Exploring the Role of Transliteration in In-Context Learning for Low-resource Languages Written in Non-Latin Scripts

Anonymous ACL submission

Abstract

Decoder-only large language models (LLMs) excel in high-resource languages across various tasks through few-shot or even zero-shot in-context learning (ICL). However, their performance often does not transfer well to low-resource languages, especially those written in non-Latin scripts. Inspired by recent work that leverages transliteration in encoder-only models, we investigate whether transliteration[1] is also effective in improving LLMs' performance for low-resource languages written in non-Latin scripts. To this end, we propose three prompt templates, where the target-language text is represented in (1) its original script ($\text{SCRIPT}_{\{\text{Orig}\}}$), (2) Latin script ($\text{SCRIPT}_{\{\text{Latn}\}}$), or (3) both ($\text{SCRIPT}_{\{\text{Combined}\}}$). We apply these methods to several representative LLMs of different sizes on various tasks including text classification and sequential labeling. Our findings show that the effectiveness of transliteration varies by task type and model size. For instance, all models benefit from transliterations for sequential labeling (with increases of up to 25%). We make our code publicly available.

1 Introduction

Decoder-only LLMs, such as LLaMA (Touvron et al., 2023), Mixtral (Jiang et al., 2024), XGLM (Lin et al., 2022), and BLOOM (Scao et al., 2023), have shown impressive capability across a wide range of tasks for high-resource languages, particularly through few-shot ICL (Brown et al., 2020). However, they often underperform in low-resource languages, especially those written in underrepresented scripts. Multiple reasons exist, such as the scarcity of low-resource languages in the training data (Team et al., 2022; Üstün et al., 2024), insufficient crosslingual alignment during pretraining (Hämmerl et al., 2024), as well as English being the only language in the instruction tuning phase

[1]We consider a special type of transliteration that converts non-Latin scripts into Latin script (also called romanization).

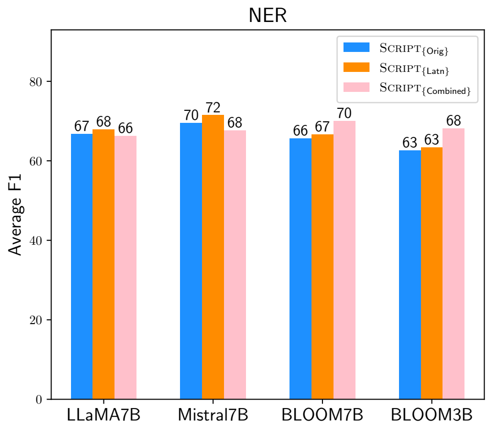

Figure 1: Results of LLaMA7B, Mistral7B, BLOOM7B and BLOOM3B on NER task. By leveraging transliteration, $\text{SCRIPT}_{\{\text{Latn}\}}$ or $\text{SCRIPT}_{\{\text{Combined}\}}$ consistently improve the performance on NER across all models.

(Chen et al., 2024). The mainstream methodology attempts to address this issue by translating the texts written in languages other than English into English using either external machine translation systems (Artetxe et al., 2023) or self-translate, i.e., translation by leveraging the few-shot translation capabilities of the model itself (Etxaniz et al., 2023). However, the quality of translations is constrained by the quality of the external systems or the LLM itself. Additionally, this type of approach is infeasible for truly low-resource languages.

Recent studies have demonstrated that leveraging transliteration into a common-script effectively improves the crosslingual transfer performance of encoder-only models on low-resource languages of non-Latin scripts (Liu et al., 2024a). This is because a common script facilitates the model to transfer knowledge through increased *lexical overlap* (Dhamecha et al., 2021; Purkayastha et al., 2023; Moosa et al., 2023). Inspired by this line of work, a natural research question is to explore whether transliteration is also effective for decoder-only LLMs, especially through their ICL capability

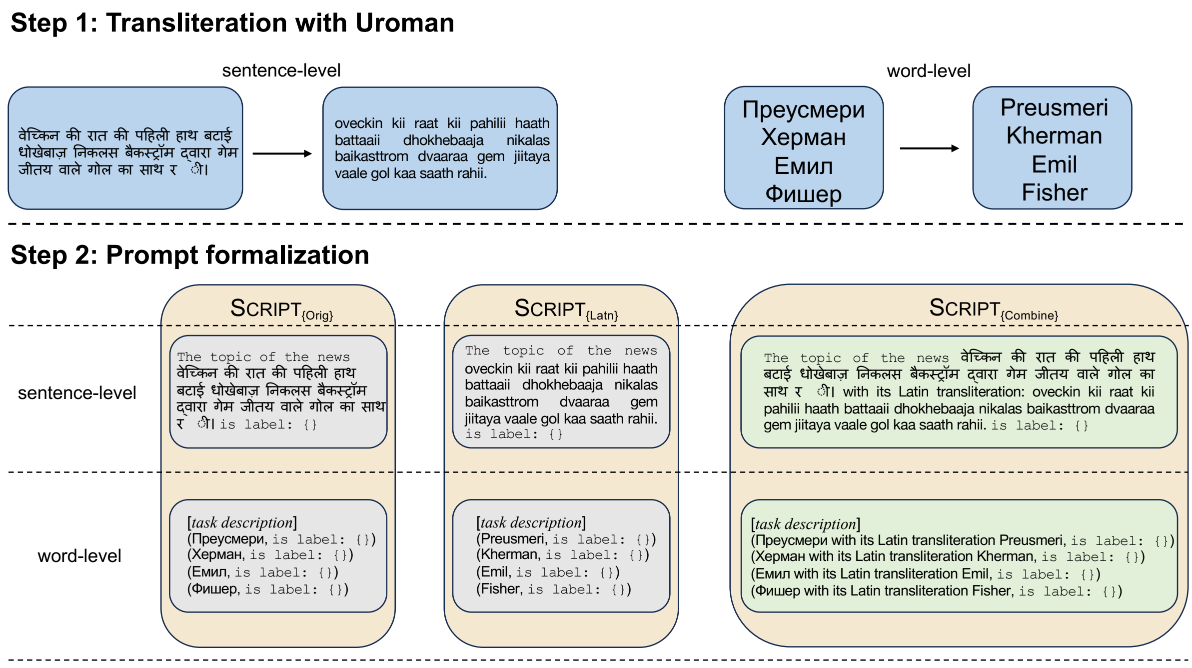

Figure 2: Illustration of our framework. We use `Uroman` (Hermjakob et al., 2018) to transliterate non-Latin texts (sentence-level for text classification, and word-level for sequential labeling). We propose three prompts: **SCRIPT**$_{\{\textbf{Orig}\}}$ (the original text is used), **SCRIPT**$_{\{\textbf{Latn}\}}$ (the Latin-script transliteration is used), and **SCRIPT**$_{\{\textbf{Combined}\}}$ (transliteration is used as an augmentation to the original text).

which does not require any parameter updates.

To this end, the paper investigates the above research question and proposes three types of prompt templates where the non-Latin target-language text is represented in (1) its original script (SCRIPT$_{\{\text{Orig}\}}$), (2) Latin script (SCRIPT$_{\{\text{Latn}\}}$), or (3) both (SCRIPT$_{\{\text{Combined}\}}$). Given that texts in different scripts convey the same semantics, the knowledge encoded in one script should complement the other. A capable model, therefore, should leverage this complementarity: when a word or an entire sentence in the original script is not well understood, the model should refer to its transliteration, and vice versa. We apply our methods to several LLMs on various tasks and observe that the effectiveness of transliteration varies by task type and model size. Transliteration is particularly helpful for sequential labeling. On other tasks, however, transliteration-augmented prompts are less effective, indicating models might have limited capacity to exploit complementary information.

Our contributions are as follows: (i) We conduct the first investigation towards the effectiveness of transliteration in ICL for decoder-only LLMs. (ii) We propose transliteration-augmented prompts that are specifically for low-resource languages in non-Latin scripts; (iii) We offer insights on when and how transliteration can enhance ICL performance.

2 Experimental Settings

Models. We experiment with six models: LLaMA2-7B (Touvron et al., 2023), Mistral-7B-Instruct (Jiang et al., 2024), and the 7B, 3B, 1B, and 560M variants of the BLOOM model (Scao et al., 2023). LLaMA2 is a model trained on 28 languages and 5 scripts (Cyrillic, Latin, Hang, Hani and Japanese). Mistral is an English-centric model trained on five languages in Latin script, while BLOOM is a multilingual LLM covering a wide range of languages in 11 scripts.[2] We select these models to compare the effectiveness of transliteration-augmented ICL on **model type** (English-centric vs multilingual models) and **model size** (different variants of BLOOM).

Methods. To investigate how transliteration impacts the ICL performance for low-resource languages in non-Latin scripts, we propose three prompt methods: (1) **SCRIPT**$_{\{\textbf{Orig}\}}$, where we feed the model with text in its original script, (2) **SCRIPT**$_{\{\textbf{Latn}\}}$, where we first transliterate the text into Latin script and only feed the transliteration into the model, and (3) **SCRIPT**$_{\{\textbf{Combined}\}}$, where we combine the text in its original script and its

[2] We check languages covered in each model's training data and consider the dominant script of each language as a script supported by the model.

Model	Size	Method	NER	SIB200	Taxi1500
LLaMA2	7B	SCRIPT$_{\{\text{Orig}\}}$	66.8	37.2	44.8
		SCRIPT$_{\{\text{Latn}\}}$	**67.9**	21.6	**46.7**
		SCRIPT$_{\{\text{Combined}\}}$	66.3	**48.5**	**46.7**
Mistral	7B	SCRIPT$_{\{\text{Orig}\}}$	69.5	**50.6**	**54.6**
		SCRIPT$_{\{\text{Latn}\}}$	**71.5**	33.2	51.1
		SCRIPT$_{\{\text{Combined}\}}$	67.7	48.6	54.3
BLOOM	7B	SCRIPT$_{\{\text{Orig}\}}$	65.6	**53.5**	**48.1**
		SCRIPT$_{\{\text{Latn}\}}$	66.7	24.3	45.7
		SCRIPT$_{\{\text{Combined}\}}$	**70.0**	53.2	47.4
	3B	SCRIPT$_{\{\text{Orig}\}}$	62.6	**48.1**	**48.0**
		SCRIPT$_{\{\text{Latn}\}}$	63.4	29.3	46.5
		SCRIPT$_{\{\text{Combined}\}}$	**68.2**	39.1	47.8
	1B	SCRIPT$_{\{\text{Orig}\}}$	51.6	42.4	50.3
		SCRIPT$_{\{\text{Latn}\}}$	56.5	22.0	**50.4**
		SCRIPT$_{\{\text{Combined}\}}$	**64.0**	**43.8**	**50.4**
	560M	SCRIPT$_{\{\text{Orig}\}}$	52.9	**41.5**	46.1
		SCRIPT$_{\{\text{Latn}\}}$	**56.7**	20.4	45.8
		SCRIPT$_{\{\text{Combined}\}}$	56.1	39.1	**46.5**

Table 1: Task performance of three prompts (**SCRIPT$_{\{\text{Orig}\}}$**, **SCRIPT$_{\{\text{Latn}\}}$**, and **SCRIPT$_{\{\text{Combined}\}}$**) for different decoder-only LLMs of various sizes, averaged by languages. Transliteration shows strong effectiveness for NER task but not for other tasks. **Bold** (underlined): best (second-best) result for each model in each task.

transliteration and feed both together into the model to solve the task. The methods are illustrated in Figure 2. For transliteration, we use `Uroman` (Hermjakob et al., 2018), a tool for universal romanization, which can be applied to any underrepresented scripts with high efficiency. Note that the task description (in English) is the same across all prompt templates. The target-language texts used for few-shot demonstrations are also transliterated in SCRIPT$_{\{\text{Latn}\}}$ and SCRIPT$_{\{\text{Combined}\}}$.

Evaluation. We consider the following tasks for evaluation: named entity recognition (**NER**), a sequence labeling task using WikiANN (Pan et al., 2017); **SIB200** (Adelani et al., 2024), a multilingual classification task covering 205 languages; and **Taxi1500** (Ma et al., 2024), a multilingual 6-class text classification dataset contains more than 1,500 languages. For each task, we only consider a subset of languages that are written in non-Latin scripts (details are shown in §A). For Taxi1500, we perform a 3-shot prompt and follow the method in Lin et al. (2024), calculating the average of word embeddings in layer 8 of the Glot500 model (Imani-Googhari et al., 2023) to retrieve semantically similar ICL samples. For NER, we perform a 3-shot prompt, since each sentence contains multiple tokens to predict and we find that 3 random demonstrations can usually cover most NER categories. We perform a 7-shot prompt for SIB200 to ensure the demonstrations cover most classes. Details of selecting the ICL demonstrations are in §B.

3 Results and Discussion

We report the average performance across all languages in Table 1 (per-language performance is in §C). In addition, we show the performance on NER averaged by script group in Table 3.

Transliteration benefits sequential labeling. Across all models, we can observe that either SCRIPT$_{\{\text{Latn}\}}$ or SCRIPT$_{\{\text{Combined}\}}$ outperforms SCRIPT$_{\{\text{Orig}\}}$ on NER. For instance, SCRIPT$_{\{\text{Combined}\}}$ increases by 12.4 compared to SCRIPT$_{\{\text{Orig}\}}$ on BLOOM-1B, which is more than 24% improvement. This demonstrates that models can make better predictions by leveraging the knowledge encoded in the Latin-script transliterations. This can be explained by the fact that NER data contains many (proper) nouns shared across languages. Transliteration enables the model to better exploit such shared vocabularies for inference.

The impact of transliteration on text classification varies across models. SCRIPT$_{\{\text{Latn}\}}$ almost always performs the worst across all models compared with its counterparts, indicating that the transliteration alone is not enough for the model to understand the sentence-level semantics. Besides, SCRIPT$_{\{\text{Combined}\}}$ performs suboptimal compared to SCRIPT$_{\{\text{Orig}\}}$ on the English-centric (Mistral) model and models trained on many multilingual

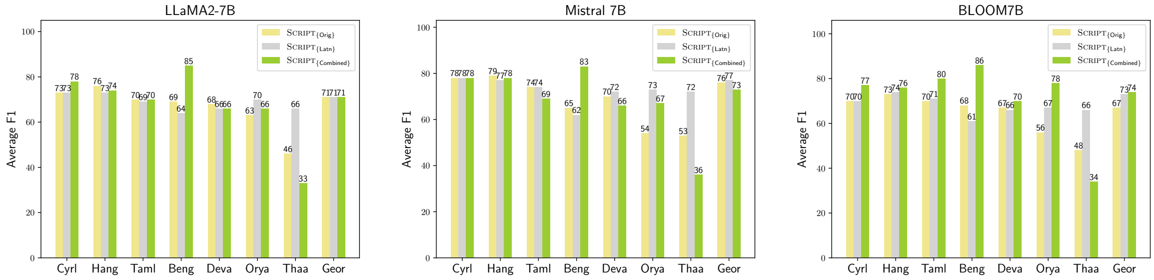

Figure 3: Performance on NER task averaged by languages of the same script. Transliterations are generally effective in improving the ICL across all models and scripts: Script$_{\{\text{Latn}\}}$ or Script$_{\{\text{Combined}\}}$ outperforms Script$_{\{\text{Orig}\}}$.

data (BLOOM), which suggests these models cannot well leverage complementary information. Instead, such information confuses the models. However, transliteration can be a good auxiliary input for good Latin-dominant models such as LLaMa (Script$_{\{\text{Combined}\}}$ achieves more than 29% and 4% on SIB200 and Taxi1500 respectively), as the model can leverage transliteration when it cannot fully understand the text in the original script.

Model performance varies by different scripts. Figure 3 shows the average macro-F1 of ten scripts on the NER task of LLaMA2-7B, Mistral-7B, and BLOOM-7B. For BLOOM-7B, Script$_{\{\text{Combined}\}}$ outperforms Script$_{\{\text{Orig}\}}$ and Script$_{\{\text{Latn}\}}$ on most scripts except Thaana, a script not seen by BLOOM-7B. Moreover, for scripts covered in the pretraining data (Tamil, Bengali, and Odia), Script$_{\{\text{Combined}\}}$ obtains the largest improvement. On the English-centric Mistral-7B, prompts containing transliteration (Script$_{\{\text{Latn}\}}$ or Script$_{\{\text{Combined}\}}$) beats Script$_{\{\text{Orig}\}}$ on 5 out of 8 scripts. For LLaMA, combining both the original text and transliteration is effective: Script$_{\{\text{Combined}\}}$ achieves the best performance on most scripts, indicating a strong ICL capability of exploring commentary information.

Model size plays an important role. Scaling up the model size usually indicates a stronger capacity from which the ICL can benefit (Zhao et al., 2023). Indeed, we observe that the performance generally increases for the BLOOM family when the model size scales up for all three prompt types across different tasks except for Taxi1500. We hypothesize this is because Taxi1500 is a relatively easy task and its data builds up on the Bible, which is part of the training data of these LLMs. In addition, the sentences in Taxi1500 contain many proper nouns whose transliterations the LLMs can easily exploit for making predictions. Therefore, we also observe good performance for Script$_{\{\text{Latn}\}}$ (comparable to the other prompts) in Taxi1500, but not in SIB200.

4 Related Work

Positive effects of transliterating data into a common script have been demonstrated in various recent works for encoder-only models (Dhamecha et al., 2021; Purkayastha et al., 2023; Moosa et al., 2023; Liu et al., 2024b). Additionally, leveraging transliteration as an auxiliary input at fine-tuning stage improves the cross-script performance (Liu et al., 2024a). To improve ICL performance for low-resource languages, demonstrations play an important role. One line of approaches replaces the target-language texts with English translations (Artetxe et al., 2023; Shi et al., 2023; Etxaniz et al., 2023). Another type of research augments the ICL demonstrations, e.g., by retrieving the most similar English texts to the target-language text (Nie et al., 2023; Li et al., 2023; Wang et al., 2023)

5 Conclusion

This study explores the effectiveness of transliteration in enhancing the ICL performance of decoder-only LLMs, focusing on low-resource languages written in non-Latin scripts. By proposing three prompt templates – using original script, Latin script, and a combination of both – we evaluate their impact across various tasks on several representative LLMs. Our findings indicate that transliteration is particularly effective for sequential labeling but its benefits for text classification tasks are less consistent. We also observe a mixed effect of transliteration related to the model type and model size. Our results highlight the potential of transliteration as a possible way to enhance LLMs' performance for low-resource languages.

Limitations

There are mainly two limitations in our work. First, we only consider models with up to 7 billion parameters due to constraints in our computing resources. Second, the evaluation data is limited in terms of the types of tasks. The major reason is the limited availability of evaluation datasets containing a variety of scripts. Nevertheless, as a pioneer study in exploring the effectiveness of transliteration for ICL involving low-resource languages in non-Latin scripts, we hope future research can leverage larger models and more datasets to explore this direction.

References

David Adelani, Hannah Liu, Xiaoyu Shen, Nikita Vassilyev, Jesujoba Alabi, Yanke Mao, Haonan Gao, and En-Shiun Lee. 2024. SIB-200: A simple, inclusive, and big evaluation dataset for topic classification in 200+ languages and dialects. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 226–245, St. Julian's, Malta. Association for Computational Linguistics.

Mikel Artetxe, Vedanuj Goswami, Shruti Bhosale, Angela Fan, and Luke Zettlemoyer. 2023. Revisiting machine translation for cross-lingual classification. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 6489–6499, Singapore. Association for Computational Linguistics.

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. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Pinzhen Chen, Shaoxiong Ji, Nikolay Bogoychev, Andrey Kutuzov, Barry Haddow, and Kenneth Heafield. 2024. Monolingual or multilingual instruction tuning: Which makes a better alpaca. In *Findings of the Association for Computational Linguistics: EACL 2024*, pages 1347–1356, St. Julian's, Malta. Association for Computational Linguistics.

Tejas Dhamecha, Rudra Murthy, Samarth Bharadwaj, Karthik Sankaranarayanan, and Pushpak Bhattacharyya. 2021. Role of Language Relatedness in Multilingual Fine-tuning of Language Models: A Case Study in Indo-Aryan Languages. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8584–8595, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Julen Etxaniz, Gorka Azkune, Aitor Soroa, Oier Lopez de Lacalle, and Mikel Artetxe. 2023. Do multilingual language models think better in english? *Preprint*, arXiv:2308.01223.

Ulf Hermjakob, Jonathan May, and Kevin Knight. 2018. Out-of-the-box universal Romanization tool uroman. In *Proceedings of ACL 2018, System Demonstrations*, pages 13–18, Melbourne, Australia. Association for Computational Linguistics.

Katharina Hämmerl, Jindřich Libovický, and Alexander Fraser. 2024. Understanding cross-lingual alignment – a survey. *Preprint*, arXiv:2404.06228.

Ayyoob ImaniGooghari, Peiqin Lin, Amir Hossein Kargaran, Silvia Severini, Masoud Jalili Sabet, Nora Kassner, Chunlan Ma, Helmut Schmid, André Martins, François Yvon, and Hinrich Schütze. 2023. Glot500: Scaling multilingual corpora and language models to 500 languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1082–1117, Toronto, Canada. Association for Computational Linguistics.

Albert Q. Jiang, Alexandre Sablayrolles, Antoine Roux, Arthur Mensch, Blanche Savary, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Emma Bou Hanna, Florian Bressand, Gianna Lengyel, Guillaume Bour, Guillaume Lample, Lélio Renard Lavaud, Lucile Saulnier, Marie-Anne Lachaux, Pierre Stock, Sandeep Subramanian, Sophia Yang, Szymon Antoniak, Teven Le Scao, Théophile Gervet, Thibaut Lavril, Thomas Wang, Timothée Lacroix, and William El Sayed. 2024. Mixtral of experts. *Preprint*, arXiv:2401.04088.

Xiaoqian Li, Ercong Nie, and Sheng Liang. 2023. Crosslingual retrieval augmented in-context learning for Bangla. In *Proceedings of the First Workshop on Bangla Language Processing (BLP-2023)*, pages 136–151, Singapore. Association for Computational Linguistics.

Peiqin Lin, Shaoxiong Ji, Jörg Tiedemann, André F. T. Martins, and Hinrich Schütze. 2024. Mala-500: Massive language adaptation of large language models. *Preprint*, arXiv:2401.13303.

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.

Yihong Liu, Chunlan Ma, Haotian Ye, and Hinrich Schütze. 2024a. Translico: A contrastive learning framework to address the script barrier in multilingual pretrained language models. *Preprint*, arXiv:2401.06620.

Yihong Liu, Chunlan Ma, Haotian Ye, and Hinrich Schütze. 2024b. Transmi: A framework to create strong baselines from multilingual pretrained language models for transliterated data. *Preprint*, arXiv:2405.09913.

Chunlan Ma, Ayyoob ImaniGooghari, Haotian Ye, Renhao Pei, Ehsaneddin Asgari, and Hinrich Schütze. 2024. Taxi1500: A multilingual dataset for text classification in 1500 languages. *Preprint*, arXiv:2305.08487.

Ibraheem Muhammad Moosa, Mahmud Elahi Akhter, and Ashfia Binte Habib. 2023. Does transliteration help multilingual language modeling? In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 670–685, Dubrovnik, Croatia. Association for Computational Linguistics.

Ercong Nie, Sheng Liang, Helmut Schmid, and Hinrich Schütze. 2023. Cross-lingual retrieval augmented prompt for low-resource languages. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8320–8340, Toronto, Canada. 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.

Sukannya Purkayastha, Sebastian Ruder, Jonas Pfeiffer, Iryna Gurevych, and Ivan Vulić. 2023. Romanization-based large-scale adaptation of multilingual language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 7996–8005, Singapore. Association for Computational Linguistics.

Teven Le Scao, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, Alexandra Sasha Luccioni, François Yvon, Matthias Gallé, et al. 2023. Bloom: A 176b-parameter open-access multilingual language model. *arXiv*.

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. 2023. Language models are multilingual chain-of-thought reasoners. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net.

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.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Weixuan Wang, Barry Haddow, and Alexandra Birch. 2023. Retrieval-augmented multilingual knowledge editing. *Preprint*, arXiv:2312.13040.

Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A survey of large language models. *Preprint*, arXiv:2303.18223.

Ahmet Üstün, Viraat Aryabumi, Zheng-Xin Yong, Wei-Yin Ko, Daniel D'souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, Freddie Vargus, Phil Blunsom, Shayne Longpre, Niklas Muennighoff, Marzieh Fadaee, Julia Kreutzer, and Sara Hooker. 2024. Aya model: An instruction finetuned open-access multilingual language model. *Preprint*, arXiv:2402.07827.

A Task Data Information

The basic information of each task dataset is shown in Table 3. The number of languages of script groups for each downstream task is shown in Table 2. We introduce the detailed hyperparameters settings for each task in the following.

For Named Entity Recognition (NER), we employ a 3-shot prompting strategy. Given that each sentence comprises multiple tokens requiring prediction, we have determined that three randomly selected demonstrations typically encompass the majority of NER categories. For SIB200, we do a 7-shot prompt. The 7 demonstrations are manually selected to cover the 7 classes of the task. For Taxi1500, we use a 3-shot prompt and adhere

Task	Method	Prompt
NER	SCRIPT{Orig}	Named Entity Recognition involves identifying and classifying named entities in a text into predefined categories such as person names, organizations, locations, and others. You will need to use the tags defined below: O means the word does'n correspond to any entity. B-PER/I-PER means the word corresponds to the beginning of/is inside a person entity. B-ORG/I-ORG means the word corresponds to the beginning of/is inside an organization entity. B-LOC/I-LOC means the word corresponds to the beginning of/is inside a location entity. Do not try to answer the question! Just tag each token in the sentence. {{'Светислав'}} ||| {{labels}}
	SCRIPT{Latn}	Named Entity Recognition involves identifying and classifying named entities in a text into predefined categories such as person names, organizations, locations, and others. You will need to use the tags defined below: O means the word doesn\u2019t correspond to any entity. B-PER/I-PER means the word corresponds to the beginning of/is inside a person entity. B-ORG/I-ORG means the word corresponds to the beginning of/is inside an organization entity. B-LOC/I-LOC means the word corresponds to the beginning of/is inside a location entity. Do not try to answer the question! Just tag each token in the sentence. {{'Svetislav'}} ||| {{labels}}
	SCRIPT{Combined}	Named Entity Recognition involves identifying and classifying named entities in a text into predefined categories such as person names, organizations, locations, and others. You will need to use the tags defined below: O means the word doesn\u2019t correspond to any entity. B-PER/I-PER means the word corresponds to the beginning of/is inside a person entity. B-ORG/I-ORG means the word corresponds to the beginning of/is inside an organization entity. B-LOC/I-LOC means the word corresponds to the beginning of/is inside a location entity. Do not try to answer the question! Just tag each token in the sentence. {{'Светислав' with its Latin transliteration Svetislav'}} ||| {{labels}}
SIB200	SCRIPT{Orig}	The topic of the news {እዚኣቶም ብጣ/ኦሚ ቀለልቲ ኣይኮኑን፣ ስለዝኮነ ኣውን ኣሰሩ ብትልሙ ክመጽእ ብትራክቲካዊ ጎቦታት ዙርያ ነዊሕ ዙረት ወሲዱ.።} is label: { }
	SCRIPT{Latn}	The topic of the news {eziaatome betaa/omi qalaleti aayekonune, selazekona aawene aasaru betelemu kematsee beteraaketikaawi gobotaate zureyaa nawihhe zurate wasidu.።} is label: { }
	SCRIPT{Combined}	The topic of the news {እዚኣቶም ብጣ/ኦሚ ቀለልቲ ኣይኮኑን፣ ስለዝኮነ ኣውን ኣሰሩ ብትልሙ ክመጽእ ብትራክቲካዊ ጎቦታት ዙርያ ነዊሕ ዙረት ወሲዱ። with its Latin transliteration: eziaatome betaa/omi qalaleti aayekonune, selazekona aawene aasaru betelemu kematsee beteraaketikaawi gobotaate zureyaa nawihhe zurate wasidu.} is label: { }
Taxi1500	SCRIPT{Orig}	The topic of the verse {既然 你們 要 按使 人 自由 的 律法 受 審判， 就 應該 按 律法 行事 為人。} is label: { }
	SCRIPT{Latn}	The topic of the verse {jiran nimen yao anshi ren ziyou de lufa shou shenpan， jiu yinggai an lufa xingshi weiren.} is label: { }
	SCRIPT{Comined}	The topic of the verse {'既然 你們 要 按使 人 自由 的 律法 受 審判， 就 應該 按 律法 行事 為人。with its Latin transliteration: jiran nimen yao anshi ren ziyou de lufa shou shenpan， jiu yinggai an lufa xingshi weiren.} is label: { }

Figure 4: Three types of prompt templates (SCRIPT$_{\{Orig\}}$, SCRIPT$_{\{Latn\}}$ and SCRIPT$_{\{Combined\}}$) that are used for each task. We follow the prompt templates in (Lin et al., 2024) for the SCRIPT$_{\{Orig\}}$, where the target-langauge text is represented in the original script. We use Latin-script transliterations obtained by `Uroman` (Hermjakob et al., 2018) for SCRIPT$_{\{Latn\}}$. SCRIPT$_{\{Combined\}}$ leverages both the original script and its Latin transliteration.

Task	|lan|								
NER	Cyrl	Arab	Hani	Deva	Geor	Hebr	Beng	other	all
	17	10	5	5	2	2	2	19	62
SIB200	Arab	Deva	Cyrl	Mymr	Beng	Tibt	Hebr		
	15	9	8	2	2	2	2	22	62
Taxi1500	Cyrl	Arab	Deva	Hani	Mymr	Beng	Orya		
	24	9	7	3	2	2	2	15	64

Table 2: The number of languages in each script group for each downstream task.

	|lan|	|rows|	#class	measure (%)
NER	62	119	7	F1 score
SIB200	62	1140	7	Accuracy
Taxi1500	64	666	6	Accuracy

Table 3: Information of evaluation tasks. |lan|: languages we select as subset to evaluate; #class: the number of the categories if it is a sequence-level or token-level classification task.

to the methodology outlined in Lin et al. (2024). Specifically, we calculate the average of contextualized word embeddings from layer 8 of the Glot500 model (ImaniGooghari et al., 2023) to identify 10 most semantically similar samples, and randomly select 3 samples as the demonstrations.

B Prompt Templates

We follow the prompt templates in (Lin et al., 2024) for SCRIPT$_{\{Orig\}}$, where the demonstrations and the query are in the original script of the target language. We employ `Uroman` (Hermjakob et al., 2018) to transliterate the target-language demonstrations and the target-language query into Latin script. SCRIPT$_{\{Latn\}}$ only uses the transliteration, while SCRIPT$_{\{Combined\}}$ leverage both the original script and its Latin transliteration.

C Full Results for All Scripts/Languages

We report the complete results for all tasks and language-scripts in Table 4 and Table 5 (**NER**), Table 6 and Table 7 (**SIB200**), and Table 8 and Table 9 (**Taxi1500**).

Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$
ara_Arab	**60.4**	56.0	41.1	ara_Arab	63.2	59.4	**82.3**	ara_Arab	69.8	63.8	**84.4**
arz_Arab	**62.4**	58.3	40.7	arz_Arab	64.1	62.4	**77.9**	arz_Arab	69.9	65.9	**81.0**
ckb_Arab	61.7	56.7	**73.6**	ckb_Arab	58.7	57.9	**77.7**	ckb_Arab	61.6	60.4	**84.1**
fas_Arab	**61.9**	59.7	59.4	fas_Arab	62.8	61.8	**82.6**	fas_Arab	66.3	63.7	**85.1**
mzn_Arab	**72.8**	68.6	64.2	mzn_Arab	67.2	70.0	**82.3**	mzn_Arab	78.4	74.0	**87.1**
pnb_Arab	53.5	**58.3**	58.2	pnb_Arab	57.1	57.5	**81.0**	pnb_Arab	66.4	65.8	**87.3**
pus_Arab	**47.2**	41.2	30.4	pus_Arab	**42.4**	39.5	28.3	pus_Arab	**49.3**	48.7	30.2
snd_Arab	**47.7**	41.5	18.3	snd_Arab	44.0	**45.4**	19.3	snd_Arab	49.7	**52.9**	19.8
uig_Arab	49.0	49.4	**50.3**	uig_Arab	49.1	50.6	**56.0**	uig_Arab	55.5	55.6	**58.7**
urd_Arab	53.8	**62.6**	39.3	urd_Arab	48.4	65.0	**71.2**	urd_Arab	47.2	69.3	**83.8**
hye_Armn	47.7	57.8	**71.3**	hye_Armn	47.3	57.4	**71.6**	hye_Armn	64.1	64.5	**78.6**
asm_Beng	51.7	49.3	**60.4**	asm_Beng	43.4	53.6	**70.7**	asm_Beng	59.3	55.1	**79.8**
ben_Beng	57.6	56.7	**66.5**	ben_Beng	60.9	57.3	**86.0**	ben_Beng	66.9	60.2	**87.0**
bak_Cyrl	64.2	**69.1**	64.6	bak_Cyrl	65.5	63.2	**69.1**	bak_Cyrl	71.3	72.2	**73.6**
bel_Cyrl	55.3	58.2	**68.7**	bel_Cyrl	56.2	55.6	**76.0**	bel_Cyrl	66.7	64.9	**75.3**
bul_Cyrl	58.5	62.7	**68.6**	bul_Cyrl	58.6	61.4	**73.7**	bul_Cyrl	69.1	70.3	**73.6**
che_Cyrl	57.2	**58.8**	42.5	che_Cyrl	59.6	**61.5**	48.4	che_Cyrl	68.9	**70.0**	52.0
chv_Cyrl	57.8	62.3	**75.6**	chv_Cyrl	56.6	61.8	**80.2**	chv_Cyrl	69.6	68.1	**85.3**
kaz_Cyrl	59.9	64.6	**69.3**	kaz_Cyrl	58.7	61.3	**70.8**	kaz_Cyrl	72.5	72.5	**75.7**
kir_Cyrl	53.5	**65.7**	62.3	kir_Cyrl	56.5	60.1	**65.1**	kir_Cyrl	70.5	**72.1**	69.7
mhr_Cyrl	54.4	55.7	**65.1**	mhr_Cyrl	51.6	54.5	**72.3**	mhr_Cyrl	62.8	65.0	**75.9**
mkd_Cyrl	58.4	**63.6**	54.7	mkd_Cyrl	59.3	**61.9**	59.2	mkd_Cyrl	69.4	**69.6**	60.1
mon_Cyrl	50.6	51.4	**70.8**	mon_Cyrl	53.9	54.6	**77.8**	mon_Cyrl	63.1	59.4	**81.7**
oss_Cyrl	56.2	59.6	**69.6**	oss_Cyrl	57.8	60.3	**71.4**	oss_Cyrl	68.3	68.0	**75.7**
rus_Cyrl	51.6	58.3	**71.2**	rus_Cyrl	55.3	55.0	**75.2**	rus_Cyrl	66.1	65.4	**79.3**
sah_Cyrl	60.3	68.3	**76.5**	sah_Cyrl	59.0	63.8	**79.8**	sah_Cyrl	70.5	71.6	**82.2**
srp_Cyrl	55.4	56.6	**69.5**	srp_Cyrl	50.3	55.1	**73.4**	srp_Cyrl	64.0	62.0	**77.2**
tat_Cyrl	58.3	61.5	**70.7**	tat_Cyrl	57.3	60.0	**73.5**	tat_Cyrl	70.9	69.8	**78.0**
tgk_Cyrl	50.6	54.7	**65.2**	tgk_Cyrl	51.7	52.6	**71.7**	tgk_Cyrl	58.6	60.1	**74.9**
ukr_Cyrl	53.8	60.1	**67.2**	ukr_Cyrl	54.8	58.3	**72.6**	ukr_Cyrl	64.7	67.3	**74.1**
bih_Deva	47.0	44.6	**49.6**	bih_Deva	45.0	45.6	**54.0**	bih_Deva	**56.1**	49.4	55.8
hin_Deva	54.0	55.6	**64.1**	hin_Deva	52.1	56.2	**79.6**	hin_Deva	62.8	60.7	**82.0**
mar_Deva	56.1	**61.6**	52.4	mar_Deva	47.6	59.1	**70.3**	mar_Deva	61.6	65.9	**77.7**
nep_Deva	50.4	**52.0**	49.1	nep_Deva	47.2	55.7	**67.5**	nep_Deva	62.2	62.0	**71.5**
san_Deva	52.0	**56.8**	40.7	san_Deva	49.7	**58.5**	50.6	san_Deva	69.5	**70.7**	55.7
amh_Ethi	41.4	54.9	**72.3**	amh_Ethi	54.2	60.0	**77.4**	amh_Ethi	63.5	63.6	**81.2**
kat_Geor	57.0	67.3	**69.9**	kat_Geor	58.1	63.8	**65.9**	kat_Geor	68.0	72.3	**73.7**
xmf_Geor	54.8	61.3	**67.8**	xmf_Geor	55.2	57.7	**69.6**	xmf_Geor	63.4	68.1	**72.1**
ell_Grek	58.0	**62.6**	59.2	ell_Grek	61.8	60.4	**66.4**	ell_Grek	**70.0**	69.6	65.1
guj_Gujr	51.4	**70.4**	24.7	guj_Gujr	31.6	**64.8**	37.8	guj_Gujr	73.3	**76.3**	47.9
pan_Guru	43.9	**57.6**	46.2	pan_Guru	33.8	58.0	**70.3**	pan_Guru	55.9	63.2	**78.5**
kor_Hang	60.7	63.6	**66.3**	kor_Hang	58.5	60.5	**70.6**	kor_Hang	69.6	70.1	**72.4**
gan_Hani	54.7	54.1	**72.6**	gan_Hani	59.3	58.7	**76.4**	gan_Hani	57.2	60.3	**72.3**
lzh_Hani	**58.4**	57.8	41.3	lzh_Hani	60.9	**62.4**	50.3	lzh_Hani	70.4	**71.7**	52.1
wuu_Hani	51.5	55.0	**71.7**	wuu_Hani	55.8	55.6	**75.6**	wuu_Hani	56.1	52.6	**72.8**
yue_Hani	**47.2**	43.2	44.1	yue_Hani	**47.7**	44.1	39.7	yue_Hani	**57.2**	53.2	50.0
zho_Hani	**44.8**	38.1	37.7	zho_Hani	**43.2**	39.0	34.2	zho_Hani	**51.4**	47.2	40.0
heb_Hebr	62.7	60.1	**67.5**	heb_Hebr	61.6	59.9	**66.0**	heb_Hebr	69.6	67.9	**72.2**
yid_Hebr	66.0	59.6	**67.3**	yid_Hebr	61.9	57.7	**69.1**	yid_Hebr	68.2	65.5	**71.2**
jpn_Jpan	33.6	**34.9**	20.0	jpn_Jpan	**38.8**	35.0	20.0	jpn_Jpan	**45.4**	45.0	25.2
khm_Khmr	47.6	46.2	**55.9**	khm_Khmr	48.1	52.4	**55.1**	khm_Khmr	51.9	56.7	**57.7**
kan_Knda	52.8	**71.6**	34.2	kan_Knda	37.4	**65.7**	61.0	kan_Knda	74.4	75.4	**76.1**
mal_Mlym	54.3	**65.9**	53.7	mal_Mlym	47.7	61.6	**70.1**	mal_Mlym	68.8	68.9	**76.9**
mya_Mymr	32.3	**52.5**	37.8	mya_Mymr	32.4	**51.4**	30.7	mya_Mymr	49.3	**59.1**	36.3
ori_Orya	35.8	**57.3**	43.4	ori_Orya	28.4	56.3	**64.9**	ori_Orya	54.8	66.8	**73.9**
sin_Sinh	44.4	56.8	**61.7**	sin_Sinh	46.5	59.1	**61.5**	sin_Sinh	62.9	64.6	**66.1**
arc_Syrc	47.4	51.9	**64.8**	arc_Syrc	51.3	50.9	**65.0**	arc_Syrc	53.1	54.8	**69.2**
tam_Taml	62.1	64.0	**65.9**	tam_Taml	57.1	62.5	**73.8**	tam_Taml	69.6	69.2	**80.2**
tel_Telu	59.4	**70.7**	61.1	tel_Telu	47.5	**67.8**	67.6	tel_Telu	73.8	75.2	**78.8**
div_Thaa	32.8	**46.6**	31.0	div_Thaa	31.4	**52.6**	31.0	div_Thaa	39.9	**58.0**	33.8
tha_Thai	17.7	**18.3**	0.7	tha_Thai	12.6	**21.7**	0.4	tha_Thai	17.5	**25.7**	0.6
bod_Tibt	61.5	53.1	**77.8**	bod_Tibt	60.5	49.9	**79.3**	bod_Tibt	59.8	51.0	**76.1**

Table 4: Macro-F1 score of NER task on BLOOM 560m, BLOOM 1B and BLOOM3B (from left to right).

Language	SCRIPT$_{\{\text{Orig}\}}$	SCRIPT$_{\{\text{Latn}\}}$	SCRIPT$_{\{\text{Combined}\}}$	Language	SCRIPT$_{\{\text{Orig}\}}$	SCRIPT$_{\{\text{Latn}\}}$	SCRIPT$_{\{\text{Combined}\}}$	Language	SCRIPT$_{\{\text{Orig}\}}$	SCRIPT$_{\{\text{Latn}\}}$	SCRIPT$_{\{\text{Combined}\}}$
ara_Arab	70.6	65.7	**86.9**	ara_Arab	69.8	65.5	**86.0**	ara_Arab	74.0	69.2	**85.9**
arz_Arab	69.8	67.7	**81.3**	arz_Arab	**73.6**	67.0	70.1	arz_Arab	78.0	73.1	**78.7**
ckb_Arab	64.0	63.0	**84.7**	ckb_Arab	65.5	64.2	**82.0**	ckb_Arab	64.3	62.2	**83.7**
fas_Arab	70.5	67.3	**87.2**	fas_Arab	73.1	68.2	**85.9**	fas_Arab	70.8	67.8	**84.5**
mzn_Arab	80.4	79.4	**87.7**	mzn_Arab	79.7	77.9	**82.3**	mzn_Arab	81.9	81.3	**83.9**
pnb_Arab	68.5	68.7	**86.9**	pnb_Arab	72.5	71.1	**90.2**	pnb_Arab	74.0	71.7	**90.9**
pus_Arab	51.1	**52.9**	31.7	pus_Arab	53.0	**58.6**	33.9	pus_Arab	55.5	**57.1**	29.4
snd_Arab	52.2	**57.5**	21.6	snd_Arab	48.8	**59.7**	25.1	snd_Arab	48.7	**59.3**	18.7
uig_Arab	57.8	60.6	**62.1**	uig_Arab	58.3	**58.5**	55.5	uig_Arab	62.3	**64.2**	60.9
urd_Arab	62.7	71.8	**83.4**	urd_Arab	74.0	70.3	**90.4**	urd_Arab	74.9	70.0	**89.6**
hye_Armn	67.5	66.6	**81.6**	hye_Armn	72.1	71.5	**80.8**	hye_Armn	71.8	73.2	**81.1**
asm_Beng	64.9	60.4	**83.3**	asm_Beng	65.9	62.0	**81.0**	asm_Beng	62.5	61.2	**78.5**
ben_Beng	72.0	62.6	**88.5**	ben_Beng	71.4	65.5	**88.7**	ben_Beng	66.7	62.3	**87.0**
bak_Cyrl	**75.2**	73.6	73.2	bak_Cyrl	**76.1**	75.9	71.9	bak_Cyrl	83.0	**83.3**	73.8
bel_Cyrl	70.1	67.4	**79.7**	bel_Cyrl	74.8	72.8	**81.1**	bel_Cyrl	77.6	78.3	**82.8**
bul_Cyrl	71.6	73.9	**78.7**	bul_Cyrl	76.2	74.7	**81.6**	bul_Cyrl	81.3	**82.1**	81.2
che_Cyrl	69.4	**73.2**	53.9	che_Cyrl	73.8	**75.3**	51.9	che_Cyrl	**76.7**	74.9	55.3
chv_Cyrl	70.0	68.8	**86.0**	chv_Cyrl	75.9	76.5	**84.8**	chv_Cyrl	79.6	80.3	**87.5**
kaz_Cyrl	73.3	74.2	**77.1**	kaz_Cyrl	77.9	76.8	**79.3**	kaz_Cyrl	**80.9**	80.2	79.9
kir_Cyrl	73.4	**74.2**	70.2	kir_Cyrl	73.1	**74.6**	69.7	kir_Cyrl	78.5	**79.3**	69.5
mhr_Cyrl	67.2	67.7	**83.0**	mhr_Cyrl	69.5	73.8	**85.3**	mhr_Cyrl	79.4	78.3	**84.9**
mkd_Cyrl	71.7	**73.0**	61.2	mkd_Cyrl	**75.7**	73.6	65.1	mkd_Cyrl	78.4	**79.0**	62.9
mon_Cyrl	66.1	66.8	**85.2**	mon_Cyrl	69.0	66.7	**83.1**	mon_Cyrl	73.9	73.3	**81.3**
oss_Cyrl	70.5	70.1	**81.2**	oss_Cyrl	73.5	72.4	**80.2**	oss_Cyrl	76.5	74.7	**80.5**
rus_Cyrl	69.0	68.8	**80.7**	rus_Cyrl	71.9	72.8	**83.4**	rus_Cyrl	78.2	76.6	**83.4**
sah_Cyrl	69.9	69.2	**82.9**	sah_Cyrl	74.7	72.5	**81.7**	sah_Cyrl	81.3	80.7	**83.9**
srp_Cyrl	66.5	68.7	**79.3**	srp_Cyrl	70.1	69.7	**82.6**	srp_Cyrl	77.2	76.8	**82.6**
tat_Cyrl	72.4	73.3	**79.7**	tat_Cyrl	76.9	74.5	**82.6**	tat_Cyrl	**81.2**	79.7	81.1
tgk_Cyrl	63.3	61.8	**76.3**	tgk_Cyrl	65.0	65.3	**74.8**	tgk_Cyrl	70.6	71.3	**78.8**
ukr_Cyrl	66.7	71.3	**78.0**	ukr_Cyrl	72.7	73.1	**82.5**	ukr_Cyrl	77.9	78.8	**80.6**
bih_Deva	**63.5**	60.7	62.3	bih_Deva	65.0	64.6	**69.0**	bih_Deva	64.9	**68.9**	57.3
hin_Deva	66.4	65.1	**82.9**	hin_Deva	69.8	65.8	**86.0**	hin_Deva	71.2	69.7	**83.2**
mar_Deva	62.3	68.1	**77.0**	mar_Deva	71.0	70.1	**78.6**	mar_Deva	76.1	75.1	**76.8**
nep_Deva	70.6	68.9	**74.5**	nep_Deva	**72.3**	64.5	60.1	nep_Deva	**72.8**	70.7	62.7
san_Deva	**71.9**	68.8	51.5	san_Deva	59.4	**65.9**	38.6	san_Deva	58.7	**71.4**	42.4
amh_Ethi	69.0	68.2	**83.5**	amh_Ethi	66.1	63.6	**76.7**	amh_Ethi	69.7	69.4	**80.4**
kat_Geor	69.7	**76.2**	74.8	kat_Geor	73.0	**73.5**	70.1	kat_Geor	77.2	**79.1**	75.0
xmf_Geor	64.7	69.2	**72.9**	xmf_Geor	68.2	68.9	**71.8**	xmf_Geor	73.8	**75.3**	72.0
ell_Grek	72.0	**74.3**	68.5	ell_Grek	**74.7**	74.4	66.6	ell_Grek	78.7	**80.8**	63.0
guj_Gujr	73.5	**77.1**	47.9	guj_Gujr	38.8	**72.5**	15.9	guj_Gujr	47.2	**77.1**	19.0
pan_Guru	58.7	64.8	**81.3**	pan_Guru	67.5	68.7	**82.5**	pan_Guru	65.3	67.4	**79.5**
kor_Hang	72.8	73.9	**75.9**	kor_Hang	**75.5**	72.7	73.9	kor_Hang	**78.6**	77.2	78.0
gan_Hani	65.3	65.1	**78.0**	gan_Hani	75.0	75.8	**85.0**	gan_Hani	68.7	72.1	**84.5**
lzh_Hani	75.0	**76.0**	52.5	lzh_Hani	**76.1**	71.5	45.2	lzh_Hani	78.6	**79.5**	51.4
wuu_Hani	68.7	62.4	**83.4**	wuu_Hani	73.3	61.4	**82.2**	wuu_Hani	69.3	66.6	**87.6**
yue_Hani	**63.6**	58.6	43.1	yue_Hani	**64.7**	61.2	43.7	yue_Hani	**68.5**	62.5	42.1
zho_Hani	**55.6**	51.5	40.5	zho_Hani	**58.9**	53.1	39.2	zho_Hani	**62.5**	55.5	40.6
heb_Hebr	**71.1**	69.8	70.5	heb_Hebr	**72.5**	71.5	71.7	heb_Hebr	**77.8**	74.5	69.8
yid_Hebr	67.2	66.3	**73.0**	yid_Hebr	**73.4**	71.2	69.2	yid_Hebr	**72.6**	72.1	71.4
jpn_Jpan	**49.7**	48.6	25.1	jpn_Jpan	**51.9**	50.1	21.5	jpn_Jpan	**55.9**	51.2	26.0
khm_Khmr	59.9	62.1	**65.7**	khm_Khmr	52.1	**62.0**	49.1	khm_Khmr	62.3	**71.4**	64.1
kan_Knda	72.0	**76.8**	74.7	kan_Knda	53.0	**76.1**	25.7	kan_Knda	71.6	**78.8**	55.8
mal_Mlym	66.9	70.3	**78.4**	mal_Mlym	**71.0**	70.0	61.3	mal_Mlym	53.1	**73.3**	34.8
mya_Mymr	50.3	**62.8**	37.9	mya_Mymr	50.6	**54.1**	42.7	mya_Mymr	56.8	**64.0**	39.8
ori_Orya	56.4	67.4	**77.9**	ori_Orya	63.4	**70.2**	66.1	ori_Orya	54.4	**73.2**	67.0
sin_Sinh	64.6	**66.1**	65.8	sin_Sinh	65.1	**67.9**	61.8	sin_Sinh	66.6	**71.7**	65.5
arc_Syrc	59.1	61.7	**77.2**	arc_Syrc	62.0	65.5	**73.7**	arc_Syrc	62.0	66.5	**79.4**
tam_Taml	70.3	70.7	**80.0**	tam_Taml	69.6	68.6	**69.7**	tam_Taml	**74.1**	73.7	68.9
tel_Telu	72.3	**78.1**	77.2	tel_Telu	56.2	**74.9**	22.6	tel_Telu	71.2	**78.7**	49.7
div_Thaa	48.1	**66.3**	33.8	div_Thaa	46.0	**65.9**	32.6	div_Thaa	53.2	**71.6**	36.1
tha_Thai	18.0	**27.0**	0.7	tha_Thai	21.8	**24.7**	0.8	tha_Thai	22.0	**26.1**	0.7
bod_Tibt	61.7	53.9	**81.4**	bod_Tibt	62.0	66.7	**84.7**	bod_Tibt	56.1	70.6	**87.6**

Table 5: Macro-F1 score of NER task on NER task on BLOOM 7B, LLaMA2-7B and Mixtral 7B (from left to right)

Language	SCRIPT$_{\{\text{Orig}\}}$	SCRIPT$_{\{\text{Latn}\}}$	SCRIPT$_{\{\text{Combined}\}}$	Language	SCRIPT$_{\{\text{Orig}\}}$	SCRIPT$_{\{\text{Latn}\}}$	SCRIPT$_{\{\text{Combined}\}}$	Language	SCRIPT$_{\{\text{Orig}\}}$	SCRIPT$_{\{\text{Latn}\}}$	SCRIPT$_{\{\text{Combined}\}}$
ace_Arab	18.1	18.6	**20.6**	ace_Arab	16.7	16.7	**21.6**	ace_Arab	**27.5**	19.6	20.6
acm_Arab	63.7	16.7	**66.7**	acm_Arab	67.6	18.1	**70.1**	acm_Arab	**77.9**	19.1	66.7
acq_Arab	63.7	16.7	**64.7**	acq_Arab	69.1	17.2	**72.1**	acq_Arab	**77.9**	21.1	64.7
aeb_Arab	**63.2**	17.6	56.9	aeb_Arab	62.7	17.2	**67.2**	aeb_Arab	**74.0**	18.6	56.9
ajp_Arab	**68.6**	18.1	67.2	ajp_Arab	71.1	18.1	**74.5**	ajp_Arab	**75.0**	25.5	67.2
apc_Arab	70.1	19.1	**71.1**	apc_Arab	74.5	17.2	**75.0**	apc_Arab	**77.9**	26.5	71.1
ars_Arab	**65.7**	15.7	65.7	ars_Arab	67.6	16.7	**72.1**	ars_Arab	**76.5**	19.6	65.7
ary_Arab	**63.7**	16.7	61.3	ary_Arab	60.8	16.7	**72.1**	ary_Arab	**76.0**	20.1	61.3
azb_Arab	**36.3**	17.6	32.4	azb_Arab	32.8	17.6	**35.8**	azb_Arab	**34.3**	22.5	32.4
ckb_Arab	**21.1**	19.1	19.6	ckb_Arab	19.6	17.2	**20.6**	ckb_Arab	**25.5**	24.0	19.6
knc_Arab	23.5	19.1	**24.0**	knc_Arab	20.6	19.6	**23.0**	knc_Arab	18.6	**28.9**	24.0
pbt_Arab	**38.7**	20.6	32.4	pbt_Arab	34.8	24.0	**35.3**	pbt_Arab	**48.0**	28.9	32.4
pes_Arab	39.7	19.1	**44.1**	pes_Arab	48.0	21.6	**52.0**	pes_Arab	**52.0**	28.4	44.1
prs_Arab	**42.2**	15.2	36.8	prs_Arab	45.6	20.1	**51.5**	prs_Arab	**52.0**	25.0	36.8
uig_Arab	**20.1**	16.7	16.7	uig_Arab	19.1	17.2	**20.6**	uig_Arab	21.6	**24.5**	16.7
hye_Armn	16.7	**25.0**	18.6	hye_Armn	15.7	**32.4**	15.7	hye_Armn	21.1	**36.3**	18.6
asm_Beng	**58.3**	12.3	42.6	asm_Beng	**43.6**	17.6	42.6	asm_Beng	**72.1**	23.5	42.6
ben_Beng	**73.5**	16.7	69.1	ben_Beng	**72.5**	16.2	69.6	ben_Beng	**77.5**	27.5	69.1
bak_Cyrl	19.6	**39.2**	24.5	bak_Cyrl	26.0	**39.2**	31.9	bak_Cyrl	37.7	**50.5**	24.5
kaz_Cyrl	23.0	**30.4**	22.1	kaz_Cyrl	29.9	**35.3**	31.4	kaz_Cyrl	37.3	**43.1**	22.1
kir_Cyrl	26.5	**34.3**	22.1	kir_Cyrl	28.9	**38.7**	34.8	kir_Cyrl	34.3	**50.0**	22.1
mkd_Cyrl	21.1	**33.8**	24.0	mkd_Cyrl	24.0	**38.2**	28.4	mkd_Cyrl	34.8	**51.5**	24.0
rus_Cyrl	24.0	**37.3**	26.0	rus_Cyrl	**43.6**	41.2	43.6	rus_Cyrl	**57.8**	51.5	26.0
srp_Cyrl	25.5	**37.3**	25.5	srp_Cyrl	25.5	**35.8**	29.4	srp_Cyrl	32.4	**54.4**	25.5
tgk_Cyrl	20.6	**25.0**	20.6	tgk_Cyrl	22.1	**29.4**	28.4	tgk_Cyrl	27.0	**45.1**	20.6
ukr_Cyrl	21.1	**39.7**	24.5	ukr_Cyrl	28.4	**41.2**	31.9	ukr_Cyrl	40.7	**51.0**	24.5
awa_Deva	**63.7**	17.6	59.8	awa_Deva	**68.6**	18.1	66.2	awa_Deva	**73.5**	26.5	59.8
bho_Deva	**69.6**	17.6	63.7	bho_Deva	64.7	19.1	**66.2**	bho_Deva	**72.5**	28.4	63.7
hin_Deva	**65.7**	14.7	63.2	hin_Deva	71.1	16.7	**72.5**	hin_Deva	**74.5**	22.5	63.2
hne_Deva	**63.2**	15.2	57.8	hne_Deva	**65.2**	18.1	64.2	hne_Deva	**73.5**	22.1	57.8
kas_Deva	43.6	23.5	**46.1**	kas_Deva	50.0	21.1	**51.5**	kas_Deva	**49.5**	35.3	46.1
mag_Deva	**66.2**	15.2	62.3	mag_Deva	**67.6**	17.2	67.2	mag_Deva	**75.5**	22.5	62.3
mai_Deva	**64.2**	15.7	59.8	mai_Deva	**63.7**	18.1	62.3	mai_Deva	**73.5**	24.5	59.8
npi_Deva	**65.7**	17.6	58.8	npi_Deva	**71.6**	20.1	68.6	npi_Deva	**65.2**	27.0	58.8
san_Deva	**57.8**	14.2	53.4	san_Deva	53.4	20.6	**56.9**	san_Deva	**59.8**	24.0	53.4
amh_Ethi	17.6	**18.1**	15.2	amh_Ethi	14.7	16.2	**17.2**	amh_Ethi	15.7	**27.9**	15.2
tir_Ethi	**20.1**	18.1	15.7	tir_Ethi	15.2	**16.2**	16.2	tir_Ethi	16.7	**27.9**	15.7
kat_Geor	21.1	**30.4**	23.5	kat_Geor	17.6	**36.3**	25.5	kat_Geor	14.7	**41.7**	23.5
ell_Grek	19.6	**27.9**	17.6	ell_Grek	16.7	**33.8**	25.0	ell_Grek	24.0	**45.6**	17.6
pan_Guru	**57.4**	17.2	54.9	pan_Guru	54.4	16.7	**56.4**	pan_Guru	**69.6**	20.6	54.9
zho_Hans	**70.1**	20.1	68.6	zho_Hans	**75.5**	17.6	73.0	zho_Hans	**73.5**	23.5	68.6
yue_Hant	**68.6**	18.1	67.2	yue_Hant	**72.1**	21.1	71.6	yue_Hant	**74.5**	25.5	67.2
zho_Hant	**74.0**	12.7	71.1	zho_Hant	**76.0**	19.1	74.0	zho_Hant	**76.5**	25.0	71.1
heb_Hebr	**21.1**	16.7	18.1	heb_Hebr	15.2	16.7	**18.6**	heb_Hebr	**21.6**	20.6	18.1
ydd_Hebr	**21.1**	18.6	15.2	ydd_Hebr	**21.6**	17.6	20.6	ydd_Hebr	16.2	**22.5**	15.2
jpn_Jpan	**63.7**	21.1	57.8	jpn_Jpan	**67.2**	18.1	66.2	jpn_Jpan	**75.0**	26.0	57.8
khm_Khmr	21.6	**26.0**	16.7	khm_Khmr	20.6	**28.9**	21.1	khm_Khmr	29.9	**37.7**	16.7
kan_Knda	**57.8**	16.2	55.4	kan_Knda	64.7	16.7	**66.2**	kan_Knda	**66.7**	27.0	55.4
lao_Laoo	22.5	**24.0**	22.1	lao_Laoo	28.9	27.0	**31.9**	lao_Laoo	28.9	**37.3**	22.1
mal_Mlym	**68.1**	16.7	49.0	mal_Mlym	67.6	18.6	**70.6**	mal_Mlym	**71.6**	21.1	49.0
mya_Mymr	17.6	**20.6**	18.1	mya_Mymr	12.7	**18.1**	15.7	mya_Mymr	19.1	**28.4**	18.1
shn_Mymr	21.1	**22.1**	18.1	shn_Mymr	27.9	**31.9**	25.0	shn_Mymr	26.5	**40.2**	18.1
nqo_Nkoo	17.2	16.7	**17.6**	nqo_Nkoo	13.2	**18.1**	14.2	nqo_Nkoo	14.2	**27.0**	17.6
sat_Olck	18.1	18.6	**20.1**	sat_Olck	**16.7**	14.7	15.7	sat_Olck	**22.5**	22.5	20.1
ory_Orya	**58.8**	19.1	58.8	ory_Orya	**67.6**	20.6	63.7	ory_Orya	**65.7**	31.4	58.8
sin_Sinh	17.2	16.2	**17.6**	sin_Sinh	13.2	**18.6**	14.7	sin_Sinh	15.2	**21.6**	17.6
tam_Taml	**76.5**	17.2	64.7	tam_Taml	**75.0**	17.2	71.1	tam_Taml	**74.5**	23.5	64.7
tel_Telu	**62.3**	15.7	53.4	tel_Telu	**67.2**	22.1	58.3	tel_Telu	**66.2**	25.0	53.4
tzm_Tfng	14.2	**16.2**	13.7	tzm_Tfng	14.2	**16.7**	12.3	tzm_Tfng	15.2	**24.0**	13.7
bod_Tibt	**19.1**	14.7	14.7	bod_Tibt	13.2	**17.2**	15.7	bod_Tibt	21.6	**25.0**	14.7
dzo_Tibt	17.2	**19.1**	14.2	dzo_Tibt	13.2	**16.2**	9.3	dzo_Tibt	14.2	**18.1**	14.2

Table 6: Accuracy of SIB200 task on BLOOM 560m, BLOOM 1B and BLOOM3B (from left to right).

Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$
ace_Arab	22.1	17.6	**24.5**	ace_Arab	17.6	11.8	**19.6**	ace_Arab	29.4	16.7	**29.9**
acm_Arab	79.9	10.3	**81.4**	acm_Arab	63.7	22.5	**70.1**	acm_Arab	**77.0**	22.5	73.0
acq_Arab	78.9	12.7	**81.9**	acq_Arab	62.3	15.2	**67.2**	acq_Arab	**77.0**	21.1	74.0
aeb_Arab	**76.5**	12.3	74.5	aeb_Arab	58.8	17.6	**65.7**	aeb_Arab	**72.1**	20.1	69.6
ajp_Arab	**82.4**	19.1	76.5	ajp_Arab	60.3	17.2	**65.7**	ajp_Arab	**72.5**	26.0	70.1
apc_Arab	79.9	17.6	**81.9**	apc_Arab	56.9	16.2	**65.2**	apc_Arab	**75.5**	24.0	73.5
ars_Arab	78.4	10.8	**78.9**	ars_Arab	61.3	15.2	**71.1**	ars_Arab	**77.5**	20.1	75.0
ary_Arab	77.0	13.2	**77.5**	ary_Arab	53.9	16.7	**62.7**	ary_Arab	**75.0**	19.6	71.1
azb_Arab	**42.6**	17.6	41.2	azb_Arab	29.9	17.2	**43.1**	azb_Arab	**56.4**	22.5	52.0
ckb_Arab	23.5	20.6	**27.0**	ckb_Arab	16.7	18.1	**22.1**	ckb_Arab	30.9	26.5	**34.8**
knc_Arab	**24.5**	21.1	20.1	knc_Arab	16.7	18.6	**20.6**	knc_Arab	23.0	24.5	**25.0**
pbt_Arab	**56.9**	30.9	50.5	pbt_Arab	25.5	22.5	**29.9**	pbt_Arab	**54.9**	32.8	50.5
pes_Arab	**68.1**	25.0	64.7	pes_Arab	53.4	21.1	**64.7**	pes_Arab	**73.5**	32.8	70.1
prs_Arab	**66.7**	21.1	62.3	prs_Arab	54.4	20.1	**62.3**	prs_Arab	**73.5**	32.8	67.6
uig_Arab	**23.0**	17.6	20.6	uig_Arab	17.2	17.2	**22.1**	uig_Arab	35.8	30.9	**39.2**
hye_Armn	20.1	**31.4**	28.4	hye_Armn	17.2	**26.0**	22.1	hye_Armn	40.2	40.2	**44.1**
asm_Beng	77.9	16.2	**78.4**	asm_Beng	26.0	25.5	**38.2**	asm_Beng	**49.5**	40.2	45.1
ben_Beng	75.0	17.6	**80.4**	ben_Beng	37.7	27.9	**50.5**	ben_Beng	**62.7**	38.2	54.4
bak_Cyrl	49.0	**51.0**	44.1	bak_Cyrl	41.7	42.2	**46.1**	bak_Cyrl	59.8	57.8	**61.3**
kaz_Cyrl	45.6	**49.0**	44.1	kaz_Cyrl	32.8	**43.1**	35.3	kaz_Cyrl	**58.8**	53.4	57.8
kir_Cyrl	45.6	**48.5**	44.6	kir_Cyrl	38.7	**41.2**	41.2	kir_Cyrl	**63.7**	56.4	62.3
mkd_Cyrl	45.1	**49.0**	48.5	mkd_Cyrl	64.7	59.8	**66.2**	mkd_Cyrl	**76.0**	68.1	76.0
rus_Cyrl	66.7	57.4	**70.6**	rus_Cyrl	73.5	66.2	**77.5**	rus_Cyrl	**83.8**	77.5	80.9
srp_Cyrl	43.1	**51.5**	49.0	srp_Cyrl	70.1	69.6	**74.0**	srp_Cyrl	**83.3**	80.9	82.4
tgk_Cyrl	33.3	**40.2**	35.8	tgk_Cyrl	25.0	**31.9**	28.9	tgk_Cyrl	49.5	**52.0**	49.5
ukr_Cyrl	52.5	50.0	**53.9**	ukr_Cyrl	74.0	55.4	**75.5**	ukr_Cyrl	80.4	71.6	**81.4**
awa_Deva	**77.9**	15.7	77.0	awa_Deva	52.0	34.8	**62.3**	awa_Deva	**64.2**	45.6	61.3
bho_Deva	**76.0**	17.2	75.5	bho_Deva	41.7	32.8	**49.5**	bho_Deva	**59.3**	45.6	57.4
hin_Deva	**79.9**	19.6	78.9	hin_Deva	52.9	41.7	**62.3**	hin_Deva	**67.6**	56.9	66.7
hne_Deva	**75.5**	17.2	74.5	hne_Deva	44.6	29.9	**54.4**	hne_Deva	60.8	42.2	**61.3**
kas_Deva	**59.8**	25.5	57.4	kas_Deva	31.4	24.0	**36.8**	kas_Deva	**50.0**	34.3	48.5
mag_Deva	77.9	15.2	**78.9**	mag_Deva	45.1	29.4	**56.4**	mag_Deva	**59.3**	42.6	55.9
mai_Deva	77.0	15.2	**77.9**	mai_Deva	45.1	34.8	**56.9**	mai_Deva	**60.3**	39.2	59.8
npi_Deva	78.4	22.1	**79.4**	npi_Deva	50.5	33.8	**52.5**	npi_Deva	**62.7**	49.5	55.9
san_Deva	**70.1**	16.7	64.7	san_Deva	39.2	33.8	**46.1**	san_Deva	**52.0**	46.1	50.5
amh_Ethi	17.2	**18.6**	14.7	amh_Ethi	14.7	**17.6**	16.2	amh_Ethi	21.6	**27.5**	25.0
tir_Ethi	18.1	**19.6**	14.7	tir_Ethi	14.7	**16.7**	15.7	tir_Ethi	21.1	23.0	**23.5**
kat_Geor	25.0	**46.6**	35.3	kat_Geor	23.5	**39.7**	28.9	kat_Geor	49.5	52.0	**56.9**
ell_Grek	32.8	**49.5**	31.4	ell_Grek	53.9	39.7	**63.2**	ell_Grek	**74.0**	60.8	69.6
pan_Guru	78.9	10.8	**79.4**	pan_Guru	16.7	19.1	**20.6**	pan_Guru	27.5	**31.9**	27.0
zho_Hans	80.9	21.6	**83.8**	zho_Hans	72.1	15.7	**78.4**	zho_Hans	**81.4**	31.4	80.9
yue_Hant	78.9	16.2	**81.4**	yue_Hant	70.1	14.7	**76.0**	yue_Hant	77.5	24.5	**78.4**
zho_Hant	82.8	14.7	**83.8**	zho_Hant	72.5	11.8	**76.0**	zho_Hant	80.4	27.9	**81.4**
heb_Hebr	**27.5**	20.1	23.0	heb_Hebr	40.7	13.2	**47.5**	heb_Hebr	**65.2**	16.2	60.8
ydd_Hebr	**23.0**	23.0	21.6	ydd_Hebr	20.6	18.6	**24.5**	ydd_Hebr	**32.8**	23.0	28.4
jpn_Jpan	**78.9**	17.6	77.5	jpn_Jpan	66.7	14.7	**76.5**	jpn_Jpan	**81.4**	25.0	77.5
khm_Khmr	**38.7**	37.3	33.3	khm_Khmr	23.0	23.5	**24.0**	khm_Khmr	**42.2**	32.8	39.2
kan_Knda	74.5	20.6	**77.5**	kan_Knda	21.1	**26.0**	25.0	kan_Knda	41.7	38.2	**42.2**
lao_Laoo	33.8	**43.1**	30.4	lao_Laoo	20.6	**26.0**	25.0	lao_Laoo	**36.3**	32.8	34.8
mal_Mlym	76.5	16.7	**80.9**	mal_Mlym	19.6	20.1	**24.5**	mal_Mlym	28.4	**32.4**	27.0
mya_Mymr	18.6	**25.0**	17.6	mya_Mymr	**20.1**	19.1	18.6	mya_Mymr	**27.9**	22.1	24.0
shn_Mymr	31.9	**39.2**	29.4	shn_Mymr	**32.4**	31.4	27.9	shn_Mymr	35.3	**39.2**	38.7
nqo_Nkoo	**15.7**	15.7	12.3	nqo_Nkoo	**16.7**	14.2	16.7	nqo_Nkoo	15.2	**18.6**	17.2
sat_Olck	15.2	**20.6**	12.7	sat_Olck	14.7	13.2	**17.2**	sat_Olck	9.8	**13.2**	11.3
ory_Orya	**78.4**	19.1	77.0	ory_Orya	17.2	**26.0**	20.6	ory_Orya	22.1	**44.1**	30.9
sin_Sinh	**18.6**	15.7	15.7	sin_Sinh	18.1	**26.0**	23.5	sin_Sinh	26.5	**35.8**	28.4
tam_Taml	**77.9**	16.2	77.9	tam_Taml	20.1	14.7	**33.8**	tam_Taml	**37.3**	23.0	32.8
tel_Telu	75.0	20.6	**76.5**	tel_Telu	18.1	**27.5**	23.0	tel_Telu	30.9	**45.6**	37.7
tzm_Tfng	**20.6**	16.2	16.7	tzm_Tfng	13.7	14.2	**16.2**	tzm_Tfng	15.7	18.6	**19.1**
bod_Tibt	**20.1**	15.7	19.1	bod_Tibt	16.2	**18.1**	16.7	bod_Tibt	22.5	17.2	**23.0**
dzo_Tibt	15.7	12.7	**16.7**	dzo_Tibt	15.2	**16.2**	15.7	dzo_Tibt	**20.6**	14.7	19.1

Table 7: Accuracy of SIB200 task on BLOOM 7B, LLaMA2-7B and Mixtral 7B (from left to right)

Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$
arb_Arab	45.9	45.9	**46.8**	arb_Arab	53.2	**54.1**	53.2	arb_Arab	49.5	46.8	**55.0**
ary_Arab	34.2	**41.4**	36.9	ary_Arab	43.2	45.0	**45.9**	ary_Arab	36.0	35.1	**44.1**
arz_Arab	35.1	35.1	**36.9**	arz_Arab	44.1	**45.9**	45.9	arz_Arab	40.5	36.9	**41.4**
azb_Arab	**43.2**	42.3	39.6	azb_Arab	46.8	**50.5**	48.6	azb_Arab	41.4	41.4	**43.2**
ckb_Arab	45.0	46.8	**47.7**	ckb_Arab	46.8	**48.6**	48.6	ckb_Arab	43.2	**44.1**	43.2
fas_Arab	**53.2**	49.5	53.2	fas_Arab	53.2	**55.0**	55.0	fas_Arab	**49.5**	49.5	49.5
pes_Arab	53.6	46.4	**55.5**	pes_Arab	50.0	51.8	**55.5**	pes_Arab	49.1	48.2	**50.0**
prs_Arab	**56.8**	54.1	55.9	prs_Arab	**60.4**	59.5	59.5	prs_Arab	56.8	53.2	**57.7**
snd_Arab	**54.1**	53.2	53.2	snd_Arab	**55.9**	54.1	55.9	snd_Arab	48.6	**49.5**	49.5
hye_Armn	45.9	**47.7**	45.9	hye_Armn	52.3	**58.6**	53.2	hye_Armn	52.3	52.3	**55.9**
asm_Beng	36.0	**37.8**	36.9	asm_Beng	**45.0**	36.9	38.7	asm_Beng	**48.6**	36.0	46.8
ben_Beng	**40.5**	37.8	40.5	ben_Beng	**47.7**	45.0	43.2	ben_Beng	**47.7**	42.3	45.0
alt_Cyrl	**48.6**	48.6	45.9	alt_Cyrl	**53.2**	52.3	46.8	alt_Cyrl	**47.7**	46.8	45.9
bak_Cyrl	**45.0**	45.0	43.2	bak_Cyrl	51.4	**55.9**	51.4	bak_Cyrl	46.8	**51.4**	46.8
bel_Cyrl	**45.9**	43.2	45.0	bel_Cyrl	51.4	**55.0**	52.3	bel_Cyrl	45.0	**45.9**	45.0
bul_Cyrl	**41.4**	36.9	37.8	bul_Cyrl	**44.1**	42.3	41.4	bul_Cyrl	**46.8**	44.1	44.1
che_Cyrl	**36.0**	36.0	36.0	che_Cyrl	40.5	**44.1**	41.4	che_Cyrl	33.3	34.2	**35.1**
chv_Cyrl	46.8	**48.6**	47.7	chv_Cyrl	50.5	**52.3**	51.4	chv_Cyrl	42.3	41.4	**45.9**
crh_Cyrl	47.7	**48.6**	46.8	crh_Cyrl	48.6	47.7	**49.5**	crh_Cyrl	**51.4**	45.9	48.6
kaz_Cyrl	45.0	44.1	**48.6**	kaz_Cyrl	50.5	**52.3**	46.8	kaz_Cyrl	**55.0**	53.2	52.3
kir_Cyrl	**62.2**	61.3	59.5	kir_Cyrl	62.2	**64.0**	63.1	kir_Cyrl	56.8	**59.5**	58.6
kjh_Cyrl	42.3	**44.1**	44.1	kjh_Cyrl	49.5	**50.5**	49.5	kjh_Cyrl	42.3	**47.7**	45.0
kmr_Cyrl	**38.7**	37.8	38.7	kmr_Cyrl	44.1	**45.0**	45.0	kmr_Cyrl	**43.2**	43.2	39.6
krc_Cyrl	**45.0**	41.4	42.3	krc_Cyrl	49.5	**55.9**	51.4	krc_Cyrl	**45.9**	45.9	45.0
mhr_Cyrl	48.2	50.9	**51.8**	mhr_Cyrl	**50.0**	49.1	44.5	mhr_Cyrl	**50.9**	44.5	49.1
mkd_Cyrl	54.1	**57.7**	55.9	mkd_Cyrl	**61.3**	61.3	57.7	mkd_Cyrl	**56.8**	53.2	55.0
myv_Cyrl	36.0	**38.7**	38.7	myv_Cyrl	**46.8**	44.1	46.8	myv_Cyrl	**45.0**	45.0	40.5
oss_Cyrl	47.7	**48.6**	48.6	oss_Cyrl	52.3	**53.2**	52.3	oss_Cyrl	46.8	**48.6**	47.7
rus_Cyrl	43.2	44.1	**45.9**	rus_Cyrl	46.8	**48.6**	48.6	rus_Cyrl	**45.0**	45.0	42.3
sah_Cyrl	48.6	**49.5**	48.6	sah_Cyrl	48.6	**56.8**	53.2	sah_Cyrl	45.9	46.8	**50.5**
tat_Cyrl	43.2	**45.0**	43.2	tat_Cyrl	**53.2**	49.5	51.4	tat_Cyrl	47.7	47.7	**50.5**
tgk_Cyrl	45.9	**48.6**	44.1	tgk_Cyrl	**54.1**	50.5	50.5	tgk_Cyrl	**47.7**	47.7	46.8
tyv_Cyrl	36.0	**39.6**	39.6	tyv_Cyrl	45.0	**47.7**	45.0	tyv_Cyrl	**47.7**	45.9	43.2
udm_Cyrl	42.3	**44.1**	43.2	udm_Cyrl	45.0	43.2	**48.6**	udm_Cyrl	43.2	**44.1**	41.4
ukr_Cyrl	**50.5**	49.5	50.5	ukr_Cyrl	49.5	**55.9**	53.2	ukr_Cyrl	48.6	**50.5**	48.6
uzn_Cyrl	43.2	**46.8**	43.2	uzn_Cyrl	48.6	49.5	**50.5**	uzn_Cyrl	43.2	**50.5**	41.4
hin_Deva	**55.0**	45.9	51.4	hin_Deva	47.7	47.7	**50.5**	hin_Deva	46.8	**50.5**	47.7
hne_Deva	**55.9**	55.0	52.3	hne_Deva	**61.3**	58.6	58.6	hne_Deva	**57.7**	55.9	55.9
mai_Deva	**45.0**	45.0	44.1	mai_Deva	52.3	**55.0**	53.2	mai_Deva	49.5	45.9	**51.4**
mar_Deva	**49.5**	44.1	48.6	mar_Deva	48.6	49.5	**51.4**	mar_Deva	**49.5**	42.3	48.6
nep_Deva	**51.4**	45.9	50.5	nep_Deva	57.7	55.9	**58.6**	nep_Deva	**54.1**	45.9	48.6
npi_Deva	**55.9**	50.5	55.9	npi_Deva	55.0	54.1	**57.7**	npi_Deva	**59.5**	49.5	52.3
suz_Deva	42.3	**45.0**	42.3	suz_Deva	47.7	46.8	**49.5**	suz_Deva	45.0	**47.7**	47.7
mdy_Ethi	46.8	**48.6**	47.7	mdy_Ethi	45.9	**49.5**	47.7	mdy_Ethi	45.0	**45.9**	42.3
tir_Ethi	37.8	35.1	**38.7**	tir_Ethi	41.4	**42.3**	38.7	tir_Ethi	31.5	**34.2**	28.8
kat_Geor	43.2	42.3	**45.0**	kat_Geor	45.0	**50.5**	46.8	kat_Geor	43.2	45.9	**49.5**
ell_Grek	44.1	44.1	**45.0**	ell_Grek	48.6	**52.3**	46.8	ell_Grek	**49.5**	48.6	45.9
guj_Gujr	**45.9**	45.9	45.0	guj_Gujr	47.7	**55.9**	52.3	guj_Gujr	**51.4**	44.1	49.5
pan_Guru	44.1	40.5	**45.0**	pan_Guru	**46.8**	42.3	44.1	pan_Guru	**46.8**	41.4	45.9
kor_Hang	48.6	**49.5**	49.5	kor_Hang	51.4	**55.9**	53.2	kor_Hang	52.3	**56.8**	53.2
cmn_Hani	44.1	40.5	**49.5**	cmn_Hani	**54.1**	49.5	54.1	cmn_Hani	54.1	43.2	**55.0**
lzh_Hani	51.4	**55.9**	53.2	lzh_Hani	55.9	48.6	**56.8**	lzh_Hani	53.2	49.5	**56.8**
yue_Hani	45.9	43.2	**52.3**	yue_Hani	**54.1**	41.4	51.4	yue_Hani	**53.2**	48.6	52.3
khm_Khmr	52.3	**55.9**	54.1	khm_Khmr	55.9	56.8	**59.5**	khm_Khmr	52.3	**53.2**	52.3
lao_Laoo	47.7	**51.4**	49.5	lao_Laoo	51.4	48.6	**53.2**	lao_Laoo	**56.8**	56.8	56.8
ksw_Mymr	39.6	**40.5**	37.8	ksw_Mymr	**49.5**	43.2	39.6	ksw_Mymr	**42.3**	40.5	40.5
mya_Mymr	**51.4**	47.7	48.6	mya_Mymr	**53.2**	50.5	51.4	mya_Mymr	41.4	41.4	**43.2**
ori_Orya	**51.4**	51.4	47.7	ori_Orya	**55.9**	51.4	49.5	ori_Orya	**54.1**	45.0	52.3
ory_Orya	**53.2**	49.5	52.3	ory_Orya	51.4	49.5	**55.9**	ory_Orya	**59.5**	52.3	58.6
sin_Sinh	41.4	43.2	**45.0**	sin_Sinh	46.8	**51.4**	45.0	sin_Sinh	42.3	**45.0**	44.1
tam_Taml	55.0	**56.8**	55.9	tam_Taml	55.0	54.1	**61.3**	tam_Taml	**60.4**	55.0	57.7
tel_Telu	38.7	36.0	**41.4**	tel_Telu	**52.3**	46.8	48.6	tel_Telu	**51.4**	41.4	49.5
tha_Thai	45.0	45.0	**45.9**	tha_Thai	46.8	46.8	**48.6**	tha_Thai	41.4	39.6	**42.3**
dzo_Tibt	42.3	41.4	**45.9**	dzo_Tibt	41.4	40.5	**44.1**	dzo_Tibt	**43.2**	43.2	39.6

Table 8: Accuracy of Taxi1500 task on BLOOM 560m, BLOOM 1B and BLOOM3B (from left to right).

Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$	Language	SCRIPT$_{\{Orig\}}$	SCRIPT$_{\{Latn\}}$	SCRIPT$_{\{Combined\}}$
arb_Arab	49.5	45.9	**51.4**	arb_Arab	43.2	**45.9**	45.9	arb_Arab	**62.2**	48.6	61.3
ary_Arab	**38.7**	30.6	38.7	ary_Arab	32.4	**35.1**	29.7	ary_Arab	**55.9**	38.7	50.5
arz_Arab	**45.0**	35.1	41.4	arz_Arab	31.5	**39.6**	34.2	arz_Arab	54.1	47.7	**58.6**
azb_Arab	47.7	43.2	**48.6**	azb_Arab	39.6	**42.3**	42.3	azb_Arab	51.4	42.3	**55.9**
ckb_Arab	45.0	**47.7**	42.3	ckb_Arab	44.1	**45.0**	42.3	ckb_Arab	**47.7**	44.1	46.8
fas_Arab	**57.7**	51.4	49.5	fas_Arab	49.5	50.5	**53.2**	fas_Arab	**66.7**	46.8	63.1
pes_Arab	**59.1**	51.8	56.4	pes_Arab	50.9	49.1	**56.4**	pes_Arab	**64.5**	49.1	62.7
prs_Arab	**55.9**	55.9	52.3	prs_Arab	50.5	55.0	**55.9**	prs_Arab	**65.8**	59.5	64.0
snd_Arab	**56.8**	50.5	52.3	snd_Arab	44.1	**46.8**	45.0	snd_Arab	**62.2**	51.4	62.2
hye_Armn	45.9	**46.8**	45.9	hye_Armn	45.9	50.5	**53.2**	hye_Armn	**55.0**	53.2	53.2
asm_Beng	55.0	43.2	**55.9**	asm_Beng	45.9	44.1	**49.5**	asm_Beng	**55.9**	50.5	53.2
ben_Beng	**52.3**	45.0	52.3	ben_Beng	40.5	**45.9**	45.0	ben_Beng	**57.7**	48.6	56.8
alt_Cyrl	45.0	**46.8**	44.1	alt_Cyrl	44.1	43.2	**48.6**	alt_Cyrl	45.9	**48.6**	45.9
bak_Cyrl	49.5	49.5	**50.5**	bak_Cyrl	45.0	**47.7**	46.8	bak_Cyrl	48.6	**52.3**	47.7
bel_Cyrl	**48.6**	39.6	45.9	bel_Cyrl	**47.7**	44.1	42.3	bel_Cyrl	55.9	**58.6**	58.6
bul_Cyrl	**48.6**	45.0	43.2	bul_Cyrl	**45.0**	45.0	44.1	bul_Cyrl	61.3	57.7	**64.0**
che_Cyrl	**36.9**	36.9	35.1	che_Cyrl	37.8	**42.3**	41.4	che_Cyrl	**42.3**	36.9	40.5
chv_Cyrl	45.0	**45.9**	45.0	chv_Cyrl	43.2	43.2	**44.1**	chv_Cyrl	45.0	50.5	**51.4**
crh_Cyrl	47.7	49.5	**51.4**	crh_Cyrl	**49.5**	47.7	49.5	crh_Cyrl	56.8	**59.5**	55.9
kaz_Cyrl	**51.4**	50.5	48.6	kaz_Cyrl	49.5	**53.2**	47.7	kaz_Cyrl	**55.0**	53.2	51.4
kir_Cyrl	47.7	**53.2**	46.8	kir_Cyrl	51.4	53.2	**56.8**	kir_Cyrl	53.2	57.7	**60.4**
kjh_Cyrl	**45.0**	43.2	41.4	kjh_Cyrl	**44.1**	42.3	43.2	kjh_Cyrl	47.7	49.5	**51.4**
kmr_Cyrl	45.0	**46.8**	40.5	kmr_Cyrl	39.6	**40.5**	38.7	kmr_Cyrl	39.6	**41.4**	39.6
krc_Cyrl	48.6	47.7	**49.5**	krc_Cyrl	**45.9**	44.1	45.0	krc_Cyrl	**55.0**	52.3	50.5
mhr_Cyrl	45.5	45.5	**46.4**	mhr_Cyrl	47.3	**51.8**	50.0	mhr_Cyrl	45.5	46.4	**50.0**
mkd_Cyrl	**56.8**	55.0	55.9	mkd_Cyrl	52.3	51.4	**53.2**	mkd_Cyrl	66.7	**72.1**	67.6
myv_Cyrl	40.5	**44.1**	38.7	myv_Cyrl	39.6	36.9	**41.4**	myv_Cyrl	45.0	**47.7**	45.9
oss_Cyrl	**49.5**	45.9	45.0	oss_Cyrl	**49.5**	45.9	48.6	oss_Cyrl	47.7	**49.5**	45.0
rus_Cyrl	50.5	**52.3**	50.5	rus_Cyrl	**49.5**	47.7	48.6	rus_Cyrl	57.7	**64.9**	64.0
sah_Cyrl	**44.1**	43.2	40.5	sah_Cyrl	40.5	**41.4**	41.4	sah_Cyrl	45.9	44.1	**47.7**
tat_Cyrl	45.9	**46.8**	45.9	tat_Cyrl	47.7	**50.5**	47.7	tat_Cyrl	**53.2**	47.7	51.4
tgk_Cyrl	48.6	**49.5**	48.6	tgk_Cyrl	42.3	44.1	**46.8**	tgk_Cyrl	55.9	**58.6**	54.1
tyv_Cyrl	43.2	44.1	**45.9**	tyv_Cyrl	38.7	**45.0**	43.2	tyv_Cyrl	**47.7**	46.8	46.8
udm_Cyrl	42.3	**45.0**	41.4	udm_Cyrl	36.9	**40.5**	38.7	udm_Cyrl	41.4	**47.7**	43.2
ukr_Cyrl	**51.4**	49.5	48.6	ukr_Cyrl	**52.3**	50.5	48.6	ukr_Cyrl	63.1	**64.0**	62.2
uzn_Cyrl	45.0	**51.4**	42.3	uzn_Cyrl	**45.9**	43.2	44.1	uzn_Cyrl	**59.5**	55.9	55.9
hin_Deva	49.5	44.1	**50.5**	hin_Deva	51.4	53.2	**54.1**	hin_Deva	**64.9**	59.5	64.0
hne_Deva	54.1	52.3	**56.8**	hne_Deva	55.9	**56.8**	55.0	hne_Deva	**61.3**	57.7	61.3
mai_Deva	**49.5**	46.8	48.6	mai_Deva	45.0	**51.4**	47.7	mai_Deva	**62.2**	51.4	58.6
mar_Deva	**53.2**	40.5	53.2	mar_Deva	49.5	**51.4**	51.4	mar_Deva	55.9	54.1	**59.5**
nep_Deva	**63.1**	49.5	57.7	nep_Deva	45.0	45.9	**46.8**	nep_Deva	**66.7**	61.3	64.9
npi_Deva	55.9	49.5	**62.2**	npi_Deva	**51.4**	50.5	51.4	npi_Deva	**66.7**	60.4	65.8
suz_Deva	**42.3**	42.3	41.4	suz_Deva	46.8	**48.6**	44.1	suz_Deva	43.2	**49.5**	48.6
mdy_Ethi	**43.2**	38.7	42.3	mdy_Ethi	39.6	**45.9**	44.1	mdy_Ethi	55.0	52.3	**57.7**
tir_Ethi	27.9	**31.5**	30.6	tir_Ethi	29.7	**36.9**	36.9	tir_Ethi	**39.6**	29.7	36.0
kat_Geor	**42.3**	41.4	41.4	kat_Geor	41.4	**44.1**	41.4	kat_Geor	45.0	**46.8**	45.9
ell_Grek	**49.5**	43.2	43.2	ell_Grek	49.5	43.2	**52.3**	ell_Grek	57.7	**62.2**	59.5
guj_Gujr	**52.3**	45.0	52.3	guj_Gujr	45.9	43.2	**49.5**	guj_Gujr	52.3	**55.9**	55.0
pan_Guru	46.8	39.6	**48.6**	pan_Guru	41.4	45.0	**47.7**	pan_Guru	45.9	**50.5**	49.5
kor_Hang	49.5	48.6	**51.4**	kor_Hang	48.6	50.5	**55.9**	kor_Hang	**72.1**	50.5	69.4
cmn_Hani	**53.2**	45.0	50.5	cmn_Hani	**48.6**	45.0	48.6	cmn_Hani	61.3	50.5	**64.0**
lzh_Hani	**54.1**	45.0	52.3	lzh_Hani	**55.0**	48.6	52.3	lzh_Hani	**65.8**	51.4	59.5
yue_Hani	**53.2**	45.0	50.5	yue_Hani	43.2	52.3	**53.2**	yue_Hani	63.1	48.6	**65.8**
khm_Khmr	48.6	47.7	**54.1**	khm_Khmr	52.3	**53.2**	53.2	khm_Khmr	**55.9**	50.5	53.2
lao_Laoo	46.8	**49.5**	46.8	lao_Laoo	45.0	49.5	**51.4**	lao_Laoo	45.0	**46.8**	46.8
ksw_Mymr	**42.3**	40.5	40.5	ksw_Mymr	44.1	**47.7**	45.0	ksw_Mymr	44.1	48.6	**49.5**
mya_Mymr	44.1	**47.7**	43.2	mya_Mymr	45.0	**51.4**	47.7	mya_Mymr	**51.4**	45.9	49.5
ori_Orya	**51.4**	46.8	50.5	ori_Orya	43.2	43.2	**44.1**	ori_Orya	50.5	**58.6**	47.7
ory_Orya	49.5	47.7	**54.1**	ory_Orya	44.1	48.6	**52.3**	ory_Orya	**57.7**	55.0	57.7
sin_Sinh	39.6	**41.4**	40.5	sin_Sinh	39.6	**52.3**	38.7	sin_Sinh	37.8	**49.5**	45.0
tam_Taml	**59.5**	51.4	58.6	tam_Taml	44.1	**49.5**	45.9	tam_Taml	**60.4**	50.5	55.9
tel_Telu	50.5	40.5	**53.2**	tel_Telu	33.3	**43.2**	36.9	tel_Telu	**54.1**	41.4	52.3
tha_Thai	**43.2**	43.2	39.6	tha_Thai	**43.2**	40.5	39.6	tha_Thai	**57.7**	43.2	52.3
dzo_Tibt	41.4	**44.1**	41.4	dzo_Tibt	45.0	**49.5**	47.7	dzo_Tibt	**45.0**	44.1	43.2

Table 9: Accuracy of Taxi1500 task on BLOOM 7B, LLaMA2-7B and Mixtral 7B (from left to right).