Prompt, Translate, Fine-Tune, Re-Initialize, or Instruction-Tune? Adapting LLMs for In-Context Learning in Low-Resource Languages

Christopher Toukmaji
University of California, Irvine
ctoukmaj@uci.edu

Jeffrey Flanigan
University of California, Santa Cruz
jmflanig@ucsc.edu

Abstract

LLMs are typically trained in high-resource languages, and tasks in lower-resourced languages tend to underperform the higher-resource language counterparts for in-context learning. Despite the large body of work on prompting settings, it is still unclear how LLMs should be adapted cross-lingually specifically for in-context learning in the low-resource target languages. We perform a comprehensive study spanning five diverse target languages, three base LLMs, and seven downstream tasks spanning over 4,100 GPU training hours (9,900+ TFLOPs) across various adaptation techniques: few-shot prompting, translate-test, fine-tuning, embedding re-initialization, and instruction fine-tuning. Our results show that the few-shot prompting and translate-test settings tend to heavily outperform the gradient-based adaptation methods. To better understand this discrepancy, we design a novel metric, *Valid Output Recall* (VOR), and analyze model outputs to empirically attribute the degradation of these trained models to catastrophic forgetting. To the extent of our knowledge, this is the largest study done on in-context learning for low-resource languages with respect to train compute and number of adaptation techniques considered. We make all our datasets and trained models available for public use.[1]

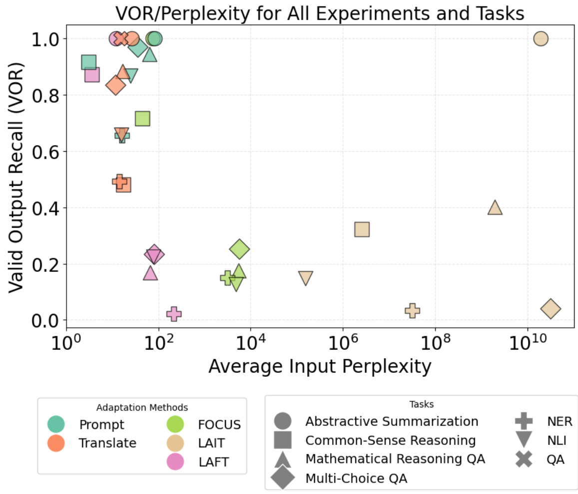

Figure 1: We report VOR scores (Valid Output Recall, the proportion of model outputs that follow the in-context labeling scheme) vs input perplexity for each adaptation method and task, averaged across target languages, and random seeds. Prompting-based methods (Prompt, Translate) demonstrate lower perplexity and higher VOR than gradient-based methods, suggesting that gradient-based methods suffer from catastrophic forgetting, degrading both linguistic ability and instruction-following alignment after training. Trained models lose the ability to learn in-context post-training while simultaneously performing worse on the target language.

1 Introduction

Large language models (LLMs) have been at the forefront of the advancements in Natural Language Processing (NLP), evidenced by state-of-the-art results on numerous benchmarks (Vaswani et al., 2017; Brown et al., 2020). LLMs are pretrained with large corpora of English text data, so the best LLMs are primarily monolingual and English-based, leaving other languages behind. Performance for tasks in non-English languages tend to underperform the same task in English for LLMs (Ahuja et al., 2023, 2024). Resource limitations prevent speakers of low-resource languages from participating in modern-day NLP since LLMs need considerable amounts of training data, and the most-capable LLMs perform poorly on low-resource languages compared to higher-resourced languages for in-context learning (Lai et al., 2023; Adelani et al., 2024a). This exclusion is a particularly crucial issue, as most languages are low-resource, and these languages have billions of speakers (Magueresse et al., 2020).

There have been several approaches to help make LLMs more multilingual. One approach

[1] https://huggingface.co/collections/ChrisToukmaji/toukmaji-flanigan-gem25

Proceedings of the Fourth Workshop on Generation, Evaluation and Metrics (GEM² 2025), pages 670–704
July 31 - August 1, 2025

Lang.	Script	Family	Speak-ers	Word Order	Tasks Evaluated	Dataset Name
hau	Latin	Afro-Asiatic-Chadic	88M	SVO	NER Mathematical Reasoning QA NLI Abstractive Summarization Multi-Choice QA	MasakhaNER (Adelani et al., 2021) AfriMGSM (Adelani et al., 2024b) AfriXNLI (Adelani et al., 2024b) XL-Sum (Hasan et al., 2021) AfriMMLU (Adelani et al., 2024b)
lug	Latin	Niger-Congo-Bantu	11M	SVO	NER Mathematical Reasoning QA NLI Multi-Choice QA	MasakhaNER (Adelani et al., 2021) AfriMGSM (Adelani et al., 2024b) AfriXNLI (Adelani et al., 2024b) AfriMMLU (Adelani et al., 2024b)
kin	Latin	Niger-Congo-Bantu	15M	SVO	NER Mathematical Reasoning QA NLI Multi-Choice QA	MasakhaNER (Adelani et al., 2021) AfriMGSM (Adelani et al., 2024b) AfriXNLI (Adelani et al., 2024b) AfriMMLU (Adelani et al., 2024b)
bur	Burmese	Sino-Tibetan-Tibeto-Burman	43M	SOV	NER NLI Abstractive Summarization Common-Sense Reasoning	Wiki-ANN (Pan et al., 2017) MyanmarXNLI (Htet and Dras, 2024) XL-Sum (Hasan et al., 2021) XStoryCloze (Lin et al., 2022)
tha	Thai	Tai-Kra-Dai	69M	SVO	NER Mathematical Reasoning QA NLI Abstractive Summarization Common-Sense Reasoning QA	Wiki-ANN (Pan et al., 2017) MGSM (Shi et al., 2022) XNLI (Conneau et al., 2018) XL-Sum (Hasan et al., 2021) XCOPA (Ponti et al., 2020) XQUAD (Artetxe et al., 2020)

Table 1: The evaluated languages (ISO 639-2 code), written script, language family, number of speakers, word order typology, and the tasks/datasets we evaluate them on.

involves pre-training an LLM from scratch on a non-English language (Martin et al., 2020; Koto et al., 2020; Wilie et al., 2020; Polignano et al., 2019; Cañete et al., 2023; Kakwani et al., 2020; Thapa et al., 2024), but this approach assumes access to a sufficiently-large corpus of text and significant computational resources. Another prevalent approach is multilingual LLMs, in which an LLM is pre-trained on many different languages (Lample and Conneau, 2019; Devlin, 2019; Conneau et al., 2019; Liu et al., 2020; Xue et al., 2021; Ogueji et al., 2021; Lin et al., 2022). However, as more languages are introduced, the monolingual and cross-lingual performance deteriorates (Conneau et al., 2019) with low-resource languages being far more vulnerable (Wu and Dredze, 2020). As a result, a large focus in the area of cross-lingual transfer has been attempting to retain the strong performance of primarily-monolingual LLMs for other non-English languages. However, these results display that the best approach fluctuates across base models, languages, and tasks (Ahuja et al., 2023).

We perform a systematic evaluation of cross-lingual transfer approaches specifically for in-context learning to identify patterns for optimal transfer settings. To the extent of our knowledge, this is the largest study (with respect to TFLOPs and GPU training hours) on cross-lingual transfer for in-context learning in low-resource languages spanning three base LLMs, five low-resource target languages, five adaptation methods, and seven NLP tasks. Our results show that the prompt and translate settings tend to heavily outperform the gradient-based adaptation methods. To better understand this discrepancy, we design *Valid Output Recall* (VOR), a novel metric, and analyze model outputs to empirically attribute the degradation of these trained models to catastrophic forgetting.

2 Related Work

The work of Tejaswi et al. (2024) is the most similar to ours. This work evaluates multilingual adaptation of LLMs for in-context learning with an emphasized study on vocabulary expansion and embedding re-initialization strategies. This study finds that that vocabulary expansion and embedding re-initialization can help bridge the gap between the performance of English and non-English languages in LLMs. Our work differs from this in that embedding re-initialization is just one of the adaptation methods that we evaluate in our study.

Ahuja et al. (2023) perform a study that evaluates on a subset of our adaptation methods - namely, translate-test and few-shot prompting. The study finds that the translate-test adaptation method outperforms few-shot prompting in most languages and tasks. This work differs from ours in that the study only considers prompt-based adaptation methods and no gradient-based approaches, like ours does. Ahuja et al. (2024) conduct an analysis

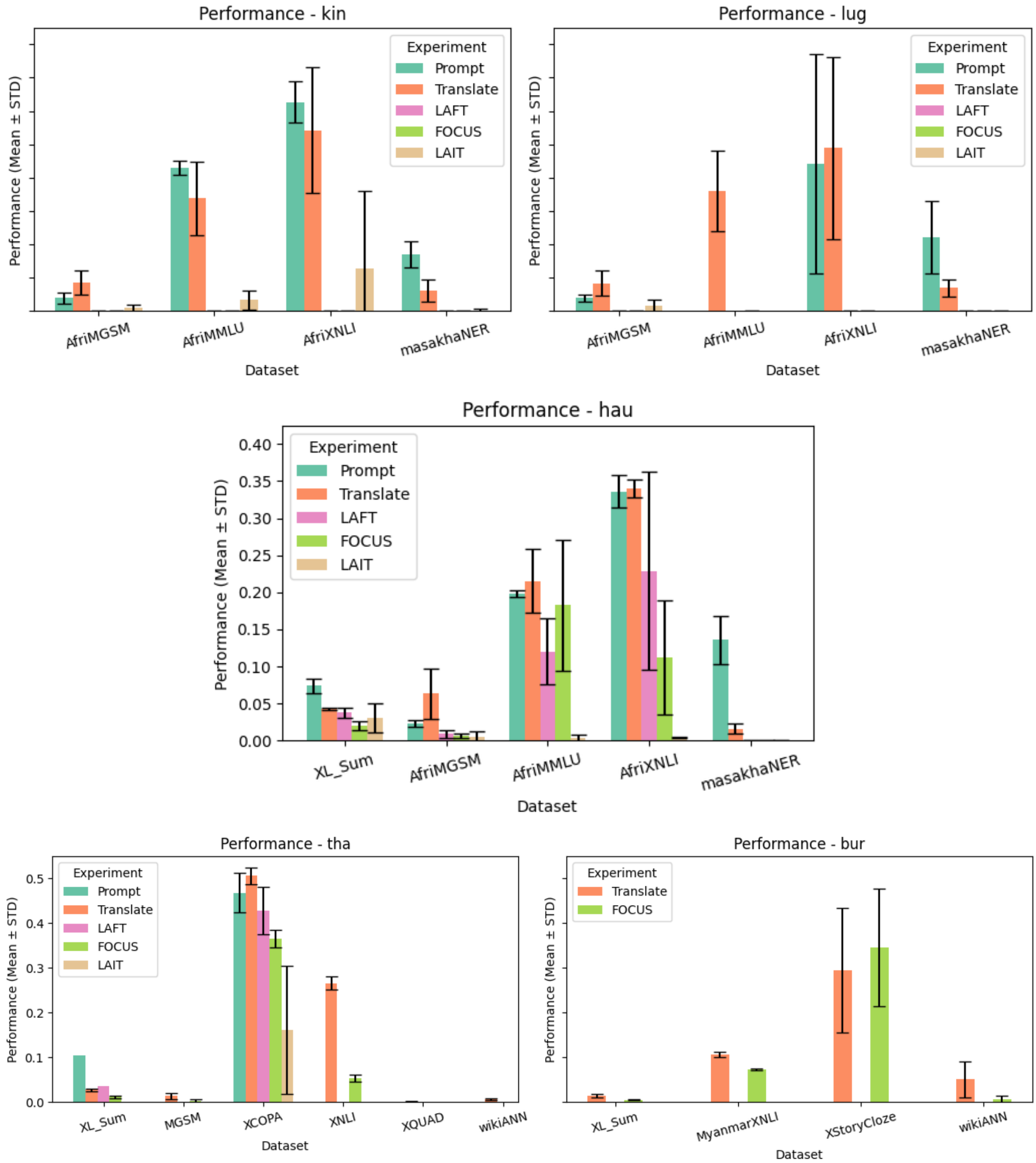

Figure 2: Few-shot downstream task performance across various adaptation methods for all five languages evaluated: Hausa (`hau`), Luganda (`lug`), Kinyarwanda (`kin`), Thai (`tha`) and Burmese (`bur`), averaged over three random seeds. We evaluate five adaptation methods: prompting-based methods (Prompt, Translate) and gradient-based methods (Language-Adaptive Fine-Tuning (LAFT), FOCUS (embedding re-initialization + LAFT), and Language-Adaptive Instruction Tuning (LAIT)). We find that prompting-based methods consistently outperform gradient-based methods across all tasks.

on few-shot prompting across many language models, tasks, and languages, but do not consider any other adaptation methods.

Others have benchmarked adaptation methods, but they differ from our study in that our main focus is on several low-resource languages (Asai et al., 2023; Toraman, 2024; Wang et al., 2025)

There is a multitude of work in cross-lingual transfer, but only the papers above have a similar emphasis of benchmarking adaptation methods for in-context learning. Other primary lines of work in cross-lingual transfer from monolingual LLMs include 1) performing further training with the target language (Alabi et al., 2022; Joshi et al., 2024; Doshi et al., 2024; Razumovskaia et al., 2024; Sani et al., 2025), 2) modifying the embedding matrix and vocabulary to better fit the target language (de Vries and Nissim, 2021; Dobler and de Melo, 2023; Remy et al., 2023; Gosal et al., 2024; Cui et al., 2024; Mundra et al., 2024; Yamaguchi et al., 2024b; Pham et al., 2024; Da Dalt et al., 2024; Vre et al., 2024; Yamaguchi et al., 2024a) or 3) instruction-tuning with the target language (Kuulmets et al., 2024) or cross-lingually (Chen et al., 2024; Ranaldi et al., 2023; Ranaldi and Pucci, 2023). These generalized cross-lingual transfer techniques are evaluated in our study.

3 Methods

We use three different multi-billion parameter base LLMs - LLaMa 2 7B (Touvron et al., 2023b), MPT-7B (MosaicAI, 2023), and Phi-2 (Javaheripi and Bubec, 2023) - for in-context learning in five diverse low-resource languages. We opt to use these models since they are primarily-monolingual, open-source, and are capable of in-context learning. We use English as our source language and evaluate on a set of five diverse low-resource target languages: Hausa (`hau`), Kinyarwanda (`kin`), Luganda (`lug`), Burmese (`bur`), and Thai (`tha`).

3.1 Evaluation Setting

We aim to evaluate scenarios where a target language speaker uses an LLM to perform a downstream task in that language. Accordingly, we only consider tasks and datasets where the task instance is in the target language. Not every task is evaluated in every language because we do not have datasets for all these tasks in each language.

During evaluation, we form our few-shot prompts with a random sample without replacement of the training split for the evaluation datasets outlined in Table 1. For all settings, each shot is prepended with a machine-translated description of the task in the target language. We use the maximum number of shots that fit within the context length for each dataset. The reported results are on the test split of the dataset for that language, and we conduct three samples with random seeds and average the performance across the test split. Some experiments were omitted due to context window or memory limitations (see Appendix I for details).

We emphasize that no task-specific fine-tuning is done at any point in any of our experiments. Our core research questions aims to answer how to transfer LLMs to new languages while remaining as general-purpose task solvers.

3.2 Datasets and Metrics

The datasets we evaluate on for each langauge are given in Table 1. We report F1-score for MasakhaNER and WikiANN, ROUGE-L for XL-Sum, and accuracy for AfriMGSM, AfriXNLI, AfriMMLU, MGSM, XCOPA, and XStoryCloze. MasakhaNER and WikiANN are language-specific datasets, whereas the others are either evaluated in a single language or are parallel.

4 Experiments

We evaluate the following five methods for adapting an LLM trained in a source language for prompting with a target language.

Few-Shot Prompting (Prompt) We prompt the LLM with the few-shot prompt in the target language and evaluate the completion. This method requires no translation, nor any gradient updates.

Translate-test (Translate) We first machine-translate the few-shot prompt from the target language to the source language. Next, the LLM is prompted in the source language. Then, the output is translated from the source language back to the target language. We use NLLB-200 3.3B (Team et al., 2022) for both translation directions. This method does not require any gradient updates (Hu et al., 2020).

Language-Adaptive Fine-Tuning (LAFT) Starting with the original LLM, we further fine-tune the LLM on a corpus of tokens in the target language using the original pre-training objective. Then, we prompt the LLM.

Vocabulary and Embedding Re-initialization (FOCUS) Following Dobler and de Melo (2023), we perform the FOCUS method; we train a new tokenizer in the target language, then use pre-trained static embeddings [2] in the target language to re-initialize semantically-similar overlapping tokens in the embedding matrix of the base LLM, and thereafter perform LAFT on the LLM. Then, we prompt the LLM.

Language-Adaptive Instruction Tuning (LAIT) We machine-translate an Instruction Tuning dataset from the source language to the target language, then we perform instruction fine-tuning on the translated dataset. Then, we prompt the LLM.

5 Results

The results in Figure 2 display few-shot prompting and translate-test adaptation methods surprisingly tend to heavily outperform the gradient-based adaptation methods. Below, we provide an empirical analysis of the LLMs' outputs, and show this disparity can be attributed to catastrophic forgetting (McCloskey and Cohen, 1989), which can occur in LLMs during continued training (Luo et al., 2025).

In order to determine whether the performance degradation is attributed to insufficient knowledge of the target language or to task forgetting, we design *Valid Output Recall* (VOR), a metric to quantify an LLM's ability to instruction-follow labels in-context. VOR is the proportion of LLM outputs of a test set that follow the same labeling scheme that was instructed and provided in-context. For example, in a binary-classification task instance where an LLM is instructed to output a label $L \in \{0, 1\}$ for test instance i in a test dataset with size N and the LLM output $\hat{y}_i$, then $VOR = \frac{1}{N} \sum_{i=1}^{N} \mathbb{I}(\hat{y}_i \in L)$. The VOR is compared with the perplexity of the inputs. Intuitively, this isolates the evaluation of an LLM's task alignment and instruction-following ability from its linguistic ability.

In Figure 1, we observe that gradient-based adaptation methods have both higher perplexities and lower VOR compared to few-shot prompting. These results empirically verify that the trained models are losing the ability to learn in-context post-training while simultaneously performing worse on the target language. This suggests that the gradient-based methods in our training environment suffer from catastrophic forgetting since both linguistic knowledge and task alignment deteriorate.

6 Conclusion

This work provides the largest comprehensive study on adapting primarily-English LLMs to low-resource languages for in-context learning. Five adaptation methods are evaluated across three base LLMs using five diverse target languages on seven downstream tasks. Few-shot prompting and translate-test worked the best in nearly all cases, but there is no trend between which of the two works best. We design a novel metric, *Valid Output Recall* (VOR), and provide an empirical analysis on LLM outputs to show that models adapted with gradient-based methods degraded due to catastrophic forgetting.

Future Work

In this work, we experiment with five diverse low-resource languages, but there are other low-resource languages that are also in need of more research. We leave this as future work, and we hope our work will help inspire research for other low-resource languages.

We used two training-free adaptation methods: few-shot prompting and translate-test. There are other training-free prompting methods such as varying design templates or demonstration selections which we leave as future work. Given that training-free adaptation methods produced the best results in our paper, we are optimistic for future work in this direction, and believe our findings provide a strong motivation for further research into training-free adaptation approaches.

Limitations

One limitation of our approach is that the translate-test setting hinges on an NMT model which could introduce translation errors and, in turn, affect performance. While this is a general issue with translation-based methods, future improvements in NMT quality could help reduce this effect.

Potential Risks

We do not anticipate any potential risks with respect to ethical or social impacts from our work. However, since a component of our contributions is the open-sourcing of the trained models, we acknowledge that LLMs are capable of generating

[2] https://fasttext.cc/docs/en/pretrained-vectors.html

text that could be harmful (Gehman et al., 2020) or non-factual (Huang et al., 2025).

Acknowledgments

We thank Brendan King, Changmao Li, Chris Liu, Brian Mak, Nilay Patel, Geetanjali Rakshit, Rongwen Zhao, Zekun Zhao, Giridhar Vadhul, Ian Lane, and Amita Misra for their insightful feedback on earlier versions of this work. We would also like to thank the anonymous reviewers and area chairs for their detailed and helpful feedback.

This work used resources available through the National Research Platform (NRP) at the University of California, San Diego. NRP has been developed, and is supported in part, by funding from National Science Foundation, from awards 1730158, 1540112, 1541349, 1826967, 2112167, 2100237, and 2120019, as well as additional funding from community partners.

References

David Ifeoluwa Adelani, Jade Abbott, Graham Neubig, Daniel D'souza, Julia Kreutzer, Constantine Lignos, Chester Palen-Michel, Happy Buzaaba, Shruti Rijhwani, Sebastian Ruder, Stephen Mayhew, Israel Abebe Azime, Shamsuddeen Muhammad, Chris Chinenye Emezue, Joyce Nakatumba-Nabende, Perez Ogayo, Anuoluwapo Aremu, Catherine Gitau, Derguene Mbaye, Jesujoba Alabi, Seid Muhie Yimam, Tajuddeen Gwadabe, Ignatius Ezeani, Rubungo Andre Niyongabo, Jonathan Mukiibi, Verrah Otiende, Iroro Orife, Davis David, Samba Ngom, Tosin Adewumi, Paul Rayson, Mofetoluwa Adeyemi, Gerald Muriuki, Emmanuel Anebi, Chiamaka Chukwuneke, Nkiruka Odu, Eric Peter Wairagala, Samuel Oyerinde, Clemencia Siro, Tobius Saul Bateesa, Temilola Oloyede, Yvonne Wambui, Victor Akinode, Deborah Nabagereka, Maurice Katusiime, Ayodele Awokoya, Mouhamadane MBOUP, Dibora Gebreyohannes, Henok Tilaye, Kelechi Nwaike, Degaga Wolde, Abdoulaye Faye, Blessing Sibanda, Orevaoghene Ahia, Bonaventure F. P. Dossou, Kelechi Ogueji, Thierno Ibrahima DIOP, Abdoulaye Diallo, Adewale Akinfaderin, Tendai Marengereke, and Salomey Osei. 2021. Masakhaner: Named entity recognition for african languages. *Preprint*, arXiv:2103.11811.

David Ifeoluwa Adelani, A. Seza Doruöz, André Coneglian, and Atul Kr. Ojha. 2024a. Comparing llm prompting with cross-lingual transfer performance on indigenous and low-resource brazilian languages. *Preprint*, arXiv:2404.18286.

David Ifeoluwa Adelani, Jessica Ojo, Israel Abebe Azime, Jian Yun Zhuang, Jesujoba O. Alabi, Xuanli He, Millicent Ochieng, Sara Hooker, Andiswa Bukula, En-Shiun Annie Lee, Chiamaka Chukwuneke, Happy Buzaaba, Blessing Sibanda, Godson Kalipe, Jonathan Mukiibi, Salomon Kabongo, Foutse Yuehgoh, Mmasibidi Setaka, Lolwethu Ndolela, Nkiruka Odu, Rooweither Mabuya, Shamsuddeen Hassan Muhammad, Salomey Osei, Sokhar Samb, Tadesse Kebede Guge, and Pontus Stenetorp. 2024b. Irokobench: A new benchmark for african languages in the age of large language models. *Preprint*, arXiv:2406.03368.

Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Maxamed Axmed, Kalika Bali, and Sunayana Sitaram. 2023. Mega: Multilingual evaluation of generative ai. *Preprint*, arXiv:2303.12528.

Sanchit Ahuja, Divyanshu Aggarwal, Varun Gumma, Ishaan Watts, Ashutosh Sathe, Millicent Ochieng, Rishav Hada, Prachi Jain, Maxamed Axmed, Kalika Bali, and Sunayana Sitaram. 2024. Megaverse: Benchmarking large language models across languages, modalities, models and tasks. *Preprint*, arXiv:2311.07463.

Jesujoba O. Alabi, David Ifeoluwa Adelani, Marius Mosbach, and Dietrich Klakow. 2022. Adapting pre-trained language models to African languages via multilingual adaptive fine-tuning. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4336–4349, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Rosana Ardila, Megan Branson, Kelly Davis, Michael Henretty, Michael Kohler, Josh Meyer, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. 2020. Common voice: A massively-multilingual speech corpus. *Preprint*, arXiv:1912.06670.

Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. 2020. On the cross-lingual transferability of monolingual representations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4623–4637, Online. Association for Computational Linguistics.

Akari Asai, Sneha Kudugunta, Xinyan Velocity Yu, Terra Blevins, Hila Gonen, Machel Reid, Yulia Tsvetkov, Sebastian Ruder, and Hannaneh Hajishirzi. 2023. Buffet: Benchmarking large language models for few-shot cross-lingual transfer. *Preprint*, arXiv:2305.14857.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. *CoRR*, abs/2005.14165.

José Cañete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. 2023. Spanish pre-trained bert model and evaluation data. *Preprint*, arXiv:2308.02976.

Yang Chen, Vedaant Shah, and Alan Ritter. 2024. Translation and fusion improves zero-shot cross-lingual information extraction. *Preprint*, arXiv:2305.13582.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised cross-lingual representation learning at scale. *CoRR*, abs/1911.02116.

Alexis Conneau, Guillaume Lample, Ruty Rinott, Adina Williams, Samuel R. Bowman, Holger Schwenk, and Veselin Stoyanov. 2018. Xnli: Evaluating cross-lingual sentence representations. *Preprint*, arXiv:1809.05053.

Yiming Cui, Ziqing Yang, and Xin Yao. 2024. Efficient and effective text encoding for chinese llama and alpaca. *Preprint*, arXiv:2304.08177.

Severino Da Dalt, Joan Llop, Irene Baucells, Marc Pamies, Yishi Xu, Aitor Gonzalez-Agirre, and Marta Villegas. 2024. FLOR: On the effectiveness of language adaptation. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 7377–7388, Torino, Italia. ELRA and ICCL.

Wietse de Vries and Malvina Nissim. 2021. As good as new. how to successfully recycle english GPT-2 to make models for other languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. Association for Computational Linguistics.

DeepSpeed. 2021. DeepSpeed ZeRO-3 Offload — deepspeed.ai. `https://www.deepspeed.ai/2021/03/07/zero3-offload.html`.

Tim Dettmers. 2022. bitsandbytes. GitHub repository.

Jacob Devlin. 2019. Bert/multilingual.md at master ů google-research/bert.

Konstantin Dobler and Gerard de Melo. 2023. FOCUS: Effective embedding initialization for monolingual specialization of multilingual models. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13440–13454, Singapore. Association for Computational Linguistics.

Meet Doshi, Raj Dabre, and Pushpak Bhattacharyya. 2024. Do not worry if you do not have data: Building pretrained language models using translationese. *Preprint*, arXiv:2403.13638.

Samuel Gehman, Suchin Gururangan, Maarten Sap, Yejin Choi, and Noah A. Smith. 2020. Realtoxicityprompts: Evaluating neural toxic degeneration in language models. *Preprint*, arXiv:2009.11462.

Gurpreet Gosal, Yishi Xu, Gokul Ramakrishnan, Rituraj Joshi, Avraham Sheinin, Zhiming, Chen, Biswajit Mishra, Natalia Vassilieva, Joel Hestness, Neha Sengupta, Sunil Kumar Sahu, Bokang Jia, Onkar Pandit, Satheesh Katipomu, Samta Kamboj, Samujjwal Ghosh, Rahul Pal, Parvez Mullah, Soundar Doraiswamy, Mohamed El Karim Chami, and Preslav Nakov. 2024. Bilingual adaptation of monolingual foundation models. *Preprint*, arXiv:2407.12869.

Tahmid Hasan, Abhik Bhattacharjee, Md. Saiful Islam, Kazi Mubasshir, Yuan-Fang Li, Yong-Bin Kang, M. Sohel Rahman, and Rifat Shahriyar. 2021. XL-sum: Large-scale multilingual abstractive summarization for 44 languages. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4693–4703, Online. Association for Computational Linguistics.

Aung Htet and Mark Dras. 2024. Myanmar xnli: Building a dataset and exploring low-resource approaches to natural language inference with myanmar.

Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. 2020. Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization. *Preprint*, arXiv:2003.11080.

Lei Huang, Weijiang Yu, Weitao Ma, Weihong Zhong, Zhangyin Feng, Haotian Wang, Qianglong Chen, Weihua Peng, Xiaocheng Feng, Bing Qin, and Ting Liu. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2):155.

Mojan Javaheripi and Sébastien Bubec. 2023. Phi-2: The surprising power of small language models — microsoft.com. `https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/`. [].

Raviraj Joshi, Kanishk Singla, Anusha Kamath, Raunak Kalani, Rakesh Paul, Utkarsh Vaidya, Sanjay Singh Chauhan, Niranjan Wartikar, and Eileen Long. 2024. Adapting multilingual llms to low-resource languages using continued pre-training and synthetic corpus. *Preprint*, arXiv:2410.14815.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 4948–4961, Online. Association for Computational Linguistics.

Fajri Koto, Afshin Rahimi, Jey Han Lau, and Timothy Baldwin. 2020. IndoLEM and IndoBERT: A benchmark dataset and pre-trained language model for Indonesian NLP. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 757–770, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Hele-Andra Kuulmets, Taido Purason, Agnes Luhtaru, and Mark Fishel. 2024. Teaching llama a new language through cross-lingual knowledge transfer. *Preprint*, arXiv:2404.04042.

Viet Dac Lai, Nghia Trung Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen. 2023. Chatgpt beyond english: Towards a comprehensive evaluation of large language models in multilingual learning. *Preprint*, arXiv:2304.05613.

Guillaume Lample and Alexis Conneau. 2019. Cross-lingual language model pretraining. *Preprint*, arXiv:1901.07291.

Xi Victoria Lin, Todor Mihaylov, Mikel Artetxe, Tianlu Wang, Shuohui Chen, Daniel Simig, Myle Ott, Naman Goyal, Shruti Bhosale, Jingfei Du, Ramakanth Pasunuru, Sam Shleifer, Punit Singh Koura, Vishrav Chaudhary, Brian O'Horo, Jeff Wang, Luke Zettlemoyer, Zornitsa Kozareva, Mona Diab, Veselin Stoyanov, and Xian Li. 2022. Few-shot learning with multilingual language models. *Preprint*, arXiv:2112.10668.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual denoising pre-training for neural machine translation. *Preprint*, arXiv:2001.08210.

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. *Preprint*, arXiv:1711.05101.

Yun Luo, Zhen Yang, Fandong Meng, Yafu Li, Jie Zhou, and Yue Zhang. 2025. An empirical study of catastrophic forgetting in large language models during continual fine-tuning. *Preprint*, arXiv:2308.08747.

Alexandre Magueresse, Vincent Carles, and Evan Heetderks. 2020. Low-resource languages: A review of past work and future challenges. *Preprint*, arXiv:2006.07264.

Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric de la Clergerie, Djamé Seddah, and Benoît Sagot. 2020. CamemBERT: a tasty French language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7203–7219, Online. Association for Computational Linguistics.

Michael McCloskey and Neal J. Cohen. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. volume 24 of *Psychology of Learning and Motivation*, pages 109–165. Academic Press.

Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. 2018. Mixed precision training. *Preprint*, arXiv:1710.03740.

MosaicAI. 2023. Introducing MPT-7B: A New Standard for Open-Source, Commercially Usable LLMs — databricks.com. `https://www.databricks.com/blog/mpt-7b`.

Nandini Mundra, Aditya Nanda Kishore, Raj Dabre, Ratish Puduppully, Anoop Kunchukuttan, and Mitesh M. Khapra. 2024. An empirical comparison of vocabulary expansion and initialization approaches for language models. *Preprint*, arXiv:2407.05841.

Kelechi Ogueji, Yuxin Zhu, and Jimmy Lin. 2021. Small data? no problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st Workshop on Multilingual Representation Learning*, pages 116–126, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017. Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1946–1958, Vancouver, Canada. Association for Computational Linguistics.

Trinh Pham, Khoi M. Le, and Luu Anh Tuan. 2024. Unibridge: A unified approach to cross-lingual transfer learning for low-resource languages. *Preprint*, arXiv:2406.09717.

Marco Polignano, Valerio Basile, Pierpaolo Basile, Marco de Gemmis, and Giovanni Semeraro. 2019. AlBERTo: Modeling italian social media language with BERT. *Italian Journal of Computational Linguistics*, 5(2):11–31.

Edoardo Maria Ponti, Goran Glavaš, Olga Majewska, Qianchu Liu, Ivan Vulić, and Anna Korhonen. 2020. XCOPA: A multilingual dataset for causal commonsense reasoning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2362–2376, Online. Association for Computational Linguistics.

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. *Preprint*, arXiv:1910.10683.

Leonardo Ranaldi and Giulia Pucci. 2023. Does the English matter? elicit cross-lingual abilities of large language models. In *Proceedings of the 3rd Workshop* *on Multi-lingual Representation Learning (MRL)*, pages 173–183, Singapore. Association for Computational Linguistics.

Leonardo Ranaldi, Giulia Pucci, and Andre Freitas. 2023. Empowering cross-lingual abilities of instruction-tuned large language models by translation-following demonstrations. *Preprint*, arXiv:2308.14186.

Evgeniia Razumovskaia, Ivan Vuli, and Anna Korhonen. 2024. Analyzing and adapting large language models for few-shot multilingual nlu: Are we there yet? *Preprint*, arXiv:2403.01929.

François Remy, Pieter Delobelle, Bettina Berendt, Kris Demuynck, and Thomas Demeester. 2023. Tik-to-tok: Translating language models one token at a time: An embedding initialization strategy for efficient language adaptation. *Preprint*, arXiv:2310.03477.

Samin Mahdizadeh Sani, Pouya Sadeghi, Thuy-Trang Vu, Yadollah Yaghoobzadeh, and Gholamreza Haffari. 2025. Extending llms to new languages: A case study of llama and persian adaptation. *Preprint*, arXiv:2412.13375.

Freda Shi, Mirac Suzgun, Markus Freitag, Xuezhi Wang, Suraj Srivats, Soroush Vosoughi, Hyung Won Chung, Yi Tay, Sebastian Ruder, Denny Zhou, Dipanjan Das, and Jason Wei. 2022. Language models are multilingual chain-of-thought reasoners. *Preprint*, arXiv:2210.03057.

Rohan Taori, Ishaan Gulrajani, Tianyi Zhang, Yann Dubois, Xuechen Li, Carlos Guestrin, Percy Liang, and Tatsunori B. Hashimoto. 2023. Stanford alpaca: An instruction-following llama model. `https://github.com/tatsu-lab/stanford_alpaca`.

NLLB Team, Marta R. Costa-jussà, James Cross, Onur Çelebi, Maha Elbayad, Kenneth Heafield, Kevin Heffernan, Elahe Kalbassi, Janice Lam, Daniel Licht, Jean Maillard, Anna Sun, Skyler Wang, Guillaume Wenzek, Al Youngblood, Bapi Akula, Loic Barrault, Gabriel Mejia Gonzalez, Prangthip Hansanti, John Hoffman, Semarley Jarrett, Kaushik Ram Sadagopan, Dirk Rowe, Shannon Spruit, Chau Tran, Pierre Andrews, Necip Fazil Ayan, Shruti Bhosale, Sergey Edunov, Angela Fan, Cynthia Gao, Vedanuj Goswami, Francisco Guzmán, Philipp Koehn, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Holger Schwenk, and Jeff Wang. 2022. No language left behind: Scaling human-centered machine translation. *Preprint*, arXiv:2207.04672.

Atula Tejaswi, Nilesh Gupta, and Eunsol Choi. 2024. Exploring design choices for building language-specific llms. *Preprint*, arXiv:2406.14670.

Prajwal Thapa, Jinu Nyachhyon, Mridul Sharma, and Bal Krishna Bal. 2024. Development of pre-trained transformer-based models for the nepali language. *Preprint*, arXiv:2411.15734.

Cagri Toraman. 2024. Adapting open-source generative large language models for low-resource languages: A case study for Turkish. In *Proceedings of the Fourth Workshop on Multilingual Representation Learning (MRL 2024)*, pages 30–44, Miami, Florida, USA. Association for Computational Linguistics.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023a. Llama: Open and efficient foundation language models. *Preprint*, arXiv:2302.13971.

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. 2023b. Llama 2: Open foundation and fine-tuned chat models. *Preprint*, arXiv:2307.09288.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *CoRR*, abs/1706.03762.

Domen Vre, Martin Boi, Alja Potonik, Toma Martini, and Marko Robnik-ikonja. 2024. Generative model for less-resourced language with 1 billion parameters. *Preprint*, arXiv:2410.06898.

Shumin Wang, Yuexiang Xie, Bolin Ding, Jinyang Gao, and Yanyong Zhang. 2025. Language adaptation of large language models: An empirical study on LLaMA2. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 7195–7208, Abu Dhabi, UAE. Association for Computational Linguistics.

Bryan Wilie, Karissa Vincentio, Genta Indra Winata, Samuel Cahyawijaya, Xiaohong Li, Zhi Yuan Lim, Sidik Soleman, Rahmad Mahendra, Pascale Fung, Syafri Bahar, and Ayu Purwarianti. 2020. IndoNLU: Benchmark and resources for evaluating Indonesian natural language understanding. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the*

10th International Joint Conference on Natural Language Processing, pages 843–857, Suzhou, China. Association for Computational Linguistics.

Shijie Wu and Mark Dredze. 2020. Are all languages created equal in multilingual BERT? In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 120–130, Online. Association for Computational Linguistics.

Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. mt5: A massively multilingual pre-trained text-to-text transformer. *Preprint*, arXiv:2010.11934.

Atsuki Yamaguchi, Aline Villavicencio, and Nikolaos Aletras. 2024a. An empirical study on cross-lingual vocabulary adaptation for efficient language model inference. *Preprint*, arXiv:2402.10712.

Atsuki Yamaguchi, Aline Villavicencio, and Nikolaos Aletras. 2024b. How can we effectively expand the vocabulary of llms with 0.01gb of target language text? *Preprint*, arXiv:2406.11477.

Appendix

A Training Details

Each model was trained using 1 NVIDIA A100-SXM4 80GB. For computational efficiency, we use DeepSpeed's Zero Redundancy Optimizer (ZeRO) at stage 3 (DeepSpeed, 2021), BF16 mixed precision training (Micikevicius et al., 2018), a block size of 1024, and a 32-bit paged AdamW optimizer (Dettmers, 2022).

A.1 Hyperparameters

We use the standard practice of selecting the maximum batch size that fits within GPU memory. This results in a batch size of 1 for LAFT and FOCUS training, and a batch size of 2 for LAIT training.

Each of the adaptation methods that required training (LAFT, FOCUS, LAIT) were trained for 6 epochs. We keep checkpoints after each epoch, and the model checkpoint with the lowest loss on the validation set is kept. We initialize training with the following hyperparameters taken from the LLaMa-2 paper (Touvron et al., 2023b): the AdamW optimizer (Loshchilov and Hutter, 2019) with $\beta_1 = 0.9, \beta_2 = 0.95, \epsilon = 10^{-5}$, a learning rate of $3e^{-4}$ with a cosine scheduler warm-up of 2000 steps, 0.1 weight decay, and gradient clipping at of 1. We report the epoch of the best-performing checkpoint in Table 2.

B Inference Details

Based on cluster availability, we use 1 of the following GPUs for inference: NVIDIA L40 (48GB), NVIDIA A4000 (16 GB), NVIDIA GeForce RTX 3090 (24GB), NVIDIA GeForce RTX 4090 (24GB), NVIDIA A10 (24GB), NVIDIA L4 (24GB). We use the standard practice of selecting the maximum batch size that fits within GPU memory. This value is one of $[4, 8, 16, 32]$ and is automatically calculated based on the size of the test instances and the GPU memory from whichever GPU the job is assigned.

C Budget

We report training runtime and TFLOPs for each trained model after all 6 epochs in Table 2.

Method	Lang.	Model	Best Epoch	Train Runtime (hours)	TFLOPs
FOCUS	bur	llama	5	206	240
FOCUS	bur	mpt	5	160	240
FOCUS	bur	phi	5	28	816
FOCUS	hau	llama	5	234	240
FOCUS	hau	mpt	3	167	240
FOCUS	hau	phi	5	27	816
FOCUS	kin	llama	4	87	116
FOCUS	kin	mpt	4	98	105
FOCUS	kin	phi	1	12	358
FOCUS	lug	llama	4	6	7
FOCUS	lug	mpt	4	17	7
FOCUS	lug	phi	4	1	23
FOCUS	tha	mpt	5	197	240
FOCUS	tha	phi	5	32	816
LAFT	bur	llama	5	217	240
LAFT	bur	mpt	5	208	240
LAFT	bur	phi	5	113	816
LAFT	hau	llama	4	256	240
LAFT	hau	mpt	5	267	240
LAFT	hau	phi	5	111	816
LAFT	kin	llama	4	219	227
LAFT	kin	mpt	4	221	211
LAFT	kin	phi	5	102	740
LAFT	lug	llama	3	17	16
LAFT	lug	mpt	3	12	15
LAFT	lug	phi	4	7	52
LAFT	tha	llama	4	200	240
LAFT	tha	mpt	5	218	240
LAFT	tha	phi	5	119	816
LAIT	bur	llama	6	30	36
LAIT	bur	mpt	6	119	36
LAIT	bur	phi	6	3	142
LAIT	hau	llama	3	20	10
LAIT	hau	mpt	6	67	10
LAIT	hau	phi	4	3	35
LAIT	kin	llama	4	22	10
LAIT	kin	mpt	6	65	9
LAIT	kin	phi	4	3	32
LAIT	lug	llama	6	20	12
LAIT	lug	mpt	6	66	11
LAIT	lug	phi	4	3	39
LAIT	tha	llama	4	28	20
LAIT	tha	mpt	5	97	21
LAIT	tha	phi	4	3	107
Total	-	-	-	**4108**	**9943**

Table 2: Used training checkpoint, final training runtime in hours, and Tera Floating Point Operations (TFLOPs) for every trained model

D Training Datasets

D.1 LAFT and FOCUS

We use the following language-specific fine-tuning corpora for LAFT and FOCUS. For Burmese, Hausa, and Thai, we use a subset of mC4 (Xue et al., 2021), a multilingual variant of the C4 pre-training corpus (Raffel et al., 2020). For Kinyarwanda and Luganda, we use a subset of CommonVoice (Ardila et al., 2020) since an mC4 split doesn't exist for these languages. For all LAFT and FOCUS experiments, we train on 25M tokens and use the provided evaluation set for validation.

D.2 LAIT

We use a professional neural machine translation system [3] to translate a random sample of 5,000 instruction-following examples from the Alpaca dataset (Taori et al., 2023). We translate the same 5,000 instruction-following examples from English to each of the target languages, and we release the translated parallel instruction-following datasets on the HuggingFace dataset hub.[4] We designate 85% of the examples for training and the remaining 15% for validation.

E Scientific Artifact Licenses

Below, we outline the scientific artifacts used (base models, training datasets, evaluation datasets) and the respective licenses.

Artifact	License
LLaMa-2 7B	LLAMA2
MPT 7B	APACHE-2.0
Phi-2	MIT
NLLB-200 3.3B	CC-BY-NA-4.0
mc4	ODC-BY
CommonVoice	CC-0
AfriMGSM	APACHE-2.0
AfriMMLU	APACHE-2.0
AfriXNLI	APACHE-2.0
MasakhaNER	CC-BY-NC-4.0
MyanmarXNLI	CC-BY-NC-4.0
MGSM	MIT
Wiki-ANN	CC-0
XCOPA	CC-BY-4.0
XNLI	CC-BY-NC-4.0
XL-Sum	CC-BY-NC-SA-4.0
XQUAD	CC-BY-SA 4.0
XStoryCloze	CC-BY-SA-4.0

Table 3: Licenses for base models, training datasets, and evaluation datasets

F Dataset Splits

We outline the size of the train and test sets in Table 4. To form the few-shot prompt, we randomly sample from the training set, or the validation set if there is no train split. We report results on the test split.

XQUAD does not natively have a train/val/test split, so we use 10% of the data for our 'train' split and the remaining 90% as the 'test' split. We use the same split for all experiments.

[3] https://cloud.google.com/translate/docs/reference/rest

[4] https://huggingface.co/collections/ChrisToukmaji/toukmaji-flanigan-gem25

Dataset and Lang.	train	eval	test
AfriMGSM (all)	8	-	250
AfriMMLU (all)	-	83	500
AfriXNLI (all)	-	450	600
MasakhaNER (hau)	1903	272	545
MasakhaNER (kin)	2110	301	604
MasakhaNER (lug)	2003	200	401
MyanmarXNLI	392,702	2,490	5,010
MGSM	8	-	250
Wiki-ANN (bur)	100	100	100
Wiki-ANN (tha)	20,000	10,000	10,000
XCOPA	-	100	500
XNLI	392,702	2,490	5,010
XL-Sum (bur)	4,569	570	570
XL-Sum (hau)	6,418	802	802
XL-Sum (tha)	6,616	826	826
XQUAD	-	1,190	-
XStoryCloze	361	-	1,511

Table 4: Evaluation dataset sizes for training, validation, and test datasets

G Prompt Selection

Our block size is 1024, and we allocate 75% of the block size (768 tokens) to context and 25% of the block size (256 tokens) for the completion. In order to determine which train/eval instances to put in context, we perform the following steps. First, for every evaluation dataset, we find the largest instance in the set (in terms of tokens). In the worst case, this determines how many tokens are left in context for the completed exemplars/shots (i.e. if the largest test instance for a given dataset is 100 tokens, we must fit the completed exemplars within 668 tokens). Then, we randomly sample from the train/eval sets to try to get k shots to fit within the remainder of the context window, where k is the desired number of shots in-context. We maximize k and stop sampling after 20 attempts. If we cannot fit even a single exemplar ($k = 1$) after 20 tries, we are unable to perform inference for this experiment (see Table 5, Table 6, and Appendix I for a list and discussion of such instances). After performing these steps, we ended with a value of $k = 1$ for all reported experiments, except for NLI tasks where we use a value of $k = 3$.

In order to perform to perform NLI faithfully, $k = 3$ is the minimum value of shots to put into context since there needs to be one exemplar for each NLI label. When sampling from the train set in NLI experiments, we enforce a constraint that there must be one exemplar for each NLI label. The order of the NLI exemplars is randomized.

For NER tasks, we enforce a constraint that the exemplar in context must have at least one named-entity. All other tasks have no constrains on training data contents sampled for in-context learning.

H Answer Extraction

We use the same cleaning procedure as outlined by Touvron et al. (2023a) for Question-Answering tasks, in which the answer is extracted from the generation by only considering content before the first line break, or the final dot/comma. For Mathematical-Reasoning QA, we extract the final space-separated integer since the output generation is Chain-of-Thought. For Multi-Choice QA, NLI, and Common-Sense Reasoning, we extract the first instance of the label set ({A,B,C,D} for Multi-Choice QA, {0,1,2} for NLI, and {1,2} for Common-Sense Reasoning). For Abstractive Summarization, we strip new line tokens. For NER, we strip out text outside the first occurrence of an opening and closing bracket, as implied by the label format in-context. The content within the brackets is filtered by only considering entity pairs with both opening and closing parentheses.

We utilize these label sets and answer extraction methods when calculating VOR. Generation tasks like abstractive summarization are free-form and do not have to adhere to strict formatting which explains why the VOR scores are near perfect for generation tasks, but much smaller for tasks with strict required outputs (i.e. NLI).

As VOR is a recall-oriented metric, instances without an extracted answer following the preprocessing steps are treated as incorrect, whereas instances with any extracted answer, regardless of its semantic correctness, are treated as correct.

I Unperformed Experiments

As outlined in Table 5 and Table 6, a few experiments were infeasible to run. The FOCUS training task for `tha` with the LLaMa-2-7B model was infeasible to train due to memory constraints (more details below). The remainder of the excluded tasks were infeasible because they were unable to fit within the partition of the block size allocated for context.

The FOCUS task for `tha` with the LLaMa-2-7B model required over 2TB of RAM to train a new tokenizer which far exceeded the 1TB RAM limits imposed on us from our compute cluster resource manager. We attempted to bypass this hurdle by renting a higher-capacity machine (with 2TB of RAM) from a popular cloud compute provider, but we were still unable to train the new tokenizer as it still exceeded the available RAM. Our study aims to emulate a compute-constrained environment and continuing to scale such an experiment to these increased levels would be in opposition to our objective. During debugging, we isolated the RAM issue as specific to the combination of the size of the mC4 Thai training split with the LLaMa-2 tokenizer.

Table 5: Few-shot downstream task performance in each training setting for Hausa (`hau`), Luganda (`lug`), and Kinyarwanda (`kin`) averaged over 3 runs for all models. We report F1-score for MasakhaNER, ROUGE-L for XL-Sum, and accuracy for AfriMGSM and AfriXNLI. The MasakhaNER dataset is specific to each language, but AfriMGSM and AfriXNLI are parallel.

	LLaMa-2 7B												
Experiment	**hau**					**lug**				**kin**			
	XL-Sum	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*
Prompt	**0.1540**	**0.0987**	0.0227	0.3356	0.1953	**0.0698**	0.0227	0.3339	**-**	**0.0672**	0.0160	0.3356	**0.2180**
Translate	0.0432	0.0165	**0.0400**	**0.3511**	**0.2253**	0.0285	**0.0467**	**0.3894**	**0.2107**	0.0249	**0.0533**	**0.3894**	0.1967
LAFT	0.0778	0.0000	0.0120	0.3294	0.1573	0.0000	0.0000	0.0000	-	0.0000	0.0000	0.0000	0.0000
FOCUS	0.0187	0.0000	0.0080	0.1844	0.0727	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LAIT	0.0159	0.0000	0.0147	-	0.0073	0.0000	0.0200	-	-	0.0034	0.0107	-	0.0180

	MPT-7B												
	hau					**lug**				**kin**			
	XL-Sum	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*
Prompt	**0.1304**	**0.1702**	0.0253	0.3356	0.1987	**0.1820**	0.0147	0.0017	**-**	**0.1820**	**0.0213**	**0.2756**	**0.2200**
Translate	0.0443	0.0085	**0.0427**	**0.3417**	0.1627	0.0265	**0.0200**	**0.0783**	**0.1013**	0.0168	0.0200	0.1467	0.1627
LAFT	0.0621	0.0000	0.0053	0.3061	0.0627	0.0000	0.0000	0.0000	-	0.0000	0.0000	0.0000	0.0000
FOCUS	0.0138	0.0000	0.0053	0.0150	**0.2713**	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LAIT	0.0652	0.0000	0.0000	-	0.0000	0.0000	0.0000	-	-	0.0000	0.0013	0.0006	0.0000

	Phi-2												
	hau					**lug**				**kin**			
	XL-Sum	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*	*MasakhaNER*	*AfriMGSM*	*AfriXNLI*	*AfriMMLU*
Prompt	**0.1720**	**0.1382**	0.0200	**0.3372**	0.2007	**0.0817**	0.0213	**0.3272**	**-**	**0.0781**	0.0213	**0.3272**	0.2053
Translate	0.0410	0.0239	**0.1080**	0.3267	**0.2587**	0.0485	**0.0587**	0.2656	**0.2280**	0.0505	**0.1080**	0.3267	**0.2133**
LAFT	0.0930	0.0000	0.0107	0.0511	0.1400	0.0000	0.0000	0.0000	-	0.0000	0.0000	0.0000	0.0000
FOCUS	0.0274	0.0000	0.0067	0.1378	0.2033	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LAIT	0.1079	0.0000	0.0013	-	0.0020	0.0000	0.0040	-	-	0.0000	0.0040	-	0.0327

Table 6: Downstream task performance in each training setting for Burmese (`bur`) and Thai (`tha`) averaged over 3 runs for all models. We report ROUGE-L for XL-Sum, F1-score for WikiANN, and accuracy for MyanmarXNLI, XStoryCloze, MGSM, XNLI, XQUAD, and XCOPA. XL-Sum and WikiANN are language specific.

	LLaMa-2 7B									
Experiment	**bur**				**tha**					
	XL-Sum	*MyanmarXNLI*	*XStoryCloze*	*WikiANN*	*XL-Sum*	*MGSM*	*XNLI*	*WikiANN*	*XQUAD*	*XCOPA*
Prompt	-	-	-	-	-	-	-	-	-	**0.5233**
Translate	**0.0154**	**0.1519**	0.3896	**0.0847**	**0.0296**	**0.0120**	**0.2774**	**0.0053**	**0.0000**	0.5013
LAFT	-	-	-	-	-	-	-	-	-	-
FOCUS	0.0050	0.0726	**0.1738**	0.0168	-	-	-	-	-	-
LAIT	-	-	-	-	-	-	-	-	-	0.2927

	MPT-7B									
Experiment	**bur**				**tha**					
	XL-Sum	*MyanmarXNLI*	*XStoryCloze*	*WikiANN*	*XL-Sum*	*MGSM*	*XNLI*	*WikiANN*	*XQUAD*	*XCOPA*
Prompt	-	-	-	-	-	-	-	-	-	0.4333
Translate	**0.0157**	**0.1419**	0.3850	**0.0395**	**0.0246**	**0.0120**	**0.2464**	**0.0059**	**0.0000**	**0.5193**
LAFT	-	-	-	-	-	-	-	-	-	0.3827
FOCUS	0.0047	-	**0.4622**	0.0000	0.0087	0.0053	0.0599	-	-	0.3560
LAIT	-	-	-	-	-	-	-	-	-	0.0307

	Phi-2									
Experiment	**bur**				**tha**					
	XL-Sum	*MyanmarXNLI*	*XStoryCloze*	*WikiANN*	*XL-Sum*	*MGSM*	*XNLI*	*WikiANN*	*XQUAD*	*XCOPA*
Prompt	-	-	-	-	-	-	-	-	-	0.4473
Translate	**0.0100**	**0.1448**	0.1090	**0.0278**	**0.0270**	**0.0147**	**0.2735**	**0.0067**	**0.0019**	**0.4960**
LAFT	-	-	-	-	-	-	-	-	-	-
FOCUS	0.0047	-	**0.4013**	0.0000	0.0136	0.0000	0.0464	-	-	0.3733
LAIT	-	-	-	-	-	-	-	-	-	-

Method	Dataset	Average Input Perplexity	VOR
Prompt	AfriMGSM	6.43e+01	0.9450
	AfriMMLU	3.59e+01	0.9703
	AfriXNLI	2.51e+01	0.8675
	XCOPA	3.05e+00	0.9164
	XL-Sum	8.39e+01	1.0000
	masakhaNER	1.61e+01	0.6567
Translate	AfriMGSM	1.77e+01	0.8990
	AfriMMLU	1.17e+01	0.8353
	AfriXNLI	1.38e+01	0.7818
	MGSM	1.40e+01	0.8382
	MyanmarXNLI	1.94e+01	0.0000
	XCOPA	2.12e+01	0.9636
	XL-Sum	2.63e+01	1.0000
	XNLI	1.68e+01	0.9487
	XQUAD	1.53e+01	1.0000
	XStoryCloze	1.37e+01	0.0000
	masakhaNER	1.48e+01	0.5415
	wikiANN	1.35e+01	0.4207
LAFT	AfriMGSM	6.56e+01	0.1684
	AfriMMLU	8.08e+01	0.2348
	AfriXNLI	7.70e+01	0.2257
	XCOPA	3.66e+00	0.8730
	XL-Sum	1.19e+01	1.0000
	masakhaNER	2.18e+02	0.0211
FOCUS	AfriMGSM	6.58e+03	0.1393
	AfriMMLU	5.69e+03	0.2526
	AfriXNLI	6.48e+03	0.1105
	MGSM	3.38e+01	0.3453
	MyanmarXNLI	8.34e+01	0.2203
	XCOPA	3.07e+01	0.6927
	XL-Sum	7.53e+01	1.0000
	XNLI	1.63e+01	0.1516
	XStoryCloze	5.47e+01	0.7326
	masakhaNER	4.20e+03	0.0375
	wikiANN	9.28e+01	0.4856
LAIT	AfriMGSM	1.94e+09	0.4010
	AfriMMLU	3.15e+10	0.0411
	AfriXNLI	1.53e+05	0.1470
	XCOPA	2.55e+06	0.3237
	XL-Sum	1.93e+10	1.0000
	masakhaNER	3.19e+07	0.0341

Table 7: Average Input Perplexity and VOR Scores

Lang.	Example Input + Output
hau	Fitar da amsar arshe kawai ga tambayar lissafi. Leah nada 32 chaculet, yar uwarta kuma 42.gudanawa suka rage musu? –> 39 Fitar da amsar arshe kawai ga tambayar lissafi. Agwagin Janet suna yin wai 16 a kullun. Tana yin karin kumallo da guda uku kowace safiya, sannan tana gasawa kawayenta guda hudu kullum. A kullum takan sayar da ragowar a kasuwar manoma akan dala 2 akan kowane wai. Dala nawa take samu a kullum a kasuwar manoma? –> **29**
kin	Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. Leah afite shokola 32 naho umuvandimwe we afite 42. Nibarya 35 bazaba basigaranye shokola zingahe zose hamwe? –> 39 Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. Igishuhe cya Jane gitera amajyi 16 ku munsi, buri mugitondo aryamo atatu kandi akora umugati winshutiye akoresheje ane, agurisha asigaye mwisoko ryabahinzi buri munsi kugichiro cya 2 kuri buri jyi. Na ngahe mumadolali yinjiza ku munsi mwisoko ryabahinzi ? –> **39**
lug	Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. Leah yalina kyokuleeti 32 ate nga muganda we ye yalina 42. Bwe baba nga baalyako 35, baasigazaawo kyokuleeti mmeka bombi omugatte? –> 39 Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. Embaata za Janet zibiika amagi 16 buli lunaku. Alya amagi asatu buli lunaku ku kyenkya n'afumbisa amalala ana g'ateeka mu bukkeeki bwa muffin bw'akolera mikwano gye. Agasigadde agatunda mu katale k'abalimi n'abalunzi buli lunaku nga buli ggi alitunda $2. Afuna ssente mmeka buli lunaku mu katale k'abalimi n'abalunzi? –> **19**

Table 8: Example few-shot prompts and their respective model outputs for the Prompt adaptation method on AfriMGSM. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	Zai zain amsa daidai: A, B, C, ko D. Wani masanin kimiyya ya auna dayamita na gashin mutum hudu. Dayamitocin, a ma'anin milimita, sune 0.091, 0.169, 0.17, da 0.023. Wanne in'ikwaliti ne ya kwatanta biyu daga dayamitocin biyu na gashin an adam? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.17 < 0.091 –> A Zai zain amsa daidai: A, B, C, ko D. Menene matsayin p a cikin 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **A**
kin	Tora igisubizo gikwiye: A, B, C, cyangwa D. Umuhanga yapimye diameter yimisatsi ine yabantu. Diameter, muri milimetero, yari 0.091, 0.169, 0.17, na 0.023. Ni ubuhe busumbane bugereranya neza diameter yimisatsi ibiri muriyo misatsi yabantu? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.169 > 0.17 –> A Tora igisubizo gikwiye: A, B, C, cyangwa D. Nakahe gaciro ka p muri 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **A**

Table 9: Example few-shot prompts and their respective model outputs for the Prompt adaptation method on AfriMMLU. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. A Tsakanin 1936 da 1940 Greece na karkashin mulkin kama karya na Ioannus Metaxas, Ana iya tunawa da sutin (a'a) da ya amsa dashi zuwa ga Mussolini ultimatum yayi mubaya'a a 1940. Tattalin arzikin Greece bai yi kyau ba a arashin mulkin kama karya na soja na Metaxas. –> 1 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. Waannan rikirkitattun na'urar kwayoyin halitta sun samo asali ne saboda zain su a haka zai iya canza yanayi su gaba aya dan haka su kwayoyin halitta suna taruwa lokacin da yanayin su na gaba aya ya haaka kuma ya canza da yanayi da suke. Duk na'urorin kwayoyin halitta suna da wahalar sha'ani. –> 2 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. masu son karatu musamman wanda suka manne a ajin karatun sanin tattalin arziki da na na'ura mai kwakwalwa basu da wani alfanu nan gaba. masu san karatu basu da wani alfanu. –> 0 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. Gaskiya bana ba tunanin sa amma na fusata sosai, kuma dai daga karshe na ige da ara yi masa magana. Ban ara masa magana ba. –> **1**
kin	Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Hagati ya 1936 na 1940 Ubugereki bwari ku butegetsi bw'igitugu bwa gisirikare bwa Ioannis Metaxas, bwibukwa kubera echi yumvikana (oya) yatanze asubiza ultimatum ya Mussolini yokwiyegurira mu 1940. Ubukungu bw'Ubugereki ntabwo bwaribumeze neza kubutegetsi bwigitugu bwa gisirikare bwa Metaxas –> 1 Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Izi nzego zo murwego rwohejuru rwibikoresho bya molekile bivuka kubera ko gutoranya bisanzwe gushobora gukora kumitungo rusange yibintu bya molekile iyo iyo mitungo rusange yongerewe imbaraga zo guhuza n'imihindagurikire y'ikirere. Ibikoresho byose bya molekile biba bgoranye –> 2 Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Aba hanga bahatamye cyane mubyubukungu n'imyijyire ya kopyuta ,ninabo bafite ukwizera gucye Aba hanga bakompyuta ntakizere bafite –> 0 Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Urebye, ntabwo nigeze ntekereza kuribyo, ariko narumiwe cyane, ndangije nongeye kumuvugisha tena Ntabwo narinongera kumuvugisha –> **1**
lug	Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Mu mutendera oguddako wansi, ddayirekita w'akabinja ka al Qaeda mu kitongole kya CIA mu kiseera ekyo yajjukira nti yali talowooza nti gwali mulimu gwe okulagira ekirina okukolebwa oba obutakolebwa. Ddayirekita w'ekitundu ekyo yali tayagala kwenyigira mu kuddukanya ekyo ekyali kikolebwa. –> 0 Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Mary Traill ajja kukikugambako. Nkimanyiiko. –> 2 Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Naye nedda, omanyi sikaati na bulawuzi oba ng'olaba kiteeteeyi wano, naye kirungi gyendi okukolera awaka kubanga mba nsobola n'okwambala engoye z'omunda. Ssambala kintu kirala kyonna okuggyako essweta bwe nkolera ewaka. –> 1 Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Kale nno, ekyo si na kye nnabadde ndowoozaako, naye olw'okuba nnabadde mu mbeera ey'okusoberwa, nnawunzise nzizeemu okwogera naye. Sinnaddamu kwogerako naye. –> **0**

Table 10: Example few-shot prompts and their respective model outputs for the Prompt adaptation method on AfriXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
tha	กำหนดสาเหตุหรือผลของสถานที่ตั้ง เอาต์พุต 0 สำหรับตัวเลือกแรก หรือ 1 สำหรับตัวเลือกที่สอง คนร้ายปล่อยตัวประกัน ผลเป็นยังไงบ้างคะ? 0: พวกเขายอมรับค่าไถ่ 1: พวกเขาหนีออกจากคุก –> 0 กำหนดสาเหตุหรือผลของสถานที่ตั้ง เอาต์พุต 0 สำหรับตัวเลือกแรก หรือ 1 สำหรับตัวเลือกที่สอง สิ่งของถูกห่อไว้ในพลาสติก ผลเป็นยังไงบ้างคะ? 0: มันบอบบาง 1: มันเล็ก –> **0**

Table 11: Example few-shot prompts and their respective model outputs for the Prompt adaptation method on XCOPA. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Samar da kanun labarai don taaitawar labarai. Sarki Abdullah na Saudi Arabia, ya yi suka kan abin da ya kira, fakewar da 'yan ta'adda ke yi da addini suna tafka ta'asa. –> Sarki Abdullah: 'Yan ta'adda na fakewa da addini Samar da kanun labarai don taaitawar labarai. Ta dai tabbata cewa maharin da ya tarwatsa kansa a gidan raye-rayen Manchester, Salman Abedi ya koma Burtaniya ne daga etare, kwanaki alilan kafin ya kai wannan farmaki. –> **alilan kafin ya kai wannan farmakiSamar da kanun labarai don taaitawar labarai. Ta dai tabbata cewa maharin da ya tarwatsa kansa a gidan raye-rayen Manchester, Salman Abedi ya koma Burtaniya ne daga etare, kwanaki alilan kafin ya kai wannan farmaki. –> alilan kafin ya kai wannan farmakiSamar da kanun labarai don taaitawar labarai. Ta dai tabbata cewa maharin da ya tarwatsa kansa a gidan raye-rayen Manchester, Salman Abedi ya koma Burtaniya ne daga etare, kwanaki alilan kafin ya kai wannan farmaki. –> alilan kafin ya kai wannan farmakiSamar da kanun labarai don taaitaw**
tha	ระบุหัวข้อข่าวสรุป ในทางการตลาด น้ำมันปลาถูกโฆษณาให้เป็นอาหารเสริมสำหรับสตรีมีครรภ์ แต่การศึกษาผู้หญิงตั้งครรภ์ 2,500 คน เป็นเวลา 10 ปี นักวิจัยในออสเตรเลียพบว่า น้ำมันปลาไม่ได้ช่วยเพิ่มระดับสติปัญญาของทารก –> น้ำมันปลาไม่ช่วยให้ทารกฉลาดขึ้น ระบุหัวข้อข่าวสรุป กรุงนิวเดลี เมืองหลวงของอินเดีย ออกมาตรการสลับวันขับรถยนต์ตามเลขทะเบียน เพื่อรับมือกับระดับมลพิษที่เพิ่มสูงจนเป็นอันตราย –> กรุงนิวเดลี เมืองหลวงของอินเดีย ออกมาตรการสลับวันขับรถยนต์ตามเลขทะเบียน เพื่อรับมือกับระดับมลพิษที่เพิ่มสูงจนเป็นอันตรายระบุหัวข้อข่าวสรุป ผู้หญิงที่มีสัตว์เลี้ยง เป็นผู้ที่มีความเสี่ยงสูงที่สุด ในการเสียชีวิต เพื่อสร้างสรรค์ และเพื่อสร้างสรรค์

Table 12: Example few-shot prompts and their respective model outputs for the Prompt adaptation method on XL-Sum. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Sai dai mai sharhi akan harkokin siyasa na kasar Delphin Kapaya ya ce yadda kotun ta gudanar da wannan sharaar shine zai nuna irin mataki nagaba da magoya bayan Beman zasu dauka . –> (Sai,O), (dai,O), (mai,O), (sharhi,O), (akan,O), (harkokin,O), (siyasa,O), (na,O), (kasar,O), (Delphin,B-PER), (Kapaya,I-PER), (ya,O), (ce,O), (yadda,O), (kotun,O), (ta,O), (gudanar,O), (da,O), (wannan,O), (sharaar,O), (shine,O), (zai,O), (nuna,O), (irin,O), (mataki,O), (nagaba,O), (da,O), (magoya,O), (bayan,O), (Beman,B-PER), (zasu,O), (dauka,O), (.,O) Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Ya kuma yaba da shawarar da bangaren al - Barnawi na Boko Haram ya yanke na sassautawa a gwagwarmayarsu . –> **[]**
kin	Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. Amazon iteganya gushora miliyari 6 , 5 zamadorali mu bikorwa byo gucururiza kuri internet , ndetse ngo ikaba izoroherwa no gukoresha internet ya Bharti ku giciro gito mu gihe ibigo byombi bizaba byemeranyije amasezerano byifuza . –> (Amazon,B-ORG), (iteganya,O), (gushora,O), (miliyari,O), (6,O), (,,O), (5,O), (zamadorali,O), (mu,O), (bikorwa,O), (byo,O), (gucururiza,O), (kuri,O), (internet,O), (,,O), (ndetse,O), (ngo,O), (ikaba,O), (izoroherwa,O), (no,O), (gukoresha,O), (internet,O), (ya,O), (Bharti,B-ORG), (ku,O), (giciro,O), (gito,O), (mu,O), (gihe,O), (ibigo,O), (byombi,O), (bizaba,O), (byemeranyije,O), (amasezerano,O), (byifuza,O), (.,O) Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. Bazwi mu cyo bise Morning Worship aho baririmba ibihangano bitandukanye byo mu gitabo bigafasha benshi . –> **['(Morning,ORG)', '(worship,ORG)', '(mu,ORG)', '(cyo,ORG)', '(bise,ORG)', '(aho,ORG)', '(baririmba,ORG)', '(ibihangano,ORG)', '(bitandukanye,ORG)', '(byo,ORG)', '(gitabo,ORG)', '(bigafasha,ORG)', '(benshi,ORG)', '(.,ORG)', '(Morning,ORG)', '(worship,ORG)', '(mu,ORG)', '(cyo,ORG)', '(bise,ORG)']**
lug	Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. Abantu abaatuwa obuyambi bampa sikaala okugenda mu Amerika okusoma diguli eyookubiri olwo bizinensi yenkoko ne ngiwa mukwano gwange Geoffrey Lwanga nga kati mu kiseera kino alina enkoko ezisoba mu 7000 ezamagi . –> (Abantu,O), (abaatuwa,O), (obuyambi,O), (bampa,O), (sikaala,O), (okugenda,O), (mu,O), (Amerika,B-LOC), (okusoma,O), (diguli,O), (eyookubiri,O), (olwo,O), (bizinensi,O), (yenkoko,O), (ne,O), (ngiwa,O), (mukwano,O), (gwange,O), (Geoffrey,B-PER), (Lwanga,I-PER), (nga,O), (kati,O), (mu,O), (kiseera,B-DATE), (kino,I-DATE), (alina,O), (enkoko,O), (ezisoba,O), (mu,O), (7000,O), (ezamagi,O), (.,O) Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. Ono ye waffe era kampeyini ze okuziyimirizaawo tujja kwesondamu ensimbi ezinamuyamba okukuba ebipande ebipande nokukola emirimu emirara , Rose Namuli akolera ku katale ka Pepsi oluvanyuma namuwa 2 , 000 . –> **['(Rose Namuli, PER PER)', '(Pepsi,LOC)']**

Table 13: Example few-shot prompts and their respective model outputs for the Prompt adaptation method on masakhaNER. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Just give the final answer to the math question. Leah chose 32 chocolates, and her sister 42. Do they have any left? –> 39 Just to give the final answer to the math question. Janet's aunt makes 16 eggs a day. She makes three for breakfast each morning, and then she brews four for her friends every day. She sells the leftovers at the farmers market for $2 per egg. How many dollars does she make a day at the farmers market? –> **16**
kin	Only the last answer to the math questions is given. Leah has 32 chocolates and her brother has 42. If there are at least 35 chocolates, how many chocolates will they have left in total? –> 39 It only comes out the last answer to the math questions. Jane's salary is 16 cents a day, she sleeps in three and makes a nice loaf of bread with four, she sells the rest at the farmers market every day for 2 cents a day. How many dollars does she make a day at the farmers market? –> **18**
lug	Write out the last answer to the number question. Leah had 32 and her brother had 42. If they ate 35, how many were left? –> 39 Give only the last answer to the math question. Janet's chickens lay 16 eggs a day. She eats three eggs a day for breakfast and cooks four more for her friends' muffin. She sells them at the farmers market every day for $2 apiece. How much money does she make a day at the farmers market? –> **12**

Table 14: Example few-shot prompts and their respective model outputs for the Translate adaptation method on AfriMGSM. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	Choose the correct answer: A, B, C, or D. A scientist measured the diameter of four human hairs. The diameters, in millimeters, are 0.091, 0.169, 0.17, and 0.023. –> A Choose the correct answer option: A, B, C, or D. What is the position of p in 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **A**
kin	The diameter, in millimeters, was 0.091, 0.169, 0.17, and 0.023. What inequality best represents the diameter of two of the hairs in a human hair? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.169 > 0.17 –> A Find the correct answer: A, B, C, or D. What is the value of p in 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **A**
lug	Choose the correct answer: A, B, C, or D. The scientist measured the width of four human hair strands. The fractional lengths are 0.091, 0.169, 0.17, and 0.023. What is the relationship between the exact values that can be used to compare the width of two human hair strands? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.17 < 0.091 –> A Choose the correct answer: A, B, C, or D. What is the value of p in 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **A**

Table 15: Example few-shot prompts and their respective model outputs for the Translate adaptation method on AfriMMLU. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	Determine if the prediction follows the theme closely. Output 0 for entailment, 1 for neutrality, or 2 for contradiction. Between 1936 and 1940 Greece was under the dictatorship of loannus Metaxas, It may be remembered for the sutin (no) he responded to Mussolini's ultimatum to capitulate in 1940. The Greek economy did not fare well under Metaxas' military dictatorship. –> 1 Determine if the prediction follows the theme closely. Exit 0 for entailment, 1 for neutrality, or 2 for contradiction. These complex molecular machines evolved because their selection could change their overall state so that their molecular assemblies when their overall state evolved and changed with the state they were in. All molecular machines are complex. –> 2 Determine if the prediction follows the theme closely. Output 0 for entailment, 1 for neutrality, or 2 for contradiction. amateur readers especially who stuck to the economics and computer science classes had no future advantages. knowledgeable readers had no advantages. –> 0 Determine if the prediction follows the topic carefully. Output 0 for entailment, 1 for neutrality, or 2 for contradiction. I honestly didn't think about it but I was very angry, and I finally snapped and spoke to him again. I didn't speak to him again. –> **1**
kin	Determine whether the hypothesis is supported. Outputs 0 for explanation, 1 for neutrality, or 2 for contradiction. Between 1936 and 1940 Greece was under the military dictatorship of Ioannis Metaxas, remembered for his eloquent (no) response to Mussolini's ultimatum to surrender in 1940. The Greek economy did not adapt well to Metaxas' military dictatorship –> 1 Determine whether the hypothesis is supported. Outputs 0 for explanation, 1 for neutrality, or 2 for contradiction. These higher-order classes of molecular properties arise because natural selection can act on the shared properties of molecular entities when those shared properties are enhanced to accommodate the changing environment. All molecular entities are complex –> 2 Determine whether the hypothesis is supported. Outputs 0 for explanation, 1 for neutrality, or 2 for contradiction. These nations are highly fragmented economically and computer-centrically, or have little faith in the computer-centricity of their nations. These nations have no confidence in the computer-centricity of their nations –> 0 Determine whether the hypothesis is supported. Outputs 0 for explanation, 1 for objection, or 2 for contradiction. Actually, I never thought about that, but I was very surprised, so I talked to him again. I never spoke to him again –> **0**
lug	Determine whether a hypothesis follows a premise logically. The output is 0 for entailment, 1 for neutral, or 2 for contradiction. Between 1936 and 1940 Greece was under the dictatorship of Ioannis Metaxas, best remembered for his 'No' response to Mussolini's offer of a hanging sentence after his defeat in 1940. The Greek economy did not fare well during the period under the dictatorship of Metaxas. –> 1 Determine whether a hypothesis follows logically from a premise. The output is 0 for entailment, 1 for neutral, or 2 for contradiction. Higher-order functions arise because the universe has the capacity to do so when it is adapted to do so. All functions are higher-order. –> 2 Determine whether the hypothesis follows logically from the premise. The output is 0 for entailment, 1 for neutral, or 2 for contradiction. People who lack social skills seek safety in economics or computer science classes, and are more likely to live in a hopeless situation. People who lack social skills are hopeless. –> 0 Decide whether the hypothesis follows logically from the premise. The output is 0 for the entailment, 1 for the neutral, or 2 for the contradiction. Well, that's not what I was thinking, but because I was in a state of confusion, I ended up talking to him again. I never spoke to him again. –> **1**

Table 16: Example few-shot prompts and their respective model outputs for the Translate adaptation method on AfriXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
tha	First, think of a step-by-step way to answer a math problem, then print out the final answer. Challenge: Leah has 32 chocolates and her sister has 42. If both of them ate 35 chocolates, how many chocolates would be left? –> 39 First, think of a step-by-step way to answer a math question, then print out the final answer. Janet's egg lays 16 pounds of eggs a day, she eats three eggs for breakfast every day, and she makes four for her friends every day, she sells the rest at the farmers market every day for $2 for a fresh egg, how much money does she make from the farmers market per day? –> **<NAN>**

Table 17: Example few-shot prompts and their respective model outputs for the Translate adaptation method on MGSM. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	Determine whether the hypothesis is capital compatible: yield 0 for correlation, 1 for neutrality, or 2 for inverse. McKim not only lost due to his many frustrations, but finished third behind Howard & Cowell. McKim was satisfied that he had finished first –> 2 We can then determine whether the concept is capital compatible, yielding 0 for coherence, 1 for neutrality, or 2 for inversion. We can be surprised that others use language in a simple way, and that it ends on our analytical side and begins on our emotional side. And I think the really interesting thing is, what can we do about this? I mean, we have to change the people who are going to represent us. And it's so boring, and we know that it's not worth changing our representation, so we shouldn't even try to change it. And then we have the inverse of the equation, which is the inverse of the equation, and we have the inverse of the equation, which is the inverse of the equation, and we have the inverse of the equation, which is the inverse of the equation, and we have the inverse of the equation, which is the inverse of the equation. –> **<NAN>**

Table 18: Example few-shot prompts and their respective model outputs for the Translate adaptation method on MyanmarXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
tha	Determine the cause and effect of the location, put 0 for the first option or 1 for the second option, the perpetrator releases the hostage, what is the result? 0: They accept the ransom: 1: They escape from prison –> 0 Determine the cause and effect of the location, put 0 for the first option, or 1 for the second option, the object is wrapped in plastic, what is the result? 0: it's thin 1: it's small –> **0**

Table 19: Example few-shot prompts and their respective model outputs for the Translate adaptation method on XCOPA. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	Give a headline for your news summary: Myanmar government forces and the KLA forces have been engaged in close combat since the morning of July 7th around the old exit road of Kokraj. Give a headline for this summary: Japanese government cancels plans to build the stadium that will host the 2020 Tokyo Olympics. –> **Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a headline for this summary: The United States and China have agreed to resume trade talks.Give a head**
hau	King Abdullah of Saudi Arabia, has criticized what he calls, the religious cover-up of terrorism in his country. She confirmed that the Manchester nightclub suicide bomber, Salman Abedi, had returned to the UK from overseas, just days before the attack. –> **2017* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Yemen* 201718 Saudi-led intervention in Y**
tha	In the commercials, fish oil is promoted as a supplement for pregnant women, but after studying 2,500 pregnant women for 10 years, researchers in Australia found that fish oil did not increase the intelligence of the baby – fish oil did not make the baby smarter. In a headline summary, India's capital New Delhi has instituted a shift in driving days to license plate numbers to deal with dangerously high pollution levels. –> **In a headline summary, a new study finds that the number of people who have died from the coronavirus in the United States is 10 times higher than the official number.In a headline summary, a new study finds that the number of people who have died from the coronavirus in the United States is 10 times higher than the official number.In a headline summary, a new study finds that the number of people who have died from the coronavirus in the United States is 10 times higher than the official number.In a headline summary, a new study finds that the number of people who have died from the coronavirus in the United States is 10 times higher than the official number.In a headline summary, a new study finds that the number of people who have died from the coronavirus in the United States is 10 times higher than the official number.In a headline summary, a new study finds that the number of people who have died from the coronavirus in the United States is 10 times higher than the official number.In a headline summary, a new study finds that the number of**

Table 20: Example few-shot prompts and their respective model outputs for the Translate adaptation method on XL-Sum. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
tha	Considering whether the assumption is logical, putting 0 for participation, 1 for neutrality or 2 for conflict, McKim was very disappointed that he not only lost but came in third behind Howard & Cauldwell. McKim was delighted because he finished first –> 2 Consider whether the assumption is rational, give it a 0 for participation, a 1 for neutrality or a 2 for conflict, others will still be just amazed at the language and wonder just where our analytical side ends and our emotional side begins. Consider whether the assumption is rational, put 0 for participation, 1 for neutrality or 2 for conflict, and that's what I think it would be really interesting is what we do about it. I mean, we have to change who represents us. I just know it's boring and not worth changing who represents us, so we shouldn't try to change – 2 Consider whether the assumption is rational, put 0 for participation, 1 for neutrality or 2 for conflict. Well, I didn't think anything of it, but I was disappointed, and, I went back to talk to him, and I didn't talk to him again. –> **<NAN>**

Table 21: Example few-shot prompts and their respective model outputs for the Translate adaptation method on XNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
tha	Considering the context, the mechanisms used to evade the adaptive immune system are complex, the simplest method being to rapidly change the non-essential antigen (amino acids and/or sugars) on the surface of the virus while protecting the important antigen. This method is called antigenic mutation, for example, HIV, which rapidly changes the shape of proteins on its viral coat that are important for entering the target cells of the host. These frequent changes in the antigen can be explained as a failure of the genes that are involved in what the virus is meant to do. This virus has been brought to the ultimate stage of uncontrolled cell detection. The same mechanism can be used to prevent changes in the immune system from itself by preventing the immune system from recognizing other cells that are not immune. In some cases, it has been proposed to hold a plea bargain for the perpetrators of rape, as in the case of Camden 28, where the accused was offered the opportunity to confess to a crime in order to avoid imprisonment. In some mass arrest situations, activists decided to use the same unity strategy so that everyone could confess with the same plea bargain, but some activists chose to confess to the crime, admitted without any plea bargain, Mahatma Gandhi confessed, and told the court, "I am here... willing to accept the maximum punishment that can be imposed on me for what I consider to be a legal crime, which was planned in advance, but which I consider to be the highest duty of the citizenry to impose on the perpetrator". –> **<NAN>**

Table 22: Example few-shot prompts and their respective model outputs for the Translate adaptation method on XQUAD. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
bur	Given the context, choose the best ending for the story: 1 or 2. This Sunday, there is a lot for Amber to do. She has made a list of places to go. She is in a hurry to get ready. She is worried about the time. 1: Amber enjoys the comfortable two-hour breakfast and lunch combination. 2: Amber left the list at home and had to work in a hurry. Given the context, choose the best ending for the story: 1 or 2. I became a fan of Law and Order in 2011. I had recovered from a stroke. When I got home, I tried to watch every episode. –> **<NAN>**

Table 23: Example few-shot prompts and their respective model outputs for the Translate adaptation method on XStoryCloze. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	Place each word in the following sentence with the NER symbol. However, the commentator on national politics Delphin Kapaya says that the way the court handled this ruling will indicate what action the Beman supporters will take. –> (Sai,O), (dai,O), (mai,O), (comment,O), (akan,O), (politik,O), (politik,O), (de,O), (negara,O), (Delphin,B-PER), (Kapaya,I-PER), (ya,O), (ce,O), (how,O), (kotun,O), (ta,O), (gover,O), (da,O), (wannan,O), (shara,O), (Oshine), (O), (Ouna,O), (O), (mat,O), (O), (magoda,O), (Bira, Place each word in the following sentence with the NER symbol. He also praised the decision of the al-Barnawi faction of Boko Haram to ease their struggle. –> **['(Delphin Kapaya,PER PER)', '(Oshine,Oshine)', '(Delphin,PER)']**
kin	Enter each word in the following sentence with the NER tag. Amazon plans to invest $6.5 billion in online retail and will be able to access Bharti's low-cost Internet service if the two companies agree to the desired deal. –> (Amazon,B-ORG), (plan,O), (invest,O), (billion,O), (6,O), (O), (O), (5,O), (zamadorali,O), (in,O), (activity,O), (that), (buy,O), (true,O), (internet,O), (), (even,O), (price,O), (Ease,O), (O), (Want,O), (U), (internet), (U), (Bharti), (B-G), (G), (ORG), (O), (O), (O), (O), (O Enter each word in the following sentence with the NER tag. They are known for their so-called Morning Worship where they sing a variety of songs from the book to help many. –> **[]**
lug	Each word in the following sentence has its own NER symbol. The sponsors gave me a scholarship to go to the United States to study for a master's degree and then the chicken business was given to my friend Geoffrey Lwanga who currently has over 7000 chickens. –> (People,O), (Give,O), (Help,O), (Give,O), (School,O), (Go,O), (In,O), (America,B-LOC), (Read,O), (Language,O), (Second,O), (Follow,O), (Business,O), (Chicken,O), (Ne,O), (Give,O), (Other), (Friend,O), (Off), (Geoff,B-PER), (Language,PER), (I), (O), (In), (Now Each word in the following sentence has its own NER symbol. This is ours and we will raise funds to support her campaigns and to help her create posters and create works of art. Rose Namuli works for Pepsi and I gave her 2, 000. –> **['(America,LOC)', '(Other,Other)', '(Off,Off)']**

Table 24: Example few-shot prompts and their respective model outputs for the Translate adaptation method on masakhaNER. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	Mark each word with its NER tag in the following sentences: husband, B-PER, husband, I-PER, ((,I-PER), husband, I-PER, (), I-PER Mark each word in the following sentence with its NER tag. Located in Alam, Yangon Province, and opened in 1963 by the Burmese timber industry. The courses that are open are: –> **['((, PER)', '(,)', '((, PER)', '(,)', '((, PER)', '(,)']**
tha	Mark each word in the following sentences with the NER tag. (,O), (B-ORG), (A,I-ORG), (R,I-ORG), (B,I-ORG), (I,I-ORG), (N,I-ORG), (I,I-ORG), (T,I-ORG), (I,I-ORG) and (I,I-ORG) and then write the following: Mark each word in the following sentences with the NER tag load load (logon A F 4) –> **['(logonAF4,logonAF4)', '(logonAF4,logonAF4)']**

Table 25: Example few-shot prompts and their respective model outputs for the Translate adaptation method on wikiANN. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Fitar da amsar arshe kawai ga tambayar lissafi. Leah nada 32 chaculet, yar uwarta kuma 42.gudanawa suka rage musu? –> 39 Fitar da amsar arshe kawai ga tambayar lissafi. Agwagin Janet suna yin wai 16 a kullun. Tana yin karin kumallo da guda uku kowace safiya, sannan tana gasawa kawayenta guda hudu kullum. A kullum takan sayar da ragowar a kasuwar manoma akan dala 2 akan kowane wai. Dala nawa take samu a kullum a kasuwar manoma? –> **39**
kin	Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. Leah afite shokola 32 naho umuvandimwe we afite 42. Nibarya 35 bazaba basigaranye shokola zingahe zose hamwe? –> 39 Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. Igishuhe cya Jane gitera amajyi 16 ku munsi, buri mugitondo aryamo atatu kandi akora umugati winshutiye akoresheje ane, agurisha asigaye mwisoko ryabahinzi buri munsi kugichiro cya 2 kuri buri jyi. Na ngahe mumadolali yinjiza ku munsi mwisoko ryabahinzi ? –> **<NAN>**
lug	Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. Leah yalina kyokuleeti 32 ate nga muganda we ye yalina 42. Bwe baba nga baalyako 35, baasigazaawo kyokuleeti mmeka bombi omugatte? –> 39 Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. Embaata za Janet zibiika amagi 16 buli lunaku. Alya amagi asatu buli lunaku ku kyenkya n'afumbisa amalala ana g'ateeka mu bukkeeki bwa muffin bw'akolera mikwano gye. Agasigadde agatunda mu katale k'abalimi n'abalunzi buli lunaku nga buli ggi alitunda $2. Afuna ssente mmeka buli lunaku mu katale k'abalimi n'abalunzi? –> **<NAN>**

Table 26: Example few-shot prompts and their respective model outputs for the LAFT adaptation method on AfriMGSM. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Zai zain amsa daidai: A, B, C, ko D. Wani masanin kimiyya ya auna dayamita na gashin mutum hudu. Dayamitocin, a ma'anin milimita, sune 0.091, 0.169, 0.17, da 0.023. Wanne in'ikwaliti ne ya kwatanta biyu daga dayamitocin biyu na gashin an adam? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.17 < 0.091 –> A Zai zain amsa daidai: A, B, C, ko D. Menene matsayin p a cikin 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **B**
kin	Tora igisubizo gikwiye: A, B, C, cyangwa D. Umuhanga yapimye diameter yimisatsi ine yabantu. Diameter, muri milimetero, yari 0.091, 0.169, 0.17, na 0.023. Ni ubuhe busumbane bugereranya neza diameter yimisatsi ibiri muriyo misatsi yabantu? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.169 > 0.17 –> A Tora igisubizo gikwiye: A, B, C, cyangwa D. Nakahe gaciro ka p muri 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **<NAN>**

Table 27: Example few-shot prompts and their respective model outputs for the LAFT adaptation method on AfriMMLU. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. A Tsakanin 1936 da 1940 Greece na karkashin mulkin kama karya na loannus Metaxas, Ana iya tunawa da sutin (a'a) da ya amsa dashi zuwa ga Mussolini ultimatum yayi mubaya'a a 1940. Tattalin arzikin Greece bai yi kyau ba a arashin mulkin kama karya na soja na Metaxas. –> 1 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. Waannan rikirkitattun na'urar kwayoyin halitta sun samo asali ne saboda zain su a haka zai iya canza yanayi su gaba aya dan haka su kwayoyin halitta suna taruwa lokacin da yanayin su na gaba aya ya haaka kuma ya canza da yanayi da suke. Duk na'urorin kwayoyin halitta suna da wahalar sha'ani. –> 2 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. masu son karatu musamman wanda suka manne a ajin karatun sanin tattalin arziki da na na'ura mai kwakwalwa basu da wani alfanu nan gaba. masu san karatu basu da wani alfanu. –> 0 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. Gaskiya bana ba tunanin sa amma na fusata sosai, kuma dai daga karshe na ige da ara yi masa magana. Ban ara masa magana ba. –> **1**
kin	Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Hagati ya 1936 na 1940 Ubugereki bwari ku butegetsi bw'igitugu bwa gisirikare bwa Ioannis Metaxas, bwibukwa kubera echi yumvikana (oya) yatanze asubiza ultimatum ya Mussolini yokwiyegurira mu 1940. Ubukungu bw'Ubugereki ntabwo bwaribumeze neza kubutegetsi bwigitugu bwa gisirikare bwa Metaxas –> 1 Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Izi nzego zo murwego rwohejuru rwibikoresho bya molekile bivuka kubera ko gutoranya bisanzwe gushobora gukora kumitungo rusange yibintu bya molekile iyo iyo mitungo rusange yongerewe imbaraga zo guhuza n'imihindagurikire y'ikirere. Ibikoresho byose bya molekile biba bgoranye –> 2 Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Aba hanga bahatamye cyane mubyubukungu n'imyijyire ya kopyuta ,ninabo bafite ukwizera gucye Aba hanga bakompyuta ntakizere bafite –> 0 Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Urebye, ntabwo nigeze ntekereza kuribyo, ariko narumiwe cyane, ndangije nongeye kumuvugisha tena Ntabwo narinongera kumuvugisha –> **<NAN>**
lug	Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Mu mutendera oguddako wansi, ddayirekita w'akabinja ka al Qaeda mu kitongole kya CIA mu kiseera ekyo yajjukira nti yali talowooza nti gwali mulimu gwe okulagira ekirina okukolebwa oba obutakolebwa. Ddayirekita w'ekitundu ekyo yali tayagala kwenyigira mu kuddukanya ekyo ekyali kikolebwa. –> 0 Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Mary Traill ajja kukikugambako. Nkimanyiiko. –> 2 Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Naye nedda, omanyi sikaati na bulawuzi oba ng'olaba kiteeteeyi wano, naye kirungi gyendi okukolera awaka kubanga mba nsobola n'okwambala engoye z'omunda. Ssambala kintu kirala kyonna okuggyako essweta bwe nkolera ewaka. –> 1 Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa 0 ku entailment, 1 ku neutral, oba 2 ku contradiction. Kale nno, ekyo si na kye nnabadde ndowoozaako, naye olw'okuba nnabadde mu mbeera ey'okusoberwa, nnawunzise nzizeemu okwogera naye. Sinnaddamu kwogerako naye. –> **<NAN>**

Table 28: Example few-shot prompts and their respective model outputs for the LAFT adaptation method on AfriXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
tha	กำหนดสาเหตุหรือผลของสถานที่ตั้ง เอาต์พุต 0 สำหรับตัวเลือกแรก หรือ 1 สำหรับตัวเลือกที่สอง คนร้ายปล่อยตัวประกัน ผลเป็นยังไงบ้างคะ? 0: พวกเขายอมรับค่าไถ่ 1: พวกเขาหนีออกจากคุก –> 0 กำหนดสาเหตุหรือผลของสถานที่ตั้ง เอาต์พุต 0 สำหรับตัวเลือกแรก หรือ 1 สำหรับตัวเลือกที่สอง สิ่งของถูกห่อไว้ในพลาสติก ผลเป็นยังไงบ้างคะ? 0: มันบอบบาง 1: มันเล็ก –> **0**

Table 29: Example few-shot prompts and their respective model outputs for the LAFT adaptation method on XCOPA. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Samar da kanun labarai don taaitawar labarai. Sarki Abdullah na Saudi Arabia, ya yi suka kan abin da ya kira, fakewar da 'yan ta'adda ke yi da addini suna tafka ta'asa. –> Sarki Abdullah: 'Yan ta'adda na fakewa da addini Samar da kanun labarai don taaitawar labarai. Ta dai tabbata cewa maharin da ya tarwatsa kansa a gidan raye-rayen Manchester, Salman Abedi ya koma Burtaniya ne daga etare, kwanaki alilan kafin ya kai wannan farmaki. –> **ungiyar ta ce, "Mun kasance mun**
tha	ระบุหัวข้อข่าวสรุป ในทางการตลาด น้ำมันปลาถูกโฆษณาให้เป็นอาหารเสริมสำหรับสตรีมีครรภ์ แต่การศึกษาผู้หญิงตั้งครรภ์ 2,500 คน เป็นเวลา 10 ปี นักวิจัยในออสเตรเลียพบว่า น้ำมันปลาไม่ได้ช่วยเพิ่มระดับสติปัญญาของทารก –> น้ำมันปลาไม่ช่วยให้ทารกฉลาดขึ้น ระบุหัวข้อข่าวสรุป กรุงนิวเดลี เมืองหลวงของอินเดีย ออกมาตรการสลับวันขับรถยนต์ตามเลขทะเบียน เพื่อรับมือกับระดับมลพิษที่เพิ่มสูงจนเป็นอันตราย –> น้ำมันปลาไม่ได้ช่วยให้ทารกฉลาดขึ้นน้ำมันปลาไม่ได้ช่วยให้ทารกฉลาดขึ้นน้ำมันปลาไม่ได้ช่วยให้ทารกฉลาดขึ้นน้ำมันปลาไม่ได้ช่วยให้ทารกฉลาดขึ้นน้ำมันปลาไม่ได้ช่วยให้ทารกฉลาดขึ้นน้ำมันปลาไม่ได้ช่วยให้ทารกฉลาดขึ้นน้ำมันปลาไ

Table 30: Example few-shot prompts and their respective model outputs for the LAFT adaptation method on XL-Sum. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Sai dai mai sharhi akan harkokin siyasa na kasar Delphin Kapaya ya ce yadda kotun ta gudanar da wannan sharaar shine zai nuna irin mataki nagaba da magoya bayan Beman zasu dauka . –> (Sai,O), (dai,O), (mai,O), (sharhi,O), (akan,O), (harkokin,O), (siyasa,O), (na,O), (kasar,O), (Delphin,B-PER), (Kapaya,I-PER), (ya,O), (ce,O), (yadda,O), (kotun,O), (ta,O), (gudanar,O), (da,O), (wannan,O), (sharaar,O), (shine,O), (zai,O), (nuna,O), (irin,O), (mataki,O), (nagaba,O), (da,O), (magoya,O), (bayan,O), (Beman,B-PER), (zasu,O), (dauka,O), (.,O) Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Ya kuma yaba da shawarar da bangaren al - Barnawi na Boko Haram ya yanke na sassautawa a gwagwarmayarsu . –> []
kin	Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. Amazon iteganya gushora miliyari 6 , 5 zamadorali mu bikorwa byo gucururiza kuri internet , ndetse ngo ikaba izoroherwa no gukoresha internet ya Bharti ku giciro gito mu gihe ibigo byombi bizaba byemeranyije amasezerano byifuza . –> (Amazon,B-ORG), (iteganya,O), (gushora,O), (miliyari,O), (6,O), (,,O), (5,O), (zamadorali,O), (mu,O), (bikorwa,O), (byo,O), (gucururiza,O), (kuri,O), (internet,O), (,,O), (ndetse,O), (ngo,O), (ikaba,O), (izoroherwa,O), (no,O), (gukoresha,O), (internet,O), (ya,O), (Bharti,B-ORG), (ku,O), (giciro,O), (gito,O), (mu,O), (gihe,O), (ibigo,O), (byombi,O), (bizaba,O), (byemeranyije,O), (amasezerano,O), (byifuza,O), (.,O) Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. Bazwi mu cyo bise Morning Worship aho baririmba ibihangano bitandukanye byo mu gitabo bigafasha benshi . –> []
lug	Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. Abantu abaatuwa obuyambi bampa sikaala okugenda mu Amerika okusoma diguli eyookubiri olwo bizinensi yenkoko ne ngiwa mukwano gwange Geoffrey Lwanga nga kati mu kiseera kino alina enkoko ezisoba mu 7000 ezamagi . –> (Abantu,O), (abaatuwa,O), (obuyambi,O), (bampa,O), (sikaala,O), (okugenda,O), (mu,O), (Amerika,B-LOC), (okusoma,O), (diguli,O), (eyookubiri,O), (olwo,O), (bizinensi,O), (yenkoko,O), (ne,O), (ngiwa,O), (mukwano,O), (gwange,O), (Geoffrey,B-PER), (Lwanga,I-PER), (nga,O), (kati,O), (mu,O), (kiseera,B-DATE), (kino,I-DATE), (alina,O), (enkoko,O), (ezisoba,O), (mu,O), (7000,O), (ezamagi,O), (.,O) Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. Ono ye waffe era kampeyini ze okuziyimirizaawo tujja kwesondamu ensimbi ezinamuyamba okukuba ebipande ebipande nokukola emirimu emirara , Rose Namuli akolera ku katale ka Pepsi oluvanyuma namuwa 2 , 000 . –> []

Table 31: Example few-shot prompts and their respective model outputs for the LAFT adaptation method on masakhaNER. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Fitar da amsar arshe kawai ga tambayar lissafi. Leah nada 32 chaculet, yar uwarta kuma 42.gudanawa suka rage musu? –> 39 Fitar da amsar arshe kawai ga tambayar lissafi. Agwagin Janet suna yin wai 16 a kullun. Tana yin karin kumallo da guda uku kowace safiya, sannan tana gasawa kawayenta guda hudu kullum. A kullum takan sayar da ragowar a kasuwar manoma akan dala 2 akan kowane wai. Dala nawa take samu a kullum a kasuwar manoma? –> **<NAN>**
kin	Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. <unk> Leah afite shokola <unk> naho umuvandimwe we afite <unk>. Nibarya <unk> bazaba basigaranye shokola zingahe zose hamwe? –> <unk>Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. <unk> Igishuhe cya Jane gitera amajyi <unk> ku munsi, buri mugitondo aryamo atatu kandi akora umugati winshutiye akoresheje ane, agurisha asigaye mwisoko ryabahinzi buri munsi kugichiro cya <unk> kuri buri jyi. Na ngahe mumadolali yinjiza ku munsi mwisoko ryabahinzi ? –> **<NAN>**
lug	Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. <unk> Leah yalina kyokuleeti <unk> ate nga muganda we ye yalina <unk>. Bwe baba nga baalyako <unk>, baasigazaawo kyokuleeti mmeka bombi omugatte? –<unk> <unk>Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. <unk> Embaata za Janet zibiika amagi <unk> buli lunaku. Alya amagi asatu buli lunaku ku kyenkya n'afumbisa amalala ana g'ateeka mu bukkeeki bwa muffin bw'akolera mikwano gye. Agasigadde agatunda mu katale k'abalimi n'abalunzi buli lunaku nga buli ggi alitunda <unk>. Afuna ssente mmeka buli lunaku mu katale k'abalimi n'abalunzi? –<unk> **<NAN>**

Table 32: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on AfriMGSM. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Zai zain amsa daidai: A, B, C, ko D. Wani masanin kimiyya ya auna dayamita na gashin mutum hudu. Dayamitocin, a ma'anin milimita, sune 0.091, 0.169, 0.17, da 0.023. Wanne in'ikwaliti ne ya kwatanta biyu daga dayamitocin biyu na gashin an adam? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.17 < 0.091 –> A Zai zain amsa daidai: A, B, C, ko D. Menene matsayin p a cikin 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **A**
kin	Tora igisubizo gikwiye: A, B, C, cyangwa D. <unk> Umuhanga yapimye diameter yimisatsi ine yabantu. Diameter, muri milimetero, yari <unk>.<unk>, <unk>.<unk>, <unk>.<unk>, na <unk>.<unk>. Ni ubuhe busumbane bugereranya neza diameter yimisatsi ibiri muriyo misatsi yabantu? <unk> A: <unk>.<unk> > <unk>.<unk> <unk> B: <unk>.<unk> < <unk>.<unk> <unk> C: <unk>.<unk> > <unk>.<unk> <unk> D: <unk>.<unk> > <unk>.<unk> –> A<unk>Tora igisubizo gikwiye: A, B, C, cyangwa D. <unk> Nakahe gaciro ka p muri <unk> = <unk>p? <unk> A: p = <unk> <unk> B: p = <unk> <unk> C: p = <unk> <unk> D: p = <unk> –> **<NAN>**
lug	Londa ekyokuddamu ekituufu: A, B, C, oba D. <unk> Munnassaayansi yapima obugazi bw'enviiri z'omuntu nnya. Obugazi mu butundutundu buli <unk>.<unk>, <unk>.<unk>, <unk>.<unk>, ne <unk>.<unk>. Bukwatane ki wakati w'emiwendo egitenkanankana egisobola okukozesebwa mu butuufu okugeraageranya obugazi bw'enviiri z'omuntu bbiri? <unk> A: <unk>.<unk> <unk> <unk>.<unk> <unk> B: <unk>.<unk> <unk> <unk>.<unk> <unk> C: <unk>.<unk> <unk> <unk>.<unk> <unk> D: <unk>.<unk> <unk> <unk>.<unk> –<unk> A<unk>Londa ekyokuddamu ekituufu: A, B, C, oba D. <unk> p erina muwendo ki mu <unk> <unk> <unk>p? <unk> A: p <unk> <unk> <unk> B: p <unk> <unk> <unk> C: p <unk> <unk> <unk> D: p <unk> <unk> –<unk> **<NAN>**

Table 33: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on AfriMMLU. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. A Tsakanin 1936 da 1940 Greece na karkashin mulkin kama karya na loannus Metaxas, Ana iya tunawa da sutin (a'a) da ya amsa dashi zuwa ga Mussolini ultimatum yayi mubaya'a a 1940. Tattalin arzikin Greece bai yi kyau ba a arashin mulkin kama karya na soja na Metaxas. –> 1 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. Waannan rikirkitattun na'urar kwayoyin halitta sun samo asali ne saboda zain su a haka zai iya canza yanayi su gaba aya dan haka su kwayoyin halitta suna taruwa lokacin da yanayin su na gaba aya ya haaka kuma ya canza da yanayi da suke. Duk na'urorin kwayoyin halitta suna da wahalar sha'ani. –> 2 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. masu son karatu musamman wanda suka manne a ajin karatun sanin tattalin arziki da na na'ura mai kwakwalwa basu da wani alfanu nan gaba. masu san karatu basu da wani alfanu. –> 0 ayyade idan hasashen ya bi jigo a hankali. Fitowa 0 don entailment, 1 don tsaka tsaki, ko 2 don sabani. Gaskiya bana ba tunanin sa amma na fusata sosai, kuma dai daga karshe na ige da ara yi masa magana. Ban ara masa magana ba. –> **<NAN>**
kin	Menya niba hypothesis ikurikiza ishingiro. Ibisohoka <unk> kubisobanuro, <unk> kubutabogamye, cyangwa <unk> kubivuguruza. <unk> Hagati ya <unk> na <unk> Ubugereki bwari ku butegetsi bw'igitugu bwa gisirikare bwa Ioannis Metaxas, bwibukwa kubera echi yumvikana (oya) yatanze asubiza ultimatum ya Mussolini yokwiyegurira mu <unk>. Ubukungu bw'Ubugereki ntabwo bwaribumeze neza kubutegetsi bwigitugu bwa gisirikare bwa Metaxas –> <unk>Menya niba hypothesis ikurikiza ishingiro. Ibisohoka <unk> kubisobanuro, <unk> kubutabogamye, cyangwa <unk> kubivuguruza. <unk> Izi nzego zo murwego rwohejuru rwibikoresho bya molekile bivuka kubera ko gutoranya bisanzwe gushobora gukora kumitungo rusange yibintu bya molekile iyo iyo mitungo rusange yongerewe imbaraga zo guhuza n'imihindagurikire y'ikirere.<unk> Ibikoresho byose bya molekile biba bgoranye –> <unk>Menya niba hypothesis ikurikiza ishingiro. Ibisohoka <unk> kubisobanuro, <unk> kubutabogamye, cyangwa <unk> kubivuguruza. <unk> Aba hanga bahatamye cyane mubyubukungu n'imyijyire ya kopyuta ,ninabo bafite ukwizera gucye Aba hanga bakompyuta ntakizere bafite –> <unk>Menya niba hypothesis ikurikiza ishingiro. Ibisohoka <unk> kubisobanuro, <unk> kubutabogamye, cyangwa <unk> kubivuguruza. <unk> Urebye, ntabwo nigeze ntekereza kuribyo, ariko narumiwe cyane, ndangije nongeye kumuvugisha tena Ntabwo narinongera kumuvugisha –> **<NAN>**
lug	Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa <unk> ku entailment, <unk> ku neutral, oba <unk> ku contradiction. <unk> Wakati wa <unk> ne <unk> Greece yali wansi w'obufuzi obwa nnakyemalira Ioannis Metaxas, ajjukirwa ennyo olw'enziramu ya 'Nedda' gye yayanukula Mussolini bwe yali amuwadde nsalessale w'okuwanika nga awanguddwa mu <unk>. Ebyenfuna bya Greece tebyatambula bulungi mu kiseera ng'eri wansi wa nnaakyemalira Metaxas. –<unk> <unk>Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa <unk> ku entailment, <unk> ku neutral, oba <unk> ku contradiction. <unk> Obusimu obw'eddaala erya waggulu busituka kubanga obutonde bubeera n'obusobozi okukikola bwe bumanyiira okukikola. Obusimu bwonna bwa ddaala lya waggulu. –<unk> <unk>Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa <unk> ku entailment, <unk> ku neutral, oba <unk> ku contradiction. <unk> Abantu abatalina bukugu mu kubeerana na balala abanoonya obubudamu mu bibiina by'amasomo g'ebyenfuna oba ebya kompyuta, ate bo basingawo mu kubeera mu mbeera y'obutaba na ssuubi. Abantu abatalina bukugu mu kubeerana na balala tebalina ssuubi. –<unk> <unk>Salawo oba endowooza (hypothesis) egoberera mu ngeri entegeerekeka (logically) ensonga (premise). Ekifulumizibwa <unk> ku entailment, <unk> ku neutral, oba <unk> ku contradiction. <unk> Kale nno, ekyo si na kye nnabadde ndowoozaako, naye olw'okuba nnabadde mu mbeera ey'okusoberwa, nnawunzise nzizeemu okwogera naye. Sinnaddamu kwogerako naye. –<unk> **<NAN>**

Table 34: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on AfriXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	အယူအဆသည် အရင်းအနှီးနှင့်ကိုက်ညီမှုရှိမရှိ ဆုံးဖြတ်ပါ။ ဆက်စပ်မှုအတွက် 0၊ ကြားနေအတွက် 1 သို့မဟုတ် ဆန့်ကျင်ဘက်အတွက် 2 ထုတ်ပေးသည်။ McKim သည် သူ၏ စိတ်ပျက်စိတ်ရှုပ်ခြင်းများစွာကြောင့် ရှုံးနိမ့်ရုံသာမက Howard & ကဧော့ထ်ဝဲလ် ၏ နောက် တတိယနေရာတွင် ရပ်တည်ခဲ့သည်။ McKim သည် သူ အရင် ပြီးဆုံးခဲ့ သဘောကြောင့် အားရကျေနပ်ဖြစ်ခဲ့သည်။ –> 2 အယူအဆသည် အရင်းအနှီးနှင့်ကိုက်ညီမှုရှိမရှိ ဆုံးဖြတ်ပါ။ ဆက်စပ်မှုအတွက် 0 ကြားနေအတွက် 1 သို့မဟုတ် ဆန့်ကျင်ဘက်အတွက် 2 ထုတ်ပေးသည်။ အခြားသူများမှာ ဘာသာစကား အသုံးပြုခြင်းကို ရိုးရိုးရှင်းရှင်း အံ့သြသွားကာ ကျွန်ုပ်တို့၏ ခွဲခြမ်းစိတ်ဖြာမှုဘက်မှ အဆုံးသတ်ပြီး ကျွန်ုပ်တို့၏ စိတ်ခံစားမှုဆိုင်ရာ ဘက်ခြမ်းက စတင်သည်ကို အံ့သြသွားမည်။ စိတ်ပိုင်းဆိုင်ရာ အယူခံဝင်မှုများ စတင်သည့် နေရာ ကို အတိအကျ ဆုံးဖြတ်ရန် ခက်ခဲနိုင်သည် ။ –> 0 အယူအဆသည် အရင်းအနှီးနှင့်ကိုက်ညီမှုရှိမရှိ ဆုံးဖြတ်ပါ။ ဆက်စပ်မှုအတွက် 0၊ ကြားနေအတွက် 1 သို့မဟုတ် ဆန့်ကျင်ဘက်အတွက် 2 ထုတ်ပေးသည်။ ပြီးတော့ ငါတွေးမိတဲ့ တကယ်စိတ်ဝင်စားဖို့ကောင်းတာက ဘာလဲဆိုတော့ ဒါနဲ့ပတ်သက်ပြီး ငါတို့ ဘာလုပ်ရမလဲ ငါဆိုလိုတာ ငါတို့ ကို ကိုယ်စားပြုမဲ့ လူတွေကို ငါတို့ ပြောင်းလဲပေးရလိမ့်မယ် အဲဒါက အရမ်းပျင်းဖို့ကောင်းတယ် ပြီးတော့ ကျွန်တော်တို့ကို ကိုယ်စားပြုတဲ့ သူတွေကို ပြောင်းလဲဖို့ မထိုက်တန်ဘူးဆိုတာ သိနတော့ ပြောင်းလဲဖို့တောင် မကိုးစားသင့်ပါဘူး။ –> 2 အယူအဆသည် အရင်းအနှီးနှင့်ကိုက်ညီမှုရှိမရှိ ဆုံးဖြတ်ပါ။ ဆက်စပ်မှုအတွက် 0 ကြားနေအတွက် 1 သို့မဟုတ် ဆန့်ကျင်ဘက်အတွက် 2 ထုတ်ပေးသည်။ ငါက ဒါတွေကိုတောင် စဉ်းစားနေခဲ့တာမဟုတ်ပမေယ့် ငါတော်တော်စိတ်ညစ်နေခဲ့ပြီး ငါသူနဲ့စကားပြန်ပြောဖြစ်ခဲ့တယ်။ ငါ သူ့ကို စကား ထပ်မပြောဖြစ်ဘူး။ –> **<NAN>**

Table 35: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on MyanmarXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	သတင်းအနှစ်ချုပ်အတွက် ခေါင်းစဉ်တစ်ခုပေးပါ။ မြန်မာအစိုးရစစ်တပ်နဲ့ ဒီကဘောီအတေပ်တွေ ဇူလိုင်လ ၇ရက်နေ့က ကဇော့ကရိတ်မှိုအြထွက် လမ်းဟောင်းတဝိုက်မှာ မနက်ကတည်းက အနီးကပ်တိုက်ပွဲတွေ ဖြစ်ပွားခဲ့ကြပါတယ်။ –> ကဇော့ကရိတ်စာသင်ကျောင်းအတွင်း မော်တာကျည်ဆံကျ သတင်းအနှစ်ချုပ်အတွက် ခေါင်းစဉ်တစ်ခုပေးပါ။ ၂၀၂၀ တိုကျို အိုလံပစ် အတွက် ပြိုင်ပွဲကျင်းပမယ့် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန် အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –> □□၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀ တိုကျို အိုလံပစ် အားကစားကွင်း တည်ဆောက်မယ့် အစီအစဉ်တွေကို ဂျပန်အစိုးရက ပယ်ဖျက်လိုက်ပါတယ်။ –>၂၀၂၀
hau	Samar da kanun labarai don taaitawar labarai. Sarki Abdullah na Saudi Arabia, ya yi suka kan abin da ya kira, fakewar da 'yan ta'adda ke yi da addini suna tafka ta'asa. –> Sarki Abdullah: 'Yan ta'adda na fakewa da addini Samar da kanun labarai don taaitawar labarai. Ta dai tabbata cewa maharin da ya tarwatsa kansa a gidan raye-rayen Manchester, Salman Abedi ya koma Burtaniya ne daga etare, kwanaki alilan kafin ya kai wannan farmaki. –> **etare: Sallar da gidan raye-raye tare da Sallar Kamaru da kuma gidan raye-rayen Manchester, wannan ita ce alamar da Sallar Kamaru da kuma gidan raye-raye tare da Sallar Kamaru. –>Salah Omar: etare na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addini. –> Sarkin Musulmi na da matukar muhimmanci ga addin**

Table 36: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on XL-Sum. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	အကြောင်းအရာကို ပေး၍ ဇာတ်လမ်းအတွက် အကောင်းဆုံးအဆုံးသတ်ကို ရွေးပါ□ 1 သို့မဟုတ် 2။ ဒီတနင်္ဂနွေတွင် အန်ဘာမှာ လုပ်စရာများစွာ ရှိသည်။ သူမသည် သွားစရာရှိသော နေရာများကို စာရင်းပြုလုပ်ခဲ့သည်။ သူမသည် အဆင်သင့်ဖြစ်ရန် အလောတကြီးလုပ်ခဲ့သည်။ သူမသည် အချိန်မလောက်မှာကို စိတ်ပူခဲ့သည်။ 1: အန်ဘာသည် စိတ်အပန်းပြေသော နှစ်နာရီ နံနက်စာနှင့်နေ့လယ်စာ ပေါင်းစားရခြင်းကို နှစ်ခြိုက်ခဲ့သည်။ 2: အန်ဘာသည် စာရင်းကို အိမ်တွင်ထားခဲ့မိ၍ အလွန်အမင်း အလောတကြီးလုပ်ခဲ့ရသည်။ –> 2 အကြောင်းအရာကို ပေး၍ ဇာတ်လမ်းအတွက် အကောင်းဆုံးအဆုံးသတ်ကို ရွေးပါ□ 1 သို့မဟုတ် 2။ ကျွန်တော်သည် ၂၀၁၁ ခုနှစ်တွင် Law and Order ၏ပရိသတ်တစ်ယောက် ဖြစ်လာခဲ့သည်။ ကျွန်တော်သည် ဦးနှောက်သွေးကြောပိတ်ခြင်းမှ ပြန်လည်သက်သာလာခဲ့သည်။ ကျွန်တော် အိမ်ပြန်ရောက်သောအခါ အပိုင်းတိုင်းကိုကြည့်ရန် ကြိုးစားခဲ့သည်။ နှစ်ပေါင်း ၂၀ ကျော်ရှိနေသည့် ရှိုးတစ်ခုကို အမီလိုက်ကြည့်ရန်မှာ ခက်ခဲခဲ့သည်။ 1: ကျွန်တော်သည် Law and Order ရှိုးကို အဆိုးရွားဆုံးရှိုးတစ်ခုဟု ထင်ခဲ့သည်။ 2: နောက်ဆုံးတွင် ကျွန်တော်သည် အပိုင်းအားလုံးကို ကြည့်ခဲ့ပါသည်။ –> <NAN>

Table 37: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on XStoryCloze. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Sai dai mai sharhi akan harkokin siyasa na kasar Delphin Kapaya ya ce yadda kotun ta gudanar da wannan sharaar shine zai nuna irin mataki nagaba da magoya bayan Beman zasu dauka . –> (Sai,O), (dai,O), (mai,O), (sharhi,O), (akan,O), (harkokin,O), (siyasa,O), (na,O), (kasar,O), (Delphin,B-PER), (Kapaya,I-PER), (ya,O), (ce,O), (yadda,O), (kotun,O), (ta,O), (gudanar,O), (da,O), (wannan,O), (sharaar,O), (shine,O), (zai,O), (nuna,O), (irin,O), (mataki,O), (nagaba,O), (da,O), (magoya,O), (bayan,O), (Beman,B-PER), (zasu,O), (dauka,O), (.,O) Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Ya kuma yaba da shawarar da bangaren al - Barnawi na Boko Haram ya yanke na sassautawa a gwagwarmayarsu . –> []
kin	Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. <unk> Amazon iteganya gushora miliyari <unk> , <unk> zamadorali mu bikorwa byo gucururiza kuri internet , ndetse ngo ikaba izoroherwa no gukoresha internet ya Bharti ku giciro gito mu gihe ibigo byombi bizaba byemeranyije amasezerano byifuza . –> (Amazon,B-ORG), (iteganya,O), (gushora,O), (miliyari,O), (<unk>,O), (,,O), (<unk>,O), (zamadorali,O), (mu,O), (bikorwa,O), (byo,O), (gucururiza,O), (kuri,O), (internet,O), (,,O), (ndetse,O), (ngo,O), (ikaba,O), (izoroherwa,O), (no,O), (gukoresha,O), (internet,O), (ya,O), (Bharti,B-ORG), (ku,O), (giciro,O), (gito,O), (mu,O), (gihe,O), (ibigo,O), (byombi,O), (bizaba,O), (byemeranyije,O), (amasezerano,O), (byifuza,O), (.,O)<unk>Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. <unk> Bazwi mu cyo bise Morning Worship aho baririmba ibihangano bitandukanye byo mu gitabo bigafasha benshi . –> []
lug	Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. <unk> Abantu abaatuwa obuyambi bampa sikaala okugenda mu Amerika okusoma diguli eyookubiri olwo bizinensi yenkoko ne ngiwa mukwano gwange Geoffrey Lwanga nga kati mu kiseera kino alina enkoko ezisoba mu <unk> ezamagi . –<unk> (Abantu,O), (abaatuwa,O), (obuyambi,O), (bampa,O), (sikaala,O), (okugenda,O), (mu,O), (Amerika,B-LOC), (okusoma,O), (diguli,O), (eyookubiri,O), (olwo,O), (bizinensi,O), (yenkoko,O), (ne,O), (ngiwa,O), (mukwano,O), (gwange,O), (Geoffrey,B-PER), (Lwanga,I-PER), (nga,O), (kati,O), (mu,O), (kiseera,B-DATE), (kino,I-DATE), (alina,O), (enkoko,O), (ezisoba,O), (mu,O), (<unk>,O), (ezamagi,O), (.,O)<unk>Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. <unk> Ono ye waffe era kampeyini ze okuziyimirizaawo tujja kwesondamu ensimbi ezinamuyamba okukuba ebipande ebipande nokukola emirimu emirara , Rose Namuli akolera ku katale ka Pepsi oluvanyuma namuwa <unk> , <unk> . –<unk> []

Table 38: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on masakhaNER. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
bur	အောက်ပါဝါကျတွင် စကားလုံးတစ်လုံးစီကို ၎င်း၏ NER တက်ဂ်ဖြင့် အမှတ်အသားပြုပါ။ ချစ်ထွန်း၊ ဦး (ဝေါ်လတာ) –> (ချစ်ထွန်း၊,B-PER), (ဦး,I-PER), ((,I-PER), (ဝေါ်လတာ,I-PER), (),I-PER) အောက်ပါဝါကျတွင် စကားလုံးတစ်လုံးစီကို ၎င်း၏ NER တက်ဂ်ဖြင့် အမှတ်အသားပြုပါ။ အလုံမြို့နယ်၊ ရန်ကုန်တိုင်းဒေသကြီး၊တွင် တည်ရှိပြီး၊၁၉၆၃ ခုနှစ် တွင် မြန်မာ့သစ်လုပ်ငန်းမှ ဖွင့်လှစ်ထားခြင်းဖြစ်သည်။ ဖွင့်လှစ်သင်ကြားနေသော သင်တန်းများမှာ - –> **['(-,-)', '(-,-)']**

Table 39: Example few-shot prompts and their respective model outputs for the FOCUS adaptation method on wikiANN. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Fitar da amsar arshe kawai ga tambayar lissafi. Leah nada 32 chaculet, yar uwarta kuma 42.gudanawa suka rage musu? –> 39 A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Fitar da amsar arshe kawai ga tambayar lissafi. Agwagin Janet suna yin wai 16 a kullun. Tana yin karin kumallo da guda uku kowace safiya, sannan tana gasawa kawayenta guda hudu kullum. A kullum takan sayar da ragowar a kasuwar manoma akan dala 2 akan kowane wai. Dala nawa take samu a kullum a kasuwar manoma? –> **42**
kin	Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. Leah afite shokola 32 naho umuvandimwe we afite 42. Nibarya 35 bazaba basigaranye shokola zingahe zose hamwe? –> 39 Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Ibisohoka gusa igisubizo cyanyuma kubibazo byimibare. Igishuhe cya Jane gitera amajyi 16 ku munsi, buri mugitondo aryamo atatu kandi akora umugati winshutiye akoresheje ane, agurisha asigaye mwisoko ryabahinzi buri munsi kugichiro cya 2 kuri buri jyi. Na ngahe mumadolali yinjiza ku munsi mwisoko ryabahinzi ? –> **22**
lug	Wansi waliwo ekiragiro ekinnyonnyola omulimu, nga kigatta n'okuyingiza ekiwa ensonga endala. Wandiika eky'okuddamu ekimaliriza okusaba mu ngeri esaanidde. Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. Leah yalina kyokuleeti 32 ate nga muganda we ye yalina 42. Bwe baba nga baalyako 35, baasigazaawo kyokuleeti mmeka bombi omugatte? –> 39 Wansi waliwo ekiragiro ekinnyonnyola omulimu, nga kigatta n'okuyingiza ekiwa ensonga endala. Wandiika eky'okuddamu ekimaliriza okusaba mu ngeri esaanidde. Fulumya ekyokuddamu ekisembayo kyokka ku kibuuzo kyokubala. Embaata za Janet zibiika amagi 16 buli lunaku. Alya amagi asatu buli lunaku ku kyenkya n'afumbisa amalala ana g'ateeka mu bukkeeki bwa muffin bw'akolera mikwano gye. Agasigadde agatunda mu katale k'abalimi n'abalunzi buli lunaku nga buli ggi alitunda $2. Afuna ssente mmeka buli lunaku mu katale k'abalimi n'abalunzi? –> **1**

Table 40: Example few-shot prompts and their respective model outputs for the LAIT adaptation method on AfriMGSM. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Zai zain amsa daidai: A, B, C, ko D. Wani masanin kimiyya ya auna dayamita na gashin mutum hudu. Dayamitocin, a ma'anin milimita, sune 0.091, 0.169, 0.17, da 0.023. Wanne in'ikwaliti ne ya kwatanta biyu daga dayamitocin biyu na gashin an adam? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.17 < 0.091 –> A A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Zai zain amsa daidai: A, B, C, ko D. Menene matsayin p a cikin 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **<NAN>**
kin	Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Tora igisubizo gikwiye: A, B, C, cyangwa D. Umuhanga yapimye diameter yimisatsi ine yabantu. Diameter, muri milimetero, yari 0.091, 0.169, 0.17, na 0.023. Ni ubuhe busumbane bugereranya neza diameter yimisatsi ibiri muriyo misatsi yabantu? A: 0.17 > 0.023 B: 0.091 < 0.023 C: 0.169 > 0.17 D: 0.169 > 0.17 –> A Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Tora igisubizo gikwiye: A, B, C, cyangwa D. Nakahe gaciro ka p muri 24 = 2p? A: p = 4 B: p = 8 C: p = 12 D: p = 24 –> **<NAN>**

Table 41: Example few-shot prompts and their respective model outputs for the LAIT adaptation method on AfriMMLU. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
kin	Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Gahunda nini ijyanye no kuvugurura igomba kurangira mu mpera za 2001 Gahunda yo kuvugurura ntabwo izakorwa neza mbere yuko umwaka wa 2000 urangira –> 0 Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Bari mururworwego, Ogle ararira Ogle yavuze ko bari hafi byumvikana –> 1 Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. byukuri ntakibazo byantera niyo baba bafite isosiyete iterwa inkunga Byambabaza cyane kumenya niba barateye inkunga isosiyete –> 2 Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Menya niba hypothesis ikurikiza ishingiro. Ibisohoka 0 kubisobanuro, 1 kubutabogamye, cyangwa 2 kubivuguruza. Urebye, ntabwo nigeze ntekereza kuribyo, ariko narumiwe cyane, ndangije nongeye kumuvugisha tena Ntabwo narinongera kumuvugisha –> **1**

Table 42: Example few-shot prompts and their respective model outputs for the LAIT adaptation method on AfriXNLI. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
tha	ด้านล่างนี้เป็นคำแนะนำที่อธิบายงาน โดยจับคู่กับอินพุตที่ให้บริบทเพิ่มเติม เขียนคำตอบที่ตอบสนองคำขอได้อย่างเหมาะสม กำหนดสาเหตุหรือผลของสถานที่ตั้ง เอาต์พุต 0 สำหรับตัวเลือกแรก หรือ 1 สำหรับตัวเลือกที่สอง คนร้ายปล่อยตัวประกัน ผลเป็นยังไงบ้างคะ? 0: พวกเขายอมรับค่าไถ่ 1: พวกเขาหนีออกจากคุก –> 0 ด้านล่างนี้เป็นคำแนะนำที่อธิบายงาน โดยจับคู่กับอินพุตที่ให้บริบทเพิ่มเติม เขียนคำตอบที่ตอบสนองคำขอได้อย่างเหมาะสม กำหนดสาเหตุหรือผลของสถานที่ตั้ง เอาต์พุต 0 สำหรับตัวเลือกแรก หรือ 1 สำหรับตัวเลือกที่สอง สิ่งของถูกห่อไว้ในพลาสติก ผลเป็นยังไงบ้างคะ? 0: มันบอบบาง 1: มันเล็ก –> **1**

Table 43: Example few-shot prompts and their respective model outputs for the LAIT adaptation method on XCOPA. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + **Output**
hau	A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Samar da kanun labarai don taaitawar labarai. Sarki Abdullah na Saudi Arabia, ya yi suka kan abin da ya kira, fakewar da 'yan ta'adda ke yi da addini suna tafka ta'asa. –> Sarki Abdullah: 'Yan ta'adda na fakewa da addini A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Samar da kanun labarai don taaitawar labarai. Ta dai tabbata cewa maharin da ya tarwatsa kansa a gidan raye-rayen Manchester, Salman Abedi ya koma Burtaniya ne daga etare, kwanaki alilan kafin ya kai wannan farmaki. –> **aya daga cikin dabarun talla, kamar yadda suke yin tafiya a cikin dabarun talla.<unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk>unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk><unk> <unk><unk><unk><unk><unk><unk>**

Table 44: Example few-shot prompts and their respective model outputs for the LAIT adaptation method on XL-Sum. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.

Lang.	Example Input + Output
hau	A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Sai dai mai sharhi akan harkokin siyasa na kasar Delphin Kapaya ya ce yadda kotun ta gudanar da wannan sharaar shine zai nuna irin mataki nagaba da magoya bayan Beman zasu dauka . –> (Sai,O), (dai,O), (mai,O), (sharhi,O), (akan,O), (harkokin,O), (siyasa,O), (na,O), (kasar,O), (Delphin,B-PER), (Kapaya,I-PER), (ya,O), (ce,O), (yadda,O), (kotun,O), (ta,O), (gudanar,O), (da,O), (wannan,O), (sharaar,O), (shine,O), (zai,O), (nuna,O), (irin,O), (mataki,O), (nagaba,O), (da,O), (magoya,O), (bayan,O), (Beman,B-PER), (zasu,O), (dauka,O), (.,O) A asa akwai umarni da ke bayyana awainiya, hae tare da shigarwar da ke ba da arin mahallin. Rubuta martani wanda ya cika bukatar da ya dace. Sanya kowace kalma a cikin jumla mai zuwa tare da alamar NER. Ya kuma yaba da shawarar da bangaren al - Barnawi na Boko Haram ya yanke na sassautawa a gwagwarmayarsu . –> []
kin	Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. Amazon iteganya gushora miliyari 6 , 5 zamadorali mu bikorwa byo gucururiza kuri internet , ndetse ngo ikaba izoroherwa no gukoresha internet ya Bharti ku giciro gito mu gihe ibigo byombi bizaba byemeranyije amasezerano byifuza . –> (Amazon,B-ORG), (iteganya,O), (gushora,O), (miliyari,O), (6,O), (,,O), (5,O), (zamadorali,O), (mu,O), (bikorwa,O), (byo,O), (gucururiza,O), (kuri,O), (internet,O), (,,O), (ndetse,O), (ngo,O), (ikaba,O), (izoroherwa,O), (no,O), (gukoresha,O), (internet,O), (ya,O), (Bharti,B-ORG), (ku,O), (giciro,O), (gito,O), (mu,O), (gihe,O), (ibigo,O), (byombi,O), (bizaba,O), (byemeranyije,O), (amasezerano,O), (byifuza,O), (.,O) Hasi ni amabwiriza asobanura umurimo, uhujwe n'igitekerezo gitanga ibindi bisobanuro. Andika igisubizo cyuzuza neza icyifuzo. Shyira buri jambo mu nteruro ikurikira hamwe na tagi ya NER. Bazwi mu cyo bise Morning Worship aho baririmba ibihangano bitandukanye byo mu gitabo bigafasha benshi . –> []
lug	Wansi waliwo ekiragiro ekinnyonnyola omulimu, nga kigatta n'okuyingiza ekiwa ensonga endala. Wandiika eky'okuddamu ekimaliriza okusaba mu ngeri esaanidde. Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. Abantu abaatuwa obuyambi bampa sikaala okugenda mu Amerika okusoma diguli eyookubiri olwo bizinensi yenkoko ne ngiwa mukwano gwange Geoffrey Lwanga nga kati mu kiseera kino alina enkoko ezisoba mu 7000 ezamagi . –> (Abantu,O), (abaatuwa,O), (obuyambi,O), (bampa,O), (sikaala,O), (okugenda,O), (mu,O), (Amerika,B-LOC), (okusoma,O), (diguli,O), (eyookubiri,O), (olwo,O), (bizinensi,O), (yenkoko,O), (ne,O), (ngiwa,O), (mukwano,O), (gwange,O), (Geoffrey,B-PER), (Lwanga,I-PER), (nga,O), (kati,O), (mu,O), (kiseera,B-DATE), (kino,I-DATE), (alina,O), (enkoko,O), (ezisoba,O), (mu,O), (7000,O), (ezamagi,O), (.,O) Wansi waliwo ekiragiro ekinnyonnyola omulimu, nga kigatta n'okuyingiza ekiwa ensonga endala. Wandiika eky'okuddamu ekimaliriza okusaba mu ngeri esaanidde. Buli kigambo mu sentensi eno wammanga giteekeko akabonero kaakyo aka NER. Ono ye waffe era kampeyini ze okuziyimirizaawo tujja kwesondamu ensimbi ezinamuyamba okukuba ebipande ebipande nokukola emirimu emirara , Rose Namuli akolera ku katale ka Pepsi oluvanyuma namuwa 2 , 000 . –> []

Table 45: Example few-shot prompts and their respective model outputs for the LAIT adaptation method on masakhaNER. We use the same prompts for all models, but the reported outputs here are from one of the random seeds in the LLaMa2-7B experiments.