Shy-hunyuan-MT at WMT25 General Machine Translation Shared Task

Mao Zheng, Zheng Li, Yang Du, Bingxin Qu, Mingyang Song

Tencent Hunyuan
moonzheng@tencent.com
https://github.com/Tencent-Hunyuan/Hunyuan-MT

Abstract

In this paper, we present our submission to the WMT25 shared task on machine translation, for which we propose **S**ynergy-en**h**anced polic**y** optimization framework, named **Shy**. This novel two-phase training framework synergistically combines knowledge distillation and fusion via reinforcement learning. In the first phase, we introduce a multi-stage training framework that harnesses the complementary strengths of multiple state-of-the-art large language models to generate diverse, high-quality translation candidates. These candidates serve as pseudo-references to guide the supervised fine-tuning of our model, Hunyuan-7B, effectively distilling the collective knowledge of multiple expert systems into a single efficient model. In the second phase, we further refine the distilled model through Group Relative Policy Optimization, a reinforcement learning technique that employs a composite reward function. By calculating reward from multiple perspectives, our model ensures better alignment with human preferences and evaluation metrics. Extensive experiments across multiple language pairs demonstrate that our model **Shy-hunyuan-MT** yields substantial improvements in translation quality compared to baselines. Notably, our framework achieves competitive performance comparable to that of state-of-the-art systems while maintaining computational efficiency through knowledge distillation and fusion.

1 Introduction

The field of machine translation has witnessed remarkable progress with the emergence of large language models; yet, challenges remain in consistently producing human-quality translations. This work addresses two key limitations: (1) the over-reliance on single-model supervision during fine-tuning, and (2) the difficulty of aligning machine outputs with nuanced human judgments during reinforcement learning.

Our method comprises three main phases. First, we collect diverse translations from state-of-the-art open-source large language models, including DeepSeek-V3 (DeepSeek-AI, 2024), DeepSeek-R1 (Guo et al., 2025), and Gemma across multiple language pairs. This ensemble approach provides a richer training signal than single-model distillation, exposing our base model Hunyuan-7B to varied translation strategies and stylistic choices. The collected translation outputs are carefully filtered and normalized before serving as supervision targets. Second, we perform Supervised Fine-Tuning (SFT) on Hunyuan-7B using the prior collected translation dataset. Crucially, we implement a dynamic weighting scheme that prioritizes higher-quality translations during SFT training, as determined by automatic metrics. Specifically, this phase enables the model to internalize the strengths of each contributor model while maintaining its own linguistic identity. The third phase applies Group Relative Policy Optimization (Shao et al., 2024), a sample-efficient Reinforcement Learning (RL) algorithm, to further refine the model. We employ XCOMET for its strong correlation with human judgments and DeepSeek-V3 for its complementary strengths in fluency assessment. The reward function combines these signals with a KL-divergence term to prevent excessive deviation from the SFT model. During RL training, we maintain multiple policy groups that explore different translation strategies, with periodic selection pressure favoring the approaches that yield the highest rewards.

Our methodology incorporates several primary techniques to ensure robust performance. Specifically, we employ a temperature-scaled sampling strategy during preference data collection to optimize the trade-off between diversity and quality. For the SFT phase, a layer-wise learning rate decay is applied to enhance model adaptation. During the RL phase, we implement a dynamic reward normalization scheme to maintain training stability.

Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 160–166
November 8–9, 2025.

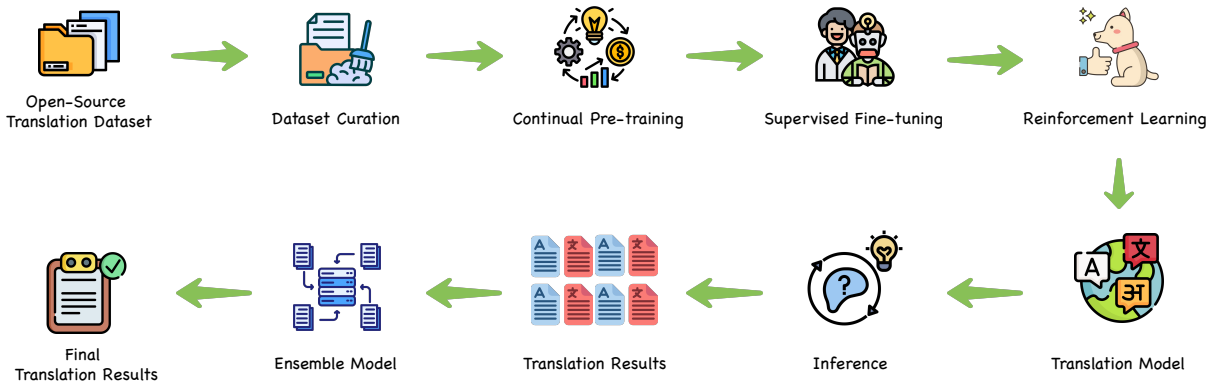

Figure 1: **Framework of the Shy-hunyuan-MT**. Firstly, we leverage an open-source translation dataset to conduct continuous pre-training on the Hunyuan-7B model. Secondly, we utilize the collected WMT translation data to perform SFT on the continuously pre-trained model. Thirdly, we sample outputs from the translation model to generate a set of diverse translation results, which are then used to train and derive the final translation result.

Model performance is rigorously evaluated on the WMT25 test sets using both automatic metrics and human assessment. Comprehensive ablation studies are conducted to validate the efficacy of each proposed component. The primary contributions are threefold:

- A novel and practical framework for the refinement of machine translation models.
- Compelling empirical evidence demonstrating the effectiveness of RL in the domain of machine translation.
- Valuable insights into the application and dynamics of multi-reward RL for complex text generation tasks.

2 Related Work

This section reviews the key research strands underlying our approach: (1) large language models for machine translation, (2) continual pre-training strategies, (3) supervised fine-tuning for translation tasks, and (4) reinforcement learning optimization for text generation.

2.1 Large Language Models for MT

The success of dense decoder-only models, such as GPT-3 (Brown et al., 2020), has revolutionized neural machine translation (NMT). Recent work has demonstrated that 7B-parameter models, such as Hunyuan, can achieve competitive performance when properly adapted. Unlike traditional encoder-decoder architectures (Vaswani et al., 2017), monolithic LLMs process translation as conditional text generation, offering advantages in zero-shot capability and multilingual transfer (Kirstain et al., 2021). Our work extends this direction by systematically investigating continual pre-training strategies for domain adaptation in translation tasks.

2.2 Continual Pre-training for MT

Domain adaptation through continual pre-training has shown promise in recent MT research. Gururangan et al. (2020) established that targeted pre-training on in-domain corpora improves downstream task performance. For WMT competitions specifically, Liao et al. (2021) demonstrates the effectiveness of iterative pre-training on parallel corpora. Our approach differs in that it employs a two-phase adaptation: first, on general bilingual corpora, and then on WMT historical data. This hierarchical adaptation strategy aligns with findings from (Pfeiffer et al., 2020) about the importance of gradual domain specialization.

2.3 Supervised Fine-tuning for MT

The transition from pre-training to task-specific fine-tuning remains a research area of active interest. Raffel et al. (2020) demonstrates that controlled fine-tuning with progressive data augmentation can prevent catastrophic forgetting. Our SFT protocol incorporates three key innovations: (1) dynamic batch sampling based on sentence complexity metrics, (2) gradient accumulation for low-frequency language pairs, and (3) temperature-annealed decoding during training. These techniques build upon curriculum learning principles first proposed by Bengio et al. (2009), but with specific adaptations for translation tasks.

2.4 Reinforcement Learning for MT

Recent advances, such as GRPO (Shao et al., 2024), provide more stable RL optimization for MT and various tasks (Guo et al., 2025; Song et al., 2025; Yang et al., 2025; Li et al., 2025; Zheng et al., 2025). To obtain better rewards, we use a combination of different metrics as the reward, and this multi-metric approach addresses limitations identified by Freitag et al. (2022) regarding single-metric optimization. The ensemble strategy further builds on the diversity-promoting techniques from Vijayakumar et al. (2016), but with novel modifications for temperature-controlled output variation.

2.5 WMT Competition Innovations

Analysis of prior WMT winning systems reveals evolving trends. The 2021 Edinburgh system (Chen et al., 2021) pioneered the use of large-scale back-translation, while Guu et al. (2020) demonstrates the effectiveness of retrieval-augmented models. Our work contributes to this lineage by showing how to effectively combine continual pre-training, reinforced fine-tuning, and learned ensemble strategies within a single LLM framework, addressing the scalability challenges noted by Tom et al. (2023) in their WMT 2023 overview.

3 Methodology

Our approach consists of two major phases: (1) a three-stage training pipeline for developing a high-quality base translation model, and (2) an ensemble strategy that leverages diversity generation and reinforcement learning to produce final translations. The overall architecture is illustrated in Figure 1.

3.1 Continual Pre-training

We leverage Hunyuan-7B, a state-of-the-art multilingual dense foundation model, as our initialization checkpoint. To effectively adapt this general-purpose model for machine translation tasks, we perform domain-adaptive continual pre-training using diverse large-scale parallel and monolingual corpora. Our training corpus encompasses multiple data sources with complementary characteristics:

- **OPUS Collection** (Tiedemann, 2012): A comprehensive multilingual parallel corpus covering over 20 language pairs across diverse domains, providing broad linguistic coverage
- **ParaCrawl** (Buck and Koehn, 2016): Large-scale web-crawled parallel data offering extensive real-world language usage patterns
- **UN Parallel Corpus** (Ziemski et al., 2016): High-quality formal documents ensuring exposure to professional and diplomatic language registers
- **C4** (Raffel et al., 2020): Cleaned English text derived from Common Crawl, contributing to robust monolingual understanding
- **WikiText** (Merity et al., 2016): Encyclopedic articles providing well-structured, factually accurate content

This diverse mixture enables our model to acquire comprehensive translation capabilities across various domains, registers, and language pairs, while maintaining the strong multilingual representations inherited from the base model.

3.2 Supervised Fine-tuning

Following domain adaptation, we conduct supervised fine-tuning using WMT benchmark datasets spanning from 2015 to 2024. Our training method incorporates several regularization techniques to mitigate catastrophic forgetting while preserving the model's acquired capabilities. Then, we optimize the standard sequence-to-sequence cross-entropy objective:

$$\mathcal{L}_{\text{SFT}} = -\sum_{t=1}^{T} \log p(y_t | y_{<t}, x; \theta) \tag{1}$$

where x denotes the source sentence, y represents the target translation, and θ encompasses the model parameters. Our training configuration employs the following strategies:

- **Learning rate scheduling**: We implement linear warmup over 5% of total training steps to ensure stable optimization dynamics.
- **Gradient regularization**: We apply gradient clipping with a maximum norm of 1.0 to prevent gradient explosion.
- **Computational efficiency**: We utilize mixed-precision training with BF16 representation to accelerate training while maintaining numerical stability.

This fine-tuning protocol enables effective task-specific adaptation while maintaining the robustness gained through domain pre-training.

3.3 Reinforcement Learning

After SFT, we apply GRPO, a sample-efficient reinforcement learning algorithm. We design the reward function by combining three metrics:

$$r = w_1 \cdot \text{BLEU} + w_2 \cdot \text{XCOMET} + w_3 \cdot \text{DeepSeek} \quad (2)$$

where $w_1 = 0.2$, $w_2 = 0.4$, and $w_3 = 0.4$ are empirically determined weights. Then we optimize the GRPO objective as follows:

$$\begin{aligned} \mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{q\sim P(Q),\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(O|q)} \\ \frac{1}{G}\sum_{i=1}^{G}\left(\min\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{\text{old}}}(o_i|q)}A_i,\right.\right. \\ \left.\left.\text{clip}\left(\frac{\pi_\theta(o_i|q)}{\pi_{\theta_{\text{old}}}(o_i|q)}, 1-\epsilon, 1+\epsilon\right)A_i\right) - \\ \beta\mathbb{D}_{\text{KL}}\left(\pi_\theta \parallel \pi_{\text{ref}}\right)\right), \end{aligned} \quad (3)$$

$$\mathbb{D}_{\text{KL}}\left(\pi_\theta||\pi_{ref}\right) = \frac{\pi_{ref}(o_i|q)}{\pi_\theta(o_i|q)} - \log\frac{\pi_{ref}(o_i|q)}{\pi_\theta(o_i|q)} - 1, \quad (4)$$

where ϵ and β are hyper-parameters, and A_i is the advantage, computed using a group of rewards $\{r_1, r_2, \ldots, r_G\}$ corresponding to the outputs within each group:

$$A_i = \frac{r_i - mean(\{r_1, r_2, \cdots, r_G\})}{std(\{r_1, r_2, \cdots, r_G\})}. \quad (5)$$

where A_i is the advantage function estimated using group-normalized rewards. To improve the translation quality of terminology and low-resource languages, we also use the methods proposed in TAT-R1 (Li et al., 2025), SSR-Zero (Yang et al., 2025), and Hunyuan-MT-7B (Zheng et al., 2025).

3.4 Synergy-based Policy Optimization

To further improve translation quality, we develop a two-phase fusion approach.

3.4.1 Diversity Generation

For each input sentence x, we generate $N = 5$ candidate translations by varying:

- Temperature ($\tau \in \{0.3, 0.5, 1.0, 1.5, 2.0\}$)
- Random seeds (5 different initializations)
- Beam search width (4-6 beams)

These settings allow our model to produce a candidate pool $Y = \{y_1, ..., y_N\}$ covering different translation properties (e.g., fluency vs. adequacy).

3.4.2 RL Training

We train another model to select/combine candidates from Y via GRPO. The framework utilizes the translation model as the base model in the first stage. The reward function remains the same weighted combination, now applied to the final output. During inference, the model can either:

1. Select the highest-scoring candidate;
2. Generate a new translation by attending to all candidates;

3.5 Implementation Details

All models are implemented in PyTorch and trained on 128 GPUs. Specifically, key hyperparameters are shown in Table 1.

Table 1: Training Hyperparameters

Parameter	Value
Batch size	64
Learning rate	5×10^{-5}
Max sequence length	8196

4 Results

Table 2 presents comprehensive evaluation results of our proposed Shy-hunyuna-MT model across 31 language directions from the WMT25 benchmark. The results of our model demonstrate exceptional performance across diverse linguistic pairs, achieving consistent state-of-the-art results on multiple automatic evaluation metrics. Our model achieves a remarkable AutoRank score of 1.0 across all 31 translation directions, indicating superior performance compared to baseline systems. This consistency across diverse language pairs, ranging from high-resource languages (e.g., English-German, English-Japanese) to low-resource ones (e.g., English-Bhojpuri, English-Maasai), demonstrates the robustness and generalization capability of Shy-hunyuna-MT.

Meanwhile, the evaluation results encompass both reference-based and reference-free metrics. On the CometKiwi-XL, our model achieves scores ranging from 0.577 to 0.720, with robust performance on the English-Estonian (0.720) and English-Korean (0.697) pairs. For GEMBA-ESA, our model consistently demonstrates excellence, with GEMBA-ESA-GPT4.1 scores predominantly

Table 2: Results of Shy-hunyuna-MT on 31 language directions of WMT25.

Direction	AutoRank↓	CometKiwi-XL↑	GEMBA-ESA-CMDA↑	GEMBA-ESA-GPT4.1↑	MetricX-24-Hybrid-XL↑	XCOMET-XL↑	chrF++↑
English-Egyptian Arabic	1.0	0.658	76.3	75.0	-5.7	0.388	-
English-Bhojpuri	11.5	-	-	-	-	-	40.6
English-Czech	1.0	0.658	83.7	89.4	-5.5	0.639	-
English-Estonian	1.0	0.72	78.8	87.8	-7.3	0.628	-
English-Icelandic	1.0	0.663	71.6	83.9	-7.5	0.543	-
English-Italian	1.0	-	84.6	88.7	-4.7	0.62	-
English-Japanese	1.0	0.687	82.2	89.6	-5.5	0.592	-
English-Korean	1.0	0.697	83.8	85.6	-4.9	0.624	-
English-Maasai	1.0	-	-	-	-	-	27.7
English-Russian	1.0	0.657	84.3	85.9	-4.9	0.652	-
English-Serbian (Cyrilics)	1.0	0.687	76.6	83.3	-4.2	0.64	-
English-Ukrainian	1.0	0.65	84.1	85.3	-5.0	0.662	-
English-Simplified Chinese	1.0	0.67	87.2	88.3	-4.0	0.576	-
Czech-Ukrainian	1.0	0.601	79.1	85.3	-5.0	0.681	-
Czech-German	1.0	0.596	78.4	88.3	-3.6	0.653	-
Japanese-Simplified Chinese	1.0	0.577	85.1	85.5	-4.2	0.629	-
English-Bengali	1.0	-	67.9	83.2	-4.8	0.449	-
English-German	1.0	-	84.3	90.6	-3.1	0.703	-
English-Greek	1.0	-	80.3	85.8	-5.3	0.601	-
English-Persian	1.0	-	80.4	84.1	-4.6	0.553	-
English-Hindi	1.0	-	77.0	82.3	-5.1	0.44	-
English-Indonesian	1.0	-	83.2	87.1	-4.4	0.677	-
English-Kannada	1.0	-	64.0	78.8	-6.0	0.446	-
English-Lithuanian	1.0	-	77.6	84.1	-6.3	0.569	-
English-Marathi	1.0	-	70.8	81.6	-5.8	0.248	-
English-Romanian	1.0	-	83.2	86.3	-5.7	0.651	-
English-Thai	1.0	-	71.3	87.9	-5.1	0.603	-
English-Serbian (Latin)	1.0	-	80.1	84.2	-3.4	0.583	-
English-Swedish	1.0	-	84.2	91.0	-4.7	0.685	-
English-Turkish	1.0	-	81.4	85.2	-7.2	0.542	-
English-Vietnamese	1.0	-	83.1	87.3	-4.5	0.623	-

above 80%, reaching peaks of 91.0% for English-Swedish and 90.6% for English-German, indicating high-quality translations that align well with human preferences. The MetricX-24-Hybrid-XL scores, while negative across all directions, remain relatively compact within the range of -3.1 to -7.5, with English-German achieving the best score of -3.1. This metric consistency indicates stable translation quality without significant degradation across different language families.

Furthermore, the proposed model Shy-hunyuna-MT exhibits powerful performance on European language pairs, with XCOMET-XL scores exceeding 0.65 for English-German (0.703), English-Swedish (0.685), and Czech-Ukrainian (0.681). For Asian languages, the results maintain competitive performance with Japanese-Simplified Chinese achieving 0.629 and English-Korean reaching 0.624 on XCOMET-XL. Notably, for low-resource languages such as English-Maasai and English-Bhojpuri, where most neural metrics are unavailable, the model still achieves chrF++ scores of 27.7 and 40.6, respectively, demonstrating its capability to handle challenging low-resource scenarios where traditional evaluation metrics fail to provide coverage. In addition, the model demonstrates effective cross-lingual transfer capabilities, as evidenced by its strong performance on non-English-centric pairs, such as Czech-Ukrainian, Czech-German, and Japanese-Simplified Chinese. These directions achieve comparable scores to English-centric pairs, with Czech-Ukrainian no-

tably achieving 0.681 on XCOMET-XL, surpassing many English-centric directions.

In summary, Shy-hunyuna-MT establishes new benchmarks across diverse translation directions, demonstrating both breadth in language coverage and depth in translation quality, making it a versatile solution for MT tasks.

5 Conclusion

In this work, we propose **Shy-hunyuan-MT**, a novel MT system built upon the **S**ynergy-en**h**anced polic**y** optimization framework (**Shy**). Our method leverages a carefully designed two-phase training paradigm that systematically transforms the open-sourced Hunyuan-7B base model into a state-of-the-art translation system. The core innovation of our method lies in the synergistic combination of three complementary training phases: domain-adaptive continual pre-training on large-scale parallel corpora, supervised fine-tuning on curated WMT datasets, and reinforcement learning through Generalized Reward Policy Optimization (GRPO) with composite reward signals. This progressive training strategy enables the model to acquire robust multilingual translation capabilities while maintaining strong performance across a diverse range of language pairs.

Our extensive experiments on 31 language directions from WMT25 demonstrate the effectiveness of Shy-hunyuan-MT. The model achieves consistent top-tier performance with an AutoRank score of 1.0 across all evaluated directions, while excelling on multiple automatic metrics, including XCOMET-XL, CometKiwi-XL, and GEMBA-ESA variants. Notably, the model demonstrates its ability to handle both high-resource and extremely low-resource language pairs with comparable proficiency, highlighting its strong generalization capabilities and effectiveness in cross-lingual transfer learning. The success of our approach validates several key design decisions: (1) the importance of domain-specific continual pre-training in adapting general-purpose LLMs for translation tasks, (2) the effectiveness of incorporating multiple COMET-based metrics as reward signals during policy optimization, and (3) the value of progressive training paradigms in building robust multilingual systems. These findings provide valuable insights for future research in neural machine translation and cross-lingual model adaptation. Looking forward, Shy-hunyuan-MT establishes a strong foundation for advancing multilingual translation technology, particularly in scenarios involving diverse language families and resource constraints. The framework's flexibility and consistent performance across varied linguistic contexts position it as a promising solution for real-world translation applications and a solid baseline for future improvements in the field.

References

Yoshua Bengio, Jérôme Louradour, Ronan Collobert, and Jason Weston. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, pages 41–48.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, and 12 others. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Christian Buck and Philipp Koehn. 2016. Findings of the wmt 2016 bilingual document alignment shared task. In *Proceedings of the First Conference on Machine Translation*, pages 554–563, Berlin, Germany. Association for Computational Linguistics.

Pinzhen Chen, Jindřich Helcl, Ulrich Germann, Laurie Burchell, Nikolay Bogoychev, Antonio Valerio Miceli-Barone, Jonas Waldendorf, Alexandra Birch, and Kenneth Heafield. 2021. The university of edinburgh's english-german and english-hausa submissions to the wmt21 news translation task. In *Proceedings of the Sixth Conference on Machine Translation*, pages 104–109.

DeepSeek-AI. 2024. Deepseek-v3 technical report. *Preprint*, arXiv:2412.19437.

Markus Freitag, Ricardo Rei, Nitika Mathur, Chi-kiu Lo, Craig Stewart, Eleftherios Avramidis, Tom Kocmi, George Foster, Alon Lavie, and André FT Martins. 2022. Results of wmt22 metrics shared task: Stop using bleu–neural metrics are better and more robust. In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 46–68.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.

Kelvin Guu, Kenton Lee, Zora Tung, Panupong Pasupat, and Mingwei Chang. 2020. Retrieval augmented language model pre-training. In *International conference on machine learning*, pages 3929–3938. PMLR.

Yuval Kirstain, Patrick Lewis, Sebastian Riedel, and Omer Levy. 2021. A few more examples may be worth billions of parameters. *arXiv preprint arXiv:2110.04374*.

Zheng Li, Mao Zheng, Mingyang Song, and Wenjie Yang. 2025. Tat-r1: Terminology-aware translation with reinforcement learning and word alignment. *Preprint*, arXiv:2505.21172.

Baohao Liao, Shahram Khadivi, and Sanjika Hewavitharana. 2021. Back-translation for large-scale multilingual machine translation. *CoRR*, abs/2109.08712.

Stephen Merity, Caiming Xiong, James Bradbury, and Richard Socher. 2016. Pointer sentinel mixture models.

Jonas Pfeiffer, Ivan Vulić, Iryna Gurevych, and Sebastian Ruder. 2020. Mad-x: An adapter-based framework for multi-task cross-lingual transfer. *arXiv preprint arXiv:2005.00052*.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *JMLR*.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *Preprint*, arXiv:2402.03300.

Mingyang Song, Mao Zheng, Zheng Li, Wenjie Yang, Xuan Luo, Yue Pan, and Feng Zhang. 2025. Fastcurl: Curriculum reinforcement learning with stage-wise context scaling for efficient training r1-like reasoning models. *Preprint*, arXiv:2503.17287.

Jörg Tiedemann. 2012. Parallel data, tools and interfaces in opus. In *Lrec*, volume 2012, pages 2214–2218.

Kocmi Tom, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, and 1 others. 2023. Findings of the 2023 conference on machine translation (wmt23): Llms are here but not quite there yet. In *WMT23-Eighth Conference on Machine Translation*, pages 198–216.

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.

Ashwin K Vijayakumar, Michael Cogswell, Ramprasath R Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. 2016. Diverse beam search: Decoding diverse solutions from neural sequence models. *arXiv preprint arXiv:1610.02424*.

Wenjie Yang, Mao Zheng, Mingyang Song, Zheng Li, and Sitong Wang. 2025. Ssr-zero: Simple self-rewarding reinforcement learning for machine translation. *Preprint*, arXiv:2505.16637.

Mao Zheng, Zheng Li, Bingxin Qu, Mingyang Song, Yang Du, Mingrui Sun, and Di Wang. 2025. Hunyuan-mt technical report. *Preprint*, arXiv:2509.05209.

Michał Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The united nations parallel corpus v1. 0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3530–3534.