

Got Compute, but No Data: Lessons From Post-training a Finnish LLM

Elaine Zosa¹ Ville Komulainen² Sampo Pyysalo²

¹Silo AI, Finland ²TurkuNLP, University of Turku, Finland
firstname.lastname@{silo.ai, utu.fi}

Abstract

As LLMs gain more popularity as chatbots and general assistants, methods have been developed to enable LLMs to follow instructions and align with human preferences. These methods have found success in the field, but their effectiveness has not been demonstrated outside of high-resource languages. In this work, we discuss our experiences in post-training an LLM for instruction-following for English and Finnish. We use a multilingual LLM to translate instruction and preference datasets from English to Finnish. We perform instruction tuning and preference optimization in English and Finnish and evaluate the instruction-following capabilities of the model in both languages. Our results show that with a few hundred Finnish instruction samples we can obtain competitive performance in Finnish instruction-following. We also found that although preference optimization in English offers some cross-lingual benefits, we obtain our best results by using preference data from both languages. We release our model, datasets, and recipes under open licenses at <https://huggingface.co/LumiOpen/Poro-34B-chat-OpenAssistant>.

1 Introduction

Foundational LLMs are language completion models that need to be finetuned after pretraining to be able to respond to user questions and follow instructions (Ouyang et al., 2022). This post-training process involves *supervised finetuning* where the model is trained to act as an assistant by training on a dataset of prompt-response pairs. *Preference optimization* further aligns the model

to human preferences such as helpfulness, harmlessness, and honesty (Bai et al., 2022). These methods have resulted in LLMs becoming more capable of answering complex questions involving reasoning, coding, math, and science (e.g., Dubey et al., 2024; Jiang et al., 2024; Team et al., 2024). The effectiveness of these methods, however, have not been demonstrated for smaller and less-resourced languages, such as Finnish.

One of the major challenges we face in post-training in smaller languages is the scarcity of training data. The situation is even more challenging for commercial settings as most of the datasets available today are generated by LLMs with restrictive licenses. The availability of evaluation benchmarks for chat models in small languages is also an issue. Popular benchmarks such as MT-Bench (Zheng et al., 2024) and IFEval (Zhou et al., 2023) are designed for English models and have not been adapted for use in a multilingual setting.¹

In this paper, we discuss our experiences in post-training an LLM in Finnish and English. We use the LLM that we want to finetune to machine-translate instruction and preference datasets into Finnish. We use a commercial machine-translation service to translate a widely-used instruction-following evaluation (IFEval) benchmark into Finnish. We experimented with different combinations of Finnish and English data in instruction tuning and preference optimization. We also experimented with different methods to improve vanilla instruction tuning.

2 Related Work

The post-training of base LLMs, popularised in InstructGPT (Ouyang et al., 2022), can be broadly divided into two categories: instruction tuning and preference optimization. Instruction tuning,

¹While revising this paper, a multilingual, multi-turn IFEval was released (He et al., 2024), but it does not include Finnish.

also known as supervised finetuning (SFT), trains a base LLM to answer questions and follow instructions by training on a dataset of prompt-response pairs with a language modeling objective. Preference optimization further improves the model’s ability to follow conversations and teaches a model to generate responses that align with human preferences by showing the model samples of desirable and undesirable responses (or a ranking of responses). Direct preference optimization (DPO; Rafailov et al. (2024)) is a reward-free preference optimization technique that optimizes directly on the preference data and does not require training a separate a reward model. It is a popular alternative to reward-based methods such as proximal policy optimization (PPO; Schulman et al. (2017)) because it is less computationally expensive and achieves promising results.

Post-training LLMs in a multilingual setting is an under-explored topic (Üstün et al., 2024; Lai et al., 2023; Martins et al., 2024). Previous studies have experimented with monolingual and multilingual instruction tuning of multilingual base LLMs (Shaham et al., 2024; Chen et al., 2024). These studies show that monolingual instruction tuning transfers some instruction-following capability to the other languages in the model but is dependent on the amount of multilingual data that the base LLM was trained on. A few studies have investigated multilingual preference optimization (Lai et al., 2023; Dang et al., 2024). Lai et al. generated synthetic preference datasets for 26 languages and performed reward-based preference optimization on BLOOM and Llama 7B models. Their results show that preference optimization offers a slight improvement over SFT. Dang et al., however, point out that these preference-optimized models still underperform compared to massively multilingual LLMs that are finetuned only with SFT.

The scarcity of instruction and preference datasets is a major challenge in post-training LLMs for smaller languages. Previous efforts to assemble finetuning datasets through machine translation, crowd-sourcing, and synthetic data generation include (Üstün et al., 2024; Lai et al., 2023; Dang et al., 2024). Evaluating opened responses of chat models in small languages is also a challenge. Previous studies have investigated using LLM-as-a-judge in multilingual settings but these studies focused on standard

NLP tasks such as summarization and question-answering (Hada et al., 2024; Ahuja et al., 2023).

3 Experimental setup

We use Poro 34B as the base LLM (Luukkonen et al., 2024). Poro is trained on 1T tokens of English, Finnish, and code, with 129B tokens for Finnish. We use the Transformer Reinforcement Learning library (TRL; (von Werra et al., 2020)) for instruction tuning and preference optimization. We finetune all of the model parameters in our experiments.²

We use 32 AMD MI250X GPUs in our experiments. For SFT, we use a micro batch size of 4 and a gradient accumulation step of 1, resulting in a global batch size of 128. We use a learning rate of $2e-5$ with a warmup rate of 0.1 and finetune for 3 epochs. For DPO, we use a global batch size of 64, learning rate of $5e-7$, warmup rate of 0.1, and train for 5 epochs.

4 Datasets

SFT We use a curated version OpenAssistant 2 (OASST2; (Köpf et al., 2024)) containing the top-ranked English conversations. This dataset has 4,692 samples.³

We translate OASST2 into Finnish using Poro with few-shot prompting. Poro has been shown to produce higher-quality Finnish translations compared to other open MT systems (Luukkonen et al., 2024). For this reason, we did not experiment with translations from other open MT models and focus our efforts on the Poro-translated dataset. We use heuristics to clean up the translations. After post-translation cleaning, our OASST2 Finnish data has 4,399 samples.

DPO We use the HelpSteer2 preference dataset (Wang et al., 2024), which consists of publicly-sourced prompts and LLM-generated completions⁴. We use the helpfulness scores included in the dataset to obtain 7,221 preference pairs (chosen and rejected responses). We also

²We experimented with LoRA finetuning (Hu et al., 2021), but our results indicated that full finetuning achieved better performance.

³The curated dataset is https://huggingface.co/datasets/sablo/oasst2_curated. The full dataset is <https://huggingface.co/datasets/OpenAssistant/oasst2>.

⁴We initially chose this dataset because it has a commercially-friendly license. Recently, however, Lambert et al. pointed out that HelpSteer2 includes ShareGPT prompts which have a questionable legal provenance.

translate this dataset into Finnish using Poro. After post-translation cleaning, we end up with 6,037 preference pairs in our Finnish HelpSteer2 dataset.

5 Evaluation

We use the Instruction Following (IFEval) benchmark to evaluate instruction-following performance (Zhou et al., 2023). IFEval has 541 prompts where a prompt contains verifiable instructions that can be checked with a deterministic program, circumventing the need of an LLM or human as judge. Examples of instructions include adding keywords to the response, formatting the response in JSON, or responding in a specified language.⁵

We translate the IFEval prompts into Finnish using DeepL⁶. IFEval has 31 prompts that require the response language to be in a language other than English. We exclude these prompts for this work due to Poro being constrained to only English and Finnish. We report the results for the remaining 510 prompts only. IFEval reports strict accuracy and loose accuracy where loose accuracy accepts minor transformations in the responses. For the sake of clarity, we report only the strict accuracy in this work.

We run evaluations through the LM Evaluation Harness (Gao et al., 2024). The translated IFEval is available at https://huggingface.co/datasets/LumiOpen/ifeval_mt.

6 Experiments

Multilingual SFT Finnish instruction data is more difficult to obtain compared to English; therefore, we want to investigate how the amount of Finnish instruction data affects performance. We construct data mixes from the English and Finnish OASST2 datasets such that we start with just the English data and gradually introduce more Finnish samples into the mix starting from 10% of the Finnish data and then going up to 100%. We call these data mixes `en-fi-Xpct` (i.e., the data mix with just the English samples is called `en-fi-0pct` while the data mix with all the English and Finnish samples is called `en-fi-100pct`). By default, we do not mask prompts during training (i.e., we incorporate the

⁵See the IFEval paper for the complete list of instructions and their descriptions.

⁶<https://www.deepl.com/>

data mix	EN (%)	FI (%)	Resp lang (%)
<code>en-fi-0pct</code>	36.39	31.41	47.45
<code>en-fi-10pct</code>	39.97	32.69	90.00
<code>en-fi-20pct</code>	37.67	28.60	93.52
<code>en-fi-40pct</code>	39.20	30.90	96.27
<code>en-fi-60pct</code>	39.20	32.95	94.90
<code>en-fi-80pct</code>	38.56	33.84	96.27
<code>en-fi-100pct</code>	39.97	34.48	95.68

Table 1: Instruction-level accuracy on English and Finnish IFEval of the SFT models trained on different data mixes. Response language refers to the proportion of responses classified as Finnish for the Finnish IFEval.

losses from the prompt and completion tokens). We train SFT models on all the data mixes using the same hyperparameters.

Improving vanilla SFT We investigate whether we can improve SFT by adding noise to the word embeddings in the instruction data (NEFTune; Jain et al. (2024)). We also experiment with *prompt masking* where the loss is calculated only on the completion tokens. Our baseline for these experiments is the SFT model trained on the `en-fi-100pct` data mix.

Multilingual DPO We opt to use DPO for preference tuning as it has been found to fare better in IFEval, in addition to being more stable and requiring less compute (Dubey et al., 2024). We tuned the β parameter of DPO with values $\{0.01, 0.05, 0.1\}$ and found $\beta = 0.05$ to be optimal. We experiment with using either the English or Finnish datasets and using both. As our baseline, we use the SFT model trained on the `en-fi-100pct` data mix.

7 Results and Discussion

Multilingual SFT In Table 1, we show the instruction-level accuracy of the SFT models on the English and Finnish IFEval. We also show the proportion of responses to Finnish IFEval that are in Finnish as classified by `langdetect`⁷. For English IFEval, the performance is comparable across the data mixes which is expected because the different data mixes contain same number of English samples. For Finnish, the best performance is from the data mix with all the English

⁷<https://pypi.org/project/langdetect/>

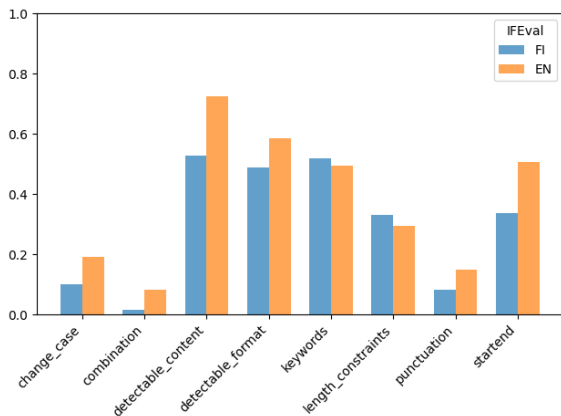


Figure 1: Accuracy by instruction category on English and Finnish IFEval of SFT model trained on the en-fi-100pct data mix.

and Finnish samples (en-fi-100pct). However, compared with the data mix with just 10% of the Finnish data (around 400 samples), the difference is less than 2 percentage points. If we finetune using only the English data, the resulting model can still follow Finnish instructions but less than half of the responses are in Finnish, which is not desirable.

For the response language, SFT models trained on data mixes containing 20% and above of the Finnish data have comparable rates of Finnish responses of over 93%. We reviewed the 22 responses from the en-fi-100pct model that were classified as *not* Finnish. We found that 19 of them are in Finnish but mixed with other languages such as English and German. We also found that the model tends to respond in a mixture of English and Finnish when asked to respond in a specific format, such as JSON or XML. This is likely because JSON tends to be treated as code in the instruction dataset and our translation pipeline did not translate code blocks which sometimes include comments in English.

Figure 1 shows the accuracy by instruction category of the en-fi-100pct model. The model struggles most with the combination category—this category includes combined instructions such as giving two responses that are separated by a given separator or repeating the prompt without modification before giving the response proper. The poor performance is probably because the instruction contains multiple steps that must be followed in order to give a correct response. For instance, if the model gives two responses but these

Model	EN (%)	FI (%)	Resp lang (%)
<i>baseline</i>	39.97	34.48	95.68
prompt masking	39.84	32.56	96.27
NEFTune	38.05	32.69	96.47
DPO-en	43.55	36.65	94.70
DPO-fi	41.76	36.01	95.49
DPO-both	44.69	37.80	95.49

Table 2: Instruction-level accuracy on English and Finnish IFEval English for the prompt masking, NEFTune, and DPO experiments. The baseline is an SFT model trained on the en-fi-100pct data mix.

responses are numbered instead of separated by an asterisk, the response is considered incorrect. The model also struggles with punctuation instructions such as avoiding the use of commas likely because texts without commas are rare in the dataset.

Overall, the results indicate that finetuning with a few hundred Finnish instruction samples achieves results close to finetuning with ten times that amount. In terms of instruction types, the model struggles with multi-step instructions and unusual instruction types. Previous work has indicated that carefully curating the instruction dataset is vital to a strong SFT model (Zhou et al., 2024). In future, we aim for a smaller but higher-quality data by, for instance, removing highly similar prompts, diversifying tasks, and curating the sources of the samples.

Prompt masking and NEFTune In Table 2 we show the accuracy of the models trained with prompt masking and NEFTune compared to the baseline model.

In our experiments, models trained with NEFTune fail to achieve better scores compared to the plain vanilla SFT baseline. As noted by Jain et al., the performance of NEFTune was found to be dataset dependent. One key difference in our study is that we train on a multi-turn dataset. We leave further examination of NEFTune and other noise augmentation techniques for future work. We find that prompt masking does not improve over the baseline. This result is in line with findings from Shi et al. where they show that incorporating the loss from the prompt is beneficial for smaller datasets such as LIMA (Zhou et al., 2024) with 1,030 examples.

Multilingual DPO Table 2 shows the results from our DPO experiments compared to the SFT model. The model optimized only on English preference data (DPO-en) improved performance on English IFEval by around 3 percentage points and also showed some improvement in Finnish IFEval. This provides further evidence that preference optimization in English benefits other languages in the model (Dang et al., 2024). The DPO model trained only on Finnish (DPO-fi), on the other hand, showed smaller improvements on both English and Finnish IFEval and, in fact, has slightly lower performance than DPO-en on the Finnish benchmark. The model trained on both languages (DPO-both) achieved the best performance on both benchmarks but compared to DPO-en, the improvements are marginal.

In terms of the response language, DPO did not improve the Finnish response rates compared to the SFT model. This might be because we optimized the model on monolingual preference pairs (the chosen and rejected responses are in the same language). Improving the response language of multilingual models through preference optimisation is an area we will explore in future work.

8 Conclusions and Future Work

In this work we share our findings from post-training Poro 34B in English and Finnish. Due to the scarcity of Finnish post-training datasets we opted to machine-translate instruction and preference datasets using Poro. To evaluate the results of our experiments, we translate IFEval, a widely-used instruction-following evaluation benchmark. We experimented with using different combinations of English and Finnish data in SFT and found that using all available data from both languages gave the best performance overall. Using only 10% of the Finnish instruction data (around 400 samples), however, still gives competitive performance. We contribute to Finnish LLM development by releasing our datasets, recipes, and model with open licenses at <https://huggingface.co/LumiOpen/Poro-34B-chat-OpenAssistant>.

In future we want to explore different ways of obtaining more Finnish data by, for instance, generating synthetic instruction and preference datasets. We will use these synthetic datasets to further investigate other alignment and finetuning methods. Additionally, we are interested on de-

veloping an evaluation benchmark for open-ended conversations in Finnish that takes cultural and linguistic nuances into account.

Acknowledgments

The authors wish to acknowledge CSC – IT Center for Science, Finland, for generous computational resources on the LUMI supercomputer. This project has received funding from the European Union’s Horizon Europe research and innovation programme under Grant agreement No 101070350. The contents of this publication are the sole responsibility of its authors and do not necessarily reflect the opinion of the European Union.

References

- Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Mohamed Ahmed, et al. 2023. Mega: Multilingual evaluation of generative ai. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 4232–4267.
- Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, et al. 2022. Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.
- Pinzhen Chen, Shaoxiong Ji, Nikolay Bogoychev, Andrey Kutuzov, Barry Haddow, and Kenneth Heafield. 2024. Monolingual or multilingual instruction tuning: Which makes a better alpaca. In *The 18th Conference of the European Chapter of the Association for Computational Linguistics*, pages 1–10. Association for Computational Linguistics.
- John Dang, Arash Ahmadian, Kelly Marchisio, Julia Kreutzer, Ahmet Üstün, and Sara Hooker. 2024. Rllhf can speak many languages: Unlocking multilingual preference optimization for llms. *CoRR*.
- Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.
- Leo Gao, Jonathan Tow, Baber Abbasi, Stella Biderman, Sid Black, Anthony DiPofi, Charles Foster, Laurence Golding, Jeffrey Hsu, Alain Le Noac’h, Haonan Li, Kyle McDonell, Niklas Muennighoff, Chris Ociepa, Jason Phang, Laria Reynolds, Hailey Schoelkopf, Aviya Skowron, Lintang Sutawika, Eric Tang, Anish Thite, Ben Wang, Kevin Wang, and Andy Zou. 2024. A framework for few-shot language model evaluation.

- Rishav Hada, Varun Gumma, Adrian Wynter, Harshita Diddee, Mohamed Ahmed, Monojit Choudhury, Kalika Bali, and Sunayana Sitaram. 2024. Are large language model-based evaluators the solution to scaling up multilingual evaluation? In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1051–1070.
- Yun He, Di Jin, Chaoqi Wang, Chloe Bi, Karishma Mandyam, Hejia Zhang, Chen Zhu, Ning Li, Tengyu Xu, Hongjiang Lv, et al. 2024. Multi-if: Benchmarking llms on multi-turn and multilingual instructions following. *arXiv preprint arXiv:2410.15553*.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- Neel Jain, Ping-yeh Chiang, Yuxin Wen, John Kirchenbauer, Hong-Min Chu, Gowthami Somepalli, Brian R Bartoldson, Bhavya Kailkhura, Avi Schwarzschild, Aniruddha Saha, et al. 2024. Nef-tune: Noisy embeddings improve instruction fine-tuning. In *The Twelfth International Conference on Learning Representations*.
- Albert Q Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, et al. 2024. Mixtral of experts. *arXiv preprint arXiv:2401.04088*.
- Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi Rui Tam, Keith Stevens, Abdullah Barhoum, Duc Nguyen, Oliver Stanley, Richárd Nagyfi, et al. 2024. Openassistant conversations-democratizing large language model alignment. *Advances in Neural Information Processing Systems*, 36.
- Viet Lai, Chien Nguyen, Nghia Ngo, Thut Nguyn, Franck Dernoncourt, Ryan Rossi, and Thien Nguyen. 2023. Okapi: Instruction-tuned large language models in multiple languages with reinforcement learning from human feedback. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 318–327.
- Nathan Lambert, Jacob Morrison, Valentina Pyatkin, Shengyi Huang, Hamish Ivison, Faeze Brahman, Lester James V Miranda, Alisa Liu Nouha Dziri Xinxi Lyu, Yuling Gu Saumya Malik Victoria Graf, Jena D Hwang, et al. 2024. Tulu 3: Pushing frontiers in open language model post-training. *arXiv preprint arXiv:2411.15124*.
- Risto Luukkonen, Jonathan Burdge, Elaine Zosa, Aarne Talman, Ville Komulainen, Väinö Hatanpää, Peter Sarlin, and Sampo Pyysalo. 2024. Poro 34b and the blessing of multilinguality. *arXiv preprint arXiv:2404.01856*.
- Pedro Henrique Martins, Patrick Fernandes, João Alves, Nuno M Guerreiro, Ricardo Rei, Duarte M Alves, José Pombal, Amin Farajian, Manuel Faysse, Mateusz Klimaszewski, et al. 2024. Eurollm: Multilingual language models for europe. *arXiv preprint arXiv:2409.16235*.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.
- Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2024. Direct preference optimization: Your language model is secretly a reward model. *Advances in Neural Information Processing Systems*, 36.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Uri Shaham, Jonathan Herzig, Roei Aharoni, Idan Szepes, Reut Tsarfaty, and Matan Eyal. 2024. Multilingual instruction tuning with just a pinch of multilinguality. *arXiv preprint arXiv:2401.01854*.
- Zhengyan Shi, Adam X. Yang, Bin Wu, Laurence Aitchison, Emine Yilmaz, and Aldo Lipani. 2024. Instruction tuning with loss over instructions.
- Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, et al. 2024. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv:2408.00118*.
- Ahmet Üstün, Viraat Aryabumi, Zheng-Xin Yong, Wei-Yin Ko, Daniel D’souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, et al. 2024. Aya model: An instruction finetuned open-access multilingual language model. *arXiv preprint arXiv:2402.07827*.
- Zhilin Wang, Yi Dong, Olivier Delalleau, Jiaqi Zeng, Gerald Shen, Daniel Egert, Jimmy J Zhang, Makesh Narsimhan Sreedhar, and Oleksii Kuchaiev. 2024. Helpsteer2: Open-source dataset for training top-performing reward models. *arXiv preprint arXiv:2406.08673*.
- Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, Shengyi Huang, Kashif Rasul, and Quentin Gallouédec. 2020. Trl: Transformer reinforcement learning. <https://github.com/huggingface/trl>.
- Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2024. Judging llm-as-a-judge with mt-bench and chatbot

arena. *Advances in Neural Information Processing Systems*, 36.

Chunting Zhou, Pengfei Liu, Puxin Xu, Srinivasan Iyer, Jiao Sun, Yuning Mao, Xuezhe Ma, Avia Efrat, Ping Yu, Lili Yu, et al. 2024. Lima: Less is more for alignment. *Advances in Neural Information Processing Systems*, 36.

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