any change in results can be attributed to the additional data. This forms our first contrastive submission. TranslateGemma was excluded from this experiment.

The external corpora are listed in Section 3; none was available for Tagin, so its training used the official data alone. We sampled with a temperature of 0.7 over both translation directions of every dataset, so that smaller datasets appear more often than their raw size would imply without the largest ones taking over. Tagin results still change between Experiments 1 and 2 despite no Tagin data being added, because Sarvam-Translate uses a single shared adapter and the additional Khasi, Mizo, and Assamese data affects it indirectly.

5.3 Experiment 3: Reinforcement learning with RLOO

We applied RLOO (Section 4) to the Experiment 1 checkpoints for Sarvam-Translate and TranslateGemma, which forms our second contrastive submission. The supervised adapter was merged into the base model and a fresh rsLoRA adapter trained on top, so the frozen reference policy π_{ref} for the KL term is the Experiment 1 checkpoint. NLLB was excluded because the RLOO implementation we used supports only decoder-only architectures.

The reward combines two translation metrics:

$$R = 0.3 \times \text{BLEU} + 0.7 \times \text{chrF++} \quad (6)$$

Both are computed with sacreBLEU on a 0–100 scale, so the weights apply to comparable ranges and no normalisation is required. We weight chrF++ more heavily because it matches human judgement better for morphologically rich and low-resource languages, and because sentence-level BLEU is frequently zero or near-zero on short sentences, giving a sparse reward with little learning signal.

Setting	Value
Completions per prompt (G)	2
KL coefficient (β)	0.05
Reward	$0.3 \text{ BLEU} + 0.7 \text{ chrF++}$
Learning rate	1×10^{-6}
Epochs	1
Sampling temperature	1.0

Table 4: RLOO configuration for Experiment 3.

We sampled 2,000 sentence pairs per language pair with seed 42, covering both translation directions. Assamese, Mizo, and Khasi used the external corpora of Experiment 2; Tagin used the official data (Section 3). Batches were built with temperature sampling at 0.7, as in Experiment 2. The RLOO configuration is given in Table 4;

6 Results

We report official test set scores as released by the shared task organizers. Ranks are stated within the relevant submission pool, primary or contrastive, for that direction. Table 3 reports BLEU and chrF++ for all three of our systems across all four language pairs (full metrics in Appendix A). We organize the discussion below by language pair, comparing our three systems to one another and to other participants within the same pair and direction.

6.1 Assamese

For the primary system, Assamese-to-English reaches 22.09 BLEU (8th of 13) against track leader *gators* at 27.26, and English-to-Assamese reaches 14.09 BLEU. Within this language pair, translating into English scores higher than translating out of it under every metric, a pattern also seen for Mizo below and consistent across every participating system in both tracks; we take this as a property of translating into a morphologically simpler target rather than evidence about the language pair’s difficulty relative to others.

Adding external data (contrastive-1) leaves Assamese essentially unchanged: -0.08 BLEU into Assamese and $+1.18$ into English, despite contributing 50,000 automatically mined Samanantar sentence pairs, more than for any other language pair in this work. RLOO (contrastive-2) degrades performance in both directions relative to the primary system, most sharply for Assamese-to-English (22.09 to 18.99 BLEU, TER rising from 68.6 to 84.35), consistent with the pattern discussed in Section 6.5.

6.2 Mizo

The primary system scores 19.02 BLEU for English-to-Mizo (6th of 9) against track leader *gators* at 21.11, and 28.07 for Mizo-to-English.

Mizo yields our strongest ranking under contrastive-1: English-to-Mizo improves to 22.85 BLEU and Mizo-to-English to 33.45, placing first among contrastive submissions in both directions, ahead of MTG-LIRA (22.72) and *gators* (21.60) into Mizo, and *gators* (33.28) into English. RLOO degrades performance relative to both prior systems, most sharply for Mizo-to-English, which falls to 20.35 BLEU with TER rising from 54.63 to 84.08, the largest RLOO regression we observe for any language pair.

6.3 Khasi

Khasi is the weakest primary result: English-to-Khasi reaches 14.47 BLEU against 29.30 for the leading system, and Khasi-to-English 7.94 against 26.28, indicating that the 25,000 official sentence pairs available for this pair were insufficient on their own.

Contrastive-1 improves Khasi more than any other pair in this work: English-to-Khasi rises to 32.46 BLEU ($+17.99$) and chrF++ to 54.77, while Khasi-to-English rises to 23.72 BLEU ($+15.78$) with TER falling from 124.56 to 67.45, placing 3rd and 4th respectively among contrastive submissions. The external data added for Khasi was 22,358 manually developed sentence pairs. We note this as a striking within-pair improvement rather than as a direct comparison to the much smaller change observed for Assamese (Section 6.1) under a much larger but automatically mined addition of external data; we return to this contrast, with appropriate caveats, in Section 6.5.

RLOO further improves on the primary system in both directions, English-to-Khasi to 20.80 BLEU and Khasi-to-English to 11.82, with TER falling by 17.32 and 13.10 points respectively though both remain below the corresponding contrastive-1 scores.

6.4 Tagin

The primary system places 3rd of 4 for English-to-Tagin, in a direction where no participant exceeded 4.81 BLEU, and 3rd of 5 for Tagin-to-English.

Contrastive-1 declines slightly on BLEU in both directions despite no Tagin data being added, consistent with the shared-adaptor effect noted in Section 5, though chrF++ improves into Tagin (19.17

to 21.76). RLOO improves on the primary system on every metric for English-to-Tagin, reaching 3.88 BLEU, 20.90 METEOR, 26.14 chrF++, and 54.86 COMET, against 2.25 / 15.46 / 19.17 / 51.52 for the primary system. Among all contrastive submissions in this direction it obtains the highest METEOR score, ahead of ANVITA (20.12) and gators (19.87), and the second-highest chrF++ behind ANVITA (28.00).

6.5 Comparative analysis

Three qualitative patterns recur across language pairs.

First, the value of additional external data appears to depend on its provenance rather than reliably on its volume. Khasi received the smallest external addition by sentence count (22,358 manually developed pairs) and showed the largest contrastive-1 improvement of any pair; Assamese received the largest addition (50,000 automatically mined pairs) and showed essentially no change. We do not claim the size of these two effects is directly comparable, BLEU behaves differently across the two languages but the direction of the pattern (manually developed data helping substantially, mined data helping negligibly) held consistently in our experiments.

This is a claim about the added external data specifically, not a claim that data volume is irrelevant in general. Holding provenance fixed and varying only the amount of official Assamese training data, a controlled ablation (Appendix B) shows that performance does still depend on training data volume: our held-out test set scores rise from 25% to 100% of the official Assamese data, with the largest single gain coming from the full dataset. The relationship is not smooth scores plateau across the 50–75% range so we read this as support for a general dependence on volume rather than for a strictly monotonic one, and this ablation covers Assamese only. Taken together, the two results suggest that within a fixed data source, more data helps, but across data sources of different provenance, quality can matter more than quantity.

Second, RLOO appears useful only where the supervised starting point was weak. On Assamese and Mizo, where fine-tuning alone had already produced competitive systems, RLOO degraded performance in every direction we measured. On Khasi and Tagin, where the primary system was weakest, RLOO improved on it in every direc-

tion we measured. The RLOO configuration used $G = 2$ completions per prompt (Table 4), so the leave-one-out baseline reduces to a single pairwise comparison and the resulting gradient estimate is high-variance; we regard this as the most likely cause of the degradation on the stronger pairs.

Third, metrics disagree with each other at the low-resource extreme. For English-to-Tagin, the RLOO system ranks 1st on METEOR and 2nd on chrF++ but only 3rd on BLEU, whose n -gram matching is too sparse to reliably separate systems scoring below 5 BLEU. This mirrors the rationale for weighting the training reward toward chrF++ (Section 5.3), and suggests that BLEU alone is a poor basis for ranking systems in this regime.

7 Conclusion

We presented our submission to the WMT 2026 Low-Resource Indic Language Translation shared task, covering four language pairs at different levels of resource availability: English↔Assamese, Mizo, Khasi, and Tagin. Two findings emerge. First, the benefit of the external data we added depends more on its provenance than on its sentence count. Roughly 22,000 manually developed Khasi sentence pairs improved English-to-Khasi by 18 BLEU, while 50,000 automatically mined Assamese pairs left performance essentially unchanged in both directions; for Mizo, external data placed our system first among contrastive submissions in both directions, this is specific to the added external data, not that data volume is irrelevant in general: a controlled ablation varying only the amount of official Assamese training data (Section 6.5, Appendix B) shows performance still depends on volume when provenance is held fixed, and typology and resource level are confounded across our four pairs in any case (Section 8). Second, reinforcement learning helped only where supervised fine-tuning had least to work with. RLOO degraded performance on Assamese and Mizo, where fine-tuning alone had already produced competitive systems, but improved over the primary system on both Khasi directions and on every metric for English-to-Tagin, obtaining the highest METEOR score of any contrastive submission in that direction. Together these results indicate that data and training strategy should be selected according to the specific resource conditions of each language rather than applied uniformly across low-resource Indic pairs.

8 Limitations

Our RLOO experiments used a single training epoch, a fixed KL coefficient of 0.05, and $G = 2$ completions per prompt. With only two completions, the leave-one-out baseline reduces to a single pairwise comparison, and the resulting gradient estimate is high-variance; we regard this as the most likely cause of the degradation observed on Assamese and Mizo. Larger G , additional epochs, and a tuned KL coefficient are the clearest directions for improvement, and we did not tune any of these under the shared-task compute budget. Because no external parallel corpus was available for Tagin, the RLOO stage for that pair reused the same official data the supervised checkpoint had already been trained on. Reward optimization on previously seen data risks sharpening the model around references it has effectively memorized rather than improving generalization, limiting the new learning signal available at every stage. Our reward function is reference-based. It combines BLEU and chrF++, both of which credit surface overlap with a single reference translation, so it relaxes rather than removes the reference-boundedness that motivated the reinforcement-learning stage in the first place.

Our four language pairs also vary in typology as well as resource level (Section 3.2), and these two factors are confounded throughout our experiments: we cannot attribute a given pair’s performance to its data volume as opposed to its typological distance from the high-resource languages our base models were pretrained on. Isolating the two would require training on matched data sizes across typologically similar and dissimilar pairs. The Assamese augmentation result (Section 6.1), where doubling the training data left performance essentially unchanged, offers indirect evidence that volume alone does not determine outcomes even for a single, typologically favorable language, but this evidence is limited to one language pair. We flag controlled, matched-size training across the four pairs as a direction for future work.

References

- Arash Ahmadian, Chris Cremer, Matthias Gallé, Marzieh Fadaee, Julia Kreutzer, Olivier Pietquin, Ahmet Üstün, and Sara Hooker. 2024. Back to basics: Revisiting reinforce-style optimization for learning from human feedback in llms. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12248–12267.
- Duarte M Alves, José Pombal, Nuno M Guerreiro, Pedro H Martins, João Alves, Amin Farajian, Ben Peters, Ricardo Rei, Patrick Fernandes, Sweta Agrawal, et al. 2024. Tower: An open multilingual large language model for translation-related tasks. *arXiv preprint arXiv:2402.17733*.
- Naveen Arivazhagan, Ankur Bapna, Orhan Firat, Dmitry Lepikhin, Melvin Johnson, Maxim Krikun, Mia Xu Chen, Yuan Cao, George Foster, Colin Cherry, et al. 2019. Massively multilingual neural machine translation in the wild: Findings and challenges. *arXiv preprint arXiv:1907.05019*.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.
- Raj Dabre, Chenhui Chu, and Anoop Kunchukuttan. 2020. A survey of multilingual neural machine translation. *ACM Computing Surveys (CSUR)*, 53(5):1–38.
- Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: Efficient finetuning of quantized llms. *Advances in neural information processing systems*, 36:10088–10115.
- Soumyadip Ghosh, Henry Lalsiam, Dorothy Marbaniang, Gracious Mary Temsen, Rahul Mishra, and Parameswari Krishnamurthy. 2025. Towards resource-rich Mizo and Khasi in NLP: Resource development, synthetic data generation and model building. In *Proceedings of the 19th Linguistic Annotation Workshop (LAW-XIX-2025)*, pages 228–239. Association for Computational Linguistics.
- Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.
- Thanh-Le Ha, Jan Niehues, and Alex Waibel. 2016. Toward multilingual neural machine translation with universal encoder and decoder. In *Proceedings of the 13th International Conference on Spoken Language Translation*.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Liang Wang, Weizhu Chen, et al. 2022. Lora: Low-rank adaptation of large language models. *Iclr*, 1(2):3.
- Nabam Kakum, Sahinur Rahman Laskar, Koj Sambyo, and Partha Pakray. 2023. Neural machine translation for limited resources english-nyishi pair. *Sādhanā*, 48(4):237.
- Damjan Kalajdzievski. 2023. A rank stabilization scaling factor for fine-tuning with lora. *arXiv preprint arXiv:2312.03732*.

- Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of machine translation summit x: papers*, pages 79–86.
- Yeskendir Koishikenov, Alexandre Berard, and Vasilina Nikoulina. 2023. [Memory-efficient NLLB-200: Language-specific expert pruning of a massively multilingual machine translation model](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3567–3585, Toronto, Canada. Association for Computational Linguistics.
- Wouter Kool, Herke Van Hoof, and Max Welling. 2019. Buy 4 reinforce samples, get a baseline for free!
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.
- Partha Pakray, Santanu Pal, Advaita Vetagiri, Reddi Krishna, Arnab Kumar Maji, Sandeep Dash, Lenin Laitonjam, Lyngdoh Sarah, and Riyanka Manna. 2024. [Findings of WMT 2024 shared task on low-resource Indic languages translation](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 654–668, Miami, Florida, USA. Association for Computational Linguistics.
- Partha Pakray, Santanu Pal, Advaita Vetagiri, Kshetri-mayum Boynao Singh, Sandeep Kumar Dash, Arnab Kumar Maji, Saralin A. Lyngdoh, Lenin Laitonjam, Anupam Jamatia, Ajit Das, Sunita Warjri, Uzzal Sharma, Koj Sambyo, and Riyanka Manna. 2026. Findings of the WMT 2026 shared task on low-resource Indic languages translation. In *Proceedings of the Eleventh Conference on Machine Translation*, Budapest, Hungary. Association for Computational Linguistics.
- Santanu Pal, Partha Pakray, Sahinur Rahman Laskar, Lenin Laitonjam, Vanlalmuansangi Khenglawt, Sunita Warjri, Pankaj Kundan Dadure, and Sandeep Kumar Dash. 2023. Findings of the wmt 2023 shared task on low-resource indic language translation. In *Proceedings of the Eighth Conference on Machine Translation*, pages 682–694.
- Gowtham Ramesh, Sumanth Doddapaneni, Aravindh Bheemaraj, Mayank Jobanputra, Raghavan AK, Ajitesh Sharma, Sujit Sahoo, Harshita Diddee, Mahalakshmi J, Divyanshu Kakwani, Navneet Kumar, Aswin Pradeep, Srihari Nagaraj, Kumar Deepak, Vivek Raghavan, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh Shantadevi Khapra. 2022. [Samanantar: The largest publicly available parallel corpora collection for 11 Indic languages](#). *Transactions of the Association for Computational Linguistics*, 10:145–162.
- Xu Tan, Jiale Chen, Di He, Yingce Xia, Tao Qin, and Tie-Yan Liu. 2019. Multilingual neural machine translation with language clustering. 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)*, pages 963–973.
- NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Hefernan, 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*.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.
- Haoran Xu, Young Jin Kim, Amr Mohamed Nabil Aly Aly Sharaf, and Hany Awadalla. 2024. A paradigm shift in machine translation: Boosting translation performance of large language models. In *International Conference on Learning Representations*, volume 2024, pages 2747–2767.

A Full Metric Results

Table 3 in the main text reports BLEU and chrF++ for all three systems. For completeness, Tables 5, 6, and 7 below give the full set of metrics (BLEU, METEOR, TER, chrF++, BERTScore, COMET) for the primary, contrastive-1, and contrastive-2 systems respectively, one table per system, matching the official shared-task submission format.

B Data Volume Ablation

To test whether fine-tuning performance depends on training data volume, we fine-tuned NLLB the strongest base model for Assamese in our primary system (Table 3) on 25%, 50%, 75%, and 100% of the official Assamese WMT training data, using a fixed random seed for each subsample and identical hyperparameters across all four runs. Since gold references for the official shared-task test set are not available to us, all four models are evaluated on a held-out test set we constructed independently, with the same references and scoring script used for every run so that scores within this table are directly comparable to one another. These scores are not comparable to the official test set scores reported elsewhere in this paper (Table 3, Appendix A), which use a different test set; this ablation should be read as a self-contained, controlled comparison.

It is noticeable that the trend is not monotonic: scores rise from 25% to 50%, are roughly flat from

Language Pair	BLEU↑	METEOR↑	TER	CHRF++↑	BERTScore↑	COMET↑
English to Assamese	14.09	35.4	73.21	47.75	85.41	80.87
Assamese to English	22.09	56.67	68.6	53.11	94.46	85.31
English to Mizo	19.02	45.81	68.81	48.07	81.44	65.29
Mizo to English	28.07	62.31	61.86	58.49	94.98	79.05
English to Khasi	14.47	36.23	85.95	36.28	76.95	62.51
Khasi to English	7.94	28.75	124.56	31.03	89.03	56.03
English to Tagin	2.25	15.46	195.42	19.17	67.91	51.52
Tagin to English	2.62	12.33	126.16	16.91	86.25	47.39

Table 5: Primary System Evaluation results across all translation directions and metrics.

Language Pair	BLEU↑	METEOR↑	TER	CHRF++↑	BERTScore↑	COMET↑
English to Assamese	14.01	34.75	69.34	46.63	84.87	81.6
Assamese to English	23.27	56.61	68.75	53.48	94.45	85.1
English to Mizo	22.85	49.16	63.32	51.02	82.39	65.35
Mizo to English	33.45	66.75	54.63	62.17	95.54	80.3
English to Khasi	32.46	57.16	48.7	54.77	84.44	68.06
Khasi to English	23.72	52.53	67.45	49.83	94.03	74.07
English to Tagin	1.65	13.63	176.5	21.76	69.46	53.37
Tagin to English	1.3	10.65	121.78	16.7	85.34	46.28

Table 6: Contrastive-1 System Evaluation results across all translation directions and metrics.

Language Pair	BLEU↑	METEOR↑	TER	CHRF++↑	BERTScore↑	COMET↑
English to Assamese	9.41	28.35	81.81	41.22	83.77	78.57
Assamese to English	18.99	53.05	84.35	50.8	93.22	81.98
English to Mizo	16.47	41.93	79.44	44.96	79.94	61.91
Mizo to English	20.35	56.23	84.08	53.32	93.44	74.27
English to Khasi	20.8	44.14	68.63	43.35	79.66	61.68
Khasi to English	11.82	35.74	111.46	36.81	90.26	59.78
English to Tagin	3.88	20.9	144.21	26.14	70.83	54.86
Tagin to English	2.28	11.87	126.74	17.64	85.79	47.26

Table 7: Contrastive-2 System Evaluation results across all translation directions and metrics.

Data fraction	en→as		as→en	
	BLEU	chrF++	BLEU	chrF++
25%	13.78	38.77	34.25	52.40
50%	14.80	39.44	35.00	52.72
75%	14.70	39.56	35.12	53.01
100%	17.48	41.78	36.66	54.12

Table 8: Effect of Assamese training data volume on NLLB fine-tuning, evaluated on a held-out test set. All four rows use the same base model, hyperparameters, and held-out test set; only the fraction of official training data used differs.

50% to 75% (en→as BLEU even dips slightly), and then rise again from 75% to 100% the largest single jump in the table, and consistent in direction across all four columns. We treat this ablation as evidence for a general dependence of performance on data volume, supporting the claim discussed in Section 6.5.