TAIWAN LLM : Bridging the Linguistic Divide with a Culturally Aligned Language Model

Yen-Ting Lin **Yun-Nung Chen**
Department of Computer Science and Information Engineering
National Taiwan University
{ytl, y.v.chen}@ieee.org
https://twllm.com

Abstract

In the realm of language models, the nuanced linguistic and cultural intricacies of Traditional Chinese, as spoken in Taiwan, have been largely overlooked. This paper introduces TAIWAN LLM , a pioneering Large Language Model that specifically caters to the Traditional Chinese language, with a focus on the variant used in Taiwan. Leveraging a comprehensive pretraining corpus and instruction-finetuning datasets, we have developed a model that not only understands the complexities of Traditional Chinese but also embodies the cultural context of Taiwan. TAIWAN LLM represents the first of its kind, a model that is not only linguistically accurate but also culturally resonant with its user base. Our evaluations demonstrate that TAIWAN LLM achieves superior performance in understanding and generating Traditional Chinese text, outperforming existing models that are predominantly trained on Simplified Chinese or English. The open-source release of TAIWAN LLM invites collaboration and further innovation, ensuring that the linguistic diversity of Chinese speakers is embraced and well-served. The model, datasets, and further resources are made publicly available to foster ongoing research and development in this field.

1 Introduction

The linguistic diversity of the Chinese language, with its various dialects and written forms, presents unique challenges in natural language processing. Traditional Chinese, used predominantly in Taiwan, Hong Kong, and Macau, has been notably underrepresented in the development of language models. This oversight has resulted in a technological divide, where speakers of Traditional Chinese lack access to high-quality language models that understand their language nuances and cultural context. To address this gap, we introduce TAIWAN LLM , the first Large Language Model (LLM) designed specifically for the Traditional Chinese language as used in Taiwan.

Recent advancements in LLMs have seen models like GPT-4 and Claude2 achieve remarkable performance in understanding and generating text. However, these models are predominantly trained on English and Simplified Chinese (zh-cn), leading to a misalignment with the linguistic intricacies of Traditional Chinese (zh-tw). This misalignment not only affects the accuracy of language generation but also overlooks the rich cultural heritage embedded within the language, which is crucial for applications such as digital humanities and culturally sensitive communication.

In this work, we focus on developing an LLM that is both linguistically and culturally aligned with Traditional Chinese speakers in Taiwan. We leverage a large-scale pretraining corpus that encompasses a wide range of text genres, from literature to colloquial dialogues, ensuring that our model captures the breadth of Traditional Chinese language use. Additionally, we employ instruction-finetuning datasets that incorporate cultural nuances and regional idiomatic expressions, further enhancing the model's alignment with Taiwanese users' expectations.

TAIWAN LLM sets a new precedent for language models by providing a culturally resonant tool for the Traditional Chinese-speaking community. Our evaluations demonstrate that TAIWAN LLM not only excels in standard language understanding and generation tasks but also shows a deep comprehension of cultural references and social norms specific to Taiwan. This breakthrough is a step towards bridging the linguistic divide and offering equitable access to language technologies for all Chinese speakers.

The contributions of this paper are manifold: we present the first LLM that is fine-tuned to the linguistic and cultural context of Traditional Chinese as spoken in Taiwan; we provide a comprehensive evaluation that showcases the model's superior performance over existing models; and we release our model, datasets, and resources to the public, fostering further research and development in this area. Models, code, and instructions are available at `https://github.com/MiuLab/Taiwan-LLaMa`.

2 Related Work

The landscape of open large language models (LLMs) has expanded rapidly, providing valuable resources for the research community. The release of models like ChatGPT and LLaMA [23] has catalyzed a wave of research into efficient fine-tuning, context management, and generative capabilities. Subsequent models such as MosaicML's MPT [16], Together AI's RedPajama-INCITE [2], TII's Falcon [17], Meta's Llama 2 [23], and Mistral 7B [12] have continued this trend, each contributing unique features and improvements. Our work with TAIWAN LLM builds upon these developments, particularly leveraging the strong performance of Llama 2 as a foundation.

The evolution of LLMs has also seen a focus on enhancing small model performance through techniques like distillation from larger counterparts. Initiatives such as the self-instruct method [25], Alpaca [21], and Vicuna [5] have paved the way for refined distillation processes. While these efforts have concentrated on distilling the supervised fine-tuning (SFT) stage, our approach encompasses both SFT and preference optimization. Notably, models like WizardLM [28] and Xwin-LM [22] have explored advanced methods beyond dSFT, including distilling preference optimization through PPO [20], which we benchmark against in our evaluations.

As LLMs have grown more powerful, so have the tools for their evaluation. Benchmarks like the LMSYS chatbot arena [29], AlpacaEval [9], and LLM-Eval [14] utilize either human judgment or other LLMs like GPT-4 and Claude to assess model outputs. MTBench [29] represents another step forward, employing GPT-4 to score responses across a variety of tasks. Additional evaluation platforms include the HuggingFace Open LLM leaderboard [3], the Chain-of-Thought Hub [10], and FastEval [1]. Our work with TAIWAN LLM is evaluated on TC-Eval [11] to ensure a comprehensive assessment of its performance.

3 Method

The objective of this work is to create a language model that is finely attuned to the linguistic subtleties and cultural context of Traditional Chinese as used in Taiwan. Our approach, depicted in Figure 1, consists of three main stages: Continue-Pretraining (cPT), Supervised Fine-Tuning (SFT), and Feedback Supervised Fine-Tuning (Feedback SFT).

Continue-Pretraining (cPT) In the continue-pretraining phase, we start with a base language model, say π_{base}, such as Llama2, and further pretrain it on a large-scale Taiwanese corpus, denoted as $\mathcal{C}_{TW}$. The objective of this phase is to maximize the log-likelihood of the Taiwanese corpus under the model, which can be expressed mathematically as:

$$\pi_{cPT} = \arg\max_{\pi} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{C}_{TW}} \log \pi(y|x)$$

where x represents a sequence of tokens and y is the next token to predict.

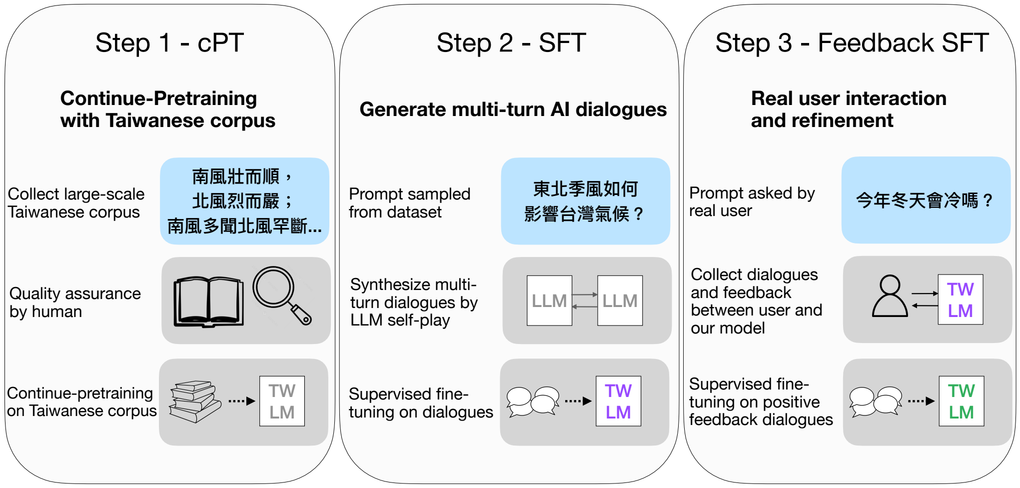

Figure 1: The three-phase methodology for TAIWAN LLM development: (1) cPT - Continue-Pretraining on a large-scale Taiwanese corpus with quality assurance checks, (2) SFT - Supervised Fine-Tuning on multi-turn dialogues through prompt datasets and LLM self-play, and (3) Feedback SFT - Enhancing model performance through real user interactions and subsequent refinement loops, leveraging native speaker insights for cultural and linguistic accuracy.

Supervised Fine-Tuning (SFT) After continue-pretraining, we fine-tune the model on a dataset of multi-turn dialogues, denoted as $\mathcal{D}_{dialogue}$. This dataset consists of pairs of prompts and responses, and the goal is to adjust the model parameters to maximize the likelihood of the responses given the prompts. This can be expressed as:

$$\pi_{SFT} = \arg\max_{\pi} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{D}_{dialogue}} \log \pi(y|x)$$

where x represents a prompt and y is the corresponding response.

The SFT phase helps the model to better understand the nuances of conversation and to generate more coherent and engaging responses. It's worth noting that the SFT phase is performed on top of the continue-pretrained model, π_{cPT}, thus leveraging the broad language understanding acquired in the cPT phase while also honing the model's ability to generate conversational responses.

Feedback Supervised Fine-Tuning (Feedback SFT) In the final stage, we refine the model based on positive feedback from real users. We collect this feedback through a user interface that allows users to interact with the model and provide binary ratings (positive or negative) on its responses. We only keep the instances where the rating is positive for feedback SFT.

This feedback can be represented as a dataset $\mathcal{F}_{pos}$ of pairs (x, y), where x is a prompt, y is the model's response, and the user's rating of the response is positive. The goal of feedback SFT is to adjust the model parameters to maximize the likelihood of the positively-rated responses given the prompts, which can be expressed as:

$$\pi_{FeedbackSFT} = \arg\max_{\pi} \mathop{\mathbb{E}}_{(x,y)\sim\mathcal{F}_{pos}} \log \pi(y|x)$$

This phase is performed on top of the supervised fine-tuned model, π_{SFT}, thus leveraging the conversational abilities acquired in the SFT phase while also incorporating user feedback to better align the model with user preferences.

In future work, we plan to explore more sophisticated training methods such as Reinforcement Learning from Human Feedback (RLHF) and Direct Preference Optimization (DPO) to further improve the model's performance based on user feedback.

4 Experiments

Our fine-tuning experiments are conducted using Llama2 [12], a state-of-the-art base language model that has demonstrated strong performance on various NLP benchmarks. We utilize the Transformer Reinforcement Learning (TRL) library for training [24], coupled with DeepSpeed ZeRO-2 [19] and FlashAttention-2 [8] to optimize memory usage and enhance training speed. All models are trained using the AdamW optimizer without weight decay. All experiments were conducted on up to 48 H100s using bfloat16 precision.

4.1 Datasets

We utilize three primary datasets for the development of TAIWAN LLM :

1. **Continue-Pretraining Corpus:** As detailed in Table 1, our continue-pretraining corpus is a comprehensive collection of documents from various sources, meticulously curated to represent the Taiwanese context. Each source of data is verified by the authors for quality assurance to primarily avoid spam and toxic content. This verification is conducted at the website level rather than the document level, allowing for a more efficient filtering process while maintaining a high standard of data quality. Future work may involve more granular document-level analysis to further enhance the quality of the training corpus.
2. **Supervised Fine-Tuning (SFT) Data:** For SFT, we compile a diverse set of instruction datasets, as shown in Table 2. We translate prompts from various instruction datasets using `gpt-3.5-turbo` and then generate responses to these prompts with the same model. Additionally, the authors have created a unique dataset, *Taiwan Instruction*, based on the experiences with the pilot model. This dataset consists of author-written conversations that are specifically designed to capture the cultural and linguistic nuances of Taiwan, ensuring that the model can handle locally relevant scenarios and idiomatic expressions.
3. **Feedback Supervised Fine-Tuning (Feedback SFT) Data:** We collect 20,000 user feedback instances from a dedicated platform (https://twllm.com) and incorporate this data into the SFT dataset. This feedback is used to further refine the model, ensuring that it aligns with user preferences and expectations.

The continue-pretraining corpus is designed to provide a broad understanding of the language, while the SFT and Feedback SFT datasets are intended to refine the model's ability to engage in dialogue and respond to instructions accurately.

Table 1: The TAIWAN LLM continue-pretraining Corpus.

Data source	Documents	Tokens	Token %
Social media	8.24 millions	16.6 billions	47.32%
News	8.60 millions	10.4 billions	29.56%
Knowledge base	3.19 millions	5.7 billions	16.29%
Books	4 thousands	2.4 billions	6.83%
Total	20.0 millions	35.1 billions	100.00%

4.2 Evaluation

For the evaluation of TAIWAN LLM , we leverage the TC-Eval benchmark suite [11], which provides a comprehensive set of tasks tailored for assessing the capabilities of Traditional Chinese language models. The suite includes benchmarks for contextual question-answering, summarization, classification, and table understanding. While we adopt the same metrics used in the TC-Eval study, such as Exact Match (EM) for question-answering tasks and ROUGE-2 for summarization, we adapt the metrics from the TC-Eval study to better suit the free-form responses generated by LLMs.

We assess TAIWAN LLM 's performance on several datasets:

Contextual QA: We use DRCD, a Traditional Chinese machine reading comprehension dataset, and FGC, which contains questions from Taiwanese news articles and government announcements.

Table 2: Supervised fine-tuning datasets used in TAIWAN LLM .

Datasets	Prompts Sourced from	Instances
SuperNI [26]	NLP datasets + Human-written Instructions	18,547
CoT [27]	NLP datasets + Human-written CoTs	35,990
Flan V2 [15]	NLP datasets + Human-written Instructions	16,782
Dolly [7]	Human-written from scratch	14,752
Open Assistant 1 [13]	Human-written from scratch	14,797
Self-instruct	Generated w/ vanilla GPT3 LM	48,409
Unnatural Instructions	Generated w/ Davinci-002	30,268
Alpaca	Generated w/ Davinci-003	41,133
Code-Alpaca [4]	Generated w/ Davinci-003	20,111
GPT4-Alpaca [18]	Generated w/ Davinci-003 + GPT4	22,472
Baize	Generated w/ ChatGPT	67,699
ShareGPT[1]	User prompts + outputs from various models	79,762
Evol Instruction	Generated w/ ChatGPT	33,798
Airoboros	Generated w/ ChatGPT	39,071
Bilingual News Corpus	Traditional Chinese-English Parallel News Texts	13,104
Taiwan Instruction	Author-written Conversations	947
twllm.com Feedback	User-written Conversations	20,00

World Knowledge: We evaluate the model's common sense abilities and knowledge of Taiwanese culture with the TTQA dataset and test its problem-solving skills with the TMMLU dataset.

Summarization: The model's ability to generate abstractive summaries is measured using the XSum-TC dataset.

Classification: We use the IMDB-TC dataset to evaluate sentiment classification capabilities.

Table Understanding: The model's understanding of tabular data is tested with the Penguins-in-a-Table-TC task.

Metrics for evaluation include Exact Match (EM) for tasks requiring precise answers and ROUGE-2 for summarization quality assessment.

TAIWAN LLM is also compared against a range of models, including those specifically fine-tuned for the Traditional Chinese language as well as larger, more general models. This comparison helps to contextualize TAIWAN LLM 's performance relative to the current state-of-the-art.

Table 3: Performance of TAIWAN LLM and baselines on TC-Eval.

Model	DRCD (EM)	FGC (EM)	TTQA (Acc)	TMMLU (Acc)	Xsum (Rouge2)	IMDB (Acc)	Table (Acc)	Avg
TAIWAN LLM 13B	87.57%	50.00%	70.87%	39.04%	5.23%	92.36%	32.89%	**53.99%**
- cPT	75.81%	38.00%	56.31%	36.28%	0.06%	93.94%	26.84%	46.75%
+ CoomonCrawl	70.08%	34.00%	77.67%	31.53%	3.92%	79.36%	26.17%	46.11%
TAIWAN LLM 7B	84.11%	46.00%	54.37%	30.03%	4.60%	86.04%	28.19%	**47.62%**
- Feedback SFT	84.51%	42.00%	55.34%	32.01%	5.21%	87.50%	25.50%	47.44%
+ CommonCrawl	69.80%	36.00%	48.54%	27.49%	4.37%	59.82%	20.13%	38.02%
`GPT-4`	96.68%	42.00%	53.40%	60.48%	4.30%	86.90%	62.42%	58.03%
`GPT-3.5 turbo`	90.75%	42.00%	55.34%	55.61%	3.71%	91.38%	38.93%	53.96%
`Claude-2.1`	93.56%	40.00%	68.93%	59.86%	3.57%	94.52%	48.99%	**58.49%**
`Claude-instant-1.2`	78.99%	34.00%	68.93%	55.28%	4.21%	91.14%	47.65%	54.31%
`Llama-2-13b-chat`	45.95%	18.00%	59.22%	35.36%	0.00%	52.40%	27.52%	34.06%
`Llama-2-7b-chat`	28.60%	12.00%	50.49%	29.34%	0.00%	51.04%	18.12%	27.08%

5 Results and Ablations

The results of our experiments, as shown in Table 3, indicate that TAIWAN LLM , with its 13 billion parameters, achieves competitive performance when compared to proprietary models such as GPT-4 and Claude-2.1. Notably, TAIWAN LLM 's average performance across all tasks is 53.99%, which is comparable to GPT-3.5 turbo's 53.96%. This demonstrates the effectiveness of our approach in creating a model that is well-suited for Traditional Chinese, despite having fewer parameters than some proprietary models.

5.1 Impact of Continue-Pretraining (cPT)

Our ablation study reveals the significant impact of continue-pretraining (cPT) on TAIWAN LLM 's performance. When we remove the cPT phase, there is a noticeable drop in performance across most tasks. For instance, the EM score on the DRCD dataset decreases from 87.57% to 75.81%, and the accuracy on the TTQA dataset drops from 70.87% to 56.31%. This highlights the importance of cPT in adapting the model to the linguistic characteristics of Traditional Chinese.

5.2 Impact of Feedback Supervised Fine-Tuning (Feedback SFT)

We also investigate the effect of incorporating user feedback through Feedback SFT. The results show that this phase contributes to the model's performance, albeit to a lesser extent than cPT. For example, the 7 billion parameter version of TAIWAN LLM without Feedback SFT achieves an average performance of 47.44%, only slightly lower than the 47.62% with Feedback SFT. This suggests that while Feedback SFT helps in fine-tuning the model to user preferences, the core capabilities are largely established during the cPT and SFT phases.

5.3 Impact of Adding Web Crawl Data

When we added approximately 9 billion tokens of CommonCrawl data, specifically filtered for Traditional Chinese content (zh-tw), the performance of TAIWAN LLM unexpectedly declined. For the 13 billion parameter model, the Exact Match (EM) score on the DRCD dataset decreased from 87.57% to 70.08%, and the accuracy on the TTQA dataset saw a slight increase to 77.67%, indicating a complex impact. The 7 billion parameter model experienced a more significant drop, with the average performance across all tasks falling from 47.62% to 38.02%.

This reduction in performance, despite the substantial volume of additional tokens, underscores the critical role of data quality over sheer quantity. The introduction of web crawl data, which may contain a mix of high-quality and lower-quality content, can introduce noise that detracts from the model's ability to accurately process and generate Traditional Chinese text. This result highlights the necessity of meticulous data curation to ensure the training material aligns with the linguistic and cultural nuances of the target language.

5.4 Comparison with Open-Source Models

When compared to the original Llama-2 models, which have not been specifically trained on Traditional Chinese, TAIWAN LLM demonstrates a clear advantage. The original Llama-2-13b-chat model achieves an average performance of 34.06%, significantly lower than TAIWAN LLM 's 53.99%. This underscores the necessity of targeted training on Traditional Chinese to ensure high-quality performance for users in Taiwan.

5.5 Comparison with Proprietary Models

Our 13 billion parameter TAIWAN LLM is on par with proprietary models like GPT-3.5 turbo, which is a testament to the effectiveness of our culturally and linguistically tailored training approach. While proprietary models like GPT-4 and Claude-2.1 still lead in some areas, TAIWAN LLM 's competitive performance suggests that with continued refinement and expansion of training data, open-source models can achieve parity with or even surpass proprietary models for specific languages and regions.

The results confirm that TAIWAN LLM effectively supports Traditional Chinese, thanks to our specialized continue-pretraining and fine-tuning for Taiwanese usage. Its release is a major advancement for underrepresented language communities.

6 Conclusion

In conclusion, TAIWAN LLM represents a significant stride forward in the development of language models tailored to the linguistic and cultural nuances of Traditional Chinese as spoken in Taiwan. Our comprehensive approach, encompassing continue-pretraining on a curated Taiwanese corpus, supervised fine-tuning with instruction datasets, and refinement through user feedback, has culminated in a model that not only understands the complexities of the language but also resonates with the cultural context of its speakers. The evaluations demonstrate that TAIWAN LLM achieves competitive performance, particularly when compared to larger, proprietary models trained on more generalized datasets. This underscores the importance of specialized training for language models to effectively serve specific linguistic communities.

The open-source release of TAIWAN LLM invites collaboration and innovation, setting a precedent for future research in language technology. By providing a model that is both linguistically accurate and culturally aligned, we aim to bridge the gap in language technology for Traditional Chinese speakers and contribute to the global effort of ensuring equitable access to high-quality language models. As we continue to refine TAIWAN LLM and expand its capabilities, we anticipate that it will not only serve as a valuable tool for the Taiwanese community but also inspire similar initiatives for other underrepresented languages around the world.

Acknowledgements

We thank Ubitus K.K. for the GPU compute grant.

Limitation

We acknowledge that while our model represents a significant advancement in language technology for Traditional Chinese speakers, it is the first step in an ongoing journey. Future work will need to address the continuous evolution of language and culture, ensuring that TAIWAN LLM remains a relevant and valuable resource for its users.

References

[1] 2023. Fasteval.

[2] Together AI. 2023. Releasing 3b and 7b redpajama-incite family of models including base, instruction-tuned and chat models.

[3] Edward Beeching, Clémentine Fourrier, Nathan Habib, Sheon Han, Nathan Lambert, Nazneen Rajani, Omar Sanseviero, Lewis Tunstall, and Thomas Wolf. 2023. Open llm leaderboard. `https://huggingface.co/spaces/HuggingFaceH4/open_llm_leaderboard`.

[4] Sahil Chaudhary. 2023. Code alpaca: An instruction-following llama model for code generation.

[5] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, Ion Stoica, and Eric P Xing. 2023. Vicuna: An Open-Source chatbot impressing GPT-4 with 90%* ChatGPT quality.

[6] Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. *See https://vicuna. lmsys. org (accessed 14 April 2023)*.

[7] Mike Conover, Matt Hayes, Ankit Mathur, Xiangrui Meng, Jianwei Xie, Jun Wan, Sam Shah, Ali Ghodsi, Patrick Wendell, Matei Zaharia, et al. 2023. Free dolly: Introducing the world's first truly open instruction-tuned llm.

[8] Tri Dao. 2023. FlashAttention-2: Faster attention with better parallelism and work partitioning.

[9] Yann Dubois, Xuechen Li, Rohan Taori, Tianyi Zhang, Ishaan Gulrajani, Jimmy Ba, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Alpacafarm: A simulation framework for methods that learn from human feedback.

[10] Yao Fu, Litu Ou, Mingyu Chen, Yuhao Wan, Hao Peng, and Tushar Khot. 2023. Chain-of-thought hub: A continuous effort to measure large language models' reasoning performance.

[11] Chan-Jan Hsu, Chang-Le Liu, Feng-Ting Liao, Po-Chun Hsu, Yi-Chang Chen, and Da shan Shiu. 2023. Advancing the evaluation of traditional chinese language models: Towards a comprehensive benchmark suite.

[12] Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, Lélio Renard Lavaud, Marie-Anne Lachaux, Pierre Stock, Teven Le Scao, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2023. Mistral 7B.

[13] Andreas Köpf, Yannic Kilcher, Dimitri von Rütte, Sotiris Anagnostidis, Zhi-Rui Tam, Keith Stevens, Abdullah Barhoum, Nguyen Minh Duc, Oliver Stanley, Richárd Nagyfi, et al. 2023. OpenAssistant Conversations–democratizing large language model alignment. *arXiv preprint arXiv:2304.07327*.

[14] Yen-Ting Lin and Yun-Nung Chen. 2023. LLM-eval: Unified multi-dimensional automatic evaluation for open-domain conversations with large language models. In *Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023)*, pages 47–58, Toronto, Canada. Association for Computational Linguistics.

[15] Shayne Longpre, Le Hou, Tu Vu, Albert Webson, Hyung Won Chung, Yi Tay, Denny Zhou, Quoc V Le, Barret Zoph, Jason Wei, et al. 2023. The flan collection: Designing data and methods for effective instruction tuning. In *Proceedings of the 40 th International Conference on Machine Learning*.

[16] Mosaic ML. 2023. Introducing mpt-7b: A new standard for open-source, commercially usable llms.

[17] Guilherme Penedo, Quentin Malartic, Daniel Hesslow, Ruxandra Cojocaru, Alessandro Cappelli, Hamza Alobeidli, Baptiste Pannier, Ebtesam Almazrouei, and Julien Launay. 2023. The refinedweb dataset for falcon llm: Outperforming curated corpora with web data, and web data only.

[18] Baolin Peng, Chunyuan Li, Pengcheng He, Michel Galley, and Jianfeng Gao. 2023. Instruction tuning with gpt-4. *arXiv preprint arXiv:2304.03277*.

[19] Samyam Rajbhandari, Jeff Rasley, Olatunji Ruwase, and Yuxiong He. 2020. Zero: Memory optimizations toward training trillion parameter models.

[20] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms.

[21] Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B Hashimoto. 2023. Alpaca: A strong, replicable instruction-following model. *Stanford Center for Research on Foundation Models. https://crfm. stanford. edu/2023/03/13/alpaca. html*, 3(6):7.

[22] Xwin-Lm Team. 2023. Xwin-LM.

[23] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi

Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open foundation and Fine-Tuned chat models.

[24] Leandro von Werra, Younes Belkada, Lewis Tunstall, Edward Beeching, Tristan Thrush, Nathan Lambert, and Shengyi Huang. 2020. TRL: Transformer reinforcement learning.

[25] Yizhong Wang, Yeganeh Kordi, Swaroop Mishra, Alisa Liu, Noah A Smith, Daniel Khashabi, and Hannaneh Hajishirzi. 2023. Self-Instruct: Aligning language models with Self-Generated instructions. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 13484–13508, Toronto, Canada. Association for Computational Linguistics.

[26] Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, et al. 2022. Super-naturalinstructions: Generalization via declarative instructions on 1600+ nlp tasks. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5085–5109.

[27] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. 2022. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837.

[28] Can Xu, Qingfeng Sun, Kai Zheng, Xiubo Geng, Pu Zhao, Jiazhan Feng, Chongyang Tao, and Daxin Jiang. Wizardlm: Empowering large language models to follow complex instructions. *arXiv preprint arXiv:2304. 12244*.

[29] Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric P Xing, Hao Zhang, Joseph E Gonzalez, and Ion Stoica. 2023. Judging LLM-as-a-Judge with MT-Bench and chatbot arena.