
***Best-of-L*: Cross-Lingual Reward Modeling for Mathematical Reasoning**

Sara Rajaei^{*1}, Rochelle Choenni², Ekaterina Shutova², Christof Monz¹

Language Technology Lab, University of Amsterdam¹

ILLC, University of Amsterdam²

Abstract

While the reasoning abilities of large language models (LLMs) continue to advance, it remains unclear how such ability varies across languages in multilingual LLMs and whether different languages produce reasoning paths that complement each other. To investigate this question, we train a reward model to rank generated responses for a given question across languages. Our results show that our cross-lingual reward model substantially improves mathematical reasoning performance compared to using reward modeling within a single language, benefiting even high-resource languages. While English often exhibits the highest performance in multilingual models, we find that cross-lingual sampling particularly benefits English under low sampling budgets. Our findings reveal new opportunities to improve multilingual reasoning by leveraging the complementary strengths of diverse languages.

1 Introduction

Recently, many studies have focused on improving reasoning ability Ranaldi and Freitas [2024], Byun et al. [2024] or discovering major factors contributing to this skill Ko et al. [2024]. Yet, reasoning research has largely centered on English models, with multilingual models receiving comparatively little attention. Among the few, Shi et al. [2023] have shown that multilingual large language models (LLMs) have strong reasoning capabilities, even for underrepresented languages. Recent work has further improved the multilingual math reasoning ability of LLMs through self-consistency Lai et al. [2025], multilingual instruction-tuning, Chen et al. [2024], Lai and Nissim [2024], and preference optimization methods She et al. [2024], Dang et al. [2024a], Yang et al. [2025]. Following the proposed studies in using reward modeling to improve the performance of math reasoning in English LLMs Cobbe et al. [2021], Shen et al. [2021], Hosseini et al. [2024], Zhang et al. [2024], Setlur et al. [2025], Hong et al. [2025] have studied the transferability of English reward models to other languages. While Wang et al. [2025] are the first to shift attention to multilingual models, they still focus on generating and scoring in-language candidate solutions, which they refer to as multilingual reward modeling. In this paper, we instead explore the potential to generate and combine solutions in multiple different languages (see Figure 1), thereby fully exploiting the multilingual capabilities of LLMs.

To this end, we first study to what extent languages could potentially complement each other’s mathematical reasoning skills. Interestingly, we find that even low-resource languages sometimes succeed where high-resource languages fail, suggesting that their reasoning signals could provide valuable complementary information (Figure 4).

Motivated by the above finding, we develop a cross-lingual outcome reward modeling (ORM) framework to harness the *Best-of-Languages* performance, for which we train a verifier to score

^{*}Corresponding author: s.rajaee@uva.nl

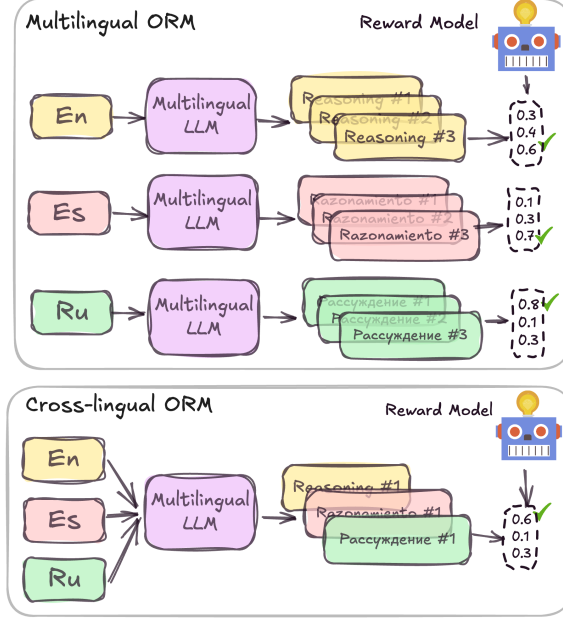


Figure 1: An illustration of multilingual outcome reward modeling (ORM) on the top, where the verifier ranks every language’s responses, and cross-lingual ORM (our framework), where the verifier ranks responses across languages for a given question.

multilingual reasoning across languages. To the best of our knowledge, we are the first to propose a cross-lingual reward model that takes advantage of complementary reasoning skills across languages. Our experimental results demonstrate that our framework improves performance by over 10% and 15% compared to the average performance of the naive multilingual RMs and the self-consistency baseline, respectively. Our analysis shows that increasing the number of languages improves the performance of our cross-lingual ORM.

Through an ablation study, we find that cross-language sampling even benefits English, especially under low-budget settings. Our analysis verifies that, while having English in the language pool of the cross-lingual ORM positively affects the performance, some selection of non-English pools outperforms other pools containing English, supporting our argument that languages have complementary reasoning skills in multilingual models.

2 Methodology

A popular approach in math reasoning tasks is utilizing reward models (also known as verifiers) to evaluate the correctness of a given answer. Based on the evaluation setup, reward models can be process-based, where the model assesses the reasoning step by step (called PRMs) Lightman et al. [2024], Luo et al. [2024], while outcome reward models (ORMs) evaluate the entire reasoning Cobbe et al. [2021], Shen et al. [2021], Hosseini et al. [2024], Zhang et al. [2024], Setlur et al. [2025]. In this work, we focus on the latter and propose a novel cross-lingual outcome reward modeling framework that leverages complementary reasoning signals across languages.

2.1 Cross-lingual Reward Modeling

Our framework is the cross-lingual version of the *Best-of-N* Lightman et al. [2024], ranking the generated answers for a given question across a set of languages and selecting the highest-scored one.

Given a math question q and a generated candidate answer a , we train a discriminative verifier to predict whether the generated reasoning is correct. More specifically, we train an LLM as the verifier using binary cross-entropy loss: $\mathcal{L}_{ORM} = -[y \cdot \log(\hat{y}) + (1 - y) \cdot \log(1 - \hat{y})]$. At inference, we use the verifier scores to rank a set of candidate answers in different languages for a given question

using the probability that the model put on the correct class, and then, we select the answer with the highest probability.

Training set generation. We use the GSM8K training set, including around 7.5k examples of high-quality grade school math problems created by human writers, Cobbe et al. [2021], to generate our verifier training set Lai and Nissim [2024]. We use Google Translate version of GSM8K in 8 languages—English(en), Spanish(es), French(fr), German(de), Russian(ru), Chinese(zh), Japanese(ja), and Thai(th)—. We then prompt 3 models—the instruction-tuned version of Aya-Expanse 8B Dang et al. [2024b], Llama3.1 8B Grattafiori et al. [2024], and Qwen2.5 7B Team [2024]—using the GSM8K training set in our 8 languages to generate responses with step-by-step reasoning. We automatically labeled the generated reasoning paths as correct or incorrect based on the correctness of the final answer. Using generations from multiple models allows us to increase the size and diversity of the training set. To make a balanced dataset, we use the same number of correct and incorrect samples for each language, resulting in a set of around 88k samples for training.

Cross-lingual-ORM We use the multilingual Qwen2.5-Instruct 3B model Team [2024] as our reward model (verifier) because it has the widest officially supported language coverage among recent multilingual models. We fine-tune the verifier using the aforementioned training set for 5 epochs, with AdamW, a learning rate of $2e-4$, and a batch size of 96. Since the main task is binary classification and to make fine-tuning efficient, we fine-tune with LoRA Hu et al. [2022] with a rank of 16 and scaling factor of 32. We use this as our cross-lingual outcome reward model for all experiments.

3 Experiments

3.1 Experimental setups

To study the chain-of-thought math reasoning ability of LLMs, we employ the MGSM (Multilingual Grade School Math) dataset Shi et al. [2023], covering 11 languages, including English(en), Spanish(es), French(fr), German(de), Russian(ru), Chinese(zh), Japanese(ja), and Thai(th) with 250 examples for each. We exclude Swahili, Telugu, and Bengali, as multilingual LLMs, including our verifier model, perform poorly on these languages. Nevertheless, our selection includes languages from diverse language families and writing scripts Lai et al. [2025]. Following the original recipe of using MGSM Shi et al. [2023], we prompt multilingual LLMs under the *Native-CoT* setting using 8-shots for all experiments².

Models. We have carried out our analysis and experiments using a wide range of instruction-tuned multilingual models, including Aya-Expanse 8B Dang et al. [2024b], Llama3.1 8B Grattafiori et al. [2024], Qwen2.5 7B Team [2024], Ministral 8B³, phi-3 7B Abdin et al. [2024], and Llama 3.2 3b⁴.

3.1.1 Baselines.

We evaluate our cross-lingual ORM against the following baselines:

Self-consistency. A simple, yet effective approach in chain-of-thought (CoT) prompting is self-consistency [Wang et al., 2023, Yao et al., 2023, Kojima et al., 2022]. This widely used baseline does the majority voting across a batch of sampled answers ($N = 8$) for each language.

Multilingual-ORM (Multi-ORM). Also known as *Best-of-N* technique, where the multilingual verifier scores N different samples within a language and selects the one with the highest score Wang et al. [2025]. We use $N = 8$, generated with a temperature sampling of $T = 0.7$, and truncated at the top-p ($p = 0.95$) for all experiments (including the self-consistency baseline).

3.2 Results and Findings

LLMs exhibit complementary mathematical reasoning skills across languages. To investigate the similarity of reasoning knowledge across languages, we employ pass@k, a well-established

²We have used the evaluation harness framework for our experiments and reported exact-match scores Gao et al. [2024].

³<https://huggingface.co/mistralai/Ministral-8B-Instruct-2410>

⁴<https://huggingface.co/meta-llama/Llama-3.2-3B-Instruct>

	En.	Avg.	SC	Multi-ORM	Cross-ORM	Pass@8-Multi	Pass@8-Cross
Aya-Expans-8b	79.6	63.4	58.3	73.3	83.2	82.4	93.2
Llama3.1-8b	80.4	64.0	71.6	76.2	84.0	86.9	92.4
Ministral-8b	82.0	65.1	70.3	76.4	87.6	84.3	93.4
Qwen2.5-7b	85.2	72.2	74.4	81.3	92.4	87.2	96.4
Phi3-7b	90.0	69.3	74.2	79.5	92.8	85.3	96.8
Llama3.2-3b	72.4	56.0	63.0	70.3	77.2	80.3	88.8

Table 1: Summary of the results across models and baselines. The leftmost columns represent the English performance and the average performance of all the languages. *SC* denotes average self-consistency accuracy. *Pass@8-cross* outperforms the average *pass@8-multi*, indicating the complementary math reasoning skills across languages. Our proposed framework, *Cross-ORM*, also exceeds the average *Multi-ORM* accuracy by a large margin.

metric used to approximate the upper-bound performance of LLMs when generating multiple answers Hosseini et al. [2024], Li et al. [2025]. This allows us to measure the degree of potential complementarity between languages in multilingual LLMs as it considers a question solved if at least one of the answers in different languages is correct. In Table 1, we report *pass@8* scores across languages (*pass@8-Cross*) and the average *pass@8* scores across different samples within languages (*pass@8-Multi*).⁵ Interestingly, we observe that *Pass@8-Cross* outperforms the performances of individual languages, suggesting that even high-resource languages can potentially benefit from other languages.

Sampling across languages is superior to sampling within a language. Building on our analysis, we employ the cross-lingual verifier described in Sec. 2.1 to see how languages benefit each other in practice. The middle part of Table 1 summarizes the accuracy of our cross-lingual ORM under within- and across-language settings. As shown, *Cross-ORM* clearly outperforms the average performance of *ORM-Multi*, with the largest benefits for non-English languages. These results suggest that leveraging cross-lingual signals is more effective than relying solely on monolingual reasoning, especially for underrepresented languages.

Increasing the pool of languages enhances the cross-lingual ORM performance. To understand the impact of language pool size, we show the average performance for all possible language combinations at different pool sizes in Figure 2. As expected, the results demonstrate that adding more languages improves cross-lingual ORM performance up to a certain point, after which the additional gains become negligible.

Sampling across languages benefits English as well. While our earlier analysis shows that cross-lingual ORM exceeds the average performance of multilingual ORM, its accuracy still lags behind that of English ORM. To better understand under what conditions other languages might benefit English, we compare the performance of English ORM (i.e., generating multiple answers in English) and cross-lingual ORM under different sampling budgets in Figure 3a. Based on the results, we observe that cross-lingual ORM outperforms English ORM at low sampling budgets. However, this advantage fades as the number of samples increases. We suspect that additional sampling from other languages becomes redundant once English samples already cover a wide range of reasoning paths.

Including English in language pools is generally helpful, yet it does not always lead to superior performance. To examine the effect of including English in the language pools for the cross-lingual ORM setup, we report the mean and standard deviation of accuracy across all possible language

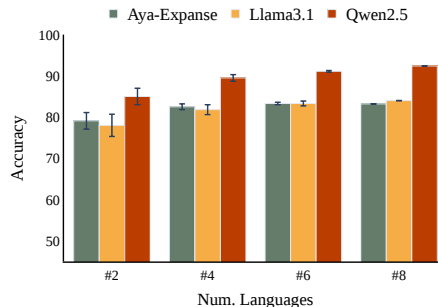


Figure 2: The mean and standard deviation cross-lingual ORM accuracy using different numbers of languages.

⁵The performance of individual languages can be found in the Appendix.

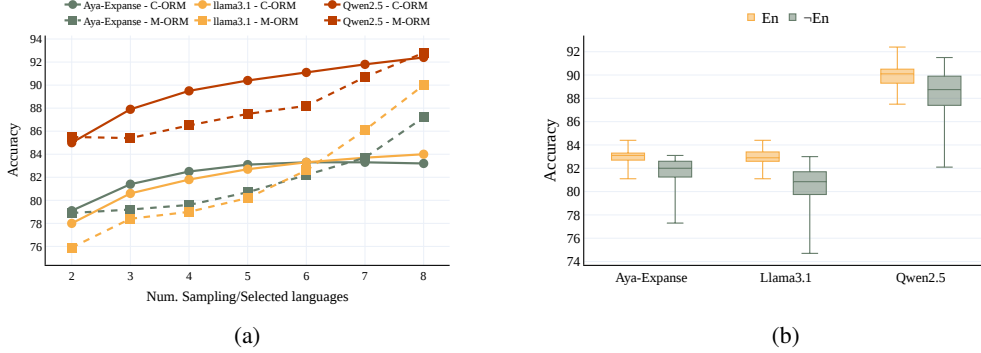


Figure 3: **(a)** Comparing the accuracy of cross-lingual ORM and multilingual ORM (M-ORM) of English using the same number of languages and samples. **(b)** Average cross-lingual ORM performance across language pools of size 2–7, with and without English. English generally helps, but some non-English sets outperform English-inclusive ones.

pools with a size of 2 to 7 with and without English in Figure 3b. As expected, including English generally improves cross-lingual ORM performance. However, this is not always the case; some language pools without English perform better than certain groups that include English, as reflected in the standard deviation of the non-English groups.

4 Conclusion

In this paper, we present a cross-lingual reward modeling framework that effectively leverages complementary mathematical reasoning skills across languages in multilingual LLMs. Our experiments show that cross-lingual reward modeling benefits even high-resource languages like English under low-budget inference settings. Furthermore, our findings show that languages mutually enhance each other’s reasoning abilities. Our results pave the way for future research into the similarities and differences of reasoning patterns across languages to improve multilingual reasoning models.

References

- Leonardo Ranaldi and Andre Freitas. Self-refine instruction-tuning for aligning reasoning in language models. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 2325–2347, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.139. URL <https://aclanthology.org/2024.emnlp-main.139/>.
- Ju-Seung Byun, Jiyun Chun, Jihyung Kil, and Andrew Perrault. ARES: Alternating reinforcement learning and supervised fine-tuning for enhanced multi-modal chain-of-thought reasoning through diverse AI feedback. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4410–4430, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.252. URL <https://aclanthology.org/2024.emnlp-main.252/>.
- Miyoung Ko, Sue Hyun Park, Joonsuk Park, and Minjoon Seo. Hierarchical deconstruction of LLM reasoning: A graph-based framework for analyzing knowledge utilization. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 4995–5027, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.288. URL <https://aclanthology.org/2024.emnlp-main.288/>.
- 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. Language models are multilingual chain-of-thought reasoners. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=fR3wGCK-IXp>.

- Huiyuan Lai, Xiao Zhang, and Malvina Nissim. Multidimensional consistency improves reasoning in language models, 2025. URL <https://arxiv.org/abs/2503.02670>.
- Nuo Chen, Zinan Zheng, Ning Wu, Ming Gong, Dongmei Zhang, and Jia Li. Breaking language barriers in multilingual mathematical reasoning: Insights and observations. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 7001–7016, Miami, Florida, USA, November 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.findings-emnlp.411. URL <https://aclanthology.org/2024.findings-emnlp.411/>.
- Huiyuan Lai and Malvina Nissim. mCoT: Multilingual instruction tuning for reasoning consistency in language models. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12012–12026, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.649. URL <https://aclanthology.org/2024.acl-long.649/>.
- Shuaijie She, Wei Zou, Shujian Huang, Wenhao Zhu, Xiang Liu, Xiang Geng, and Jiajun Chen. MAPO: Advancing multilingual reasoning through multilingual-alignment-as-preference optimization. In Lun-Wei Ku, Andre Martins, and Vivek Srikumar, editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 10015–10027, Bangkok, Thailand, August 2024. Association for Computational Linguistics. doi: 10.18653/v1/2024.acl-long.539. URL <https://aclanthology.org/2024.acl-long.539/>.
- John Dang, Arash Ahmadian, Kelly Marchisio, Julia Kreutzer, Ahmet Üstün, and Sara Hooker. RLHF can speak many languages: Unlocking multilingual preference optimization for LLMs. In Yaser Al-Onaizan, Mohit Bansal, and Yun-Nung Chen, editors, *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 13134–13156, Miami, Florida, USA, November 2024a. Association for Computational Linguistics. doi: 10.18653/v1/2024.emnlp-main.729. URL <https://aclanthology.org/2024.emnlp-main.729/>.
- Wen Yang, Junhong Wu, Chen Wang, Chengqing Zong, and Jiajun Zhang. Language imbalance driven rewarding for multilingual self-improving. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=Kak2ZH5Itp>.
- Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, Christopher Hesse, and John Schulman. Training verifiers to solve math word problems, 2021. URL <https://arxiv.org/abs/2110.14168>.
- Jianhao Shen, Yichun Yin, Lin Li, Lifeng Shang, Xin Jiang, Ming Zhang, and Qun Liu. Generate & rank: A multi-task framework for math word problems. In Marie-Francine Moens, Xuanjing Huang, Lucia Specia, and Scott Wen-tau Yih, editors, *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2269–2279, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics. doi: 10.18653/v1/2021.findings-emnlp.195. URL <https://aclanthology.org/2021.findings-emnlp.195/>.
- Arian Hosseini, Xingdi Yuan, Nikolay Malkin, Aaron Courville, Alessandro Sordoni, and Rishabh Agarwal. V-Star: Training verifiers for self-taught reasoners. In *First Conference on Language Modeling*, 2024. URL <https://openreview.net/forum?id=stmqBSW2dV>.
- Lunjun Zhang, Arian Hosseini, Hritik Bansal, Mehran Kazemi, Aviral Kumar, and Rishabh Agarwal. Generative verifiers: Reward modeling as next-token prediction. In *The 4th Workshop on Mathematical Reasoning and AI at NeurIPS’24*, 2024. URL <https://openreview.net/forum?id=CxHRoTLmPX>.
- Amrith Setlur, Chirag Nagpal, Adam Fisch, Xinyang Geng, Jacob Eisenstein, Rishabh Agarwal, Alekh Agarwal, Jonathan Berant, and Aviral Kumar. Rewarding progress: Scaling automated process verifiers for LLM reasoning. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=A6Y7AqlzLW>.

- Jiwoo Hong, Noah Lee, Rodrigo Martínez-Castaño, César Rodríguez, and James Thorne. Cross-lingual transfer of reward models in multilingual alignment. In Luis Chiruzzo, Alan Ritter, and Lu Wang, editors, *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers)*, pages 82–94, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-190-2. URL <https://aclanthology.org/2025.naacl-short.8/>.
- Weixuan Wang, Minghao Wu, Barry Haddow, and Alexandra Birch. Demystifying multilingual chain-of-thought in process reward modeling. *arXiv preprint arXiv:2502.12663*, 2025.
- Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let’s verify step by step. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=v8L0pN6EOi>.
- Liangchen Luo, Yinxiao Liu, Rosanne Liu, Samrat Phatale, Harsh Lara, Yunxuan Li, Lei Shu, Yun Zhu, Lei Meng, Jiao Sun, and Abhinav Rastogi. Improve mathematical reasoning in language models by automated process supervision. *arXiv preprint arXiv:2406.06592*, 2024.
- John Dang, Shivalika Singh, Daniel D’souza, Arash Ahmadian, Alejandro Salamanca, Madeline Smith, Aidan Peppin, Sungjin Hong, Manoj Govindassamy, Terrence Zhao, Sandra Kublik, Meor Amer, Viraat Aryabumi, Jon Ander Campos, Yi-Chern Tan, Tom Kocmi, Florian Strub, Nathan Grinsztajn, Yannis Flet-Berliac, Acyr Locatelli, Hangyu Lin, Dwarak Talupuru, Bharat Venkitesh, David Cairuz, Bowen Yang, Tim Chung, Wei-Yin Ko, Sylvie Shang Shi, Amir Shukayev, Sammie Bae, Aleksandra Piktus, Roman Castagné, Felipe Cruz-Salinas, Eddie Kim, Lucas Crawhall-Stein, Adrien Morisot, Sudip Roy, Phil Blunsom, Ivan Zhang, Aidan Gomez, Nick Frosst, Marzieh Fadaee, Beyza Ermis, Ahmet Üstün, and Sara Hooker. Aya expand: Combining research breakthroughs for a new multilingual frontier, 2024b. URL <https://arxiv.org/abs/2412.04261>.
- Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan,

Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vitor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenber, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkan Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippus Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabza, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta, Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. The

- llama 3 herd of models, 2024. URL <https://arxiv.org/abs/2407.21783>.
- Qwen Team. Qwen2.5: A party of foundation models, September 2024. URL <https://qwenlm.github.io/blog/qwen2.5/>.
- Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations*, 2022. URL <https://openreview.net/forum?id=nZeVKeeFYf9>.
- Leo Gao, Jonathan Tow, Baber Abbasi, Stella Biderman, Sid Black, Anthony DiPofi, Charles Foster, Laurence Golding, Jeffrey Hsu, Alain Le Noac’h, Haonan Li, Kyle McDonell, Niklas Muennighoff, Chris Ociepa, Jason Phang, Laria Reynolds, Hailey Schoelkopf, Aviya Skowron, Lintang Sutawika, Eric Tang, Anish Thite, Ben Wang, Kevin Wang, and Andy Zou. The language model evaluation harness, 07 2024. URL <https://zenodo.org/records/12608602>.
- Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, Weizhu Chen, Yen-Chun Chen, Yi-Ling Chen, Hao Cheng, Parul Chopra, Xiyang Dai, Matthew Dixon, Ronen Eldan, Victor Fragoso, Jianfeng Gao, Mei Gao, Min Gao, Amit Garg, Allie Del Giorno, Abhishek Goswami, Suriya Gunasekar, Emman Haider, Junheng Hao, Russell J. Hewett, Wenxiang Hu, Jamie Huynh, Dan Iter, Sam Ade Jacobs, Mojan Javaheripi, Xin Jin, Nikos Karampatziakis, Piero Kauffmann, Mahoud Khademi, Dongwoo Kim, Young Jin Kim, Lev Kurilenko, James R. Lee, Yin Tat Lee, Yanzhi Li, Yunsheng Li, Chen Liang, Lars Liden, Xihui Lin, Zeqi Lin, Ce Liu, Liyuan Liu, Mengchen Liu, Weishung Liu, Xiaodong Liu, Chong Luo, Piyush Madan, Ali Mahmoudzadeh, David Majercak, Matt Mazzola, Caio César Teodoro Mendes, Arindam Mitra, Hardik Modi, Anh Nguyen, Brandon Norick, Barun Patra, Daniel Perez-Becker, Thomas Portet, Reid Pryzant, Heyang Qin, Marko Radmilac, Liliang Ren, Gustavo de Rosa, Corby Rosset, Sambudha Roy, Olatunji Ruwase, Olli Saarikivi, Amin Saied, Adil Salim, Michael Santacrose, Shital Shah, Ning Shang, Hiteshi Sharma, Yelong Shen, Swadheen Shukla, Xia Song, Masahiro Tanaka, Andrea Tupini, Praneetha Vaddamanu, Chunyu Wang, Guanhua Wang, Lijuan Wang, Shuohang Wang, Xin Wang, Yu Wang, Rachel Ward, Wen Wen, Philipp Witte, Haiping Wu, Xiaoxia Wu, Michael Wyatt, Bin Xiao, Can Xu, Jiahang Xu, Weijian Xu, Jilong Xue, Sonali Yadav, Fan Yang, Jianwei Yang, Yifan Yang, Ziyi Yang, Donghan Yu, Lu Yuan, Chenruidong Zhang, Cyril Zhang, Jianwen Zhang, Li Lyna Zhang, Yi Zhang, Yue Zhang, Yunan Zhang, and Xiren Zhou. Phi-3 technical report: A highly capable language model locally on your phone, 2024. URL <https://arxiv.org/abs/2404.14219>.
- Xuezhi Wang, Jason Wei, Dale Schuurmans, Quoc V Le, Ed H. Chi, Sharan Narang, Aakanksha Chowdhery, and Denny Zhou. Self-consistency improves chain of thought reasoning in language models. In *The Eleventh International Conference on Learning Representations*, 2023. URL <https://openreview.net/forum?id=1PL1NIMMrw>.
- Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik R Narasimhan. Tree of thoughts: Deliberate problem solving with large language models. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL <https://openreview.net/forum?id=5Xc1ecx0lh>.
- Takeshi Kojima, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa. Large language models are zero-shot reasoners. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL <https://openreview.net/forum?id=e2TBb5y0yFf>.
- Zhong-Zhi Li, Duzhen Zhang, Ming-Liang Zhang, Jiaxin Zhang, Zengyan Liu, Yuxuan Yao, Haotian Xu, Junhao Zheng, Pei-Jie Wang, Xiuyi Chen, Yingying Zhang, Fei Yin, Jiahua Dong, Zhijiang Guo, Le Song, and Cheng-Lin Liu. From system 1 to system 2: A survey of reasoning large language models, 2025. URL <https://arxiv.org/abs/2502.17419>.

A Full Results

Question:

Melanie is a door-to-door saleswoman. She sold a third of her vacuum cleaners at the green house, 2 more to the red house, and half of what was left at the orange house. If Melanie has 5 vacuum cleaners left, how many did she start with?

English Response:

Melanie sold a third of her vacuum cleaners at the green house. That means she sold $1/3$ of her vacuum cleaners. $1/3$ of 5 is $5/3$, which is 1.67. Melanie sold 1.67 vacuum cleaners at the green house. She had 5 left, so she had $5 + 1.67 = 6.67$ vacuum cleaners before the green house. She sold 2 more at the red house, so she had $6.67 - 2 = 4.67$ vacuum cleaners before the red house. She sold half of what was left at the orange house. Half of 4.67 is 2.335. Melanie had $4.67 - 2.335 = 2.335$ vacuum cleaners left. Melanie had 5 vacuum cleaners left, so she must have had $5 + 2.335 = 7.335$ vacuum cleaners before the orange house. Melanie had 7.335 vacuum cleaners before the orange house. **The answer is 7.335.**

Japanese Response:

緑の家で3分の1の掃除機を売った場合、 $3/1 \times 3 = 9$ 個の掃除機を売ったことになります。赤の家で2台を売ったので、合計 $9+2=11$ 台を売ったことになります。オレンジの家で残りの半分を売った場合、残りの半分は $1/2 \times 5 = 2.5$ となり、オレンジの家で 2.5 台を売ったことになります。合計で、メラニーは $11+2.5=13.5$ 台の掃除機を売りました。掃除機は整数でなければならないため、彼女は 13 台の掃除機を売りました。彼女は 5 台の掃除機を手元に残したため、彼女は $13+5=18$ 台の掃除機を持っていました。答えは18です。

Google translation of Japanese Response:

If she sells one-third of the vacuums at the green house, she has sold $3/1 \times 3 = 9$ vacuums. She sells two at the red house, so she has sold a total of $9+2=11$ vacuums. If she sells the other half at the orange house, the remaining half is $1/2 \times 5 = 2.5$, so she has sold 2.5 at the orange house. In total, Melanie has sold $11+2.5=13.5$ vacuums. Because vacuums must be whole numbers, she has sold 13 vacuums. She has 5 vacuums left, so she has $13+5=18$ vacuums. **The answer is 18.**

Figure 4: A cherry-picked example of Llama 3.1’s responses to the same question in English and Japanese, where the English answer is incorrect but the Japanese response is correct, illustrating the complementary reasoning skills across languages.

	en	fr	es	de	ru	zh	ja	th	avg.
Aya-Expanse-8b									
CoT	76.6 ±1.1	65.1 ±2.0	73.0 ±1.8	68.8 ±1.9	67.6 ±2.5	63.7 ±0.8	57.7 ±1.2	19.7 ±1.8	61.5
SC	84.0	70.8	76.8	76.8	73.6	71.2	67.2	23.2	58.3
Llama3.1-8b									
CoT	75.0 ±1.7	60.4 ±1.4	66.8 ±1.9	59.6 ±3.8	61.1 ±2.8	57.5 ±2.3	46.9 ±2.6	47.4 ±1.1	59.4
SC	84.8	72.4	80.0	74.0	74.0	69.6	58.8	59.2	71.6
Ministral-8b									
CoT	77.4 ±1.5	65.1 ±0.7	71.4 ±1.3	64.9 ±1.8	65.4 ±0.8	58.8 ±1.7	43.8 ±1.9	44.5 ±1.1	61.4
SC	87.2	72.0	80.0	74.0	75.2	67.6	55.2	54.0	70.3
Qwen2.5-7b									
CoT	84.3 ±1.0	70.2 ±1.8	76.3 ±2.3	63.8 ±1.2	69.0 ±1.7	67.8 ±1.7	63.7 ±1.4	51.8 ±1.3	68.4
SC	89.2	73.6	82.8	70.8	73.6	75.6	70.0	59.6	74.4
Phi3-7b									
CoT	87.6 ±1.8	77.0 ±2.3	83.7 ±1.7	77.3 ±0.9	71.6 ±1.8	69.2 ±2.3	58.0 ±1.4	18.2 ±1.3	67.8
SC	92.8	84.4	88.4	83.6	82.8	74.4	65.6	23.6	74.2
Llama3.2-3b									
CoT	65.6 ±1.1	51.0 ±2.7	56.4 ±1.0	52.4 ±1.1	53.1 ±2.4	48.8 ±2.1	32.0 ±1.7	44.1 ±1.9	50.4
SC	79.2	60.4	70.8	67.2	65.6	62.8	44.4	55.6	63.0

Table 2: Vanilla Chain-of-thought(CoT) performance and self-consistency (SC) on MGSM.

	en	fr	es	de	ru	zh	ja	th	avg.
Aya-Expans-8b	94.4	85.6	90.8	89.2	87.6	84.4	82.4	44.4	82.3
Llama3.1-8b	94.8	85.2	92.4	89.2	92.0	86.8	74.8	80.0	86.9
Ministral-8b	94.4	85.6	91.6	85.2	86.4	86.4	73.2	71.6	84.3
Qwen2.5-7b	95.6	84.0	94.4	85.2	87.6	91.2	84.0	75.6	81.3
phi3-3b	96.4	92.0	95.6	90.8	92.0	88.4	84.8	42.0	85.3
Llama3.2-3b	91.6	78.8	85.6	83.2	82.8	80.8	66.0	73.6	80.3

Table 3: Comparison of Pass@8-Multi across different languages on the MGSM task.

	en	fr	es	de	ru	zh	ja	th	avg.
Aya-Expans-8b	87.2	76.4	83.2	76.8	78.4	76.0	71.2	37.2	73.3
Llama3.1-8b	90.0	77.6	84.0	60.4	80.0	78.8	68.4	70.0	76.2
Ministral-8b	91.2	76.0	87.6	77.6	77.2	78.0	60.4	62.8	76.4
Qwen2.5-7b	92.8	80.0	86.4	74.8	82.4	85.2	76.0	72.4	81.3
Phi3-7b	94.0	86.8	91.2	86.0	84.0	83.2	75.6	35.2	79.5
Llama3.2-3b	83.2	70.0	76.4	62.4	75.2	70.8	56.8	67.2	70.3

Table 4: Comparison of Multi-ORM across different languages on the MGSM task.

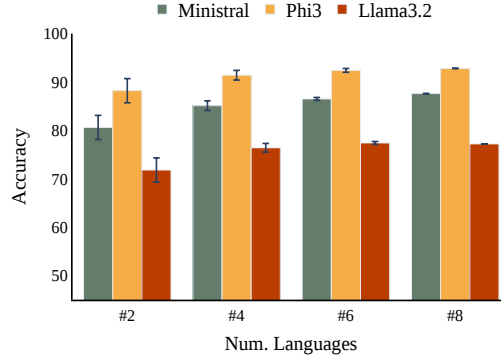


Figure 5: The average and standard deviation cross-lingual ORM performance using different numbers of languages.

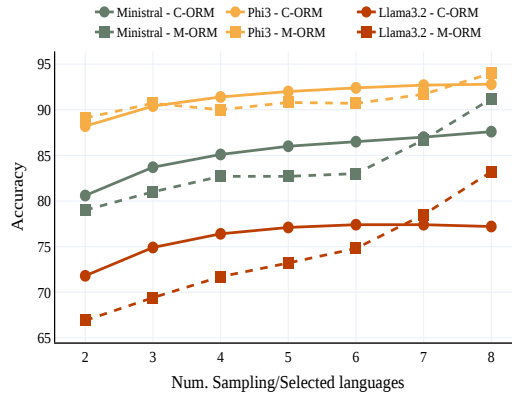


Figure 6: Comparing the accuracy of cross-lingual ORM and multilingual ORM of English using the same number of languages and samples.

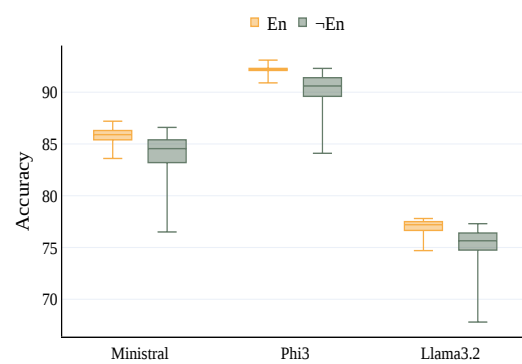


Figure 7: Comparing the role of English on the performance.