

KrenTransV0.1: Multi-Model Routing and Reranking for Low-Resource Northeast Indian Translation

Badal Nyalang
MWire Labs
Shillong, Meghalaya, India
nyalang@mwirelabs.com

Riya Tekcham
MWire Labs
Shillong, Meghalaya, India

Biman Debbarma
Department of Kokborok
Tripura University
Tripura, India
bimandblg@gmail.com

Abstract

KrenTransV0-1 is MWire Labs’ submission to the WMT 2026 Low-Resource Indic MT Shared Task. We achieve #1 primary rank on both English→Khasi and English→Kokborok. The system covers all 11 Northeast Indian language pairs using a hybrid architecture combining fine-tuned IndicTrans2 for Indic-script languages, NLLB-200 1.3B mono fine-tuned models for Latin-script languages, and M2M-100 418M fine-tuned models for three severely data-starved languages that exhibited complete output collapse under NLLB. All submissions are primary, covering the English-to-target direction only. We additionally report a negative reranking result: neither BharatGen PARAM-1 (2.9B) perplexity reranking nor NE-BERT fluency reranking improved over a beam-search baseline, and all final submissions use the baseline decoding configuration.

1 Introduction

Northeast India is home to over 220 indigenous languages, the overwhelming majority of which have no machine translation coverage and little to no benchmarked NLP support. MWire Labs, a research lab based in Shillong, Meghalaya, submitted to the WMT 2026 Low-Resource Indic Languages Translation shared task (Pakray et al., 2026) under the team name KrenTransV0-1, covering all 11 registered language pairs: Khasi (kha), Mizo (lus), Nyishi (njz), Kokborok (trp), Karbi (mjw), Tagin (tgj), Nagamese (nag), Bodo (brx), Assamese (as), Manipuri/Meitei (mni, Bengali and Meitei Mayek scripts).

Our system is deliberately heterogeneous: rather than a single multilingual model, we route each language to one of three backbones based on script family and, for Latin-script languages, empirically observed NLLB fine-tuning collapse, and select the fine-tuning regime per language based on which configuration performed best on held-

out data. Among primary submissions, this approach achieves #1 rank on both English→Khasi and English→Kokborok. We also report a reranking ablation that produced a clean negative result, which we believe is useful for other low-resource WMT participants to know about before investing engineering time in it.

2 Task and Data

All 11 language pairs use officially provided WMT 2026 parallel training data only; no external parallel or monolingual data was used in the final submitted models. Table 1 summarizes the language pairs by script family and training pair count.

Nyishi (njz) was subsequently withdrawn by the shared task organizers and is excluded from the official results discussed in Section 6.

3 Related Work

The WMT Low-Resource Indic Languages Translation shared task has run annually since 2023, progressively expanding language coverage: the 2023 edition (Pal et al., 2023) covered four language pairs – Assamese, Mizo, Khasi, and Manipuri – and the 2024 edition (Pakray et al., 2024) covered the same four pairs without further expansion. The 2025 edition (Pakray et al., 2025) expanded to seven pairs, adding Nyishi as a Category 1 pair alongside Bodo and Kokborok as Category 2 (severely low-resource) pairs, and introduced the enriched IndicNE-Corp1.0 dataset. The 2026 edition further extends the task with Karbi, Nagamese, and Tagin, and adds a separate Meitei Mayek script track; Nyishi carries over from 2025 but was withdrawn by the organizers during the 2026 edition (Section 2). Across prior editions, participating systems have relied primarily on fine-tuning existing multilingual translation backbones – IndicTrans2 (Gala et al., 2023) and NLLB-200 (NLLB Team et al., 2022) – with LoRA-based adaptation, back-

Language	Script	Pairs	Backbone (Section 3)
Assamese (as)	Indic (Bengali-Assamese)	47,572	IndicTrans2 200M, fine-tuned
Bodo (brx)	Indic (Devanagari)	12,749	IndicTrans2 200M, fine-tuned
Manipuri (mni)	Indic (Bengali)	22,541	IndicTrans2 200M, fine-tuned
Meitei Mayek (mni_Mtei)	Indic (Meitei Mayek)	16,666	IndicTrans2 200M, fine-tuned
Mizo (lus)	Latin	47,841	NLLB-200 1.3B, mono fine-tuned
Khasi (kha)	Latin	25,793	NLLB-200 1.3B, mono fine-tuned
Nyishi (njz)	Latin	49,173	NLLB-200 1.3B, mono fine-tuned (withdrawn)
Kokborok (trp)	Latin	2,263	NLLB-200 1.3B, mono fine-tuned
Nagamese (nag)	Latin	1,983	M2M-100 418M, fine-tuned (collapse fix target, §6)
Tagin (tgj)	Latin	4,326	M2M-100 418M, fine-tuned (collapse fix target, §6)
Karbi (mjw)	Latin	898	M2M-100 418M, fine-tuned (collapse fix target, §6)

Table 1: Language pairs, script family, WMT 2026 official training pair counts, and the backbone assigned to each language.

translation, and data augmentation as the dominant strategies for improving low-resource pairs, rather than training new architectures from scratch (SivabHAVANI et al., 2025). Our system follows this same fine-tuning paradigm but differs in explicitly routing languages to different backbones by script family and by empirically observed collapse behavior, rather than fine-tuning a single backbone across all pairs.

4 System Description

Model selection was driven by script type and, for Latin-script languages, empirically observed NLLB fine-tuning collapse, rather than a single unified architecture. Figure 1 summarizes the routing logic: each language is assigned a backbone based on its script family and, for Latin-script languages, whether NLLB fine-tuning produced usable (non-collapsed) output.

4.1 Indic-Script Languages

For Assamese, Bodo, and Manipuri (both Bengali and Meitei Mayek script variants), we fine-tuned IndicTrans2 200M (ai4bharat/indictrans2-en-indic-dist-200M) (Gala et al., 2023) separately per language on the official WMT training data, following the same per-language fine-tuning approach used for the Latin-script backbones (Section 4.2).

4.2 Latin-Script Languages

For Mizo, Khasi, and Kokborok, we fine-tuned NLLB-200 1.3B (facebook/nllb-200-1.3B) (NLLB Team et al., 2022) separately per language. Mono (single-language-pair) fine-tuning consistently outperformed multilingual fine-tuning across a shared checkpoint in our internal evaluation, so all Latin-script submissions use dedicated

per-language checkpoints rather than one shared multilingual model.

4.3 Data-Starved Languages: M2M-100 Fallback

Nagamese, Tagin, and Karbi produced collapsed output under NLLB fine-tuning (off-target language generation, source-copy passthrough, and code-switched output respectively). For these three languages we instead fine-tuned M2M-100 418M (facebook/m2m100_418M) (Fan et al., 2021), adding custom target-language tokens initialized from the mean of existing Latin-script token embeddings in the M2M-100 vocabulary, following the FOCUS approach of initializing new-language embeddings from the average of related-script embeddings rather than random initialization (Dobler and de Melo, 2023). Latin-script tokens were identified using M2M-100’s language tag metadata for its supported Latin-script languages.

4.4 Training Details

Training was carried out on single NVIDIA A40 GPUs (48GB) in bf16 precision using the Hugging Face Transformers Seq2SeqTrainer. For the NLLB-200 1.3B mono fine-tuning runs, we used the AdamW optimizer with a cosine learning-rate schedule, peak learning rate 1×10^{-5} (reduced from an initial 2×10^{-5} to improve stability at this model size), effective batch size 32, 500 warmup steps, weight decay 0.01, max gradient norm 1.0, and early stopping (patience 3) on validation loss, evaluating and checkpointing every epoch with the best three checkpoints retained. For the highest-resource pairs where a curated gold-standard subset was available (distinct from the official shared task test set), we followed a two-stage recipe: an initial fine-tuning stage on the full training corpus,

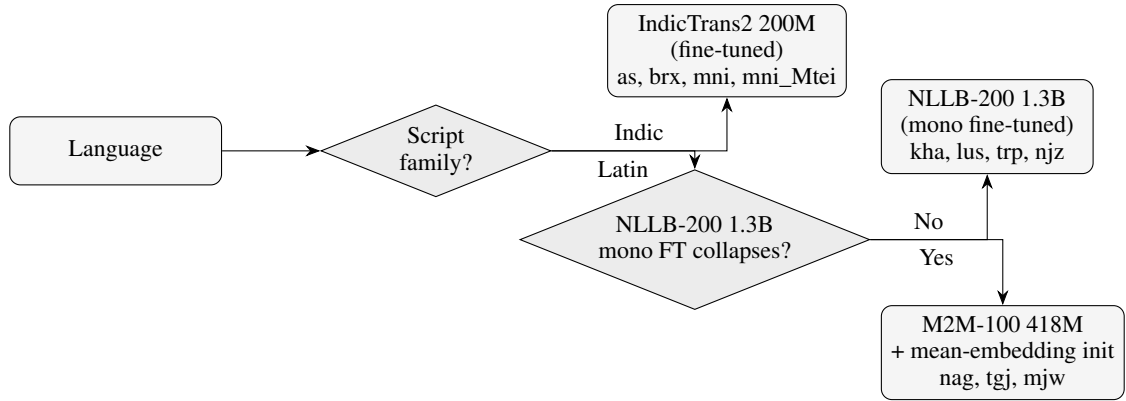


Figure 1: Backbone routing logic. Languages are assigned to a backbone by script family; Latin-script languages that exhibited output collapse under NLLB-200 mono fine-tuning were routed to M2M-100 418M with mean-embedding target-token initialization instead.

followed by a lower learning-rate (5×10^{-6}) polishing stage on this gold-standard subset, using a held-out slice of the same gold data for validation, with more frequent evaluation to guard against overfitting on the smaller gold set. IndicTrans2 200M fine-tuning used 5 epochs, batch size 16, learning rate 3×10^{-4} , fp16, and 200 warmup steps. M2M-100 418M fine-tuning used batch size 8, learning rate 5×10^{-4} , fp16, and 100 warmup steps, with 5 epochs for Nagamese and Tagin and 10 epochs for Karbi given its smaller training set; best checkpoints were selected by epoch based on output coherence.

5 Reranking Ablation

We tested two rerankers on top of beam-search output on our two most complete internal ablation pairs (Khasi, Assamese): (1) perplexity-based reranking using BharatGen PARAM-1, a 2.9B-parameter Hindi-English decoder-only model (Pundalik et al., 2025), and (2) fluency-based reranking using our in-house NE-BERT encoder (Nyalang, 2026), both individually and combined with minimum Bayes risk (chrF utility) candidate selection. Neither reranker produced a consistent BLEU or chrF improvement over the beam-search baseline on internal held-out evaluation; in several configurations both metrics decreased slightly. We therefore did not use reranking in any final submitted system. All primary submissions use plain beam search with num_beams=4, repetition_penalty=1.8, and no_repeat_ngram_size=3.

We report this as a negative result: for our data regime, decoding-time reranking did not compensate for training-data scarcity, and the more effec-

tive lever was backbone/architecture selection per language (Section 3) rather than post-hoc candidate reranking. Table 2 reports representative internal ablation numbers, evaluated with sacreBLEU (Post, 2018). These internal development-set scores (e.g. 37.93 BLEU for Khasi) are higher than the official shared task score reported in Table 3 (29.3 BLEU); this gap is expected and reflects the usual difference between an internal held-out split and the organizers’ blind, independently curated test set, rather than any change to the submitted model between internal evaluation and official scoring.

6 Results

All submissions cover English-to-target only and are entered as primary systems. Table 3 reports official shared task results for all ten scored directions (Nyishi was withdrawn by the organizers due to unforeseen circumstances and has no official result).

Khasi and Kokborok are our strongest results, ranking #1 among primary submissions in their respective directions. We note that both top-ranked directions use the NLLB-200 backbone; the multi-backbone strategy as a whole is best understood as a fallback for collapse cases rather than a demonstrated uniform improvement. Assamese ranks #3 of 15 primary systems. Mizo, both Manipuri variants, and Bodo place in the middle of the field.

Nagamese, Tagin, and Karbi remain at or near collapse in the official results (0.67, 0.0, and 3.47 BLEU respectively). For these three languages, which exhibited complete output collapse under NLLB fine-tuning, we switched to M2M-100 418M with mean-embedding initialization of the new target-language tokens (Dobler and de Melo,

Language	Decoding strategy	BLEU	chrF
Khasi (kha)	Beam search baseline	37.93	60.79
Khasi (kha)	+ NE-BERT fluency reranking	37.86	–
Assamese (as)	Beam search baseline	46.45	–
Assamese (as)	+ BharatGen PARAM-1 perplexity reranking	45.29	–

Table 2: Internal reranking ablation on held-out development data. Neither reranking strategy improved over the beam search baseline; both are marginally worse, so neither is used in the final submitted system.

Direction	BLEU	METEOR	TER↓	chrF++	BERTScore	COMET	Rank (primary)
en→kha	29.3	54.21	50.44	53.01	83.89	66.98	#1 of 8
en→trp	6.1	26.17	80.58	34.91	74.73	60.84	#1 of 4
en→as	16.64	41.72	63.59	51.22	86.63	82.9	#3 of 15
en→lus	19.06	45.25	67.81	47.38	80.94	64.08	#5 of 9
en→mni (Bengali)	8.16	17.42	105.67	37.92	83.98	67.19	#8 of 14
en→mni_Mtei	3.56	18.86	122.00	38.99	96.81	85.13	#7 of 9
en→brx	9.2	30.78	79.10	44.06	82.7	59.64	#9 of 10
en→mjw	3.47	15.03	92.41	22.06	62.44	40.45	#2 of 2 (last)
en→nag	0.67	10.95	108.24	16.4	67.56	44.16	#5 of 5 (last)
en→tgj	0.0	3.73	98.38	8.98	68.57	47.32	#4 of 4 (last)

Table 3: Official WMT 2026 primary-system results by direction, ranked among primary submissions only (contrastive systems excluded). Nyishi (njz) was withdrawn by the organizers.

2023). Official results show that this change did not recover performance (0.67 / 0.0 / 3.47 BLEU). Data volume alone does not fully explain this pattern: Kokborok (2,263 pairs) achieves our best result under NLLB-200, while Tagin (4,326 pairs) collapses completely under both NLLB-200 and M2M-100. We interpret the persistent collapse across two backbone families as evidence that architecture selection alone is not the fix, while noting that data volume, quality, domain, and linguistic characteristics likely interact in ways this study cannot fully disentangle.

7 Discussion: Persistent Collapse as a Research Signal

Nagamese, Tagin, and Karbi did not respond to a backbone swap (NLLB → M2M-100) plus FOCUS-style mean-embedding initialization (Dobler and de Melo, 2023), which suggests the failure mode for these three languages is not primarily an architecture or initialization problem – a pattern consistent with long-documented low-resource NMT failure modes (Koehn and Knowles, 2017). We see three distinct signals worth separating out for future work rather than treating "collapse" as a single phenomenon:

Karbi and Tagin (898 and 4,326 training pairs): both are near or below what appears to be a practical viability floor for fine-tuning a pretrained multilingual backbone at all, regardless of which

backbone is chosen. This is consistent with prior work on building parallel corpora for Northeast India’s very low-resource languages, which reports similar data-volume constraints across the region (Tonja et al., 2023). Tagin’s 0.0 BLEU with source-copy-like behavior and Karbi’s low but non-zero score are consistent with a data-volume floor rather than a script or embedding-initialization problem; addressing this likely requires either substantially more monolingual/parallel data, cross-lingual transfer from a linguistically related pivot language, or a fundamentally different low-shot adaptation method (e.g. in-context / retrieval-augmented MT) rather than further backbone fine-tuning.

Nagamese (1,983 training pairs): the more informative case, because Nagamese’s output tends toward Hindi-English code-switched text rather than random noise. Nagamese is itself a contact language / Creole with an English and Assamese lexifier base and a fluid, code-switched register in everyday use. Standard bitext sequence-to-sequence MT implicitly assumes a single consistent target register; if Nagamese speakers’ own text data is itself code-switched, the model may be reproducing an accurate register rather than failing outright, which would not be well captured by BLEU/chrF++ against a single reference style. Dedicated Creole MT efforts such as Kreyòl-MT (Robinson et al., 2024) suggest that Creole and contact languages benefit from treatment as a distinct MT subproblem rather than as ordinary low-resource bitext pairs;

we flag this as a concrete direction for the WMT low-resource Indic community: code-switching-aware modeling and multi-reference or register-aware evaluation for Nagamese, rather than continued backbone search under a standard MT framing.

More broadly, we think persistent collapse across two different backbones and initialization strategies is a stronger and more actionable research signal than a single failed run, and we would encourage other low-resource WMT participants encountering similar collapse to report it explicitly rather than omitting sub-1-BLEU results, since the collapse pattern itself (which languages, at what data volume, in what direction) is informative for the shared task series going forward.

8 Conclusion

We presented KrenTransV0-1, a script- and collapse-aware multi-backbone system for WMT 2026 Low-Resource Indic MT, achieving #1 primary rank on Khasi and Kokborok. We report a clean negative result on decoding-time reranking, and a second, more consequential negative result on architecture-swap-as-collapse-fix for Nagamese, Tagin, and Karbi, which we argue is itself a useful signal: it points toward data-volume floors and code-switching-aware modeling, rather than backbone selection, as the more promising directions for these languages. Code and models are released for reproducibility.

Limitations

Official results show Nagamese, Tagin, and Karbi at or near output collapse despite a backbone swap and improved embedding initialization intended to address exactly this failure mode (Section 6); we do not yet have a validated fix for these three languages. All results are constrained to the officially provided WMT training data; no monolingual back-translation or externally sourced parallel data was used in the final submitted models, which likely leaves performance on the table for languages with usable public monolingual text.

Reproducibility

Code: <https://github.com/BadalNyalang/KrenTransV0-1>

Models: <https://huggingface.co/MWirelabs/krentransv0-1>

Acknowledgments

We thank the WMT 2026 Low-Resource Indic Languages Translation shared task organizers for their work compiling and releasing training and evaluation data for these language pairs.

References

- Konstantin Dobler and Gerard de Melo. 2023. [FOCUS: Effective embedding initialization for monolingual specialization of multilingual models](#). In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13440–13454, Singapore. Association for Computational Linguistics.
- 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, Michael Auli, and Armand Joulin. 2021. [Beyond english-centric multilingual machine translation](#). *Journal of Machine Learning Research*, 22(107):1–48.
- Jay Gala, Pranjal A. Chitale, AK Raghavan, Varun Gumma, Sumanth Doddapaneni, Aswanth Kumar M, Janki Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, Pratyush Kumar, Mitesh M. Khapra, Raj Dabre, and Anoop Kunchukuttan. 2023. [Indic-Trans2: Towards high-quality and accessible machine translation models for all 22 scheduled Indian languages](#). *arXiv preprint arXiv:2305.16307*.
- Philipp Koehn and Rebecca Knowles. 2017. [Six challenges for neural machine translation](#). In *Proceedings of the First Workshop on Neural Machine Translation*, pages 28–39, Vancouver, Canada. Association for Computational Linguistics.
- NLLB Team, Marta R. Costa-jussà, et al. 2022. [No language left behind: Scaling human-centered machine translation](#). *arXiv preprint arXiv:2207.04672*.
- Badal Nyalang. 2026. [NE-BERT: A multilingual language model for nine Northeast Indian languages](#). In *Proceedings of the Second Workshop on Language Models for Low-Resource Languages (LoResLM 2026)*, pages 1–12, Rabat, Morocco. Association for Computational Linguistics.
- Partha Pakray, Reddi Krishna, Santanu Pal, Advaita Vetagiri, Sandeep Dash, Arnab Kumar Maji, Saralin A. Lyngdoh, Lenin Laitonjam, Anupam Jamatia, Koj Sambyo, Ajit Das, and Riyanka Manna. 2025. [Findings of WMT 2025 shared task on low-resource Indic languages translation](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 532–553, Suzhou, China. Association for Computational Linguistics.
- 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 WMT 2026 shared task on low-resource Indic languages translation. In *Proceedings of the Eleventh Workshop on Machine Translation (WMT), EMNLP 2026*, Budapest, Hungary.
- 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. Association for Computational Linguistics.
- Matt Post. 2018. [A call for clarity in reporting BLEU scores](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium. Association for Computational Linguistics.
- Kundeshwar Pundalik, Piyush Sawarkar, Nihar Sahoo, Abhishek Shinde, Prateek Chanda, Vedant Goswami, Ajay Nagpal, Atul Singh, Viraj Thakur, Vijay Dewane, Aamod Thakur, Bhargav Patel, Smita Gautam, Bhagwan Panditi, Shyam Pawar, Madhav Kotcha, Suraj Racha, Saral Sureka, Pankaj Singh, and 3 others. 2025. [PARAM-1 BharatGen 2.9B model](#). *arXiv preprint arXiv:2507.13390*.
- Nathaniel R. Robinson, Raj Dabre, Ammon Shurtz, Rasul Dent, Onenamiyi Onesi, Claire Bizon Monroc, Loïc Grobol, Hasan Muhammad, Ashi Garg, Naome A. Etori, Vijay Murari Tiyyala, Olanrewaju Samuel, Matthew Dean Stutzman, Bismarck Bamfo Odoom, Sanjeev Khudanpur, Stephen D. Richardson, and Kenton Murray. 2024. [Kreyòl-mt: Building mt for latin american, caribbean and colonial african creole languages](#). In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pages 3083–3110, Mexico City, Mexico. Association for Computational Linguistics.
- J. Sivabhavani, Daneshwari Kankanwadi, Abhinav Mishra, and Biswajit Paul. 2025. [ANVITA: A multi-pronged approach for enhancing machine translation of extremely low-resource indian languages](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 1240–1247, Suzhou, China. Association for Computational Linguistics.
- Atnafu Lambebo Tonja, Melkamu Mersha, Ananya Kalita, Olga Kolesnikova, and Jugal Kalita. 2023. [First attempt at building parallel corpora for machine translation of Northeast India’s very low-resource languages](#). In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 534–539, Goa University, Goa, India. NLP Association of India (NLP AI).