

Jiutian-MT-8B at WMT26: Faithfulness-First Training for Chinese–Southeast Asian Translation

Meng Su Jian Liu Youxuan Fang

China Mobile Jiutian Artificial Intelligence Technology (Beijing) Co., Ltd.
{sumeng,liujian,fangyouxuan}@chinamobile.com

Abstract

This paper describes Jiutian-MT-8B, our submission to the WMT 2026 Chinese–Southeast Asian Multilingual Machine Translation shared task. The system adapts the 8-billion-parameter Jiutian-8B-Instruct checkpoint to fourteen translation directions between Chinese and Thai, Vietnamese, Lao, Burmese, Khmer, Indonesian, and Malay. We formulate translation as a tightly controlled instruction-following task and use a four-stage full-parameter training recipe: (i) supervised multilingual adaptation on released parallel data, (ii) faithfulness-first confidence-gated multi-pair preference alignment (F2MTA), (iii) direction-coupled, evidence-filtered self-training from task monolingual data, and (iv) a final clean-data calibration stage. The preference stage uses a partial order: a candidate must gain source adequacy without materially sacrificing target fluency, rather than merely win a scalarized quality score. The recipe directly updates all Jiutian-8B parameters with sharded bf16 training while retaining authentic data anchors to limit synthetic-data drift. We evaluate the submitted system on the WMT26 test set and obtain a macro-average of 35.34 sacreBLEU and 82.68 COMET across fourteen directions.

1 Introduction

The WMT 2026 Chinese–Southeast Asian multilingual MT task evaluates efficient, bidirectional translation between Chinese and seven Southeast Asian languages: Thai (TH), Vietnamese (VI), Lao (LO), Burmese (MY), Khmer (KM), Indonesian (ID), and Malay (MS) (WMT 2026 Chinese–Southeast Asian Task Organizers, 2026). It explicitly values both translation quality and throughput, and limits submitted systems to 20B total parameters. These conditions make multilingual adaptation of a compact instruction model an appealing alternative to training a separate model for every direction.

We build Jiutian-MT-8B on Jiutian-8B-Instruct, introduced by Hao et al. (2025). Although that

work targets mathematical reasoning, its multi-stage optimization perspective is useful here: different data sources serve different objectives and should not be mixed indiscriminately. Our adaptation is deliberately translation-specific. It does *not* use chain-of-thought or reinforcement learning; instead, it progressively moves from clean parallel supervision to filtered monolingual augmentation and back to clean, reference-style translation examples.

Our design has three goals: (1) one model covers all fourteen directions; (2) its decoder produces only the requested translation, not explanations or chat markup; (3) monolingual augmentation helps without allowing noisy synthetic examples to dominate genuine bitext. The contributions of this paper are:

- a direction-explicit multilingual instruction format for adapting Jiutian-8B to Chinese–Southeast Asian translation;
- a full-parameter four-stage training recipe that combines parallel data, preference alignment, and task-provided monolingual text;
- F2MTA, a confidence-gated, faithfulness-first partial-order multi-pair objective that exposes partial omissions without rewarding merely fluent targets;
- a macro-averaged test-set evaluation over all fourteen directions using sacreBLEU and COMET under the shared-task evaluation protocol.

2 Related Work

Back-translation converts target-language monolingual data into synthetic parallel examples and is a standard way to exploit sparse bitext (Sennrich et al., 2016). We use its basic idea but filter pseudo-pairs through bidirectional agreement

before adding them to training. This avoids depending on unfiltered model outputs in the lowest-resource directions.

Large language models can be adapted to translation through controlled instruction templates, but their conversational prior makes target-only decoding essential. ALMA shows that staged language-model adaptation can substantially improve LLM translation (Xu et al., 2024a). Contrastive preference optimization (CPO) uses preferences to move beyond pure reference-based SFT (Xu et al., 2024b). More recently, M2PO identifies a failure mode of QE-driven preference data—fluent but partial translations—and uses separate fidelity and fluency signals together with multiple ranked pairs (Wang et al., 2026). We incorporate these two principles in F2MTA while using a confidence-gated partial order: fidelity determines whether a preference constraint is admissible, rather than being traded off blindly against fluency in a single scalar ranking. The same bidirectional evidence is then reused to filter pseudo-bitext. We use full-parameter, rather than adapter-only, optimization. Finally, Hao et al. (2025) demonstrate a multi-stage curriculum for specializing an 8B model; we adapt that high-level idea to MT data quality and directionality rather than mathematical reasoning.

3 Proposed Framework

3.1 Task Formulation and Translation Interface

Our backbone is Jiutian-8B-Instruct (Hao et al., 2025). Each parallel example is turned into one instruction-response instance:

```
Translate the following text from
<SOURCE_LANG> to <TARGET_LANG>. Return
only the translation.
Source: <SOURCE_TEXT>
Translation:
```

Language names are written in English in the prompt and are accompanied by ISO codes in an internal direction field. We use one canonical prompt at training and inference time; prompt variants are not treated as extra data. The target is the reference sentence only. This simple contract separates translation from explanation, transliteration, and language identification, three common failure modes of chat-tuned models.

At decoding time, we use deterministic beam search with a fixed prompt, temperature 0, and no sampling.

3.2 Four-Stage Full-Parameter Adaptation

Stage 1: clean bidirectional supervised warm-up. We first update *all* Jiutian-8B parameters on the cleaned authentic parallel corpus in both directions. This stage establishes strict source/target control before synthetic data or preference signals are introduced. Each language-direction example is assigned a sampling weight

$$p(d, k) \propto (n_{d,k} + \epsilon)^\alpha, \quad (1)$$

where d is a direction, k is a domain, $n_{d,k}$ is its retained number of examples, $\epsilon > 0$ prevents zero mass, and α smooths imbalances caused by filtering. The smoothing and per-domain cap are fixed for all stages. All directions share one model and one optimizer, but are balanced at the minibatch level. A frozen copy of this stage-1 checkpoint, π_{ref} , is retained as the reference policy for the preference stage.

Stage 2: Faithfulness-First Multi-Pair Translation Alignment (F2MTA). Pure SFT treats the human reference as the only learning signal even when the model produces several plausible but imperfect alternatives. Following CPO (Xu et al., 2024b) and M2PO (Wang et al., 2026), we generate a candidate list $\mathcal{Y}(x) = \{y, \hat{y}_1, \dots, \hat{y}_K\}$ for each authentic source x , where y is the human reference. Candidates are produced by diverse decoding from the frozen stage-1 checkpoint π_{ref} .

The central design choice in F2MTA is to avoid a fixed weighted sum of faithfulness and fluency. Such a sum can rank a fluent omission above a less polished but complete translation. For every candidate y_i , we compute a source-adequacy score a_i , a bidirectional-consistency score b_i , and a target-fluency score g_i . The first is produced by a source-conditioned quality estimator; the second compares a reverse translation of y_i with x ; and the third combines target-language identity, normalized target-side likelihood, and well-formedness checks. All scorers are frozen before stage 2 and their thresholds are fixed before stage 2. We define faithfulness conservatively as $f_i = \min(a_i, b_i)$: a candidate is not treated as adequate when either direct source evidence or reverse evidence is weak.

F2MTA then constructs a confidence-gated partial order rather than a total ranking. Let $v_i \in \{0, 1\}$ denote a hard language, script, and length-validity gate. We retain a preference pair only when the preferred candidate has a meaningful faithful-

ness advantage and does not incur a material fluency loss:

$$\mathcal{P}(x) = \{(i, j) : v_i = 1, \quad f_i - f_j \geq \delta_F, \quad g_i \geq g_j - \delta_G\}. \quad (2)$$

The two margins δ_F and δ_G suppress near ties and make the asymmetry explicit: fidelity creates a constraint, whereas fluency prevents the model from learning visibly degraded targets. The human reference remains in every candidate set and is additionally protected by the SFT anchor below.

Instead of retaining one best-worst comparison, F2MTA uses all admissible pairs in $\mathcal{P}(x)$. Their weights w_{ij} increase with the clipped faithfulness margin, so clear omissions receive more emphasis than arbitrary micro-rankings. The reference-anchored preference and SFT objectives are jointly optimized as

$$\begin{aligned} \Delta_{ij} &= \log \pi_\theta(y_i|x) - \log \pi_\theta(y_j|x) \\ &\quad - \log \pi_{\text{ref}}(y_i|x) + \log \pi_{\text{ref}}(y_j|x), \\ \mathcal{L}_{\text{stage2}}(x) &= - \sum_{(i,j) \in \mathcal{P}(x)} w_{ij} \log \sigma(\beta \Delta_{ij}) \\ &\quad + \lambda_{\text{sft}} \mathcal{L}_{\text{NLL}}(x, y). \end{aligned} \quad (3)$$

This gives partially incomplete outputs a clear negative role without forcing the model to imitate any one noisy candidate ordering.

Stage 3: filtered bidirectional self-training. Stage 3 reuses the evidence decomposition from F2MTA rather than treating all generated bitext as equally reliable. For every monolingual sentence y in a Southeast Asian language, the stage-2 reverse model generates a Chinese candidate $\hat{x}$ and then reconstructs y from $\hat{x}$. The pseudo-pair $(\hat{x}, y)$ is considered jointly for Chinese→SEA and SEA→Chinese training, but is retained only when it passes both a format gate and an evidence gate:

1. output-script and language checks agree with the requested target;
2. normalized source and generated text are nonempty and satisfy the same length-ratio rule as authentic bitext;
3. direct adequacy and reconstructed-source consistency both exceed per-direction calibrated thresholds; and
4. target fluency is acceptable in both orientations of the pair.

Track	Directions
ZH → SEA	ZH→TH, ZH→VI, ZH→LO, ZH→MY, ZH→KM, ZH→ID, ZH→MS
SEA → ZH	TH→ZH, VI→ZH, LO→ZH, MY→ZH, KM→ZH, ID→ZH, MS→ZH

Table 1: The fourteen directions covered by Jiutian-MT-8B. SEA denotes the seven Southeast Asian languages in the WMT26 task.

This direction-coupled acceptance rule is important for the shared task: it prevents an easy-to-generate Chinese side from licensing a pseudo-pair whose Southeast Asian realization is incomplete or in the wrong script. The model is then continued on a mixture of authentic and retained synthetic examples, with a cap that keeps authentic bitext dominant.

Stage 4: clean-data calibration. Finally, we train for a short pass on authentic parallel data only, preserving the same direction and domain mixture as stage 1. This stage reduces overfitting to the style or occasional artifacts of synthetic translations. Early stopping is based on the mean of direction-level validation quality scores; a checkpoint that regresses on any language by more than a preset margin is not selected without manual error analysis.

3.3 Full-Parameter Optimization

All Jiutian-8B parameters are trainable in every stage. We use bf16 arithmetic, activation checkpointing, and FSDP parameter/gradient/optimizer-state sharding to make full-parameter training feasible without changing the submitted 8B architecture. The SFT anchor in stage 2 and clean-data calibration in stage 4 limit preference- and synthetic-data drift while every parameter remains trainable.

4 Experiments

4.1 Task, Data, and Evaluation Setting

The shared task contains Chinese→Southeast Asian and Southeast Asian→Chinese translation tracks. We train one model for all directions and encode the source and target language in every training prompt. Table 1 lists the covered directions.

We use the task training and validation resources released by the organizers. The task description reports 20K manually aligned parallel training pairs per Southeast Asian language (140K total), 100K

- Junlan Feng. 2025. [Jt-math: A multi-stage framework for advanced mathematical reasoning in large language models](#). *arXiv preprint arXiv:2507.19748*.
- Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Improving neural machine translation models with monolingual data. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 86–96.
- Hao Wang, Linlong Xu, Heng Liu, Yangyang Liu, Xiaohu Zhao, Bo Zeng, Liangying Shao, Yichen Dong, Xinwei Wu, Jiang Zhou, Tianyu Dong, Xiangxiang Zeng, Longyue Wang, and Weihua Luo. 2026. [M2PO: Multi-perspective multi-pair preference optimization for machine translation](#). In *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10315–10336, San Diego, California, United States. Association for Computational Linguistics.
- WMT 2026 Chinese–Southeast Asian Task Organizers. 2026. Wmt 2026 chinese–southeast asian multilingual machine translation task. <https://www2.statmt.org/wmt26/chinese-southeast-asian.html>. Accessed: 2026-08-01.
- Haoran Xu, Young Jin Kim, Amr Mohamed Nabil Aly Aly Sharaf, and Hany Awadalla. 2024a. A paradigm shift in machine translation: Boosting translation performance of large language models. In *International Conference on Learning Representations*.
- Haoran Xu, Amr Sharaf, Yunmo Chen, Weiting Tan, Lingfeng Shen, Benjamin Van Durme, Kenton Murray, and Young Jin Kim. 2024b. [Contrastive preference optimization: Pushing the boundaries of LLM performance in machine translation](#). *arXiv preprint arXiv:2401.08417*.