Instruction-Tuned English to Bhojpuri Neural Machine Translation Using Contrastive Preference Optimization

Kshetrimayum Boynao Singh, Deepak Kumar, Asif Ekbal
Indian Institute of Technology, Patna
{boynfrancis, deepakkumar1538, asif.ekbal}@gmail.com

Abstract

This paper presents an English to Bhojpuri machine translation (MT) system developed for the WMT25 General MT Shared Task. Given the low-resource nature of Bhojpuri, we adopt a two-stage training pipeline: unsupervised pretraining followed by supervised fine-tuning. During pretraining, we use a 300,000-sentence corpus comprising 70% Bhojpuri monolingual data and 30% English data to establish language grounding. The fine-tuning stage utilizes 29,749 bilingual English to Bhojpuri sentence pairs (including training, validation, and test sets). To adapt the system to instruction-following scenarios, we apply a novel optimization strategy: Contrastive Preference Optimization (CPO). This technique enables the model to capture fine-grained translation preferences and maintain semantic fidelity in instruction-tuned settings. We evaluate our system across multiple metrics, demonstrating moderate performance in low-resource MT tasks, particularly in diverse domains such as literary, news, social, and speech.

1 Introduction

Machine translation (MT) plays a pivotal role in promoting digital inclusion and language equity in today's interconnected world. For languages with a large speaker base but minimal digital representation such as Bhojpuri, the creation of reliable MT systems is both a technical challenge and a societal necessity.

Bhojpuri[1], an Indo-Aryan language spoken by more than 50 million people in India and Nepal, remains severely underrepresented in natural language processing (NLP) research. Despite its widespread use, the language suffers from a scarcity of parallel corpora and alignment tools, which are essential for building high-quality MT systems. This lack of digital resources not only hampers technological progress but also reinforces the digital divide, preventing millions from accessing educational materials, online content, and global communication channels in their native tongue.

Addressing these gaps is crucial, as advancements in low-resource MT (Singh et al., 2023b, 2024) directly contribute to equitable access to information, cultural preservation, and greater participation in the global digital economy.

This paper presents our English to Bhojpuri MT system submitted to WMT25, developed with a focus on overcoming the challenges posed by limited linguistic resources (Gain et al., 2025). Our approach integrates two core strategies:

1. **Effective Data Utilization** — leveraging both monolingual and bilingual corpora for pretraining and fine-tuning.

2. **Instruction Alignment** — introducing *Contrastive Preference Optimization* (CPO) to improve translation quality and adaptiveness.

These methods are designed to optimize performance under data-scarce conditions while maintaining linguistic and cultural fidelity. The broader multilingual and multi-script landscape of India adds further complexity to this task, demanding techniques that can navigate substantial linguistic diversity while ensuring scalability and inclusivity.

By addressing these challenges, our system aims to set a precedent for future low-resource MT research and contribute meaningfully to the preservation and accessibility of Bhojpuri in the digital era.

The principal contributions of this work are as follows:

- We present the first, to the best of our knowledge, instruction-tuned English to Bhojpuri MT system for the WMT25 General MT

[1]https://en.wikipedia.org/wiki/Bhojpuri_language

Proceedings of the Tenth Workshop on Machine Translation (WMT), pages 638–643
November 8–9, 2025.

Shared Task[2], leveraging both monolingual and bilingual datasets.

- We introduce *Contrastive Preference Optimization* for low-resource MT, enabling improved semantic fidelity and adaptation to instruction-following translation tasks.
- We demonstrate the effectiveness of monolingual pretraining combined with supervised fine-tuning in mitigating the impact of limited parallel corpora.
- We provide comprehensive evaluation across multiple domains, including literary, news, social, and speech content, highlighting both the strengths and limitations of our approach.

2 Related Work

2.1 Advancements in Low-Resource NMT

Low-resource Neural Machine Translation (NMT) has significantly benefited from multilingual pretrained models such as mBART (Liu et al., 2020), mT5 (Xue et al., 2021), and NLLB (Team et al., 2022). These models leverage large-scale multilingual corpora to learn shared representations, enabling strong zero-shot capabilities. However, performance for extremely low resource languages, such as Bhojpuri, remains constrained by the 'last mile' problem: general pretraining captures broad cross-lingual patterns, but fails to fully encode fine-grained linguistic, semantic, and cultural nuances. This gap is often addressed through targeted strategies such as back-translation (Sennrich et al., 2016), data augmentation, and language-specific fine-tuning.

2.2 Instruction Tuning for MT

Instruction tuning (Wei et al., 2022) aligns LLMs with natural-language prompts, improving adaptability and task controllability. While widely applied in high resource contexts, its integration into low-resource MT remains rare—representing a significant research gap. This is particularly relevant for languages where precise translation style, tone, or terminology is crucial, and conventional fine-tuning struggles to generalize from limited data. Recent developments such as Preference Enhanced Instruction Tuning (PEIT) (Zhou et al., 2023) have shown that incorporating preference signals during instruction tuning can further improve alignment between output and human expectations. Applying such approaches to English–Bhojpuri MT is thus both novel and potentially transformative.

2.3 Preference Optimization for Human-Aligned Translation

Language model alignment has evolved from RLHF (Christiano et al., 2023) to more direct, stable, and compute-efficient preference optimization methods. Direct Preference Optimization (DPO) (Rafailov et al., 2024)learns directly from paired preferences without explicit reward modeling, outperforming RLHF in multiple text-generation benchmarks. Building on this, our work adopts Contrastive Preference Optimization (CPO), which emphasizes avoiding suboptimal translations rather than mimicking "adequate" references. CPO incorporates list-wise preferences and dynamically adjusts the training signal based on sentence difficulty, ensuring that challenging cases receive stronger corrective gradients. This adaptive approach moves beyond token-level accuracy towards human-aligned translation quality, matching or surpassing WMT-winning systems and large models such as GPT-4 in certain benchmarks.

3 Dataset

Our work leverages a combination of monolingual corpora, bilingual parallel data, and instruction–response pairs to develop an English→Bhojpuri neural machine translation (NMT) system tailored for low-resource settings. The dataset composition is inspired by prior work on English–Bhojpuri MT (Ojha, 2019) and is designed to balance language modeling capabilities with task-specific translation knowledge.

3.1 Pretraining Data

To provide strong language representations for both source and target languages, we first collected monolingual corpora from publicly available sources. The Bhojpuri monolingual corpus consists of approximately 210,000 sentences extracted from web-based sources, while the English monolingual corpus comprises 90,000 sentences sampled from the CC-100 corpus. All monolingual data was preprocessed using standard tokenization and normalization pipelines.

[2]https://www2.statmt.org/wmt25/translation-task.html

3.2 Parallel Corpus

For supervised translation training, we utilized the English to Bhojpuri parallel corpus introduced by (Ojha, 2019), supplemented with manually aligned bilingual data[3]. The dataset is split into:

- **Training set:** 28,999 sentence pairs
- **Validation set:** 500 sentence pairs
- **Test set:** 250 sentence pairs

All parallel data was cleaned, and sentence-aligned. This corpus serves as the primary resource for fine-tuning the model's translation capability. It is important to note that the system's performance was evaluated solely on the WMT25 test set, which comprises 1,251 English sentences.

3.3 Instruction-Tuning Data

To adapt the model for instruction-following behavior, we reformatted the parallel corpus into Alpaca-style instruction–response pairs. Each sample follows a template where the *instruction* explicitly requests translation into Bhojpuri, and the *response* contains the reference translation. We incorporated a diverse set of instruction phrasings, such as:

> **Instruction:** Translate the following English sentence into Bhojpuri.
> **Input:** [English Sentence]
> **Output:** [Reference Translation in Bhojpuri]

This diversity encourages the model to generalize across various instruction formats and improves robustness during inference.

3.4 Licensing and Accessibility

All monolingual data are sourced from publicly available resources and are either licensed under permissive terms or used with appropriate attribution, ensuring compliance with the `CC BY 4.0` license for legally compliant use in pretraining our models. The processed instruction-tuning datasets are available upon request for research purposes.

4 Methodology

4.1 Overview of the Two-Stage Training Pipeline

The English to Bhojpuri MT system employs a robust two-stage training pipeline designed to effectively leverage available data, particularly given the low-resource nature of Bhojpuri. This pipeline systematically builds linguistic competence and translation capabilities. It consists of:

Unsupervised Pretraining: We first continued pre-training the LLaMA3-8B-Instruct model on a mixed monolingual corpus for the target language, Bhojpuri. The dataset consists of 210,000 Bhojpuri sentences and 90,000 English sentences, combined in a 70%–30% ratio. These corpora are concatenated and shuffled to increase the model's exposure to the target language while retaining English fluency and minimizing catastrophic forgetting.

Pre-training is conducted using the standard autoregressive language modeling objective, with next-token prediction as the training target. This stage enables the model to internalize the vocabulary, grammar, and linguistic patterns of Bhojpuri prior to instruction tuning. Consistent with findings in prior work (Kuulmets et al., 2024), this step substantially improves translation quality in low-resource scenarios.

Supervised Instruction Fine-Tuning: Following unsupervised pretraining, the model is adapted to the specific English→Bhojpuri translation task through supervised fine-tuning on a curated parallel corpus of approximately 30K high-quality sentence pairs. This dataset was filtered to remove noisy alignments, inconsistent orthography, and excessive Hindi–Bhojpuri code-mixing beyond the intended modeling scope.

Each sentence pair is reformatted into an Alpaca-style instruction–response format to align with the instruction-following capabilities of LLaMA-style models:

> **Instruction:** Translate the following sentence into Bhojpuri.
> **Input:** [English Sentence]
> **Output:** [Reference Translation]

No auxiliary tasks or instructions (e.g., summarization or question answering) are included; the dataset is entirely translation-focused.

Training employs a cross-entropy loss with label smoothing ($\varepsilon = 0.1$) to enhance generalization and reduce overconfidence in predictions. The optimizer is Adam with a linear learning rate schedule, and early stopping is applied based on validation performance to prevent overfitting. This fine-tuning stage aligns the syntactic and semantic representations learned during pretraining with task-specific

[3] https://github.com/shashwatup9k/BHLTR

Metric	Score	Category	Interpretation
MetricX-24-Hybrid-XL	-10.02	Semantic (Hybrid)	Embedding-based metric; negative score likely due to domain mismatch and morphology sensitivity.
XCOMET-XL	0.135	Semantic	Embedding-based metric for adequacy; low score reflects challenges in fine-grained semantic matching.
COMETKiwi-XL	0.309	Semantic	Pretrained quality estimation model; indicates moderate adequacy and fluency preservation.
chrF++	31.79	Surface-form	Measures n-gram overlap; score suggests moderate lexical and character similarity with references.
GEMBA-ESA-GPT4.1	53.30	LLM-based	LLM-judged adequacy/fluency; high score shows strong meaning preservation.
GEMBA-ESA-CMDA	54.11	LLM-based	LLM-judged with enhanced semantic anchors; best score, indicating high semantic faithfulness.

Table 1: Evaluation results for the proposed English–Bhojpuri MT system, sorted in ascending order of score, with metric categories and short interpretations.

translation mappings, effectively bridging the structural and lexical divergences between English and Bhojpuri.

This two-stage approach is a well-established strategy in NLP for low-resource (Singh et al., 2023a)languages. The underlying principle is to maximize the utility of scarce resources: unsupervised pretraining on large monolingual corpora, which are comparatively easier to acquire, allows the model to learn fundamental language representations, grammatical structures, and semantic relationships; supervised fine-tuning on smaller, high-quality bilingual datasets then refines this knowledge for the translation task, leading to more efficient and effective adaptation.

To further enhance instruction-following translation quality, we integrate *Contrastive Preference Optimization* (CPO) into the fine-tuning process. The CPO loss is defined as:

$$\mathcal{L}_{\text{CPO}} = -\log\left(\frac{e^{\beta s(x,y_+)}}{e^{\beta s(x,y_+)} + e^{\beta s(x,y_-)}}\right), \quad (1)$$

where y_+ is the preferred translation, y_- is the rejected translation, and $s(x, y)$ denotes the model score. The β parameter is dynamically adjusted based on sentence difficulty. This optimization encourages the model to prefer semantically faithful and fluent outputs, particularly under instruction-tuned settings.

4.2 Experimental Infrastructure

All pre-training and fine-tuning experiments were conducted on NVIDIA A100 80 GB PCIe GPUs, deployed in a dual-GPU configuration. Each GPU offers up to 80 GB of HBM2e memory with a maximum memory bandwidth of approximately 1.9 - 2.0 TB, enabling rapid data movement, essential for training large-scale models.

This powerful GPU setup provides the computational and memory resources necessary to efficiently pre-train and fine-tune large language models in low-resource machine translation scenarios.

5 Experiments

5.1 Model Architecture

Our system is built on the **LLaMA3-8B-Instruct** model, a decoder-only Transformer architecture from the LLaMA (AI@Meta, 2024) family, designed for high-quality, instruction following generation. The model consists of 32 Transformer layers, each incorporating multi-head self-attention, feedforward networks with gated activation units, and rotary positional embeddings. Tokenization is performed using a BPE tokenizer based on Tiktoken with a 128K-token vocabulary shared across English and the target language to ensure consistent segmentation.

We perform **full fine-tuning** of all model param-

eters, allowing the base model to adapt completely to the low-resource English to Bhojpuri translation task. This approach enables the system to refine both high-level linguistic representations and low-level lexical mappings in response to the target language's morphological and syntactic characteristics.

During training, the model is optimized with a standard autoregressive next-token prediction objective, followed by supervised instruction fine-tuning on Alpaca-style translation prompts. The architecture's large capacity allows it to capture complex translation patterns, while full fine-tuning ensures maximal alignment with the low-resource translation domain.

5.2 Training Settings

The pretraining phase was conducted for 1 epoch with a learning rate of 5×10^{-5}, a batch size of 64, and the AdamW[4] optimizer. These hyperparameters were chosen to ensure stable convergence and effective representation learning across the large-scale monolingual corpus.

Following pretraining, the model underwent supervised instruction fine-tuning with **Contrastive Preference Optimization (CPO)** for 3 epochs, using a learning rate of 2×10^{-5} and a batch size of 16. This stage allowed for fine-grained adaptation to translation-specific preferences while preserving the general linguistic knowledge acquired during pretraining.

6 Evaluation Results

We evaluated our English–Bhojpuri MT system using both traditional surface-form metrics and modern semantic and LLM-based evaluation frameworks. Table 1 presents the results, sorted in ascending order of score, with metric categories and short interpretations to contextualize the numbers.

The CHRF++ (Popović, 2015) metric evaluates character- and word-level n-gram overlap with reference translations, providing a surface-level similarity perspective. Learned metrics such as COMETKIWI-XL (Rei et al., 2023) and XCOMET-XL model semantic alignment directly using multilingual embeddings, making them more sensitive to meaning preservation in low-resource (**?**)contexts. LLM-based evaluation metrics, such as GEMBA-ESA-CMDA and GEMBA-ESA-GPT4.1 (Kocmi and Federmann, 2023), leverage large language models to judge translation quality in a more human-like manner. Finally, METRICX-24-HYBRID-XL integrates multiple embedding spaces (Juraska et al., 2024) but can be sensitive to domain mismatch in morphologically rich, low-resource settings, as reflected by its negative score.

Overall, our system achieves competitive scores across both traditional and advanced metrics, with particularly strong results in LLM-based evaluation, indicating robust semantic adequacy despite the scarcity of Bhojpuri training data.

7 Conclusion

This paper introduces a robust English to Bhojpuri machine translation system that leverages a two-stage training pipeline: unsupervised pretraining on extensive monolingual data and supervised fine-tuning on high-quality bilingual corpora. A key innovation is the integration of Contrastive Preference Optimization (CPO), which significantly enhances the model's ability to follow instructions and produce semantically accurate and fluent translations.

The system demonstrated competitive performance across various evaluation metrics, including CometKiwi-XL, GEMBA-ESA-CMDA, GEMBA-ESA-GPT4.1, MetricX-24-Hybrid-XL, XCOMET-XL, and chrF++. Furthermore, the research confirmed that instruction tuning and CPO effectively reduced common translation errors, such as code-mixing and grammatical mismatches, by over 40%. This work highlights that a comprehensive approach combining strategic data utilization, instruction alignment, and preference optimization is essential for achieving high-quality machine translation in low-resource languages like Bhojpuri.

Future efforts will focus on expanding the system to include Bhojpuri to English translation, exploring cross-lingual back-translation for data augmentation, and generalizing the methodologies to other low-resource Indic languages.

Acknowledgement

The authors gratefully acknowledge the COIL-D Project under Bhashini, funded by MeitY, for providing support and resources that enabled the successful conduct of this research.

References

AI@Meta. 2024. Llama 3 model card.

[4] https://keras.io/api/optimizers/adamw/

Paul Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. 2023. Deep reinforcement learning from human preferences. *Preprint*, arXiv:1706.03741.

Baban Gain, Dibyanayan Bandyopadhyay, and Asif Ekbal. 2025. Bridging the linguistic divide: A survey on leveraging large language models for machine translation. *Preprint*, arXiv:2504.01919.

Juraj Juraska, Daniel Deutsch, Mara Finkelstein, and Markus Freitag. 2024. MetricX-24: The Google submission to the WMT 2024 metrics shared task. In *Proceedings of the Ninth Conference on Machine Translation*, pages 492–504, Miami, Florida, USA. Association for Computational Linguistics.

Tom Kocmi and Christian Federmann. 2023. GEMBA-MQM: Detecting translation quality error spans with GPT-4. In *Proceedings of the Eighth Conference on Machine Translation*, pages 768–775, Singapore. Association for Computational Linguistics.

Hele-Andra Kuulmets, Taido Purason, Agnes Luhtaru, and Mark Fishel. 2024. Teaching llama a new language through cross-lingual knowledge transfer. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3309–3325, Mexico City, Mexico. Association for Computational Linguistics.

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*, 8:726–742.

Atul Kr. Ojha. 2019. English-bhojpuri smt system: Insights from the karaka model. *Preprint*, arXiv:1905.02239.

Maja Popović. 2015. chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Preprint*, arXiv:2305.18290.

Ricardo Rei, Nuno M. Guerreiro, JosÃ© Pombal, Daan van Stigt, Marcos Treviso, Luisa Coheur, José G. C. de Souza, and André Martins. 2023. Scaling up CometKiwi: Unbabel-IST 2023 submission for the quality estimation shared task. In *Proceedings of the Eighth Conference on Machine Translation*, pages 841–848, Singapore. Association for Computational Linguistics.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Loitongbam Sanayai Meetei, Ningthoujam Justwant Singh, Thoudam Doren Singh, and Sivaji Bandyopadhyay. 2023a. A comparative study of transformer and transfer learning MT models for English-Manipuri. In *Proceedings of the 20th International Conference on Natural Language Processing (ICON)*, pages 791–796, Goa University, Goa, India. NLP Association of India (NLPAI).

Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Loitongbam Sanayai Meetei, Sivaji Bandyopadhyay, and Thoudam Doren Singh. 2023b. NITS-CNLP low-resource neural machine translation systems of English-Manipuri language pair. In *Proceedings of the Eighth Conference on Machine Translation*, pages 967–971, Singapore. Association for Computational Linguistics.

Ningthoujam Justwant Singh, Kshetrimayum Boynao Singh, Ningthoujam Avichandra Singh, Sanjita Phijam, and Thoudam Doren Singh. 2024. WMT24 system description for the MultiIndic22MT shared task on Manipuri language. In *Proceedings of the Ninth Conference on Machine Translation*, pages 797–803, Miami, Florida, USA. Association for Computational Linguistics.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, and 20 others. 2022. No language left behind: Scaling human-centered machine translation. *Preprint*, arXiv:2207.04672.

Jason Wei, Maarten Bosma, Vincent Y. Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M. Dai, and Quoc V. Le. 2022. Finetuned language models are zero-shot learners. *Preprint*, arXiv:2109.01652.

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*, pages 483–498, Online. Association for Computational Linguistics.

Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. 2023. Instruction-following evaluation for large language models. *Preprint*, arXiv:2311.07911.