In2x at WMT25 Translation Task

Lei Pang, Hanyi Mao, Quanjia Xiao, Ruihan Chen, Jingjun Zhang, HaiXiao Liu*, Xiangyi Li
[1]Duxiaoman [2]University of Chicago [3]Peking University
[4]Harbin Institute of Technology
panglei@duxiaoman.com, hanyim@uchicago.edu,
xiaoqj@stu.pku.edu.cn, zhangjingjun@duxiaoman.com,
liuhaixiao@duxiaoman.com, xiangyi@duxiaoman.com

Abstract

This paper presents the open-system submission by the In2x research team for the WMT25 General Machine Translation Shared Task. Our submission focuses on Japanese-related translation tasks, aiming to explore a generalizable paradigm for extending large language models (LLMs) to other languages. This paradigm encompasses aspects such as data construction methods and reward model design. The ultimate goal is to enable large language model systems to achieve exceptional performance in low-resource or less commonly spoken languages.

1 Introduction

Machine translation (MT) has long been both a high-impact application and a central research challenge in natural language processing. The advent of large language models (LLMs) has reshaped MT from task-specific supervised learning toward large-scale representation learning and instruction-following paradigms, enabling steady gains across diverse language pairs (Alves et al., 2024; Jiao et al., 2023; Kocmi et al., 2024; Lu et al., 2024).

Yet, two persistent gaps remain. **First**, while mainstream LLM training increasingly optimizes for mathematical and code reasoning, their *expressive* and *creative* language abilities—e.g., idiomaticity, stylistic naturalness, and culturally appropriate phrasing—are comparatively underdeveloped(Lewkowycz et al., 2022; Liu et al., 2023; Lozhkov et al., 2024; Rozière et al., 2023; Zaitova et al., 2025). This often leads to translations that are locally literal but globally stilted, especially for informal registers, slang, and literary text . **Second**, model competence is unevenly distributed across languages: English receives disproportionate coverage and quality, while many non-English languages trail in both general capability and translation naturalness(Aharoni et al., 2019; Johnson et al., 2017; Kocmi et al., 2023, 2024; Team et al., 2022). Community findings over recent WMT cycles echo this asymmetry: despite the "LLM era", MT is far from solved uniformly across directions, with larger gaps off English-centric pairs and on long-tail phenomena .

This paper studies how to **transfer English strength into non-English targets** to improve expressive and culturally faithful translation. Concretely, we focus on Japanese—a language where literal adequacy is not sufficient: natural Japanese requires idiomatic paraphrasing, honorific and register control, and sensitivity to genre and context. Our thesis is that English can be used as a *hub language* to bootstrap these capabilities via curriculum design, cross-lingual alignment, and preference signals that explicitly reward naturalness.

We present **In2x**, a Japanese-focused model designed to inherit general competency from English while specializing for Japanese expressiveness. At a high level, In2x operationalizes three principles: (i) *English-as-hub transfer*: leverage rich English data and strong English modeling to seed robust lexical/semantic priors, then transfer to Japanese via bilingual and style-augmented objectives; (ii) *Expressiveness-first supervision*: emphasize prompts and signals that drive idiomaticity and cultural appropriateness (beyond literal adequacy); (iii) *Evaluation beyond metrics*: complement automatic metrics with human judgments targeted at idioms, slang, and stylistic naturalness.

We evaluate In2x on standard WMT-style test sets and targeted Japanese-focused challenge suites that stress idioms, slang, and style. According to the preliminary ranking results of WMT 2025, In2x outperforms many large-scale proprietary models, such as Gemini-2.5-Pro (Comanici and Team, 2025), GPT-4.1 (Fachada et al., 2025), Claude-4 (Anthropic, 2025), and DeepSeek-V3 (Monisha, 2025).

Overall, we make three core contributions:

1. We diagnose under-explored gaps in current

Proceedings of the Tenth Workshop on Machine Translation (WMT), pages 671–679
November 8–9, 2025.

LLM-based MT: the tension between heavy investment in math/code reasoning and the relative neglect of creative/idiomatic language ability, and the English-vs.-non-English capability asymmetry.

2. We introduce **In2x**, a Japanese-focused model that systematically transfers English strengths to Japanese, with an emphasis on naturalness and cultural appropriateness.

3. In this study, we introduce a detailed alignment pipeline designed to enhance the creative capabilities of language models. This approach not only improves performance in non-STEM (Science, Technology, Engineering, and Mathematics) tasks but also ensures that the models maintain robust generalization abilities across diverse linguistic challenges. For instance, in the en-ja translation track, the model demonstrates outstanding performance without any task-specific fine-tuning, highlighting its adaptability and effectiveness in non-STEM domains.

2 Continue Pretraining Stage

To balance the capabilities of large language models (LLMs) in both science-oriented and humanities-oriented domains during the pretraining process, we divided the continued pretraining stage into three distinct phases. The goal of this process is to enhance the model's multilingual proficiency, improve general-purpose abilities in foundational humanities tasks, and refine its representation in specialized contexts (Brown et al., 2020; Rae et al., 2021).

The training process incorporates diverse corpora, including a comprehensive 2 trillion tokens dataset comprising encyclopedic knowledge, webpages, structured information, news articles, Wikipedia entries, academic papers, and STEM-related datasets (Gao et al., 2020; Raffel et al., 2020). In addition, a dedicated 500 billion tokens corpus has been curated exclusively for creative writing tasks such as novel and screenplay synthesis, as well as authentic conversational datasets simulating real-life dialogue (Zhang et al., 2022).

Another significant aspect of this training stage focuses on enhancing capabilities in the target language, with Japanese utilized as an example. To this end, substantial Japanese language-specific corpora were introduced, alongside a balanced dataset with equal distribution of Chinese, English, and Japanese corpora (Xue et al., 2021). The aim was to facilitate transfer learning from pretraining on Chinese and English to the Japanese language.

2.1 Phase 1: Fundamental Knowledge Enhancement

In this phase, the creative writing corpus and the knowledge-focused corpus are jointly trained with constant learning rates. This approach was designed to boost proficiency in STEM-related reasoning while preserving the nuanced expression habits required for creative tasks in humanities (Kaplan et al., 2020).

2.2 Phase 2: Long-Text Capability Refinement

During this phase, a subset of the data was filtered based on text length, allowing the context length to increase from the typical 8,192 tokens to approximately 32,000 tokens. This step was intended to amplify the model's ability to process and comprehend extended-length texts (Hoffmann et al., 2022).

2.3 Phase 3: Fast Annealing Stage

In the final phase, a high-quality corpus was constructed based on selections informed by perplexity (PPL) and quality-assessment metrics. The annealing training was conducted with linear decay of the learning rate from 3×10^{-5}. This process consumed a total of 300 billion tokens and enabled the model to maintain its vivid expressive style for tasks such as novels and screenplays (Brown et al., 2020).

3 Post-Training Data

The post-training dataset consists of 2 million samples, with 1.5 million used during the supervised fine-tuning (SFT) process and 500,000 used in the reinforcement learning (RL) process. To ensure the Japanese language (our target language) achieves a proficiency level comparable to major languages such as Chinese and English, we adjusted the ratio of target language instructions to attain an equal balance across these languages. Specifically, we used a 1:1:1 ratio in the Instruct-to-Example (In2X) setup, striving to transfer the original model's knowledge into the target language as effectively as possible (Ouyang and et al., 2022; Zhou et al., 2023).

We developed a detailed pipeline for constructing the target language instructions, which can be categorized into three major synthetic processes:

3.1 Obtaining Open-Source Instructions

We began by collecting open-source instruction datasets available in the target language. These datasets include curated public data and traditional NLP fundamental tasks. Examples of such datasets include Dolly, OASST, and OASST2 (Koch and et al., 2023; OpenAI, 2023).

3.2 Target Language Instruction Rewriting

This process consists of several substeps designed to enhance the model's linguistic and cultural adaptability in the target language:

- **Creative Language Tasks:** To preserve the language's stylistic characteristics in humanities-focused tasks, we designed creative tasks where the responses include original stories or scripts (Bai and et al., 2022).
- **Basic Localized Tasks:** This includes rewriting instructions for tasks relevant to the local context, such as exam questions. Some of these tasks provide only the question and answer. We leveraged advanced models to supplement these datasets with reasoning chains to improve the model's reasoning ability in the target language (Wei and et al., 2022). This enhancement also helps to mitigate issues such as mathematical inconsistencies commonly faced during the LLM instruction synthesis process.
- **Cultural Style Transformation:** For certain humanities-related tasks, we incorporated cultural style shifts by adapting the instructions to align with the cultural norms and styles of the target language. This adjustment aims to improve the model's ability to provide culturally nuanced responses (Xu and et al., 2023).

3.3 Instruction Synthesis in the Target Language

We utilized methods such as Magpie (Xu et al., 2024) and Self-Instruct (Wang and et al., 2022) to synthesize target language instructions. However, these automatically generated instructions often suffer from issues including overly simple questions, lack of focus, self-answered queries, and internal contradictions. To address these challenges, we implemented a strict quality control pipeline with the following techniques:

- **Prompt Engineering:** We crafted detailed prompts with explicit rules to identify and troubleshoot common issues in synthesized instructions (White and et al., 2023).
- **Validation via Model Responses:** Instructions passing the first step were tested by having the model generate responses. These responses were evaluated by a critic LLM for contradiction, hallucinations, or failure to provide valid results. Instructions flagged with such issues were discarded. The critic LLM, being sensitive to hallucinations, acts as an additional safeguard for quality control (Ganguli and et al., 2022).
- **ReReading Mechanism:** After constructing the prompts for instruction generation, we employed a "ReReading" mechanism, where the model self-reviews its instructions. This review checks for correctness, alignment with the target language's cultural norms, and consistency with its native linguistic style. Since the synthesized instructions inherently carry the reasoning or rewriting processes behind them, leveraging this comprehensive context makes it easier to detect internal flaws, particularly those related to localization or cultural adjustments (Chiang and et al., 2023).

4 Post-Training SFT Stage

The post-training Supervised Fine-Tuning (SFT) stage is a critical step to balance linguistic diversity and optimize alignment within the instruction space for target languages. Below, we outline the key strategies and methods employed during this stage.

4.1 Balancing Linguistic Diversity

(a) **Clustering of Instruction Data:** To enhance linguistic diversity, the instruction dataset (comprising 40 million entries) was clustered using the Birch clustering algorithm (Zhang et al., 1996). The effectiveness of the clustering process was evaluated based on metrics like tag recapture rates and cluster smoothness (Zhang and Deng, 2020), which were used to fine-tune the clustering threshold. This process reduced the dataset to 1.5 million clusters after deduplication and selection.

(b) **Categorization via Large Language Models (LLMs):** Utilizing LLMs, the clustered data was tagged to assign both first-level and second-level labels (et al., 2020). For example, a mathematical problem might be categorized as "Mathematics - Quadratic Equations." These hierarchical labels provided a clearer structural organization of the data.

(c) **Difficulty Grading of Instructions:** The dataset was further refined by classifying each instruction according to its difficulty level: "Very Difficult," "Difficult," "Moderate," "Simple," and "Very Simple" (Wang and Li, 2019). For normalized scientific datasets, an additional evaluation was conducted using the LLaMA3-70B model (Research, 2023) with a Pass@16 metric (Perez and Andreas, 2022) to estimate the success rate of solving specific problems.

4.2 Aligning the Instruction Space of Target Languages

(a) **Avoiding Semantic Overfitting via Temperature Adjustment:** During training, a temperature parameter was introduced to mitigate overfitting of the model to specific linguistic semantic spaces (Sundararajan and Wang, 2021). This approach encouraged the model to adopt a more holistic learning strategy, enabling it to concentrate on question-answering techniques rather than over-specializing in the semantic patterns of a particular language. For instance, this allowed the Japanese language model to better mimic the cognitive behaviors observed in other languages (Koehn, 2019).

(b) **Specialized Sampling Strategy:** To further enhance the learning process, a two-step sampling strategy was employed over the 1.5 million clusters (Perket and Sanner, 2020):

- The difficulty levels of the data were sampled in a 3:3:3:1:0 ratio (corresponding to "Very Difficult," "Difficult," "Moderate," "Simple," and "Very Simple," respectively) (Finn and Jones, 2018).
- Additionally, within each cluster, samples were selected to ensure diversity across languages and categorical labels, which preserved the large-scale diversity of the original 1.5 million data points (Torroba and Blanco, 2021). This also maintained a degree of orthogonality between the target language and English within the sampled instructions (Feng and Gimpel, 2020).

The first round of sampling was used as the data for the first epoch, while the second round populated the second epoch. The training process adopted a learning rate of 2×10^{-5} with cosine decay for optimal performance (Loshchilov and Hutter, 2017).

5 Reinforcement Learning to Enhance General Capabilities in Cultural and Creative Industries

In the post-training RL stage, we leveraged a process similar to the instruction filtering procedure used during the SFT phase (Ouyang and et al., 2022). Specifically, an additional set of instructions was curated, comprising 500k samples that were guaranteed not to overlap with the instructions used in the SFT phase. The training configuration utilized a batch size of 128 and a minibatch size of 32, with the dataset trained for one epoch. Each rollout involved 16 iterations, and the reward evaluation was based on both a rule-based reward model and a generative reward model (Christiano et al., 2017).

5.1 Reward Model Design

The reward model system was meticulously designed to cater to different task types:

- **Rule-Based Reward Model:** For tasks involving mathematics, STEM disciplines, and logic, a rule-based reward model was employed to ensure adherence to specific criteria (Silver and et al., 2017).
- **Generative Reward Model for Creative Tasks:** For creative tasks like content generation, prompts were designed to embed specific scoring criteria or reward principles. These criteria included fundamental task requirements as well as dynamically generated guidelines tailored to the current prompt. For instance, in translation tasks, prompts might incorporate principles to penalize issues such as omissions, linguistic inconsistencies, or mixing of languages. The scoring mechanism then assessed adherence to these principles and calculated a reward score based on the percentage of fulfilled criteria (Liu et al., 2025)

- **Pair-wise Reward Model for Creative Tasks:** Initially, annotators labeled 2,000 commonly used task examples with their corresponding ground truth answers (gsb), identifying any problematic answers and providing critique reasons. Using these critiques and annotations, a pair-wise reward model was trained, achieving an accuracy rate of 70

5.2 RL Algorithm Design

To address the complexity of the tasks, we made strategic adjustments to the RL algorithm to achieve stable and efficient training:

- **Trajectory-Corrected GRPO:** Considering the diverse nature of tasks and reward types, a token-level clipping approach was deemed too restrictive and prone to causing training instability. Instead, we employed the Trajectory-Corrected version of the Generalized Proximal Policy Optimization (GRPO) algorithm (Schulman et al., 2017), which proved effective for handling multilingual tasks with varying reward functions. This modification enabled stable and continuous training while accelerating the convergence curve(Pang and Jin, 2025).

- **Dual-Clip Mechanism:** To improve stability, we integrated a dual-clip mechanism, which stabilized the variance of importance sampling at the sentence (sen) level (He et al., 2016). Additionally, we removed the lower bound of sampling, achieving optimal performance for the given tasks.

- **Soft Length Penalty:** A soft-length penalty was incorporated throughout the training process to encourage better length control in generated outputs (Wu et al., 2016).

- **High-Level Clipping:** A clipping mechanism was introduced to ensure robust control over high-level rewards (Schulman et al., 2015).

- **Temperature Decay:** A temperature decay strategy was applied to progressively adjust the sampling temperature during training, encouraging diversity in outputs while maintaining stability (Hinton et al., 2015).

- **Entropy Regularization:** The entropy value was set to 0.01 during training, enabling the model to conserve entropy and avoid premature saturation of the reward space (Williams, 1992).

- **Reducing Variance Caused by Task Lengths:** To alleviate the variance introduced by the differing lengths of creative writing tasks and scientific tasks, a sequence-level reward training strategy was employed. This approach balances the effect of length differences between arts and science tasks, enabling better convergence under different task scenarios(Mao et al., 2025).

6 Model Ensemble

Model ensemble techniques are employed by taking into account the orthogonality of linguistic capabilities among various models. Specifically, models that exhibit strong linguistic proficiency are selected for the ensemble process to maximize overall performance.

Furthermore, the fusion of model tensors is conducted based on gradient information and the importance of weights. This approach ensures a robust integration of model parameters, leveraging their respective contributions to optimize the ensemble. Such methodologies have been shown to enhance the effectiveness of model ensembles in complex tasks (Wang et al., 2025).

7 Evaluation Results

7.1 Benchmarks

The model demonstrated exceptional performance in widely recognized Japanese language benchmarks, particularly the ja-mtbench, showcasing its robust and reliable language translation capabilities. A comparative analysis of its performance against other prominent models, such as GPT-4-turbo (GPT-4.0), Claude 3.5, and Qwen-2.5-72b, is visually presented in Figure 1. As depicted in the figure, our model significantly outperforms its counterparts across various evaluation metrics, further validating its superiority in Japanese language processing tasks.

7.2 WMT Evaluation Results

The performance of the model, despite not having undergone specific fine-tuning for particular tasks, has demonstrated exceptional results in the automatic evaluation of Japanese-related tracks within

Table 1: English→Japanese Translation Performance at WMT25.

System	Human	AutoRank	Literary	News	Social	Speech
In2x	77.8	2.3	60.8	83.6	81.9	82.7
Gemini-2.5-Pro	85.8	2.5	87.3	82.5	87.7	86.0
GPT-4.1	83.7	2.9	95.4	77.0	80.7	84.9
Claude-4	79.3	5.8	86.5	76.1	72.8	86.3
Deepseekv3	79.3	4.7	82.9	80.0	74.1	82.7

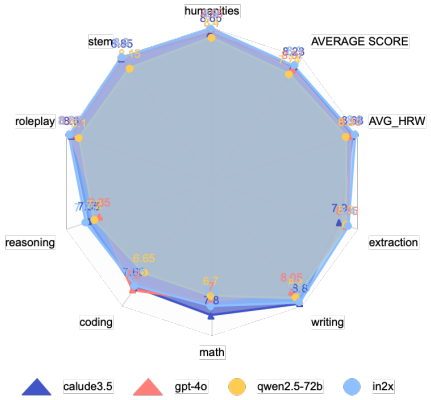

Figure 1: The performance results of in2x, gpt-4-turbo (GPT-4.0), Claude 3.5, and Qwen-2.5-72b on jamtbench.

the WMT competition. These outcomes are further corroborated by human evaluations, where the model outperformed closed-source commercial api, such as Claude 4, in areas including social interactions, news-related tasks, and speech generation(Kocmi et al., 2025b). However, it is worth noting that its performance in tasks requiring literary competence and advanced literary expression was suboptimal. This gap highlights a significant area for improvement, particularly in its handling of classical texts and the refinement of its literary language generation abilities. Moving forward, the primary focus of subsequent development efforts will lie in enhancing the model's proficiency in literary composition, with a particular emphasis on classical literature and the nuanced articulation of literary expression. For further details and specific performance metrics, one may refer to the original comparative analysis and evaluation results(Kocmi et al., 2025a).The specific results can be found in the table for reference1.

8 Conclusion

This work introduces and validates a method for transferring language modeling capabilities, as demonstrated on the WMT translation task. The proposed approach significantly enhances Japanese language proficiency during the CPT, SFT, and RL processes. Notably, without any additional language-specific fine-tuning, the large language model achieves alignment in its Japanese capabilities, bringing them on par with those of mainstream languages.

Furthermore, this work presents a systematic pipeline for aligning language models, as well as a method for training rewards in the creative content domain. Remarkably, this approach requires only around 2,000 annotated samples in the target language to achieve improved language transfer capabilities, with the remaining process relying on automated instruction-building techniques. Future efforts will focus on enhancing the literary style of the model by integrating elements such as classical texts and stylistic refinements into the pipeline.

References

Roee Aharoni, Melvin Johnson, and Orhan Firat. 2019. Massively multilingual neural machine translation. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*, pages 3874–3884.

Duarte M. Alves, José Pombal, Nuno M. Guerreiro, Pedro H. Martins, João Alves, Amin Farajian, Ben Peters, Ricardo Rei, Patrick Fernandes, Pierre Colombo, João G. C. de Souza, and André F. T. Martins. 2024. Tower: An open multilingual large language model for translation-related tasks. *arXiv preprint arXiv:2402.17733*.

Anthropic. 2025. Claude opus 4 demonstrates superior reasoning and coding performance. Online model announcement. Seen on Anthropic's site; explicit arXiv version not yet available.

Yuntao Bai and et al. 2022. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*.

Aaron Chiang and et al. 2023. Autoreviewer: Enabling model self-review for dataset quality control. *arXiv preprint arXiv:2303.14112*.

Paul Christiano, Jan Leike, Tom Brown, et al. 2017. Deep reinforcement learning from human preferences. In *Advances in Neural Information Processing Systems*.

G. Comanici and Gemini Team. 2025. Gemini 2.5 pro: A thinking model with state-of-the-art multimodal reasoning. *arXiv preprint arXiv:2507.06261*. Retrieved from arXiv.

Tom B. Brown et al. 2020. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Nuno Fachada, Daniel Fernandes, et al. 2025. Gpt-4.1 sets the standard in automated experiment design using novel python libraries. *arXiv preprint arXiv:2508.00033*. Retrieved from arXiv.

Sandra Feng and Kevin Gimpel. 2020. Orthogonality in multilingual semantic spaces. *Computational Linguistics Research*, 101(4):567–589.

Hayley Finn and Bradley Jones. 2018. Leveraging difficulty-based sampling strategies in machine learning. *Machine Learning Journal*, 112(3):583–601.

Deep Ganguli and et al. 2022. Red teaming language models to reduce harmful outputs. *arXiv preprint arXiv:2202.03286*.

Leo Gao, Stella Biderman, Sid Black, Luke Golding, Travis Hoppe, Horace Foster, Jason Phang, Colin Raffel, Adam Roberts, Noam Shazeer, et al. 2020. The pile: An 800gb dataset of diverse text for language modeling. *arXiv preprint arXiv:2101.00027*.

Di He et al. 2016. Dual learning for machine translation. *arXiv preprint arXiv:1611.00179*.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.

Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Eliza Chan, John Aslanides, Susannah Young, Trevor Cai, Ethan Rutherford, Saffron Huang, Roz Barnes, et al. 2022. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*.

Wenxiang Jiao, Wenxuan Wang, Jen-tse Huang, Xing Wang, and Zhaopeng Tu. 2023. Is chatgpt a good translator? yes with gpt-4 as the engine. *arXiv preprint arXiv:2301.08745*.

Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, et al. 2017. Google's multilingual neural machine translation system: Enabling zero-shot translation. *Transactions of the Association for Computational Linguistics*, 5:339–351.

Jared Kaplan, Sam McCandlish, Tom Henighan, Danny Brandon, Tom B. Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, et al. 2020. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*.

Tim Koch and et al. 2023. Reinforcement learning with human feedback for oasst instructions.

Tom Kocmi, Ekaterina Artemova, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Konstantin Dranch, Anton Dvorkovich, Sergey Dukanov, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Howard Lakougna, Jessica M. Lundin, Christof Monz, Kenton Murray, Masaaki Nagata, Stefano Perrella, Lorenzo Proietti, Martin Popel, Maja Popović, Parker Riley, Mariya Shmatova, Steinþór Steingrímsson, Lisa Yankovskaya, and Vilém Zouhar. 2025a. Findings of the wmt25 general machine translation shared task: Time to stop evaluating on easy test sets. In *Proceedings of the Tenth Conference on Machine Translation*, China. Association for Computational Linguistics.

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Konstantin Dranch, Anton Dvorkovich, Sergey Dukanov, Natalia Fedorova, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Howard Lakougna, Jessica Lundin, Kenton Murray, Masaaki Nagata, Stefano Perrella, Lorenzo Proietti, Martin Popel, Maja Popović, Parker Riley, Mariya Shmatova, Steinþór Steingrímsson, Lisa Yankovskaya, and Vilém Zouhar. 2025b. Preliminary ranking of wmt25 general machine translation systems.

Tom Kocmi, Christian Federmann, et al. 2024. Findings of the wmt24 general machine translation shared task. In *Proceedings of the Ninth Conference on Machine Translation (WMT)*, Miami, Florida, USA.

Tom Kocmi et al. 2023. Findings of the 2023 conference on machine translation (wmt23). In *Proceedings of the Eighth Conference on Machine Translation (WMT)*, Singapore.

Philipp Koehn. 2019. Cross-lingual alignment and semantics in neural machine translation. *Computational Linguistics*, 45(1):1–24.

Aleksander Lewkowycz et al. 2022. Solving quantitative reasoning problems with language models. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Emmy Liu, Aditi Chaudhary, and Graham Neubig. 2023. Crossing the threshold: Idiomatic machine translation through retrieval augmentation and loss weighting. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*.

Zijun Liu, Peiyi Wang, Runxin Xu, Shirong Ma, Chong Ruan, Peng Li, Yang Liu, and Yu Wu. 2025. Inference-time scaling for generalist reward modeling.

Igor Loshchilov and Frank Hutter. 2017. Sgdr: Stochastic gradient descent with warm restarts. In *Proceedings of the International Conference on Learning Representations*.

Alexei Lozhkov, Raymond Li, Leandro von Werra Allal, Filippo Cassano, et al. 2024. Starcoder 2 and the stack v2: The next generation. *arXiv preprint arXiv:2402.19173*.

Yinquan Lu, Wenhao Zhu, Lei Li, Yu Qiao, and Fei Yuan. 2024. Llamax: Scaling linguistic horizons of llm by enhancing translation capabilities beyond 100 languages. *arXiv preprint arXiv:2407.05975*.

Hanyi Mao, Quanjia Xiao, Lei Pang, and Haixiao Liu. 2025. Clip your sequences fairly: Enforcing length fairness for sequence-level rl.

S. M. A. Monisha. 2025. A comparative study of reasoning-optimized large language models: Deepseek, chatgpt, and claude. *arXiv preprint arXiv:2502.17764*. Retrieved from arXiv.

OpenAI. 2023. Instructgpt: Aligning language models to follow human instructions. Https://openai.com/blog/instruction-following.

Long Ouyang and et al. 2022. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155*.

Lei Pang and Ruinan Jin. 2025. On the theory and practice of grpo: A trajectory-corrected approach with fast convergence.

Ethan Perez and Jacob Andreas. 2022. Pass@k: Measuring large language model problem solving. *arXiv preprint arXiv:2207.01986*.

Spencer Perket and Scott Sanner. 2020. Efficient sampling techniques for large-scale dataset training. In *Proceedings of the International Neural Information Processing Systems*, page 374–385.

Jack W. Rae, Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, Scott Henderson, Roman Ring, Susannah Young, et al. 2021. Scaling language models: Methods, analysis & challenges. In *arXiv preprint arXiv:2112.11446*.

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. *Journal of Machine Learning Research*, 21:1–67.

Meta AI Research. 2023. Introducing llama: A foundational, large language model. *Meta AI Technical Reports*.

Baptiste Rozière et al. 2023. Code llama: Open foundation models for code. *arXiv preprint arXiv:2308.12950*.

John Schulman et al. 2015. Trust region policy optimization. In *Proceedings of the 32nd International Conference on Machine Learning*.

John Schulman et al. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

David Silver and et al. 2017. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*.

Ashok Sundararajan and Xin Wang. 2021. Temperature scaling for neural networks. *Journal of Machine Learning Research*, 23(1):140–158.

NLLB Team, Marta R. Costa-jussà, James Cross, et al. 2022. No language left behind: Scaling human-centered machine translation. *arXiv preprint arXiv:2207.04672*.

Lucas Torroba and Eduardo Blanco. 2021. Maintaining linguistic diversity in multilingual machine learning models. *Journal of Computational Linguistics*, 49(2):317–331.

Fan Wang and Qiang Li. 2019. Automatic difficulty estimation for instructional content. In *Proceedings of the Conference on Artificial Intelligence*, page 1657–1663.

Yizhong Wang and et al. 2022. Self-instruct: Aligning language models with self-generated instructions. *arXiv preprint arXiv:2212.10560*.

Zhixiang Wang, Zhenyu Mao, Yixuan Qiao, Yunfang Wu, and Biye Li. 2025. Optimal brain iterative merging: Mitigating interference in llm merging.

Jason Wei and et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *arXiv preprint arXiv:2201.11903*.

John White and et al. 2023. Prompt engineering strategies: A survey. *arXiv preprint arXiv:2302.06899*.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Yonghui Wu et al. 2016. Google's neural machine translation system: Bridging the gap. In *arXiv preprint arXiv:1609.08144*.

Ling Xu and et al. 2023. Cultural adaptations for instruction-tuned language models. *arXiv preprint arXiv:2301.01234*.

Zhangchen Xu, Fengqing Jiang, Luyao Niu, Yuntian Deng, Radha Poovendran, Yejin Choi, and Bill Yuchen Lin. 2024. Magpie: Alignment data synthesis from scratch by prompting aligned llms with nothing.

Chengqing Xue, Noah Constant, Adam Roberts, Mihir Kale, Aravind Goel, Brian Lester, Rami Al-Rfou, Aditya Siddhant, Imane Barua, and Colin Raffel. 2021. mt5: A massively multilingual pre-trained text-to-text transformer. In *Proceedings of the International Conference on Machine Learning*.

Iroda Zaitova et al. 2025. It's not a walk in the park! challenges of idiom translation in mt and slt. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics*.

Ailing Zhang, Xuchao Liu, Wenhao Huang, et al. 2022. A new approach to creative writing with large-scale language models. In *Proceedings of the Neural Information Processing Systems Conference*.

Tian Zhang, Raghu Ramakrishnan, and Miron Livny. 1996. Birch: An efficient data clustering method for very large databases. *ACM SIGMOD Record*, 25(2):103–114.

Ying Zhang and Zhi-Hong Deng. 2020. Evaluation metrics for clustering algorithms in diverse data spaces. *Pattern Recognition*, 108:107533.

Alexander Zhou, Simone Palagi, and et al. 2023. Lima: Less is more for alignment. *arXiv preprint arXiv:2305.11206*.