

CUNI-MH and CUNI-EdUKate at the WMT26 General MT Shared Task

Miroslav Hrabal, Josef Jon, Ondřej Bojar

Charles University, Faculty of Mathematics and Physics

{hrabal, jon, bojar}@ufal.mff.cuni.cz

Abstract

We describe our constrained CUNI-MH-v3 (based on Qwen3-14B) and unconstrained CUNI-EdUKate (based on Granite 4.1 8B) submissions for the WMT26 General MT shared task. Both submissions use a three-stage training procedure: QLoRA supervised fine-tuning on synthetic sentence pairs, continued supervised fine-tuning on a smaller mixture of paragraph-level and instruction-oriented examples, and iterative preference optimization with automatically generated and scored preference pairs. We document the data construction, training setup, checkpoint selection, and evaluation. We also share the training data and model checkpoints for our constrained system.

1 Introduction

The WMT General MT shared task evaluates translation across diverse language pairs. Our submissions focus on adapting instruction-tuned multilingual large language models with parameter-efficient fine-tuning on just four of them: English–Czech, Czech–Vietnamese, Czech–German and Czech–Ukrainian. We used the same general training recipe across all language pairs and put most of our focus on the new Czech–Vietnamese pair. Table 1 summarizes our submissions and the supported language pairs. Although our adapters train bidirectionally, the shared task evaluates only one direction for each pair. The constrained CUNI-MH-v3 covers the four language pairs and uses Qwen-14B (Team, 2025) as a base model. CUNI-EdUKATE covers just Czech–German and Czech–Ukrainian and uses Granite-4.1-8B (IBM Research, 2026) as base model. Since it was trained on additional educational domain EdUKate data that cannot be redistributed, it falls into the unconstrained category.

2 Data

Our training data is organized around the three different training stages. This year, we choose to focus

on using synthetic parallel data created using back-translation.

2.1 Stage 1: Synthetic Sentence Data

We took monolingual documents from HPLTv3 (Oepen et al., 2026), split them into sentences, removed exact duplicates and applied deterministic shuffle. We backtranslated the monolingual sentences with `cyankiwi/gemma-4-31B-it-AWQ-4bit`, an AWQ-quantized (Lin et al., 2026) version of the instruction-tuned Gemma 4 31B model (Team et al., 2026). The same model also assigned an integer 0–100 direct-assessment score to each resulting pair. For this stage, we retained pairs with a score of at least 75. We also used the forward translation by swapping source and target, combined both orientations and shuffled the result.

2.2 Stage 2: Data Augmented Mixture

The second supervised stage uses a smaller mixture intended to adapt the model for the types of segments we expected to see in the WMT26 test set. We deterministically shuffle each source, normalize the configured mixture weights, and construct the largest mixture that does not require oversampling any source. The resulting datasets contain between 8 943 and 14 399 examples depending on the language pair.

The components, in Table 2 order, are: (1) all synthetic sentence pairs scoring at least 90; (2) the deterministically selected first shard of scored paragraph pairs, with no additional threshold; (3) different paragraph layouts¹ generated from a separate sample of 500 paragraph pairs scoring at least 95; (4) glossary examples using either the fixed English instruction “Use the following glossary”

¹By paragraph layouts, we mean that the source and target paragraphs consist of several aligned segments and these segments are separated by different whitespace separators, for example newline, double newline, and space. Different combinations of separators mean different paragraph layouts in this context.

System	Backbone	cs-de	cs-uk	cs-vi	en-cs
CUNI-MH-v3	Qwen3-14B	✓	✓	✓	✓
CUNI-EdUKATE	Granite-4.1-8B	✓	✓	–	–

Table 1: Submitted systems and WMT26 evaluation directions.

Stage-2 source category	Share
Filtered synthetic sentences	25.81%
Scored paragraphs	12.90%
Generated/aligned paragraphs	12.90%
Glossary instructions	25.81%
Inline tags	12.90%
Do-not-translate spans	6.45%
Lowercase inputs	3.23%

Table 2: Target composition of the Stage-2 supervised mixture.

or LLM-generated paraphrases; (5) variants containing aligned HTML, placeholder, or randomized XML-like tags; (6) up to five do-not-translate (DNT) variants that copy a source span wrapped in `<ufaldnt>` unchanged into the target; and (7) lowercased sources paired with targets in their original casing.

Components 4–6 start from deterministic samples of 500 synthetic pairs per language pair scoring at least 90. We used `cyankiwi/gemma-4-31B-it-AWQ-4bit` for generation or bilingual alignment where it was required, for example when proposing alternative glossary terms and applying them to the segment in question.

2.3 Stage 3: Preference Pairs

Preference pairs are generated online rather than generated in advance. For CUNI-MH-v3, each direction uses a pool containing 75% Stage-2 examples and 25% examples derived from WMT25. We included the WMT25 data in order to expose the models to shared-task-style instructions. We therefore do not use WMT25 as an independent development set.

For CUNI-EdUKATE, we use a 50/25/25 mixture of Stage-2, WMT25, and EdUKate examples. The final EdUKate sampling pools have the same sizes as the corresponding general pools because the available WMT25 portion is the limiting source.

3 Methods

All stages use the same chat representation. A system message simply states "You are a helpful as-

sistant."² The user message gives the source and target language, an optional context (translation instructions) field, and the source text. The exact format can be found in Appendix A. The reference translation is the assistant response.

We train QLoRA (Dettmers et al., 2023; Hu et al., 2022) adapters while loading the base model in 4-bit NF4 format with double quantization and bfloat16 computation. Qwen adapters use rank 128 and α 256. Granite adapters use rank 256 and α 512. In both cases, LoRA adapter is applied to all linear modules with zero dropout. Table 3 summarizes the main settings.

3.1 Supervised fine-tuning

Stage 1 starts from either Qwen/Qwen3-14B or `ibm-granite/granite-4.1-8b`. We train for at most 10,000 steps with sequence length 512, batch size 16, a paged 8-bit AdamW optimizer, cosine scheduling, and a learning rate of approximately 6.36×10^{-5} . The first 512 records are reserved for evaluation loss.

Stage 2 continues the same adapter on the structured mixture described in Section 2. We increase the maximum sequence length to 1,024, enable packing and assistant-only loss, and train for 500 steps with batch size 16 and learning rate 2×10^{-5} . We again reserve the first 512 records for evaluation. Stage 2 is deliberately short: its purpose is instruction and structure coverage rather than another pass over millions of sentence pairs.

3.2 Iterative preference optimization

Unlike our submission from previous year (Hrabal et al., 2025), we chose not to pre-generate the preference pairs but train in several iterations and generating the pairs using the latest such version of the model, similarly to Sun et al. (2025).

Stage 3 consists of multiple iterations. At iteration i , we use reservoir sampling for 500 records

²We included this system message primarily to unify and override system messages of some models we initially experimented. An extreme case of this is for example `granite-3.3-8b-instruct`, whose chat template dynamically inserts current date to system prompt, which could potentially have hard to predict effects on reproducibility when not overridden.

Family/stage	Backbone or parent	LoRA r	Steps/rounds	Max. length	Learning rate
MH-v3 Stage 1	Qwen3-14B	128	10,000 steps	512	6.36×10^{-5}
MH-v3 Stage 2	Stage-1 adapter	128	500 steps	1,024	2×10^{-5}
MH-v3 Stage 3	Stage-2 adapter	128	10×500 samples	1,024	10^{-5}
EdUKate Stage 1	Granite-4.1-8B	256	10,000 steps	512	6.36×10^{-5}
EdUKate Stage 2	Stage-1 adapter	256	500 steps	1,024	2×10^{-5}
EdUKate Stage 3	Stage-2 adapter	256	10×500 samples	1,024	10^{-5}

Table 3: Training configuration. Stage-3 values for Steps/rounds denote the number of rounds times sampled inputs. Candidate generation may produce multiple pairs per input, and filtering by score-difference then determines the size of the training-set, so the actual number of training steps of each such round is variable. The selected MH-v3 en-cs run uses learning rate 5×10^{-6} .

from the corresponding preference pool using seed $4243 + i$. We serve the current adapter with vLLM and translate each sampled source greedily. When a reference is available, it forms an alternative candidate. For the CUNI-MH-v3 Czech–German and Czech–Vietnamese runs and both EdUKate runs, Gemma-4-31B-it was also used for one additional candidate through a simple DSPy translation prompt. The MH-v3 Czech–Ukrainian and English–Czech runs use no extra candidate model.

Unique candidates are assigned scores from 0 to 100 by `cyankiwi/gemma-4-31B-it-AWQ-4bit` through a DSPy direct-assessment prompt. We order candidates by this score and retain chosen/rejected pairs whose score differs by at least two points.

We then continue the current adapter on the retained pairs. The objective combines sigmoid DPO (Rafailov et al., 2024), BCO-pair (Jung et al., 2025), and supervised fine-tuning losses with weights 0.8, 0.2, and 1.0, respectively. Training uses a per-device batch size of 2 and gradient accumulation over 16 steps, giving an effective batch size of 32. The learning rate is 10^{-5} for most runs. The selected English–Czech run uses 5×10^{-6} . Each subsequent iteration starts from the adapter produced by the preceding iteration.

3.3 Submitted checkpoints

Table 4 records the selected stage and iteration for each submitted direction.

4 Internal Evaluation

We compare submitted checkpoints with their unadapted backbones, intermediate supervised checkpoints, prior CUNI WMT25 systems, and Gemma-4-E2B and E4B system baselines. “Direct” denotes the prompt we use during training together with greedy decoding; Prior systems are

System	Direction	Selected model
EdUKate	cs-de	Stage 3, iteration 1
EdUKate	cs-uk	Stage 3, iteration 1
MH-v3	cs-de	Stage 3, iteration 1
MH-v3	cs-uk	Stage 3, iteration 2
MH-v3	cs-vi	Stage 3, iteration 1
MH-v3	en-cs	Stage 3, iteration 2

Table 4: Checkpoints used for the final submissions.

re-evaluated here rather than represented by their official WMT25 scores.

We evaluated checkpoints internally on FLORES+ dev for Czech–Vietnamese, Czech–Ukrainian, and English–Czech; on WMT23 for English–Czech; and on an EdUKate development set for Czech–Ukrainian. We report corpus BLEU (Papineni et al., 2002), chrF2 (Popović, 2015), and direct assessment by Gemma-4-31B. WMT25 is excluded because it contributes examples to the Stage-3 pools.

We report corpus BLEU, chrF2, and reference-free direct assessment (DA) by the quantized Gemma-4-31B. The evaluator receives source, hypothesis, language pair, and optional context and produces an integer 0–100 segment score; tables report the mean. FLORES+ dev contains 997 examples per direction, WMT23 contains 2,074, and EdUKate dev contains 3,770.

Tables 5 to 8 group the systems into baselines, submitted checkpoints, and later preference-optimization checkpoints. Czech–German and Czech–Vietnamese runs ended after iteration 6, and the EdUKate runs ended after iteration 1; consequently, iteration 10 exists only for the MH-v3 Czech–Ukrainian and English–Czech runs.

The submitted MH-v3 system performs closest to the strongest listed baseline on Czech–Vietnamese: its BLEU is close to Gemma-4-E4B with DSPy, while its DA score is higher. On the three transla-

System	FLORES+ dev		
	BLEU	chrF2	DA
<i>Baselines</i>			
WMT25 CUNI-MH-v2	27.92	58.37	93.78
Gemma-4-E2B, direct	9.32	46.17	62.13
Gemma-4-E2B, DSPy	24.94	54.72	83.87
Gemma-4-E4B, direct	22.32	55.77	88.60
Gemma-4-E4B, DSPy	27.72	57.10	90.75
<i>Training progression</i>			
Qwen3-14B, base	24.94	55.56	91.21
CUNI-MH-v3, Stage 1	26.98	56.64	92.89
CUNI-MH-v3, Stage 2	26.68	56.49	93.70
Granite-4.1-8B, base	23.70	55.34	87.84
CUNI-EdUKate, Stage 1	26.77	56.62	90.52
CUNI-EdUKate, Stage 2	26.25	56.16	89.78
<i>Submitted checkpoints</i>			
CUNI-MH-v3, it. 1	26.74	56.55	93.68
CUNI-EdUKate, it. 1	26.34	56.41	90.55

Table 5: Czech–German development results. No iteration-10 checkpoint was produced; the MH-v3 run ended at iteration 6 and the EdUKate run at iteration 1.

tion directions shared with WMT25, however, the previous systems remain stronger. The gap is largest for English–Czech and on the EdUKate development set. The iteration-10 checkpoints do not seem to improve over the early submitted checkpoints on Czech–Ukrainian or English–Czech. On Czech–Vietnamese, later iterations seem to trade lower BLEU for a higher Gemma DA score.

Overall, we hypothesize that the underwhelming performance on the English–Czech and Czech–Ukrainian language pairs may be at least partly to our tuning of the training hyperparameters being overly focused on Czech–Vietnamese, particularly with respect to the relatively short stage 1 and stage 2 training. Unfortunately, by the time we identified this issue, there was insufficient time to rerun the experiments before the WMT26 submission deadline. One possible explanation is that differences between the language pairs, including differences in morphological complexity, may affect the amount of training required. We leave verification of this hypothesis to future experiments.

Manual experiments with the final checkpoints before submission suggested that the model is sensitive to changes in the provided instructions and that these instructions could, in some cases, degrade translation quality. We therefore chose to disable the context field for our final submissions by ignoring the instruction field in the WMT26 evaluation set. However, because WMT25 data had been used during training and we did not have another

authentic development set containing comparable instructions, we were unable to quantify the effect reliably before submission. Consequently, this decision was based on qualitative observations rather than a systematic evaluation.

5 Inference and Tools

We use Transformers (Wolf et al., 2020) for model loading, PEFT (Mangrulkar et al., 2022) for LoRA adapters, TRL (von Werra et al., 2020) for supervised and preference trainers, bitsandbytes for quantized loading and optimization, and vLLM (Kwon et al., 2023) for all model inference. We DSPy (Khatab et al., 2022, 2024) for the direct-assessment tool. Preliminary QLoRA hyperparameter experiments use Optuna (Akiba et al., 2019) with a seeded multivariate TPE sampler.

For submission inference, we serve each selected adapter through vLLM with model thinking disabled and greedy decoding. Source text is divided into chunks of at most 128 Qwen tokenizer tokens. Longer inputs are split hierarchically at paragraph, newline, sentence, and finally whitespace or token boundaries; the recorded separators are restored after translation. The outputs are then reinserted into the shared-task JSONL structure.

6 Model Weights and Training Data

We publish the weights of the CUNI-MH-v3 model on HuggingFace as separate LoRA adapters per bidirectional language pair: en-cs³ cs-vi⁴ cs-uk⁵ cs-de.⁶

The training data for CUNI-MH-v3 SFT stages will also be published.⁷

7 Limitations

The Gemma-based DA results scores be interpreted cautiously because variants of the same model family were also used for synthetic-data generation, filtering, preference-pair construction, and candidate generation.

³<https://huggingface.co/hrabalm/CUNI-MH-v3-encs>

⁴<https://huggingface.co/hrabalm/CUNI-MH-v3-csvi>

⁵<https://huggingface.co/hrabalm/CUNI-MH-v3-csuk>

⁶<https://huggingface.co/hrabalm/CUNI-MH-v3-csde>

⁷<https://huggingface.co/datasets/hrabalm/wmt26-train>

System	FLORES+ dev			EdUKate dev		
	BLEU	chrF2	DA	BLEU	chrF2	DA
<i>Baselines</i>						
WMT25 CUNI-EdUKate	23.51	53.08	96.39	37.98	63.70	89.58
Gemma-4-E2B, direct	9.43	42.46	67.37	7.90	40.97	58.25
Gemma-4-E2B, DSPy	14.27	39.53	69.91	21.65	41.55	60.83
Gemma-4-E4B, direct	21.24	51.26	91.65	30.95	58.72	81.76
Gemma-4-E4B, DSPy	22.05	51.57	92.23	33.36	59.01	80.33
<i>Training progression</i>						
Qwen3-14B, base	18.49	48.51	87.96	31.57	57.58	81.90
CUNI-MH-v3, Stage 1	20.10	50.33	91.77	34.79	60.92	86.49
CUNI-MH-v3, Stage 2	20.14	50.23	91.95	34.15	60.37	86.59
Granite-4.1-8B, base	15.80	46.52	79.57	11.92	45.92	67.88
CUNI-EdUKate, Stage 1	19.48	49.92	90.70	32.33	58.77	82.50
CUNI-EdUKate, Stage 2	19.38	49.73	88.99	31.21	57.85	82.36
<i>Submitted checkpoints</i>						
CUNI-MH-v3, it. 2	20.11	50.07	92.06	33.88	60.12	86.51
CUNI-EdUKate, it. 1	19.72	49.96	89.45	31.07	58.03	82.89
<i>Iteration 10</i>						
CUNI-MH-v3, it. 10	19.89	49.95	91.60	33.88	59.91	86.07

Table 6: Czech–Ukrainian development results.

8 Conclusion

We described two CUNI submissions submitted to the WMT26 General MT shared task. Both use a large synthetic sentence-level SFT stage, a smaller SFT stage on mixture of sentence-level, paragraph-level, and augmented data, targeting terminology translation, Followed by iterative preference optimization of QLoRA adapters. CUNI-MH-v3 uses Qwen3-14B for four directions, while CUNI-EdUKATE uses Granite-4.1-8B and additional EdUKate data for Czech–German and Czech–Ukrainian. For the selected checkpoints we evaluated internally, preference optimization changed the metric trade-off but did not consistently improve every automatic metric.

Acknowledgments

The authors acknowledge the support of the National Recovery Plan funded project MPO 60273/24/21300/21000 CEDMO 2.0 NPO.

This work has been supported by the grants CZ.02.01.01/00/23_020/0008518 (Jazykověda, umělá inteligence a jazykové a řečové technologie: od výzkumu k aplikacím),

The work has also been supported by the Ministry of Education, Youth and Sports of the Czech Republic, Project No. LM2023062 LINDAT/CLARIAH-CZ.

System	FLORES+ dev		
	BLEU	chrF2	DA
<i>Baselines</i>			
Gemma-4-E2B, direct	21.68	49.27	79.11
Gemma-4-E2B, DSPy	18.38	34.60	52.08
Gemma-4-E4B, direct	31.83	52.93	92.37
Gemma-4-E4B, DSPy	33.41	53.41	92.85
<i>Training progression</i>			
Qwen3-14B, base	29.77	51.80	93.59
CUNI-MH-v3, Stage 1	31.44	52.82	94.85
CUNI-MH-v3, Stage 2	31.80	52.87	94.51
<i>Submitted checkpoint</i>			
CUNI-MH-v3, it. 1	31.69	53.03	94.85
<i>Last available PO checkpoint</i>			
CUNI-MH-v3, it. 6	30.99	53.31	95.66

Table 7: Czech–Vietnamese development results. We ended the run at iteration 6, so no iteration-10 checkpoint exists.

System	FLORES+ dev			WMT23		
	BLEU	chrF2	DA	BLEU	chrF2	DA
<i>Baselines</i>						
WMT25 CUNI-MH-v2	33.19	61.07	93.45	33.51	60.58	87.78
Gemma-4-E2B, direct	9.53	44.70	51.93	6.88	39.14	39.78
Gemma-4-E2B, DSPy	22.93	51.25	67.92	25.31	51.39	65.99
Gemma-4-E4B, direct	26.41	55.03	81.70	27.24	55.65	78.39
Gemma-4-E4B, DSPy	27.16	55.07	82.23	30.13	56.09	78.85
<i>Training progression</i>						
Qwen3-14B, base	25.16	53.44	81.57	27.85	54.17	77.87
CUNI-MH-v3, Stage 1	27.95	55.66	86.64	31.41	56.94	84.03
CUNI-MH-v3, Stage 2	27.51	55.44	86.37	31.33	57.03	83.42
<i>Submitted checkpoint</i>						
CUNI-MH-v3, it. 2	27.47	55.24	86.53	31.12	56.83	83.58
<i>Iteration 10</i>						
CUNI-MH-v3, it. 10	26.67	54.56	86.33	30.71	56.18	83.48

Table 8: English–Czech development results.

AI Use Declaration

We used OpenAI GPT models as code assistants during the development of the experiment code and to refine the phrasing of portions of the text.

References

- Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. 2019. Optuna: A next-generation hyperparameter optimization framework. In *The 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2623–2631.
- Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2023. Qlora: efficient finetuning of quantized llms. In *Proceedings of the 37th International Conference on Neural Information Processing Systems, NIPS ’23*, Red Hook, NY, USA. Curran Associates Inc.
- Miroslav Hrabal, Josef Jon, Martin Popel, and Ondřej Bojar. 2025. [CUNI at WMT25 general translation task](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 680–687, Suzhou, China. Association for Computational Linguistics.
- Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. [LoRA: Low-rank adaptation of large language models](#). In *International Conference on Learning Representations*.
- IBM Research. 2026. Granite 4.1 language models. <https://huggingface.co/blog/ibm-granite/granit-4-1>. Accessed: 2026-04-28.
- Seungjae Jung, Gunsoo Han, Daniel Wontae Nam, and Kyoung-Woon On. 2025. [Binary classifier optimization for large language model alignment](#). *Preprint*, arXiv:2404.04656.
- Omar Khattab, Keshav Santhanam, Xiang Lisa Li, David Hall, Percy Liang, Christopher Potts, and Matei Zaharia. 2022. Demonstrate-search-predict: Composing retrieval and language models for knowledge-intensive NLP. *arXiv preprint arXiv:2212.14024*.
- Omar Khattab, Arnav Singhvi, Paridhi Maheshwari, Zhiyuan Zhang, Keshav Santhanam, Sri Vardhamanan, Saiful Haq, Ashutosh Sharma, Thomas T. Joshi, Hanna Moazam, Heather Miller, Matei Zaharia, and Christopher Potts. 2024. Dspy: Compiling declarative language model calls into self-improving pipelines.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.
- Ji Lin, Jiaming Tang, Haotian Tang, Shang Yang, Wei-Ming Chen, Wei-Chen Wang, Guangxuan Xiao, Xingyu Dang, Chuang Gan, and Song Han. 2026. [Awq: Activation-aware weight quantization for llm compression and acceleration](#). *Preprint*, arXiv:2306.00978.
- Sourab Mangrulkar, Sylvain Gugger, Lysandre Debut, Younes Belkada, Sayak Paul, Benjamin Bossan, and Marian Tietz. 2022. PEFT: State-of-the-art parameter-efficient fine-tuning methods. <https://github.com/huggingface/peft>.
- Stephan Oepen, Nikolay Arefev, Mikko Aulamo, Marta Bañón, Maja Buljan, Laurie Burchell, Lucas Charpentier, Pinzhen Chen, Mariya Fedorova, Ona de Gibert, Barry Haddow, Jan Hajič, Jindřich Helcl, Andrey Kutuzov, Veronika Laippala, Zihao Li, Risto Luukkonen, Bhavitvya Malik, Vladislav Mikhailov, and 13 others. 2026. [Hplt 3.0: Very large-scale multilingual resources for llms and mt. mono- and bi-lingual](#)

data, multilingual evaluation, and pre-trained models. *Preprint*, arXiv:2511.01066.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a method for automatic evaluation of machine translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

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.

Haoxiang Sun, Ruize Gao, Pei Zhang, Baosong Yang, and Rui Wang. 2025. [Enhancing machine translation with self-supervised preference data](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 23916–23934, Vienna, Austria. Association for Computational Linguistics.

Gemma Team, Sherif El Abd, Vaibhav Aggarwal, Robin Algayres, Alek Andreev, Olivier Bachem, Ian Ballantyne, Cormac Brick, Victor Cărbune, Michelle Casbon, Mayank Chaturvedi, Aditya Chawla, Victor Cotruta, Alice Coucke, Phil Culliton, Robert Dadashi, Lucas Dixon, Mohamed Elhawaty, Utku Evci, and 304 others. 2026. [Gemma 4 technical report](#). *Preprint*, arXiv:2607.02770.

Qwen Team. 2025. [Qwen3 technical report](#). *Preprint*, arXiv:2505.09388.

Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Galouédec. 2020. [TRL: Transformers Reinforcement Learning](#).

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, and 3 others. 2020. [Transformers: State-of-the-art natural language processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

A Training Prompt

The user portion of the common training prompt is:

```
### Instruction
Translate Input from {source
    language} to {target language}
### Context
{context}
### Input
{source}
### Response
```

The reference translation is supplied as the assistant response.