MILU: A Multi-task Indic Language Understanding Benchmark

Sshubam Verma[1] **Mohammed Safi Ur Rahman Khan**[1,2]
Vishwajeet Kumar[3] **Rudra Murthy**[3] **Jaydeep Sen**[3]
[1]Nilekani Centre at AI4Bharat [2]Indian Institute of Technology, Madras
[3]IBM Research, India
Correspondence: {sshubamverma, safikhan}@ai4bharat.org, {vishk024, rmurthyv}@in.ibm.com

https://huggingface.co/datasets/ai4bharat/MILU

https://github.com/AI4Bharat/MILU

Abstract

Evaluating Large Language Models (LLMs) in low-resource and linguistically diverse languages remains a significant challenge in NLP, particularly for languages using non-Latin scripts like those spoken in India. Existing benchmarks predominantly focus on English, leaving substantial gaps in assessing LLM capabilities in these languages. We introduce MILU—**M**ulti-task **I**ndic **L**anguage **U**nderstanding Benchmark—a comprehensive evaluation benchmark designed to address this gap. MILU spans 8 domains and 41 subjects across 11 Indic languages, reflecting both general and culturally specific knowledge. With an India-centric design, MILU incorporates material from regional and state-level examinations, covering topics such as local history, arts, festivals, and laws, alongside standard subjects like science. We evaluate over 42 LLMs, and find that current LLMs struggle with MILU, with GPT-4o achieving the highest average accuracy at 74%. Open multilingual models outperform language-specific fine-tuned models, which perform only slightly better than random baselines. Models also perform better in high-resource languages as compared to low-resource ones. Domain-wise analysis indicates that models perform poorly in culturally relevant areas like Arts & Humanities and Law & Governance compared to general fields like STEM. To the best of our knowledge, MILU is the first of its kind benchmark focused on Indic languages, serving as a crucial step towards comprehensive cultural evaluation. All code, benchmarks, and artifacts will be made publicly available to foster open research.

1 Introduction

Recent advancements in Large Language Models (LLMs) have reshaped the field of NLP, by enabling these models to perform a variety of tasks across diverse domains (Doddapaneni et al., 2023; OpenAI et al., 2023; Team et al., 2024a; Anthropic).

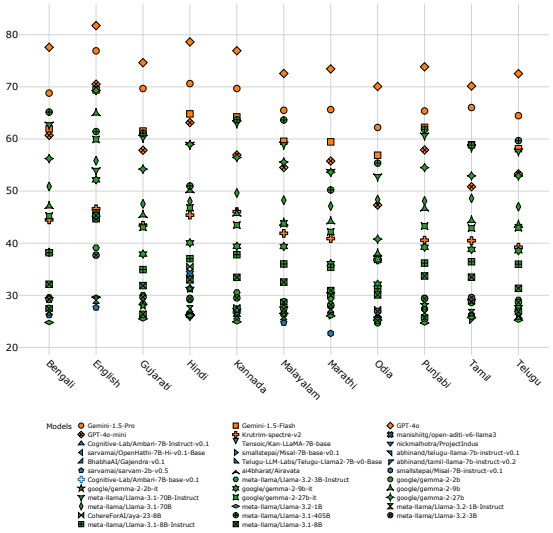


Figure 1: Average performance of all the evaluated models on MILU. The closed models are shown in Orange, the open models are shown in Green, and the language-specific models are shown in Blue.

While many LLMs now claim to support multiple languages, there is still a huge discrepancy in their performance in English and other languages (Liu et al., 2024). Particularly languages using non-Latin scripts, such as those in India, are affected the most by this discrepancy (Ahuja et al., 2023). One key reason for this is the absence of high-quality benchmarks for these languages. Well-designed benchmarks are crucial in driving model development by revealing limitations and guiding improvements (McIntosh et al., 2024). However, most existing benchmarks focus primarily on English, *leaving significant gaps in evaluating LLMs capability in low-resource and linguistically diverse languages.*

India's diversity presents a unique challenge for these models. With over 1.4 billion people speaking more than 120 languages and around 19,500 dialects across 28 states (Javed et al., 2024), many

Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 10076–10132

of which are underrepresented in NLP research, the need for not just linguistic but culturally appropriate benchmarks becomes urgent (Doddapaneni et al., 2023; Singh et al., 2024b). Standard benchmarks like MMLU (Hendrycks et al., 2021) and AGIEVAL (Zhong et al., 2023), while useful for evaluating general world knowledge, falls short in capturing these intricacies of India. Each state in India has its own history, traditions, festivals, and art forms, forming a rich cultural mosaic. Translating existing English benchmarks into Indian languages fails to capture this knowledge, which is required for real-world applications. Given the increasing deployment of LLMs in tasks that directly impact local populations, *the need for a benchmark that evaluates both linguistic competence and cultural understanding has become more pressing than ever.*

In this work, we introduce MILU-**M**ulti-task **I**ndic **L**anguage **U**nderstanding Benchmark, a comprehensive evaluation dataset designed to address these gaps. MILU spans 8 domains and 41 subjects across 11 Indic languages, reflecting both general and culturally specific knowledge. We designed MILU with an India-first perspective by collecting questions from various national, state, and regional exams. These questions include culturally relevant subjects such as local history, arts, festivals, and laws, alongside traditional academic subjects like science. Following previous efforts (Hendrycks et al., 2021; Zhong et al., 2023), we create this benchmark by collecting questions from over 1500 competitive exams from India. We focus on region-specific exams to authentically capture local knowledge in the respective language.

We evaluate 45 different LLMs - a mix of closed proprietary, open-source, and language-specific models- on MILU. Our findings suggest that models struggle with MILU, with GPT-4o achieving the highest average accuracy at 74%. Interestingly, open multilingual models outperform language-specific models, which only achieve slightly better than random scores. Our analysis of in-context learning reveals that adding more examples improves performance in base models, but the effect on instruct models remains inconclusive. We also explore how performance scales with the number of parameters, finding significant improvements as model size increases. Our domain-wise analysis reveals that models perform poorly in culturally relevant areas, such as Arts & Humanities and Social Sciences, compared to more general fields like STEM. All the artifacts will be released publicly.

2 Related Works

Large Language Models (LLMs): Recent LLMs, both proprietary-GPT-4o, GPT-4o-mini, Claude-3, and Gemini (Team et al., 2024a)-and open-source-the Llama-3 (Dubey et al., 2024) and Gemma (Team et al., 2024b) series-have demonstrated significant improvements across various tasks and benchmarks (Chiang et al., 2024; Wang et al., 2024). While these models are primarily trained in English, many claim to include a reasonable amount of multilingual data in their pretraining corpus (Team et al., 2024b; Dubey et al., 2024; Aryabumi et al., 2024). Additionally, significant progress has been made in developing language-specific models (Ustun et al., 2024; Nguyen et al., 2023), including for Indic languages (Gala et al., 2024; Balachandran, 2023; Kohli et al., 2023). Most of these models are built on top of stronger English base models by (optionally) either continually pretraining on smaller, language-specific datasets (Zhao et al., 2024), or by language-specific instruction fine-tuning (Ghosh et al., 2024). In this work, we conduct a comprehensive evaluation of the performance of these models on MILU.

LLM Evaluation Benchmarks: Over the years, various benchmarks have been developed to evaluate the performance of large language models (LLMs). Recent benchmarks such as MMLU (Hendrycks et al., 2021), MMLUPRO (Wang et al., 2024), AGIEVAL (Zhong et al., 2023), BIGGEN-BENCH (Kim et al., 2024), and HELLASWAG (Zellers et al., 2019) assess these models across a wide range of tasks. However, these benchmarks primarily focus on English, and progress on multilingual benchmarks has been comparatively slower (Doddapaneni et al., 2023; Kumar et al., 2022). Some popular multilingual benchmarks include OMGEVAL (Liu et al., 2024), XTREME (Hu et al., 2020), and XQUAD (Artetxe et al., 2020), though these are largely limited to simple natural language understanding (NLU) tasks. As a result, previous studies often rely on translations of popular English benchmarks to evaluate performance in other languages (Dubey et al., 2024; Gala et al., 2024). This approach is suboptimal, as it fails to account for cultural nuances and concepts unique to specific languages.

Indic LLM Evaluation Benchmarks: The first major Indic language benchmarks, INDICGLUE (Kakwani et al., 2020) and INDICNLG-

BENCHMARK (Kumar et al., 2022), cover 11 languages, focusing on various language understanding and generation tasks, respectively. INDICXTREME (Doddapaneni et al., 2023) extends these efforts to include all 22 scheduled Indian languages for NLU evaluation. More recently, INDICGENBENCH (Singh et al., 2024b) offers a more comprehensive evaluation of multilingual language generation by consolidating multiple generation tasks. In addition, the INDIC-QA BENCHMARK (Singh et al., 2024a) assesses context-based question-answering (QA) performance in 11 Indian languages, while L3CUBE-INDICQUEST (Rohera et al., 2024) explores regional knowledge through a translation-based approach to data creation. More recently, PARIKSHA (Watts et al., 2024) evaluated more than 25 Indic models and created dynamic leaderboards by collecting more than 90,000 human preferences. Other initiatives, such as AIRAVATA (Gala et al., 2024) and the INDIC-LLM-LEADERBOARD (Kolavi, 2024), focus on translating existing English benchmarks into Indian languages. In contrast, MILU addresses this limitation by centering on India-specific topics and questions, providing a more culturally relevant and context-aware evaluation.

MMLU	MILU
Which principle was established by the Supreme Court's decision in Marbury v. Madison?	Which of the decisions was taken by Maharashtra government on 24th November 2001 that had a far-reaching impact on the education field?
At breakfast, lunch, and dinner, Joe randomly chooses with equal probabilities either an apple, an orange, or a banana to eat. On a given day, what is the probability that Joe will eat at least two different kinds of fruit?	Shyam's monthly income is Rs. 12,000. He saves Rs. 1200. Find the percent of his savings and his expenditure.
The Space Shuttle orbits 300 km above Earth's surface; Earth's radius is 6,400 km. What is the gravitational acceleration experienced by the Space Shuttle?	FASTag at the Toll Gates uses which waves?
The primary goal of the Gramm-Rudman Acts of 1985 and 1987 was to	Under the Pradhan Mantri Kisan Samman Nidhi Scheme, the financial help provided to the small and marginal farmers is _____ .
A guitar string creates a sound wave of known frequency. Which of the following describes a correct and practical method of measuring the wavelength of the sound wave with a meterstick?	In a Sitar, which type of sound vibrations are produced?

Table 1: Comparison of some question from MMLU (Hendrycks et al., 2021) with similar questions from MILU. The Red highlights the global concepts covered in MMLU, while Green highlights the India-centric concepts covered in MILU.

3 MILU: The IndicMMLU Benchmark

In this section, we describe the collection process (§3.1), the cleaning and filtering process (§3.2), and present the analysis (§3.3).

3.1 Questions Curation

MILU is a large, multi-domain test set containing multiple-choice based questions (MCQs) taken from over 41 subjects with an emphasis on India-specific knowledge. This benchmark covers many domains, including Science, Social Sciences, Humanities, Arts, Business Studies, and Law, among others. MILU is designed as a culturally relevant benchmark to assess general problem-solving abilities and India-specific knowledge. These questions were sourced following an approach similar to AGIEVAL (Zhong et al., 2023), collecting the questions from various public exams taken by individuals intending to either pursue higher studies or seek career advancements, such as qualification tests and national and state-level civil services exams, among others.

We gathered exam-specific questions by scraping various online exam portals that offer previously released question papers from various exams in multiple different languages. These portals typically tag questions manually with topic names and language details, and subject experts ensure the accuracy of the answers. Our benchmark includes questions from over 40 different types of exams conducted both at the national and state levels over recent years. Regional state exams are particularly valuable as they cover various state-level topics and emphasize the official language of each state. Further details about the different exams included in our collection are provided in the Appendix A.

In total, we collected more than 150K questions across 11 Indian Languages- Bengali (*bn*), Gujarati (*gu*), Hindi (*hi*), Kannada (*kn*), Malayalam (*ml*), Marathi (*mr*), Odia (*or*), Punjabi (*pn*), Tamil (*ta*), Telugu (*te*), and English (*en*)-spanning 41 diverse subjects. English questions are also included as these often address Indian culture-specific content, which is notably missing from existing popular benchmarks. Table 1 presents a brief comparison of the type of questions present in MMLU (Hendrycks et al., 2021) and MILU.

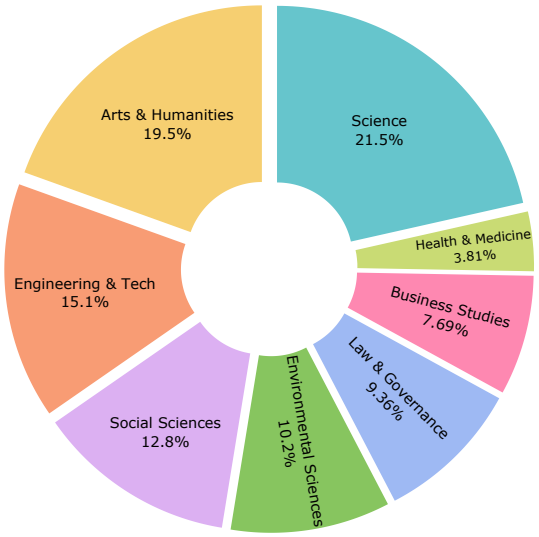


Figure 2: Distribution of the number of questions across different domains, averaged across all languages. Refer to Section (§3.3) for more details.

lang	Total Qs	Total Translated Qs	Avg Words Per Q
bn	6638	1601	15.12
gu	4827	2755	16.12
hi	14837	115	20.61
kn	6234	1522	12.42
ml	4321	3354	12.39
mr	6924	1235	18.76
or	4525	3100	14.96
pa	4099	3411	19.26
ta	6372	1524	13.14
te	7304	1298	15.71
en	13536	-	22.07
total	**79617**	**19915**	**16.41**

Table 2: Overall statistics of MILU. Refer to Section (§3.3) for more details.

3.2 Data Cleaning and Filtering

Despite our best efforts to maintain the quality of questions collected, some amount of noise or errors may still be present. To address potential noise in the questions, we employ multiple layers of manual and automated cleaning filters. Initially, we manually review a large sample of questions to detect and eliminate potential sources of noise. During the collection process, we exclude any reading-comprehension-style questions, images-based questions, and those with more than four answer options to ensure uniformity and consistency. To remove incorrect language entries, we utilize a combination of IndicLID (Madhani et al., 2023) and Unicode-based filtering (Khan et al., 2024), ensuring that the questions are in the correct language. To further refine the dataset, we remove any duplicate questions to retain only the unique ones. As a final step, we manually verify a sample of questions from each language to ensure accuracy and correct any remaining errors.

Upon examination, we found that approximately 45% of questions were accurately labeled with a topic name, while the remaining questions lacked this information. To address this issue, we first translate the untagged questions into English using IndicTrans2 (Gala et al., 2023) and then prompt GPT-4o-mini model to assign an appropriate topic name to the question. Finally, in total, we get around 20K tags. However, these tags are highly fine-grained, often having a heavy overlap. To organize them, we embed the tags using the NV-Embed-v2 (Lee et al., 2024) model and apply K-means clustering to group tags into 50 clusters. We manually review these clusters and assign appropriate subject labels. Following the manual merging of related clusters, we determine 41 distinct subject names, which fall into eight main domains: Arts and Humanities, Social Sciences, Environmental Sciences, Law and Governance, Health and Medicine, Science, Engineering and Technology, and Business Studies. The final distribution of domains within MILU is shown in Figure 2. Detailed analysis of question distribution by topic & language is provided in Appendix B.

Finally, we observed that some topics in certain languages had less than 100 questions. To ensure thorough evaluation across all subjects and languages, we aimed to have at least 100 questions per subject in each language. For subjects with insufficient questions, we sampled questions from the English set from that subject and translated them into the required language using GPT-4o. We chose GPT-4o over specialized translation models for their ability to remain task-aware during translation (Ahuja et al., 2024), ensuring the translated content aligns with the intent of the question.

In total, we release around 79K questions across 41 subjects across 8 domains in 11 languages, capping each subject-language pair at 500 questions for feasible evaluations.

3.3 Data Analysis

Table 2 shows the overall statistics of MILU. Of the total 79K questions, only 25% of questions are translated from English, with the remainder

preserved in their original source languages. The average question length varies across languages, with languages such as *kn*, *ml*, and *ta* having shorter word counts due to their agglutinative nature.

As shown in Figure 2, MILU consists of eight domains.

Arts & Humanities: This domain covers topics such as Indian art, literature, dance, festivals, and architecture. Its content varies significantly by language, as it is highly dependent on regional culture.

Science: Includes topics in physics, chemistry, biology, with references to Ayurveda, and ancient as well as modern scientific achievements.

Health & Medicine: Covers public health policies, modern healthcare, traditional Ayurveda and Unani medicine, and health-related government initiatives.

Business Studies: Focuses on entrepreneurship, trade, and economic policies such as Make in India. It also covers topics like taxation, MSMEs, and global trade.

Law & Governance: Includes topics on the constitution, governance, judiciary, public administration, fundamental rights, and various governmental schemes, including local governance policies.

Environmental Sciences: Covers topics related to biodiversity, environmental policies, and both local and national environmental initiatives.

Social Sciences: Explores topics such as history, geography, and politics. The content in this domain is region-specific and varies by language.

Engineering & Technology: Includes discussions on modern developments in India, such as IT, telecommunications, infrastructure, and space technology, with a focus on government policies.

Figure 2 shows the domain-wise statistics of MILU. We see a fairly large number of questions in the culturally relevant domains of Arts & Humanities and Social Sciences. Detailed statistics per language can be found in the Appendix B.

4 Experimental Setup

We evaluate 42 different models on MILU, including large proprietary models, open-source multilingual models, and popular fine-tuned models specific to Indic languages. Both the base versions and instruction fine-tuned variants of these models, wherever applicable, are evaluated to measure the improvements gained from fine-tuning. All models, except for proprietary models and LLAMA-3.1-405B, are tested under 0-shot, 1-shot, and 5-shot setups. We maintain a separate validation set of approximately 9,000 questions to serve as examples for few-shot evaluations.

For non-API-based models, we use the LM-EVALUATION-HARNESS (Gao et al., 2024; Biderman et al., 2024) to ensure clean and reproducible evaluations. We use the log-likelihood method, where the probability of a given output string is computed by conditioning it on some provided input (Brown et al., 2020). Specifically, the log-likelihood of an answer (a) given the question (x), i.e., $logP(a|x)$, is calculated by concatenating the answer (a) with question (x), and then summing up the log probabilities, of each target token. For multiple choice questions, given k possible answer strings, we select the answer string (a_i) with the highest conditional log probability, i.e., $argmax(logP(a_1|x), ..., logP(a_k|x))$.

The API-based models are evaluated using the generative approach due to the lack of support for prompt log probabilities. We explicitly prompt these models to generate the correct response in a structured JSON format to simplify response parsing. Due to the high costs involved, these models are evaluated only in the zero-shot setup.

5 Results and Discussions

In this section, we discuss the results and our findings across all the experiments conducted.

5.1 How do models supporting all languages perform?

Table 3 shows the performance of various models supporting all the languages in the 5-shot setup. Among these, GPT-4O emerges as the strongest model, consistently outperforming all competitors across languages, with an average accuracy of 74.74%. Closed proprietary models, generally achieve higher performance than the open-source alternatives. Among the open-source models, LLAMA-3.1-70B shows the best performance.

Performance is notably higher in high-resource languages- *en*, *hi*, *bn* -as compared to the low-resource ones. Interestingly, the models also show low accuracies in agglutinative languages, such as *ta*, *te*. Notably, models specifically trained for Indian languages such as KRUTRIM-SPECTRE-V2 and SARVAM-2B-V0.5, perform slightly below their English counterparts of comparable scale. Additionally, while several models achieve better results in English, their performance in other lan-

Model	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*	*avg*
Closed Models												
GPT-4o	77.59	81.75	74.64	78.62	76.93	72.57	73.44	70.07	73.84	70.15	72.53	74.74
GPT-4o-mini	60.69	70.52	57.84	63.14	56.92	54.52	55.76	47.31	57.89	50.84	53.31	57.16
Gemini-1.5-Pro	68.8	76.91	69.68	70.63	69.68	65.5	65.63	62.21	65.37	66.03	64.47	67.72
Gemini-1.5-Flash	61.93	69.42	61.52	64.81	64.22	59.56	59.44	56.88	62.23	58.89	58.13	61.55
Krutrim-spectre-v2	44.48	46.59	43.43	45.39	46.05	41.89	40.89	36.8	40.57	40.48	39.22	42.34
Open Multilingual Models												
$<$ 4B Models												
meta-llama/Llama-3.2-1B	25.37	34.35	25.46	27.01	25.51	25.27	26.5	25.02	24.52	25.56	25.79	26.4
meta-llama/Llama-3.2-1B-Instruct	25.17	25.84	25.19	25.24	25.91	25.02	25.82	24.4	24.74	26.8	26.11	25.48
sarvamai/sarvam-2b-v0.5	29.83	34.73	29.5	31.37	29.03	25.94	28.96	27.09	29.06	27.68	28.57	29.25
sarvamai/sarvam-1	33.31	30.3	29.83	33.05	33.29	30.8	27.9	29.33	28.76	27.64	28.5	30.25
google/gemma-2-2b	29.72	50.86	28.82	32.32	29.77	28.72	30.99	25.28	30.64	30.1	28.46	31.43
google/gemma-2-2b-it	30.6	52.4	29.25	34.33	28.52	28.81	32.18	26.32	29.01	30.46	29.22	31.92
meta-llama/Llama-3.2-3B	32.75	50.25	31.32	34.63	31.75	29.07	32.41	28.07	30.42	30.93	29.75	32.85
meta-llama/Llama-3.2-3B-Instruct	32.9	31.74	32.82	30.55	31.83	30.36	28.54	26.48	32.35	31.32	31.19	30.92
nvidia/Nemotron-4-Mini-Hindi-4B-Base	31.41	51.37	25.85	51.42	27.4	27.01	39.37	26.9	26.47	27.89	28.25	33.03
7B to 27B Models												
Telugu-LLM-Labs/Navarasa-2.0	36.43	51.69	37.44	39.86	40.01	36.40	37.00	35.36	36.86	37.18	36.86	38.64
neulab/Pangea-7B	42.06	56.41	36.81	42.21	33.83	33.86	36.73	32.18	36.45	32.72	36.43	38.15
CohereForAI/aya-23-8B	27.85	45.76	26.93	36.31	27.53	28.4	30.96	27.51	26.98	28.88	26.34	30.31
meta-llama/Llama-3.1-8B	37.89	59.66	35.26	42.69	37.09	34.71	37.65	32.86	36.33	36.25	34.94	38.67
meta-llama/Llama-3.1-8B-Instruct	34.53	46.22	33.77	35.73	36.04	34.37	33.28	30.08	33.33	34.15	33.6	35.01
google/gemma-2-9b	53.12	69.8	51.19	56.24	53.99	49.55	49.81	40.42	49.7	50.05	48.78	52.06
google/gemma-2-9b-it	46.04	63.61	38.82	44.82	43.47	40.8	37.62	34.19	40.42	38.83	39.44	42.55
google/gemma-2-27b	63.47	74.82	59.29	64.33	63.43	59.62	59.66	44.86	59.94	58.6	57.94	60.54
google/gemma-2-27b-it	59.82	73.7	55.17	61.25	58.23	53.37	54.78	44.99	55.55	55.6	54.29	56.98
$>$ 27B Models												
CohereForAI/aya-23-35B	32.9	53.09	30.35	43.53	31.39	32.42	37.22	31.27	27.27	35	29.34	34.89
meta-llama/Llama-3.1-70B	67.37	74.59	62.42	66.97	68.37	62.69	62.2	57.48	64.6	60.7	60.64	64.37
meta-llama/Llama-3.1-70B-Instruct	62.77	54.00	60.26	58.90	63.02	58.94	53.71	52.81	60.67	58.41	57.66	58.29
meta-llama/Llama-3.1-405B	65.15	61.42	61.11	50.98	63.69	63.64	50.21	55.35	61.84	58.81	59.67	59.26

Table 3: Evaluation results of all the Closed and Open Multilingual models supporting all languages on MILU. We report 5-shot accuracies for all open models (except for LLAMA-3.1-70B-INSTRUCT & LLAMA-3.1-405B for which we report 0-shot accuracy) with the accuracy averaged across all the domains per language. Higher values indicate better model performance for the given language. Refer Section (§5.1) for more details.

guages is often nearly at the random baseline.

5.2 How do language-specific fine-tuned models perform?

We evaluate around 16 Indic language LLMs on MILU. These models are primarily built by adapting English LLMs, such as LLAMA-2-7B, by first continually pretraining on small amount Indic language data, followed by optionally instruction finetuning them. As seen from Table 4, across languages, the models exhibit average performance comparable to random baselines, with minimal variations among them.

Among the evaluated models, the ARYABHATTA-GEMMAGENZ-VIKAS model stands out, performing better than its counterparts. When analyzed across domains, the models generally perform worse in Arts, Humanities, and Social Sciences than in STEM subjects.

5.3 How do the models respond to In-Context examples?

We compare the performance of different Base and Instruct models across zero, one, and five-shot setups. As shown in Figure 3, the performance of base models consistently improves with an increasing number of in-context examples, with the 5-shot setup yielding the best results. In contrast, Instruct models exhibit more varied behavior, where models either stagnate or even degrade in performance. This also aligns with expectations, as Instruct models are specifically fine-tuned as conversation assistants and may not respond well to the few-shot in-context examples format. This also correlates with results in Table 3 where most instruction fine-tuned variants performs worse on English itself. This essentially implies the inferior performance is less about the language shift but due to the change of tasks during instruction finetuning and inference

lang	Model	Business Studies	Engg. & Tech	Social Sciences	Env. Sciences	Law & Governance	Science	Arts & Humanities	Health & Medicine	Avg
hi	OpenHathi-7B-Hi	29.72	28.59	28.81	27.74	30.96	28.53	28.28	30.76	29.17
	Airavata	28.03	28.45	28.92	27.14	27.59	28.95	27.54	32.77	28.67
	ProjectIndus	26.21	25.83	26.95	25.03	22.93	24.91	25.13	24.41	25.18
	Gajendra-v0.1	29.30	28.36	25.54	29.30	27.75	29.18	28.38	30.26	28.51
	open-aditi-v6-llama3	29.30	31.65	29.99	30.15	34.78	32.57	32.43	37.12	32.25
	AryaBhatta-GemmaGenZ-Merged	**39.42**	**37.88**	**37.59**	**41.14**	**39.60**	**41.25**	**35.71**	**48.49**	**40.14**
te	telugu-llama-7b-instruct-v0.1	**30.19**	**27.20**	**26.27**	24.03	27.16	26.41	26.81	24.63	**26.59**
	Telugu-Llama2-7B-v0-Base	22.96	27.10	24.80	**27.01**	**26.14**	**27.13**	25.42	27.53	26.01
	Telugu-Llama2-7B-v0-Instruct	26.03	24.00	24.45	23.68	24.48	26.82	**27.17**	**28.01**	25.58
ka	Kan-LLaMA-7B-base	29.00	27.42	**30.37**	27.42	**28.47**	26.92	**29.28**	30.50	28.67
	Ambari-7B-base-v0.1	**31.49**	27.52	29.08	**28.70**	27.15	**27.82**	27.93	**31.00**	**28.84**
	Ambari-7B-Instruct-v0.1	28.17	**28.02**	24.83	27.78	23.84	27.56	27.42	27.00	26.83
mr	Misal-7B-base-v0.1	**26.96**	**26.65**	**25.02**	**27.19**	**27.28**	**26.11**	**26.82**	**24.05**	**26.26**
	Misal-7B-instruct-v0.1	25.05	21.44	21.55	22.89	21.82	24.59	23.36	22.16	22.86
ta	tamil-llama-7b-instruct-v0.2	22.22	24.12	25.32	24.89	28.69	24.98	25.07	22.79	24.76
ml	malayalam-llama-7b-instruct-v0.1	30.00	27.92	23.42	27.88	25.00	27.57	23.33	20.00	25.64

Table 4: Evaluation results of all the language-specific fine-tuned models on MILU. We report domain level 5-shot accuracies for all the models on the language supported by the model. Higher values indicate better model performance for the given domain. Refer to Section (§5.2) for more details.

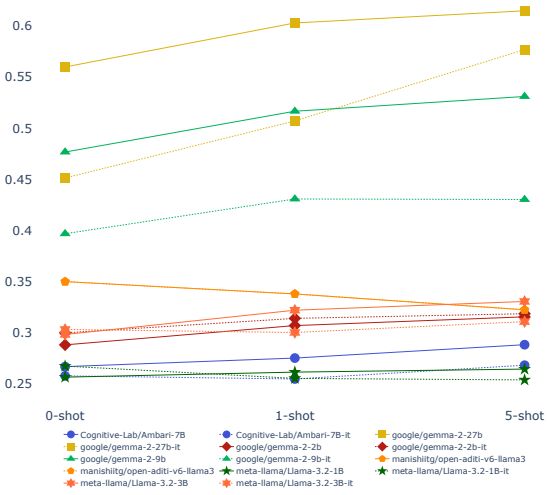


Figure 3: Comparison of Base and Instruct models averaged across all languages for varying number of in-context examples. We plot the average accuracies of the GEMMA and LLAMA series of models, highlighting the performance trend as the number of in-context examples increases. Refer to Section (§5.3) for more details.

task for MILUevaluation. Understandably, the instruction finetuned variants performs poor on the overall average across languages as well.

5.4 How does the performance vary with scale?

We evaluate the LLAMA and GEMMA family of models, ranging from 1B to 405B parameters, to analyze how performance scales with model size. Figure 5 shows that the model performance improves significantly with increasing scale. Notably, instruction-tuned models in the LLAMA family show more substantial improvements as compared to those in the GEMMA family.

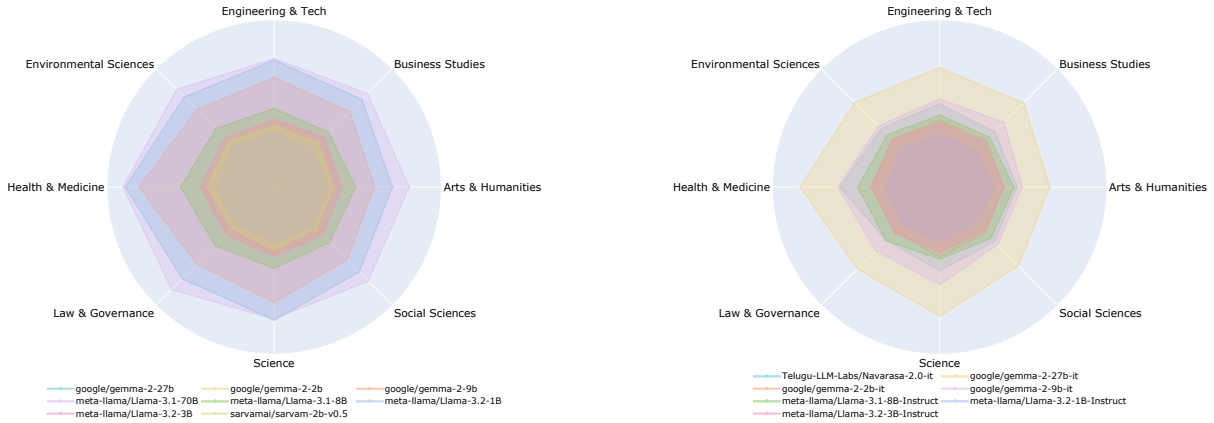
5.5 How do the models perform on different domains?

We analyze the performance of various base and instruct models across multiple domains and languages. Similar trends to those in Section (§5.2) are observed where the open models perform poorly in domains specific to Indian culture—such as Arts & Humanities, Social Sciences, and Law & Governance—but demonstrate higher performance in STEM fields. This suggests that the training corpora for these models lack sufficient culturally specific data. Bridging this gap requires a more inclusive data distribution that ensures equitable representation of all cultures and languages.

5.6 Does language adaptation on models help?

As most Indic LLMs are built on English base models like LLAMA-2-7B, we assess the impact of language adaptation on their performance. Table 5 compares language-specific models with the original LLAMA-2-7B, and instruction-tuned models with LLAMA-2-7B-CHAT. Our findings show minimal gains, with some models even underperforming post-adaptation.

Given the varied training data, this is not a di-



(a) Base Models

(b) Instruct Models

Figure 4: Evaluation results of Base models (4a) and Instruct models (4b) on the different domains supported in MILU. The plot shows the average 5-shot accuracies across all languages for various models. Refer to Section (§5.5)

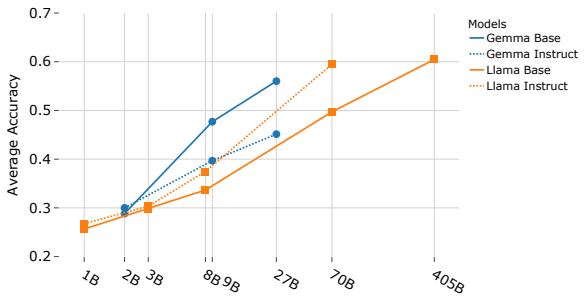


Figure 5: Comparison of performance of GEMMA and LLAMA family of models across different parameter scales. We plot the zero-shot average accuracies of all models across languages. Refer to Section (§5.4) for more details.

lang	Model	CPT	IFT	0	1	5
hi	OpenHathi	Y	N	-0.28%	-0.25%	0.58%
	Airavata	Y	Y	**1.70%**	**2.40%**	3.11%
	Gajendra-v0.1	Y	Y	0.73%	0.92%	**3.22%**
kn	Ambari-base	Y	N	**0.72%**	1.17%	2.46%
	Ambari-Instruct	Y	Y	0.60%	0.49%	1.87%
	Kan-7B-base	Y	N	**1.13%**	**0.84%**	**1.30%**
te	telugu-llama-instruct	Y	Y	**1.13%**	**0.84%**	**1.30%**
	Telugu-7B-v0-Base	Y	N	0.32%	-0.52%	0.88%
	Telugu-7B-v0-Instruct	Y	Y	-0.98%	-0.12%	0.29%
mr	Misal-7B-base-v0.1	Y	N	**0.16%**	**-0.60%**	**-0.93%**
	Misal-7B-instruct-v0.1	Y	Y	-2.55%	-2.17%	-2.38%
ta	tamil-llama-instruct	Y	Y	-0.06%	-0.04%	-0.19%
ml	malayalam-llama-instruct	Y	Y	0.77%	1.09%	0.33%

Table 5: Performance differences of Indian language models compared to the LLAMA-2-7B baseline at different in-context example settings. Each value shows the relative improvement or decline (% difference) over the baseline, with a +ve value indicating improvement and a -ve value indicating a drop in the performance. Refer to Section (§5.6) for more details.

rect "apples-to-apples" comparison but highlights key challenges. We conjecture that limited performance gains may result from small language-specific datasets and reliance on parameter-efficient methods like LoRA (Hu et al., 2022). Another contributing factor could be the lack of diversity in instruction fine-tuning datasets. Models like AIRAVATA, which utilize more diverse data (Gala et al., 2024), exhibit noticeably better performance. Further investigation is required to fully understand the limitations and opportunities in this area.

6 Conclusion

In this paper, we introduced MILU—**M**ultilingual **I**ndic **L**anguage **U**nderstanding Benchmark-a comprehensive benchmark specifically designed to evaluate LLMs across 11 Indic languages, spanning diverse domains and culturally relevant subjects. We evaluate 45 different LLMs and find that the majority of LLMs struggle on MILU, with GPT-4o achieving the highest average accuracy. The analysis also shows that models perform significantly better in high-resource languages than low-resource ones, highlighting the need for more robust multilingual strategies. Additionally, the domain-specific analysis indicates that models perform better in general fields such as STEM while facing challenges in culturally relevant subjects like Arts, Humanities, and Law, highlighting the lack of this knowledge in the current models and datasets. To foster open research, we release MILU along-

side all code, and other artifacts. As LLMs continue to become pivotal in modern applications, we hope that MILU offers a foundational benchmark for developing more inclusive, culturally aware models that perform well across both general and culturally relevant domains.

Limitations

This work has a few limitations. First, we restricted our study to the top 11 languages due to the lack of readily available questions in low-resource languages, which we aim to address in future work. Second, limited computational resources prevented a thorough evaluation of larger models, such as LLAMA-3.1-70B-INSTRUCT and LLAMA-3.1-405B. Third, the scarcity of questions necessitated translating a portion of the dataset. Finally, our evaluation primarily relies on the log-likelihood approach, which may yield different results compared to other established methods, such as generation-based evaluation and chain-of-thought (CoT) prompting.

Acknowledgements

We would like to thank EkStep Foundation and Nilekani Philanthropies for their generous grant towards building datasets, models, tools and other resources for Indian languages. We are also immensely grateful to the volunteers from the AI4Bharat team for their motivation and meticulous efforts in conducting manual audits.

Ethics

All data described in this work was scraped from publicly available resources. The datasets used in this paper will be made available under permissible licenses. Additionally, the code used for our evaluations will be made publicly available under the MIT License. We only used ChatGPT for assistance purely with the language of the paper, e.g., paraphrasing, spell-checking, or polishing the author's original content, without suggesting new content.

References

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

Sanchit Ahuja, Kumar Tanmay, Hardik Hansrajbhai Chauhan, Barun Patra, Kriti Aggarwal, Luciano Del Corro, Arindam Mitra, Tejas Indulal Dhamecha, Ahmed Awadallah, Monojit Choudhary, et al. 2024. sphinx: Sample efficient multilingual instruction fine-tuning through n-shot guided prompting. *arXiv preprint arXiv:2407.09879*.

Anthropic. Introducing the next generation of claude. https://www.anthropic.com/news/claude-3-family. Accessed: 2024-10-14.

Mikel Artetxe, Sebastian Ruder, and Dani Yogatama. 2020. On the cross-lingual transferability of monolingual representations. *ACL*.

Viraat Aryabumi, John Dang, Dwarak Talupuru, Saurabh Dash, David Cairuz, Hangyu Lin, Bharat Venkitesh, Madeline Smith, Jon Ander Campos, Yi Chern Tan, Kelly Marchisio, Max Bartolo, Sebastian Ruder, Acyr Locatelli, Julia Kreutzer, Nick Frosst, Aidan Gomez, Phil Blunsom, Marzieh Fadaee, Ahmet Üstün, and Sara Hooker. 2024. Aya 23: Open weight releases to further multilingual progress. *arXiv preprint arXiv: 2405.15032*.

Abhinand Balachandran. 2023. Tamil-llama: A new tamil language model based on llama 2. *arXiv preprint arXiv: 2311.05845*.

Stella Biderman, Hailey Schoelkopf, Lintang Sutawika, Leo Gao, Jonathan Tow, Baber Abbasi, Alham Fikri Aji, Pawan Sasanka Ammanamanchi, Sid Black, Jordan Clive, Anthony DiPofi, Julen Etxaniz, Benjamin Fattori, Jessica Zosa Forde, Charles Foster, Mimansa Jaiswal, Wilson Y. Lee, Haonan Li, Charles Lovering, Niklas Muennighoff, Ellie Pavlick, Jason Phang, Aviya Skowron, Samson Tan, Xiangru Tang, Kevin A. Wang, Genta Indra Winata, Franccois Yvon, and Andy Zou. 2024. Lessons from the trenches on reproducible evaluation of language models. *ArXiv*, abs/2405.14782.

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*.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E. Gonzalez, and Ion Stoica. 2024. Chatbot arena: An open platform for evaluating llms by human preference.

Sumanth Doddapaneni, Rahul Aralikatte, Gowtham Ramesh, Shreya Goyal, Mitesh M. Khapra, Anoop Kunchukuttan, and Pratyush Kumar. 2023. Towards leaving no Indic language behind: Building monolingual corpora, benchmark and models for Indic languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12402–12426, Toronto, Canada. Association for Computational Linguistics.

Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, 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, 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, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, 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, 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, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, 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, 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, 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, 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, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaoqing Ellen Tan, 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, Aaron Grattafiori, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alex Vaughan, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Franco, 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, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, Danny Wyatt, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Firat Ozgenel, Francesco Caggioni, Francisco Guzmán, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Govind Thattai, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, 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, Karthik Prasad, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kun Huang, Kunal Chawla, Kushal Lakhotia, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng

Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Maria Tsimpoukelli, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, 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, Nikolay Pavlovich Laptev, Ning Dong, Ning Zhang, 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, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Rohan Maheswari, Russ Howes, Ruty Rinott, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, 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, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Kohler, 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, Vítor Albiero, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaofang Wang, Xiaojian Wu, Xiaolan Wang, Xide Xia, Xilun Wu, Xinbo Gao, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yuchen Hao, Yundi Qian, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, and Zhiwei Zhao. 2024. The llama 3 herd of models. *arXiv preprint arXiv: 2407.21783*.

Jay Gala, Thanmay Jayakumar, Jaavid Aktar Husain J, Aswanth Kumar M, Mohammed Safi Ur Rahman Khan, Diptesh Kanojia, Ratish Puduppully, Mitesh M Khapra, Raj Dabre, Rudra Murthy, and Anoop Kunchukuttan. 2024. Airavata: Introducing hindi instruction-tuned LLM.

Jay P. Gala, Pranjal A. Chitale, AK Raghavan, Varun Gumma, Sumanth Doddapaneni, M. AswanthKumar, J. Nawale, Anupama Sujatha, Ratish Puduppully, Vivek Raghavan, Pratyush Kumar, Mitesh M. Khapra, Raj Dabre, and Anoop Kunchukuttan. 2023. Indictrans2: Towards high-quality and accessible machine translation models for all 22 scheduled indian languages. *Trans. Mach. Learn. Res.*

Leo Gao, Jonathan Tow, Baber Abbasi, Stella Biderman, Sid Black, Anthony DiPofi, Charles Foster, Laurence Golding, Jeffrey Hsu, Alain Le Noac'h, Haonan Li, Kyle McDonell, Niklas Muennighoff, Chris Ociepa, Jason Phang, Laria Reynolds, Hailey Schoelkopf, Aviya Skowron, Lintang Sutawika, Eric Tang, Anish Thite, Ben Wang, Kevin Wang, and Andy Zou. 2024. A framework for few-shot language model evaluation.

Sreyan Ghosh, Chandra Kiran Reddy Evuru, Sonal Kumar, Ramaneswaran S., Deepali Aneja, Zeyu Jin, Ramani Duraiswami, and Dinesh Manocha. 2024. A closer look at the limitations of instruction tuning. In *Forty-first International Conference on Machine Learning, ICML 2024, Vienna, Austria, July 21-27, 2024*. OpenReview.net.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2021. Measuring massive multitask language understanding. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. Lora: Low-rank adaptation of large language models. In *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022*. OpenReview.net.

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

Tahir Javed, Janki Nawale, Eldho George, Sakshi Joshi, Kaushal Bhogale, Deovrat Mehendale, Ishvinder Sethi, Aparna Ananthanarayanan, Hafsah Faquih, Pratiti Palit, Sneha Ravishankar, Saranya Sukumaran, Tripura Panchagnula, Sunjay Murali, Kunal Gandhi, Ambujavalli R, Manickam M, C Vaijayanthi, Krishnan Karunganni, Pratyush Kumar, and Mitesh Khapra. 2024. IndicVoices: Towards building an inclusive multilingual speech dataset for Indian languages. In *Findings of the Association for Computational Linguistics ACL 2024*, pages 10740–10782, Bangkok, Thailand and virtual meeting. Association for Computational Linguistics.

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

Mohammed Safi Ur Rahman Khan, Priyam Mehta, Ananth Sankar, Umashankar Kumaravelan, Sumanth Doddapaneni, Suriyaprasaad G, Varun Balan G,

Sparsh Jain, Anoop Kunchukuttan, Pratyush Kumar, Raj Dabre, and Mitesh M. Khapra. 2024. Indicllmsuite: A blueprint for creating pre-training and fine-tuning datasets for indian languages. *arXiv preprint arXiv: 2403.06350*.

Seungone Kim, Juyoung Suk, Ji Yong Cho, Shayne Longpre, Chaeeun Kim, Dongkeun Yoon, Guijin Son, Yejin Cho, Sheikh Shafayat, Jinheon Baek, Sue Hyun Park, Hyeonbin Hwang, Jinkyung Jo, Hyowon Cho, Haebin Shin, Seongyun Lee, Hanseok Oh, Noah Lee, Namgyu Ho, Se June Joo, Miyoung Ko, Yoonjoo Lee, Hyungjoo Chae, Jamin Shin, Joel Jang, Seonghyeon Ye, Bill Yuchen Lin, Sean Welleck, Graham Neubig, Moontae Lee, Kyungjae Lee, and Minjoon Seo. 2024. The biggen bench: A principled benchmark for fine-grained evaluation of language models with language models. *arXiv preprint arXiv: 2406.05761*.

Guneet Singh Kohli, Shantipriya Parida, Sambit Sekhar, Samirit Saha, Nipun B Nair, Parul Agarwal, Sonal Khosla, Kusumlata Patiyal, and Debasish Dhal. 2023. Building a llama2-finetuned llm for odia language utilizing domain knowledge instruction set. *arXiv preprint arXiv: 2312.12624*.

Adithya S Kolavi. 2024. Indic llm leaderboard. https://huggingface.co/spaces/Cognitive-Lab/indic_llm_leaderboard.

Aman Kumar, Himani Shrotriya, Prachi Sahu, Raj Dabre, Ratish Puduppully, Anoop Kunchukuttan, Amogh Mishra, Mitesh M. Khapra, and Pratyush Kumar. 2022. Indicnlg benchmark: Multilingual datasets for diverse nlg tasks in indic languages. *arXiv preprint arXiv: 2203.05437*.

Chankyu Lee, Rajarshi Roy, Mengyao Xu, Jonathan Raiman, Mohammad Shoeybi, Bryan Catanzaro, and Wei Ping. 2024. Nv-embed: Improved techniques for training llms as generalist embedding models. *arXiv preprint arXiv: 2405.17428*.

Yang Liu, Meng Xu, Shuo Wang, Liner Yang, Haoyu Wang, Zhenghao Liu, Cunliang Kong, Yun Chen, Yang Liu, Maosong Sun, and Erhong Yang. 2024. Omgeval: An open multilingual generative evaluation benchmark for large language models. *arXiv preprint arXiv: 2402.13524*.

Yash Madhani, Mitesh M. Khapra, and Anoop Kunchukuttan. 2023. Bhasa-abhijnaanam: Native-script and romanized language identification for 22 Indic languages. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 816–826, Toronto, Canada. Association for Computational Linguistics.

Timothy R. McIntosh, Teo Susnjak, Nalin Arachchilage, Tong Liu, Paul Watters, and Malka N. Halgamuge. 2024. Inadequacies of large language model benchmarks in the era of generative artificial intelligence. *arXiv preprint arXiv: 2402.09880*.

Xuan-Phi Nguyen, Wenxuan Zhang, Xin Li, Mahani Aljunied, Zhiqiang Hu, Chenhui Shen, Yew Ken Chia, Xingxuan Li, Jianyu Wang, Qingyu Tan, Liying Cheng, Guanzheng Chen, Yue Deng, Sen Yang, Chaoqun Liu, Hang Zhang, and Lidong Bing. 2023. Seallms - large language models for southeast asia. *arXiv preprint arXiv: 2312.00738*.

OpenAI, Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, Red Avila, Igor Babuschkin, Suchir Balaji, Valerie Balcom, Paul Baltescu, Haiming Bao, Mohammad Bavarian, Jeff Belgum, Irwan Bello, Jake Berdine, Gabriel Bernadett-Shapiro, Christopher Berner, Lenny Bogdonoff, Oleg Boiko, Madelaine Boyd, Anna-Luisa Brakman, Greg Brockman, Tim Brooks, Miles Brundage, Kevin Button, Trevor Cai, Rosie Campbell, Andrew Cann, Brittany Carey, Chelsea Carlson, Rory Carmichael, Brooke Chan, Che Chang, Fotis Chantzis, Derek Chen, Sully Chen, Ruby Chen, Jason Chen, Mark Chen, Ben Chess, Chester Cho, Casey Chu, Hyung Won Chung, Dave Cummings, Jeremiah Currier, Yunxing Dai, Cory Decareaux, Thomas Degry, Noah Deutsch, Damien Deville, Arka Dhar, David Dohan, Steve Dowling, Sheila Dunning, Adrien Ecoffet, Atty Eleti, Tyna Eloundou, David Farhi, Liam Fedus, Niko Felix, Simón Posada Fishman, Juston Forte, Isabella Fulford, Leo Gao, Elie Georges, Christian Gibson, Vik Goel, Tarun Gogineni, Gabriel Goh, Rapha Gontijo-Lopes, Jonathan Gordon, Morgan Grafstein, Scott Gray, Ryan Greene, Joshua Gross, Shixiang Shane Gu, Yufei Guo, Chris Hallacy, Jesse Han, Jeff Harris, Yuchen He, Mike Heaton, Johannes Heidecke, Chris Hesse, Alan Hickey, Wade Hickey, Peter Hoeschele, Brandon Houghton, Kenny Hsu, Shengli Hu, Xin Hu, Joost Huizinga, Shantanu Jain, Shawn Jain, Joanne Jang, Angela Jiang, Roger Jiang, Haozhun Jin, Denny Jin, Shino Jomoto, Billie Jonn, Heewoo Jun, Tomer Kaftan, Łukasz Kaiser, Ali Kamali, Ingmar Kanitscheider, Nitish Shirish Keskar, Tabarak Khan, Logan Kilpatrick, Jong Wook Kim, Christina Kim, Yongjik Kim, Jan Hendrik Kirchner, Jamie Kiros, Matt Knight, Daniel Kokotajlo, Łukasz Kondraciuk, Andrew Kondrich, Aris Konstantinidis, Kyle Kosic, Gretchen Krueger, Vishal Kuo, Michael Lampe, Ikai Lan, Teddy Lee, Jan Leike, Jade Leung, Daniel Levy, Chak Ming Li, Rachel Lim, Molly Lin, Stephanie Lin, Mateusz Litwin, Theresa Lopez, Ryan Lowe, Patricia Lue, Anna Makanju, Kim Malfacini, Sam Manning, Todor Markov, Yaniv Markovski, Bianca Martin, Katie Mayer, Andrew Mayne, Bob McGrew, Scott Mayer McKinney, Christine McLeavey, Paul McMillan, Jake McNeil, David Medina, Aalok Mehta, Jacob Menick, Luke Metz, Andrey Mishchenko, Pamela Mishkin, Vinnie Monaco, Evan Morikawa, Daniel Mossing, Tong Mu, Mira Murati, Oleg Murk, David Mély, Ashvin Nair, Reiichiro Nakano, Rajeev Nayak, Arvind Neelakantan, Richard Ngo, Hyeonwoo Noh, Long Ouyang, Cullen O'Keefe, Jakub Pachocki, Alex Paino, Joe Palermo, Ashley Pantuliano, Giambattista Parascandolo, Joel Parish, Emy Parparita, Alex

Passos, Mikhail Pavlov, Andrew Peng, Adam Perelman, Filipe de Avila Belbute Peres, Michael Petrov, Henrique Ponde de Oliveira Pinto, Michael, Pokorny, Michelle Pokrass, Vitchyr H. Pong, Tolly Powell, Alethea Power, Boris Power, Elizabeth Proehl, Raul Puri, Alec Radford, Jack Rae, Aditya Ramesh, Cameron Raymond, Francis Real, Kendra Rimbach, Carl Ross, Bob Rotsted, Henri Roussez, Nick Ryder, Mario Saltarelli, Ted Sanders, Shibani Santurkar, Girish Sastry, Heather Schmidt, David Schnurr, John Schulman, Daniel Selsam, Kyla Sheppard, Toki Sherbakov, Jessica Shieh, Sarah Shoker, Pranav Shyam, Szymon Sidor, Eric Sigler, Maddie Simens, Jordan Sitkin, Katarina Slama, Ian Sohl, Benjamin Sokolowsky, Yang Song, Natalie Staudacher, Felipe Petroski Such, Natalie Summers, Ilya Sutskever, Jie Tang, Nikolas Tezak, Madeleine B. Thompson, Phil Tillet, Amin Tootoonchian, Elizabeth Tseng, Preston Tuggle, Nick Turley, Jerry Tworek, Juan Felipe Cerón Uribe, Andrea Vallone, Arun Vijayvergiya, Chelsea Voss, Carroll Wainwright, Justin Jay Wang, Alvin Wang, Ben Wang, Jonathan Ward, Jason Wei, CJ Weinmann, Akila Welihinda, Peter Welinder, Jiayi Weng, Lilian Weng, Matt Wiethoff, Dave Willner, Clemens Winter, Samuel Wolrich, Hannah Wong, Lauren Workman, Sherwin Wu, Jeff Wu, Michael Wu, Kai Xiao, Tao Xu, Sarah Yoo, Kevin Yu, Qiming Yuan, Wojciech Zaremba, Rowan Zellers, Chong Zhang, Marvin Zhang, Shengjia Zhao, Tianhao Zheng, Juntang Zhuang, William Zhuk, and Barret Zoph. 2023. Gpt-4 technical report. *PREPRINT*.

Pritika Rohera, Chaitrali Ginimav, Akanksha Salunke, Gayatri Sawant, and Raviraj Joshi. 2024. L3cube-indicquest: A benchmark questing answering dataset for evaluating knowledge of llms in indic context. *arXiv preprint arXiv: 2409.08706*.

Abhishek Kumar Singh, Rudra Murthy, Vishwajeet kumar, Jaydeep Sen, and Ganesh Ramakrishnan. 2024a. Indic qa benchmark: A multilingual benchmark to evaluate question answering capability of llms for indic languages. *arXiv preprint arXiv: 2407.13522*.

Harman Singh, Nitish Gupta, Shikhar Bharadwaj, Dinesh Tewari, and Partha Talukdar. 2024b. Indicgenbench: A multilingual benchmark to evaluate generation capabilities of llms on indic languages. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2024, Bangkok, Thailand, August 11-16, 2024*, pages 11047–11073. Association for Computational Linguistics.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M. Dai, Anja Hauth, Katie Millican, David Silver, Melvin Johnson, Ioannis Antonoglou, Julian Schrittwieser, Amelia Glaese, Jilin Chen, Emily Pitler, Timothy Lillicrap, Angeliki Lazaridou, Orhan Firat, James Molloy, Michael Isard, Paul R. Barham, Tom Hennigan, Benjamin Lee, Fabio Viola, Malcolm Reynolds, Yuanzhong Xu, Ryan Doherty, Eli Collins, Clemens Meyer, Eliza Rutherford, Erica Moreira, Kareem Ayoub, Megha Goel, Jack Krawczyk, Cosmo Du, Ed Chi, Heng-Tze Cheng, Eric Ni, Purvi Shah, Patrick Kane, Betty Chan, Manaal Faruqui, Aliaksei Severyn, Hanzhao Lin, YaGuang Li, Yong Cheng, Abe Ittycheriah, Mahdis Mahdieh, Mia Chen, Pei Sun, Dustin Tran, Sumit Bagri, Balaji Lakshminarayanan, Jeremiah Liu, Andras Orban, Fabian Güra, Hao Zhou, Xinying Song, Aurelien Boffy, Harish Ganapathy, Steven Zheng, HyunJeong Choe, Ágoston Weisz, Tao Zhu, Yifeng Lu, Siddharth Gopal, Jarrod Kahn, Maciej Kula, Jeff Pitman, Rushin Shah, Emanuel Taropa, Majd Al Merey, Martin Baeuml, Zhifeng Chen, Laurent El Shafey, Yujing Zhang, Olcan Sercinoglu, George Tucker, Enrique Piqueras, Maxim Krikun, Iain Barr, Nikolay Savinov, Ivo Danihelka, Becca Roelofs, Anaïs White, Anders Andreassen, Tamara von Glehn, Lakshman Yagati, Mehran Kazemi, Lucas Gonzalez, Misha Khalman, Jakub Sygnowski, Alexandre Frechette, Charlotte Smith, Laura Culp, Lev Proleev, Yi Luan, Xi Chen, James Lottes, Nathan Schucher, Federico Lebron, Alban Rrustemi, Natalie Clay, Phil Crone, Tomas Kocisky, Jeffrey Zhao, Bartek Perz, Dian Yu, Heidi Howard, Adam Bloniarz, Jack W. Rae, Han Lu, Laurent Sifre, Marcello Maggioni, Fred Alcober, Dan Garrette, Megan Barnes, Shantanu Thakoor, Jacob Austin, Gabriel Barth-Maron, William Wong, Rishabh Joshi, Rahma Chaabouni, Deeni Fatiha, Arun Ahuja, Gaurav Singh Tomar, Evan Senter, Martin Chadwick, Ilya Kornakov, Nithya Attaluri, Iñaki Iturrate, Ruibo Liu, Yunxuan Li, Sarah Cogan, Jeremy Chen, Chao Jia, Chenjie Gu, Qiao Zhang, Jordan Grimstad, Ale Jakse Hartman, Xavier Garcia, Thanumalayan Sankaranarayana Pillai, Jacob Devlin, Michael Laskin, Diego de Las Casas, Dasha Valter, Connie Tao, Lorenzo Blanco, Adrià Puigdomènech Badia, David Reitter, Mianna Chen, Jenny Brennan, Clara Rivera, Sergey Brin, Shariq Iqbal, Gabriela Surita, Jane Labanowski, Abhi Rao, Stephanie Winkler, Emilio Parisotto, Yiming Gu, Kate Olszewska, Ravi Addanki, Antoine Miech, Annie Louis, Denis Teplyashin, Geoff Brown, Elliot Catt, Jan Balaguer, Jackie Xiang, Pidong Wang, Zoe Ashwood, Anton Briukhov, Albert Webson, Sanjay Ganapathy, Smit Sanghavi, Ajay Kannan, Ming-Wei Chang, Axel Stjerngren, Josip Djolonga, Yuting Sun, Ankur Bapna, Matthew Aitchison, Pedram Pejman, Henryk Michalewski, Tianhe Yu, Cindy Wang, Juliette Love, Junwhan Ahn, Dawn Bloxwich, Kehang Han, Peter Humphreys, Thibault Sellam, James Bradbury, Varun Godbole, Sina Samangooei, Bogdan Damoc, Alex Kaskasoli, Sébastien M. R. Arnold, Vijay Vasudevan, Shubham Agrawal, Jason Riesa, Dmitry Lepikhin, Richard Tanburn, Srivatsan Srinivasan, Hyeontaek Lim, Sarah Hodkinson, Pranav Shyam, Johan Ferret, Steven Hand, Ankush Garg, Tom Le Paine, Jian Li, Yujia Li, Minh Giang, Alexander Neitz, Zaheer Abbas, Sarah York, Machel Reid, Elizabeth Cole, Aakanksha Chowdhery, Dipanjan Das, Dominika Rogozińska, Vitaliy Nikolaev, Pablo Sprechmann, Zachary Nado, Lukas Zilka, Flavien Prost, Luheng He, Marianne Monteiro, Gaurav Mishra, Chris Welty, Josh Newlan, Dawei Jia, Miltiadis Allamanis, Clara Huiyi Hu,

Raoul de Liedekerke, Justin Gilmer, Carl Saroufim, Shruti Rijhwani, Shaobo Hou, Disha Shrivastava, Anirudh Baddepudi, Alex Goldin, Adnan Ozturel, Albin Cassirer, Yunhan Xu, Daniel Sohn, Devendra Sachan, Reinald Kim Amplayo, Craig Swanson, Dessie Petrova, Shashi Narayan, Arthur Guez, Siddhartha Brahma, Jessica Landon, Miteyan Patel, Ruizhe Zhao, Kevin Villela, Luyu Wang, Wenhao Jia, Matthew Rahtz, Mai Giménez, Legg Yeung, James Keeling, Petko Georgiev, Diana Mincu, Boxi Wu, Salem Haykal, Rachel Saputro, Kiran Vodrahalli, James Qin, Zeynep Cankara, Abhanshu Sharma, Nick Fernando, Will Hawkins, Behnam Neyshabur, Solomon Kim, Adrian Hutter, Priyanka Agrawal, Alex Castro-Ros, George van den Driessche, Tao Wang, Fan Yang, Shuo yiin Chang, Paul Komarek, Ross McIlroy, Mario Lučić, Guodong Zhang, Wael Farhan, Michael Sharman, Paul Natsev, Paul Michel, Yamini Bansal, Siyuan Qiao, Kris Cao, Siamak Shakeri, Christina Butterfield, Justin Chung, Paul Kishan Rubenstein, Shivani Agrawal, Arthur Mensch, Kedar Soparkar, Karel Lenc, Timothy Chung, Aedan Pope, Loren Maggiore, Jackie Kay, Priya Jhakra, Shibo Wang, Joshua Maynez, Mary Phuong, Taylor Tobin, Andrea Tacchetti, Maja Trebacz, Kevin Robinson, Yash Katariya, Sebastian Riedel, Paige Bailey, Kefan Xiao, Nimesh Ghelani, Lora Aroyo, Ambrose Slone, Neil Houlsby, Xuehan Xiong, Zhen Yang, Elena Gribovskaya, Jonas Adler, Mateo Wirth, Lisa Lee, Music Li, Thais Kagohara, Jay Pavagadhi, Sophie Bridgers, Anna Bortsova, Sanjay Ghemawat, Zafarali Ahmed, Tianqi Liu, Richard Powell, Vijay Bolina, Mariko Iinuma, Polina Zablotskaia, James Besley, Da-Woon Chung, Timothy Dozat, Ramona Comanescu, Xiance Si, Jeremy Greer, Guolong Su, Martin Polacek, Raphaël Lopez Kaufman, Simon Tokumine, Hexiang Hu, Elena Buchatskaya, Yingjie Miao, Mohamed Elhawaty, Aditya Siddhant, Nenad Tomasev, Jinwei Xing, Christina Greer, Helen Miller, Shereen Ashraf, Aurko Roy, Zizhao Zhang, Ada Ma, Angelos Filos, Milos Besta, Rory Blevins, Ted Klimenko, Chih-Kuan Yeh, Soravit Changpinyo, Jiaqi Mu, Oscar Chang, Mantas Pajarskas, Carrie Muir, Vered Cohen, Charline Le Lan, Krishna Haridasan, Amit Marathe, Steven Hansen, Sholto Douglas, Rajkumar Samuel, Mingqiu Wang, Sophia Austin, Chang Lan, Jiepu Jiang, Justin Chiu, Jaime Alonso Lorenzo, Lars Lowe Sjösund, Sébastien Cevey, Zach Gleicher, Thi Avrahami, Anudhyan Boral, Hansa Srinivasan, Vittorio Selo, Rhys May, Konstantinos Aisopos, Léonard Hussenot, Livio Baldini Soares, Kate Baumli, Michael B. Chang, Adrià Recasens, Ben Caine, Alexander Pritzel, Filip Pavetic, Fabio Pardo, Anita Gergely, Justin Frye, Vinay Ramasesh, Dan Horgan, Kartikeya Badola, Nora Kassner, Subhrajit Roy, Ethan Dyer, Víctor Campos Campos, Alex Tomala, Yunhao Tang, Dalia El Badawy, Elspeth White, Basil Mustafa, Oran Lang, Abhishek Jindal, Sharad Vikram, Zhitao Gong, Sergi Caelles, Ross Hemsley, Gregory Thornton, Fangxiaoyu Feng, Wojciech Stokowiec, Ce Zheng, Phoebe Thacker, Çağlar Ünlü, Zhishuai Zhang, Mohammad Saleh, James Svensson, Max Bileschi, Piyush Patil, Ankesh Anand, Roman Ring, Katerina Tsihlas, Arpi Vezer, Marco Selvi, Toby Shevlane, Mikel Rodriguez, Tom Kwiatkowski, Samira Daruki, Keran Rong, Allan Dafoe, Nicholas FitzGerald, Keren Gu-Lemberg, Mina Khan, Lisa Anne Hendricks, Marie Pellat, Vladimir Feinberg, James Cobon-Kerr, Tara Sainath, Maribeth Rauh, Sayed Hadi Hashemi, Richard Ives, Yana Hasson, Eric Noland, Yuan Cao, Nathan Byrd, Le Hou, Qingze Wang, Thibault Sottiaux, Michela Paganini, Jean-Baptiste Lespiau, Alexandre Moufarek, Samer Hassan, Kaushik Shivakumar, Joost van Amersfoort, Amol Mandhane, Pratik Joshi, Anirudh Goyal, Matthew Tung, Andrew Brock, Hannah Sheahan, Vedant Misra, Cheng Li, Nemanja Rakićević, Mostafa Dehghani, Fangyu Liu, Sid Mittal, Junhyuk Oh, Seb Noury, Eren Sezener, Fantine Huot, Matthew Lamm, Nicola De Cao, Charlie Chen, Sidharth Mudgal, Romina Stella, Kevin Brooks, Gautam Vasudevan, Chenxi Liu, Mainak Chain, Nivedita Melinkeri, Aaron Cohen, Venus Wang, Kristie Seymore, Sergey Zubkov, Rahul Goel, Summer Yue, Sai Krishnakumaran, Brian Albert, Nate Hurley, Motoki Sano, Anhad Mohananey, Jonah Joughin, Egor Filonov, Tomasz Kępa, Yomna Eldawy, Jiawern Lim, Rahul Rishi, Shirin Badiezadegan, Taylor Bos, Jerry Chang, Sanil Jain, Sri Gayatri Sundara Padmanabhan, Subha Puttagunta, Kalpesh Krishna, Leslie Baker, Norbert Kalb, Vamsi Bedapudi, Adam Kurzrok, Shuntong Lei, Anthony Yu, Oren Litvin, Xiang Zhou, Zhichun Wu, Sam Sobell, Andrea Siciliano, Alan Papir, Robby Neale, Jonas Bragagnolo, Tej Toor, Tina Chen, Valentin Anklin, Feiran Wang, Richie Feng, Milad Gholami, Kevin Ling, Lijuan Liu, Jules Walter, Hamid Moghaddam, Arun Kishore, Jakub Adamek, Tyler Mercado, Jonathan Mallinson, Siddhinita Wandekar, Stephen Cagle, Eran Ofek, Guillermo Garrido, Clemens Lombriser, Maksim Mukha, Botu Sun, Hafeezul Rahman Mohammad, Josip Matak, Yadi Qian, Vikas Peswani, Pawel Janus, Quan Yuan, Leif Schelin, Oana David, Ankur Garg, Yifan He, Oleksii Duzhyi, Anton Älgmyr, Timothée Lottaz, Qi Li, Vikas Yadav, Luyao Xu, Alex Chinien, Rakesh Shivanna, Aleksandr Chuklin, Josie Li, Carrie Spadine, Travis Wolfe, Kareem Mohamed, Subhabrata Das, Zihang Dai, Kyle He, Daniel von Dincklage, Shyam Upadhyay, Akanksha Maurya, Luyan Chi, Sebastian Krause, Khalid Salama, Pam G Rabinovitch, Pavan Kumar Reddy M, Aarush Selvan, Mikhail Dektiarev, Golnaz Ghiasi, Erdem Guven, Himanshu Gupta, Boyi Liu, Deepak Sharma, Idan Heimlich Shtacher, Shachi Paul, Oscar Akerlund, François-Xavier Aubet, Terry Huang, Chen Zhu, Eric Zhu, Elico Teixeira, Matthew Fritze, Francesco Bertolini, Liana-Eleonora Marinescu, Martin Bölle, Dominik Paulus, Khyatti Gupta, Tejasi Latkar, Max Chang, Jason Sanders, Roopa Wilson, Xuewei Wu, Yi-Xuan Tan, Lam Nguyen Thiet, Tulsee Doshi, Sid Lall, Swaroop Mishra, Wanming Chen, Thang Luong, Seth Benjamin, Jasmine Lee, Ewa Andrejczuk, Dominik Rabiej, Vipul Ranjan, Krzysztof Styrc, Pengcheng Yin, Jon Simon, Malcolm Rose Harriott, Mudit Bansal, Alexei Robsky, Geoff Bacon, David Greene, Daniil Mirylenka, Chen Zhou, Obaid Sarvana, Abhimanyu Goyal, Samuel Andermatt, Patrick Siegler, Ben Horn, Assaf Is-

rael, Francesco Pongetti, Chih-Wei "Louis" Chen, Marco Selvatici, Pedro Silva, Kathie Wang, Jackson Tolins, Kelvin Guu, Roey Yogev, Xiaochen Cai, Alessandro Agostini, Maulik Shah, Hung Nguyen, Noah Ó Donnaile, Sébastien Pereira, Linda Friso, Adam Stambler, Adam Kurzrok, Chenkai Kuang, Yan Romanikhin, Mark Geller, ZJ Yan, Kane Jang, Cheng-Chun Lee, Wojciech Fica, Eric Malmi, Qijun Tan, Dan Banica, Daniel Balle, Ryan Pham, Yanping Huang, Diana Avram, Hongzhi Shi, Jasjot Singh, Chris Hidey, Niharika Ahuja, Pranab Saxena, Dan Dooley, Srividya Pranavi Potharaju, Eileen O'Neill, Anand Gokulchandran, Ryan Foley, Kai Zhao, Mike Dusenberry, Yuan Liu, Pulkit Mehta, Ragha Kotikalapudi, Chalence Safranek-Shrader, Andrew Goodman, Joshua Kessinger, Eran Globen, Prateek Kolhar, Chris Gorgolewski, Ali Ibrahim, Yang Song, Ali Eichenbaum, Thomas Brovelli, Sahitya Potluri, Preethi Lahoti, Cip Baetu, Ali Ghorbani, Charles Chen, Andy Crawford, Shalini Pal, Mukund Sridhar, Petru Gurita, Asier Mujika, Igor Petrovski, Pierre-Louis Cedoz, Chenmei Li, Shiyuan Chen, Niccolò Dal Santo, Siddharth Goyal, Jitesh Punjabi, Karthik Kappaganthu, Chester Kwak, Pallavi LV, Sarmishta Velury, Himadri Choudhury, Jamie Hall, Premal Shah, Ricardo Figueira, Matt Thomas, Minjie Lu, Ting Zhou, Chintu Kumar, Thomas Jurdi, Sharat Chikkerur, Yenai Ma, Adams Yu, Soo Kwak, Victor Ähdel, Sujeevan Rajayogam, Travis Choma, Fei Liu, Aditya Barua, Colin Ji, Ji Ho Park, Vincent Hellendoorn, Alex Bailey, Taylan Bilal, Huanjie Zhou, Mehrdad Khatir, Charles Sutton, Wojciech Rzadkowski, Fiona Macintosh, Konstantin Shagin, Paul Medina, Chen Liang, Jinjing Zhou, Pararth Shah, Yingying Bi, Attila Dankovics, Shipra Banga, Sabine Lehmann, Marissa Bredesen, Zifan Lin, John Eric Hoffmann, Jonathan Lai, Raynald Chung, Kai Yang, Nihal Balani, Arthur Bražinskas, Andrei Sozanschi, Matthew Hayes, Héctor Fernández Alcalde, Peter Makarov, Will Chen, Antonio Stella, Liselotte Snijders, Michael Mandl, Ante Kärrman, Paweł Nowak, Xinyi Wu, Alex Dyck, Krishnan Vaidyanathan, Raghavender R, Jessica Mallet, Mitch Rudominer, Eric Johnston, Sushil Mittal, Akhil Udathu, Janara Christensen, Vishal Verma, Zach Irving, Andreas Santucci, Gamaleldin Elsayed, Elnaz Davoodi, Marin Georgiev, Ian Tenney, Nan Hua, Geoffrey Cideron, Edouard Leurent, Mahmoud Alnahlawi, Ionut Georgescu, Nan Wei, Ivy Zheng, Dylan Scandinaro, Heinrich Jiang, Jasper Snoek, Mukund Sundararajan, Xuezhi Wang, Zack Ontiveros, Itay Karo, Jeremy Cole, Vinu Rajashekhar, Lara Tumeh, Eyal Ben-David, Rishub Jain, Jonathan Uesato, Romina Datta, Oskar Bunyan, Shimu Wu, John Zhang, Piotr Stanczyk, Ye Zhang, David Steiner, Subhajit Naskar, Michael Azzam, Matthew Johnson, Adam Paszke, Chung-Cheng Chiu, Jaume Sanchez Elias, Afroz Mohiuddin, Faizan Muhammad, Jin Miao, Andrew Lee, Nino Vieillard, Jane Park, Jiageng Zhang, Jeff Stanway, Drew Garmon, Abhijit Karmarkar, Zhe Dong, Jong Lee, Aviral Kumar, Luowei Zhou, Jonathan Evens, William Isaac, Geoffrey Irving, Edward Loper, Michael Fink, Isha Arkatkar, Nanxin Chen, Izhak Shafran, Ivan Petrychenko, Zhe Chen, Johnson Jia, Anselm Levskaya, Zhenkai Zhu, Peter Grabowski, Yu Mao, Alberto Magni, Kaisheng Yao, Javier Snaider, Norman Casagrande, Evan Palmer, Paul Suganthan, Alfonso Castaño, Irene Giannoumis, Wooyeol Kim, Mikołaj Rybiński, Ashwin Sreevatsa, Jennifer Prendki, David Soergel, Adrian Goedeckemeyer, Willi Gierke, Mohsen Jafari, Meenu Gaba, Jeremy Wiesner, Diana Gage Wright, Yawen Wei, Harsha Vashisht, Yana Kulizhskaya, Jay Hoover, Maigo Le, Lu Li, Chimezie Iwuanyanwu, Lu Liu, Kevin Ramirez, Andrey Khorlin, Albert Cui, Tian LIN, Marcus Wu, Ricardo Aguilar, Keith Pallo, Abhishek Chakladar, Ginger Perng, Elena Allica Abellan, Mingyang Zhang, Ishita Dasgupta, Nate Kushman, Ivo Penchev, Alena Repina, Xihui Wu, Tom van der Weide, Priya Ponnapalli, Caroline Kaplan, Jiri Simsa, Shuangfeng Li, Olivier Dousse, Fan Yang, Jeff Piper, Nathan Ie, Rama Pasumarthi, Nathan Lintz, Anitha Vijayakumar, Daniel Andor, Pedro Valenzuela, Minnie Lui, Cosmin Paduraru, Daiyi Peng, Katherine Lee, Shuyuan Zhang, Somer Greene, Duc Dung Nguyen, Paula Kurylowicz, Cassidy Hardin, Lucas Dixon, Lili Janzer, Kiam Choo, Ziqiang Feng, Biao Zhang, Achintya Singhal, Dayou Du, Dan McKinnon, Natasha Antropova, Tolga Bolukbasi, Orgad Keller, David Reid, Daniel Finchelstein, Maria Abi Raad, Remi Crocker, Peter Hawkins, Robert Dadashi, Colin Gaffney, Ken Franko, Anna Bulanova, Rémi Leblond, Shirley Chung, Harry Askham, Luis C. Cobo, Kelvin Xu, Felix Fischer, Jun Xu, Christina Sorokin, Chris Alberti, Chu-Cheng Lin, Colin Evans, Alek Dimitriev, Hannah Forbes, Dylan Banarse, Zora Tung, Mark Omernick, Colton Bishop, Rachel Sterneck, Rohan Jain, Jiawei Xia, Ehsan Amid, Francesco Piccinno, Xingyu Wang, Praseem Banzal, Daniel J. Mankowitz, Alex Polozov, Victoria Krakovna, Sasha Brown, MohammadHossein Bateni, Dennis Duan, Vlad Firoiu, Meghana Thotakuri, Tom Natan, Matthieu Geist, Ser tan Girgin, Hui Li, Jiayu Ye, Ofir Roval, Reiko Tojo, Michael Kwong, James Lee-Thorp, Christopher Yew, Danila Sinopalnikov, Sabela Ramos, John Mellor, Abhishek Sharma, Kathy Wu, David Miller, Nicolas Sonnerat, Denis Vnukov, Rory Greig, Jennifer Beattie, Emily Caveness, Libin Bai, Julian Eisenschlos, Alex Korchemniy, Tomy Tsai, Mimi Jasarevic, Weize Kong, Phuong Dao, Zeyu Zheng, Frederick Liu, Fan Yang, Rui Zhu, Tian Huey Teh, Jason Sanmiya, Evgeny Gladchenko, Nejc Trdin, Daniel Toyama, Evan Rosen, Sasan Tavakkol, Linting Xue, Chen Elkind, Oliver Woodman, John Carpenter, George Papamakarios, Rupert Kemp, Sushant Kafle, Tanya Grunina, Rishika Sinha, Alice Talbert, Diane Wu, Denese Owusu-Afriyie, Cosmo Du, Chloe Thornton, Jordi Pont-Tuset, Pradyumna Narayana, Jing Li, Saaber Fatehi, John Wieting, Omar Ajmeri, Benigno Uria, Yeongil Ko, Laura Knight, Amélie Héliou, Ning Niu, Shane Gu, Chenxi Pang, Yeqing Li, Nir Levine, Ariel Stolovich, Rebeca Santamaria-Fernandez, Sonam Goenka, Wenny Yustalim, Robin Strudel, Ali Elqursh, Charlie Deck, Hyo Lee, Zonglin Li, Kyle Levin, Raphael Hoffmann, Dan Holtmann-Rice, Olivier Bachem, Sho

Arora, Christy Koh, Soheil Hassas Yeganeh, Siim Põder, Mukarram Tariq, Yanhua Sun, Lucian Ionita, Mojtaba Seyedhosseini, Pouya Tafti, Zhiyu Liu, Anmol Gulati, Jasmine Liu, Xinyu Ye, Bart Chrzaszcz, Lily Wang, Nikhil Sethi, Tianrun Li, Ben Brown, Shreya Singh, Wei Fan, Aaron Parisi, Joe Stanton, Vinod Koverkathu, Christopher A. Choquette-Choo, Yunjie Li, TJ Lu, Abe Ittycheriah, Prakash Shroff, Mani Varadarajan, Sanaz Bahargam, Rob Willoughby, David Gaddy, Guillaume Desjardins, Marco Cornero, Brona Robenek, Bhavishya Mittal, Ben Albrecht, Ashish Shenoy, Fedor Moiseev, Henrik Jacobsson, Alireza Ghaffarkhah, Morgane Rivière, Alanna Walton, Clément Crepy, Alicia Parrish, Zongwei Zhou, Clement Farabet, Carey Radebaugh, Praveen Srinivasan, Claudia van der Salm, Andreas Fidjeland, Salvatore Scellato, Eri Latorre-Chimoto, Hanna Klimczak-Plucińska, David Bridson, Dario de Cesare, Tom Hudson, Piermaria Mendolicchio, Lexi Walker, Alex Morris, Matthew Mauger, Alexey Guseynov, Alison Reid, Seth Odoom, Lucia Loher, Victor Cotruta, Madhavi Yenugula, Dominik Grewe, Anastasia Petrushkina, Tom Duerig, Antonio Sanchez, Steve Yadlowsky, Amy Shen, Amir Globerson, Lynette Webb, Sahil Dua, Dong Li, Surya Bhupatiraju, Dan Hurt, Haroon Qureshi, Ananth Agarwal, Tomer Shani, Matan Eyal, Anuj Khare, Shreyas Rammohan Belle, Lei Wang, Chetan Tekur, Mihir Sanjay Kale, Jinliang Wei, Ruoxin Sang, Brennan Saeta, Tyler Liechty, Yi Sun, Yao Zhao, Stephan Lee, Pandu Nayak, Doug Fritz, Manish Reddy Vuyyuru, John Aslanides, Nidhi Vyas, Martin Wicke, Xiao Ma, Evgenii Eltyshev, Nina Martin, Hardie Cate, James Manyika, Keyvan Amiri, Yelin Kim, Xi Xiong, Kai Kang, Florian Luisier, Nilesh Tripuraneni, David Madras, Mandy Guo, Austin Waters, Oliver Wang, Joshua Ainslie, Jason Baldridge, Han Zhang, Garima Pruthi, Jakob Bauer, Feng Yang, Riham Mansour, Jason Gelman, Yang Xu, George Polovets, Ji Liu, Honglong Cai, Warren Chen, XiangHai Sheng, Emily Xue, Sherjil Ozair, Christof Angermueller, Xiaowei Li, Anoop Sinha, Weiren Wang, Julia Wiesinger, Emmanouil Koukoumidis, Yuan Tian, Anand Iyer, Madhu Gurumurthy, Mark Goldenson, Parashar Shah, MK Blake, Hongkun Yu, Anthony Urbanowicz, Jennimaria Palomaki, Chrisantha Fernando, Ken Durden, Harsh Mehta, Nikola Momchev, Elahe Rahimtoroghi, Maria Georgaki, Amit Raul, Sebastian Ruder, Morgan Redshaw, Jinhyuk Lee, Denny Zhou, Komal Jalan, Dinghua Li, Blake Hechtman, Parker Schuh, Milad Nasr, Kieran Milan, Vladimir Mikulik, Juliana Franco, Tim Green, Nam Nguyen, Joe Kelley, Aroma Mahendru, Andrea Hu, Joshua Howland, Ben Vargas, Jeffrey Hui, Kshitij Bansal, Vikram Rao, Rakesh Ghiya, Emma Wang, Ke Ye, Jean Michel Sarr, Melanie Moranski Preston, Madeleine Elish, Steve Li, Aakash Kaku, Jigar Gupta, Ice Pasupat, Da-Cheng Juan, Milan Someswar, Tejvi M., Xinyun Chen, Aida Amini, Alex Fabrikant, Eric Chu, Xuanyi Dong, Amruta Muthal, Senaka Buthpitiya, Sarthak Jauhari, Nan Hua, Urvashi Khandelwal, Ayal Hitron, Jie Ren, Larissa Rinaldi, Shahar Drath, Avigail Dabush, Nan-Jiang Jiang, Harshal Godhia, Uli Sachs, Anthony Chen, Yicheng Fan, Hagai Taitelbaum, Hila Noga, Zhuyun Dai, James Wang, Chen Liang, Jenny Hamer, Chun-Sung Ferng, Chenel Elkind, Aviel Atias, Paulina Lee, Vít Listík, Mathias Carlen, Jan van de Kerkhof, Marcin Pikus, Krunoslav Zaher, Paul Müller, Sasha Zykova, Richard Stefanec, Vitaly Gatsko, Christoph Hirnschall, Ashwin Sethi, Xingyu Federico Xu, Chetan Ahuja, Beth Tsai, Anca Stefanoiu, Bo Feng, Keshav Dhandhania, Manish Katyal, Akshay Gupta, Atharva Parulekar, Divya Pitta, Jing Zhao, Vivaan Bhatia, Yashodha Bhavnani, Omar Alhadlaq, Xiaolin Li, Peter Danenberg, Dennis Tu, Alex Pine, Vera Filippova, Abhipso Ghosh, Ben Limonchik, Bhargava Urala, Chaitanya Krishna Lanka, Derik Clive, Yi Sun, Edward Li, Hao Wu, Kevin Hongtongsak, Ianna Li, Kalind Thakkar, Kuanysh Omarov, Kushal Majmundar, Michael Alverson, Michael Kucharski, Mohak Patel, Mudit Jain, Maksim Zabelin, Paolo Pelagatti, Rohan Kohli, Saurabh Kumar, Joseph Kim, Swetha Sankar, Vineet Shah, Lakshmi Ramachandruni, Xiangkai Zeng, Ben Bariach, Laura Weidinger, Tu Vu, Alek Andreev, Antoine He, Kevin Hui, Sheleem Kashem, Amar Subramanya, Sissie Hsiao, Demis Hassabis, Koray Kavukcuoglu, Adam Sadovsky, Quoc Le, Trevor Strohman, Yonghui Wu, Slav Petrov, Jeffrey Dean, and Oriol Vinyals. 2024a. Gemini: A family of highly capable multimodal models.

Gemma Team, Morgane Riviere, Shreya Pathak, Pier Giuseppe Sessa, Cassidy Hardin, Surya Bhupatiraju, Léonard Hussenot, Thomas Mesnard, Bobak Shahriari, Alexandre Ramé, Johan Ferret, Peter Liu, Pouya Tafti, Abe Friesen, Michelle Casbon, Sabela Ramos, Ravin Kumar, Charline Le Lan, Sammy Jerome, Anton Tsitsulin, Nino Vieillard, Piotr Stanczyk, Sertan Girgin, Nikola Momchev, Matt Hoffman, Shantanu Thakoor, Jean-Bastien Grill, Behnam Neyshabur, Olivier Bachem, Alanna Walton, Aliaksei Severyn, Alicia Parrish, Aliya Ahmad, Allen Hutchison, Alvin Abdagic, Amanda Carl, Amy Shen, Andy Brock, Andy Coenen, Anthony Laforge, Antonia Paterson, Ben Bastian, Bilal Piot, Bo Wu, Brandon Royal, Charlie Chen, Chintu Kumar, Chris Perry, Chris Welty, Christopher A. Choquette-Choo, Danila Sinopalnikov, David Weinberger, Dimple Vijaykumar, Dominika Rogozińska, Dustin Herbison, Elisa Bandy, Emma Wang, Eric Noland, Erica Moreira, Evan Senter, Evgenii Eltyshev, Francesco Visin, Gabriel Rasskin, Gary Wei, Glenn Cameron, Gus Martins, Hadi Hashemi, Hanna Klimczak-Plucińska, Harleen Batra, Harsh Dhand, Ivan Nardini, Jacinda Mein, Jack Zhou, James Svensson, Jeff Stanway, Jetha Chan, Jin Peng Zhou, Joana Carrasqueira, Joana Iljazi, Jocelyn Becker, Joe Fernandez, Joost van Amersfoort, Josh Gordon, Josh Lipschultz, Josh Newlan, Ju yeong Ji, Kareem Mohamed, Kartikeya Badola, Kat Black, Katie Millican, Keelin McDonell, Kelvin Nguyen, Kiranbir Sodhia, Kish Greene, Lars Lowe Sjoesund, Lauren Usui, Laurent Sifre, Lena Heuermann, Leticia Lago, Lilly McNealus, Livio Baldini Soares, Logan Kilpatrick, Lucas Dixon, Luciano Martins, Machel Reid, Manvinder Singh, Mark Iverson, Mar-

tin Görner, Mat Velloso, Mateo Wirth, Matt Davidow, Matt Miller, Matthew Rahtz, Matthew Watson, Meg Risdal, Mehran Kazemi, Michael Moynihan, Ming Zhang, Minsuk Kahng, Minwoo Park, Mofi Rahman, Mohit Khatwani, Natalie Dao, Nenshad Bardoliwalla, Nesh Devanathan, Neta Dumai, Nilay Chauhan, Oscar Wahltinez, Pankil Botarda, Parker Barnes, Paul Barham, Paul Michel, Pengchong Jin, Petko Georgiev, Phil Culliton, Pradeep Kuppala, Ramona Comanescu, Ramona Merhej, Reena Jana, Reza Ardeshir Rokni, Rishabh Agarwal, Ryan Mullins, Samaneh Saadat, Sara Mc Carthy, Sarah Perrin, Sébastien M. R. Arnold, Sebastian Krause, Shengyang Dai, Shruti Garg, Shruti Sheth, Sue Ronstrom, Susan Chan, Timothy Jordan, Ting Yu, Tom Eccles, Tom Hennigan, Tomas Kocisky, Tulsee Doshi, Vihan Jain, Vikas Yadav, Vilobh Meshram, Vishal Dharmadhikari, Warren Barkley, Wei Wei, Wenming Ye, Woohyun Han, Woosuk Kwon, Xiang Xu, Zhe Shen, Zhitao Gong, Zichuan Wei, Victor Cotruta, Phoebe Kirk, Anand Rao, Minh Giang, Ludovic Peran, Tris Warkentin, Eli Collins, Joelle Barral, Zoubin Ghahramani, Raia Hadsell, D. Sculley, Jeanine Banks, Anca Dragan, Slav Petrov, Oriol Vinyals, Jeff Dean, Demis Hassabis, Koray Kavukcuoglu, Clement Farabet, Elena Buchatskaya, Sebastian Borgeaud, Noah Fiedel, Armand Joulin, Kathleen Kenealy, Robert Dadashi, and Alek Andreev. 2024b. Gemma 2: Improving open language models at a practical size. *arXiv preprint arXiv: 2408.00118*.

A. Ustun, 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. *Annual Meeting of the Association for Computational Linguistics*.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, and Wenhu Chen. 2024. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark (published at neurips 2024 track datasets and benchmarks). *arXiv preprint arXiv: 2406.01574*.

Ishaan Watts, Varun Gumma, Aditya Yadavalli, Vivek Seshadri, Manohar Swaminathan, and Sunayana Sitaram. 2024. PARIKSHA: A large-scale investigation of human-LLM evaluator agreement on multilingual and multi-cultural data. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 7900–7932, Miami, Florida, USA. Association for Computational Linguistics.

Rowan Zellers, Ari Holtzman, Yonatan Bisk, Ali Farhadi, and Yejin Choi. 2019. Hellaswag: Can a machine really finish your sentence? *ACL*.

Jun Zhao, Zhihao Zhang, Luhui Gao, Qi Zhang, Tao Gui, and Xuanjing Huang. 2024. Llama beyond english: An empirical study on language capability transfer. *arXiv preprint arXiv: 2401.01055*.

Wanjun Zhong, Ruixiang Cui, Yiduo Guo, Yaobo Liang, Shuai Lu, Yanlin Wang, Amin Saied, Weizhu Chen, and Nan Duan. 2023. Agieval: A human-centric benchmark for evaluating foundation models. *arXiv preprint arXiv: 2304.06364*.

Appendix

A Details about different Exams

We collected our questions from over 40 exam types ranging from various National and state level civil service examinations to examinations conducted by various government and private organizations. Tables 6, 7 and 8

B Details about subject and language distribution.

Detailed analysis of topic and language distribution across languages can be found in Table 9 and Figure 6

C Details about the various models evaluated

Model details about the different models evaluated in this work is present in Table 10.

Organization	Years Covered
Railway Recruitment Board	2018-2019
Railway Recruitment Cell	2018
Dedicated Freight Corridor Corporation of India Limited	2016-2021
Intelligence Bureau	2012-2021
Union Public Service Commission	2010-2023

Table 6: Overview of various national-level exams and the corresponding years of coverage considered in MILU.

Organization	Years Covered
Punjab Police	2016-2022
Punjab State Power Corporation Limited	2018
Chandigarh Police	2018
Telangana Police	2015-2022
Andhra Pradesh Police	2016-2018
Tamil Nadu Public Service Commission	2010-2021
Tamil Nadu Uniformed Services Recruitment Board	2010-2022
Odisha Police	2022
Karnataka State Police	2019-2023
Madhya Pradesh Police	2016-2023
Delhi Police	2014-2022
Haryana Police	2021
Lucknow Metro Rail Corporation	2018
Delhi Metro Rail Corporation	2017-2020
Punjab Subordinate Service Selection Board	2022
Punjab State Transmission Corporation Limited	2016
Gujarat Metro Rail Corporation	2022
Kolkata Police	2023
Maharashtra Police	2021
Jharkhand Police	2015-2021

Table 7: Overview of various government and private organization exams and the corresponding years of coverage considered in MILU.

Organization	Years Covered
Staff Selection Commission	2017-2023
West Bengal Civil Service	2015-2021
Bihar Public Service Commission	2015-2020
Maharashtra Public Service Commission	2013-2023
Rajasthan Subordinate and Ministerial Services Selection Board	2016-2022
Odisha Public Service Commission	2016-2022
Uttar Pradesh Public Service Commission	2012-2023
Haryana Public Service Commission	2014-2022
Andhra Pradesh Public Service Commission	2017-2019
Chhattisgarh Public Service Commission	2014-2022
Jammu & Kashmir Public Service Commission	2021
Himachal Pradesh Public Service Commission	2015-2022
Jharkhand Public Service Commission	2021
Delhi Subordinate Services Selection Board	2019-2022

Table 8: Overview of various State-level civil services exams and the corresponding years of coverage considered in MILU.

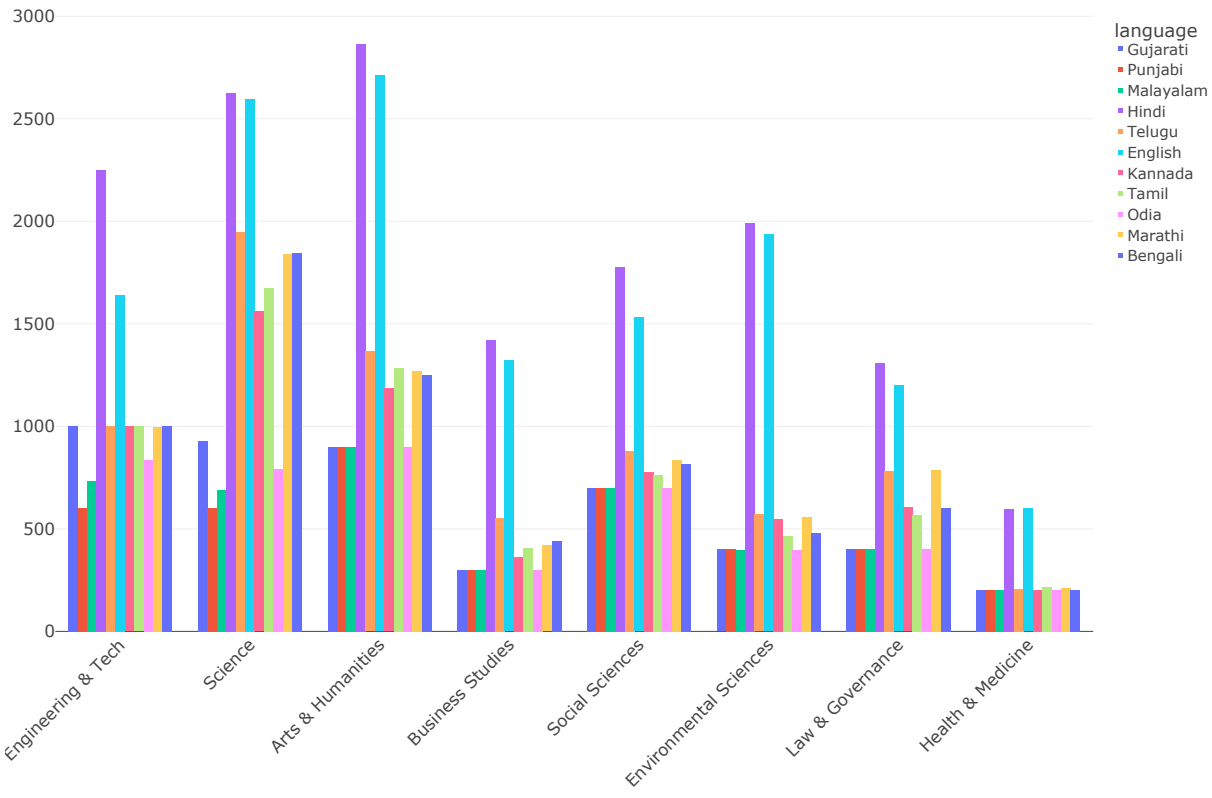


Figure 6: Distribution of tags across languages in MILU.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*	Total
Agriculture	100	497	99	497	100	99	100	100	100	96	99	1887
Anthropology	100	100	100	99	100	100	100	100	100	98	100	1097
Architecture and Design	100	100	100	100	100	100	100	100	100	100	100	1100
Arts and Culture	165	500	100	499	163	100	183	100	100	157	199	2266
Astronomy and Astrophysics	100	100	100	128	100	100	100	100	100	100	100	1128
Biology	284	499	100	499	244	99	335	100	100	314	318	2892
Business and Management	100	323	100	427	100	100	100	100	100	100	106	1656
Chemistry	364	499	116	499	306	100	379	104	100	337	401	3205
Computer Science	192	500	100	500	169	100	169	100	100	172	223	2325
Defense and Security	100	100	100	211	100	100	100	100	100	100	100	1211
Earth Sciences	100	440	100	499	100	100	100	99	100	100	100	1838
Economics	139	498	100	496	100	100	142	100	100	126	167	2068
Education	100	314	100	338	100	100	157	100	100	99	100	1608
Energy and Power	100	148	100	295	100	100	100	100	100	100	100	1343
Engineering	500	500	500	500	500	234	500	336	100	499	500	4669
Environmental Science	100	499	100	499	128	100	190	99	100	99	179	2093
Ethics and Human Rights	100	100	100	100	100	100	100	100	100	100	100	1100
Finance and Investment	202	500	100	500	162	100	177	100	100	179	280	2400
Food Science	100	100	100	99	100	100	100	100	100	100	100	1099
Geography	179	500	100	498	219	99	169	100	100	171	192	2327
Health and Medicine	100	499	100	499	100	100	112	100	100	115	107	1932
History	375	498	100	496	273	100	269	100	100	267	391	2969
Information Technology	100	100	100	179	100	100	100	100	100	100	100	1179
International Relations	100	120	100	257	100	100	100	100	100	100	100	1277
Language Studies	100	500	100	391	129	100	149	100	100	210	151	2030
Law and Ethics	148	500	100	499	157	100	232	100	100	127	160	2223
Literature and Linguistics	111	499	100	495	120	100	113	100	100	154	124	2016
Logical Reasoning	407	500	154	499	241	100	359	100	100	251	404	3115
Materials Science	100	263	100	458	100	100	100	100	100	100	100	1621
Media and Communication	100	100	100	147	100	100	100	100	100	100	100	1147
Music and Performing Arts	100	100	100	151	100	100	100	100	100	97	100	1148
Physics	500	500	359	500	500	190	500	288	100	499	500	4436
Politics and Governance	255	499	100	498	247	100	356	100	100	241	424	2920
Psychology	100	100	100	129	100	100	100	100	100	99	100	1128
Public Administration	100	99	99	99	99	100	99	100	99	100	100	1094
Religion and Spirituality	100	100	100	247	100	100	100	100	100	100	100	1247
Social Welfare and Development	100	112	100	193	100	100	100	100	100	99	100	1204
Sociology	115	500	100	500	100	100	159	100	100	110	169	2053
Sports and Recreation	202	500	100	500	178	100	177	100	100	156	210	2323
Technology and Innovation	100	500	100	500	100	100	100	100	100	100	100	1900
Transportation and Logistics	100	130	100	317	99	100	98	99	100	100	100	1343
TOTAL	6638	13536	4827	14837	6234	4321	6924	4525	4099	6372	7304	**79617**

Table 9: Detailed subject level statistics of MILU across different languages.

Model Name	Type	#Params	Link
GPT-4o	Instruct	-	Link
GPT-4o-mini	Instruct	-	Link
Gemini-1.5-Pro	Instruct	-	Link
Gemini-1.5-Flash	Instruct	-	Link
Krutrim-spectre-v2	Instruct	-	Link
meta-llama/Llama-3.2-1B	Base	1B	Link
meta-llama/Llama-3.2-1B-Instruct	Instruct	1B	Link
sarvamai/sarvam-2b-v0.5	Base	2B	Link
google/gemma-2-2b	Base	2B	Link
google/gemma-2-2b-it	Instruct	2B	Link
meta-llama/Llama-3.2-3B	Base	3B	Link
meta-llama/Llama-3.2-3B-Instruct	Instruct	3B	Link
Telugu-LLM-Labs/Indic-gemma-7b-finetuned-sft-Navarasa-2.0	Instruct	7B	Link
CohereForAI/aya-23-8B	Base	8B	Link
meta-llama/Llama-3.1-8B	Base	8B	Link
meta-llama/Llama-3.1-8B-Instruct	Instruct	8B	Link
google/gemma-2-9b	Base	9B	Link
google/gemma-2-9b-it	Instruct	9B	Link
google/gemma-2-27b	Base	27B	Link
google/gemma-2-27b-it	Instruct	27B	Link
CohereForAI/aya-23-35B	Base	35B	Link
meta-llama/Llama-3.1-70B	Base	70B	Link
meta-llama/Llama-3.1-70B-Instruct	Instruct	70B	Link
meta-llama/Llama-3.1-405B	Base	405B	Link
sarvamai/OpenHathi-7B-Hi-v0.1-Base	Base	7B	Link
ai4bharat/Airavata	Instruct	7B	Link
BhabhaAI/Gajendra-v0.1	Instruct	7B	Link
manishiitg/open-aditi-v6-llama3	Instruct	8B	Link
GenVRadmin/AryaBhatta-GemmaGenZ-Vikas-Merged	Instruct	7B	Link
nickmalhotra/ProjectIndus	Base	1B	Link
abhinand/telugu-llama-7b-instruct-v0.1	Instruct	7B	Link
Telugu-LLM-Labs/Telugu-Llama2-7B-v0-Base	Base	7B	Link
Telugu-LLM-Labs/Telugu-Llama2-7B-v0-Instruct	Instruct	7B	Link
Tensoic/Kan-LLaMA-7B-base	Base	7B	Link
Cognitive-Lab/Ambari-7B-base-v0.1	Base	7B	Link
Cognitive-Lab/Ambari-7B-Instruct-v0.1	Instruct	7B	Link
smallstepai/Misal-7B-base-v0.1	Base	7B	Link
smallstepai/Misal-7B-instruct-v0.1	Instruct	7B	Link
abhinand/tamil-llama-7b-instruct-v0.2	Instruct	7B	Link
abhinand/malayalam-llama-7b-instruct-v0.1	Instruct	7B	Link

Table 10: Details about the different models evaluated on MILU.

Subject	ta
Agriculture	29.00/ 28.00/ 26.00
Anthropology	21.00/ 23.00/ 26.00
Architecture and Design	19.00/ 24.00/ 20.00
Arts and Culture	31.52/ 32.73/ 29.70
Astronomy and Astrophysics	22.00/ 32.00/ 32.00
Biology	24.20/ 19.43/ 24.52
Business and Management	36.00/ 28.00/ 24.00
Chemistry	23.74/ 24.04/ 23.44
Computer Science	28.49/ 25.00/ 26.74
Defense and Security	29.00/ 24.00/ 32.00
Earth Sciences	28.00/ 25.00/ 28.00
Economics	24.60/ 20.63/ 19.05
Education	28.00/ 24.00/ 26.00
Energy and Power	29.00/ 20.00/ 17.00
Engineering	26.60/ 26.00/ 25.20
Environmental Science	23.00/ 19.00/ 20.00
Ethics and Human Rights	31.00/ 27.00/ 29.00
Finance and Investment	27.93/ 27.93/ 23.46
Food Science	27.00/ 21.00/ 21.00
Geography	22.86/ 26.29/ 24.57
Health and Medicine	22.41/ 24.14/ 24.14
History	26.18/ 24.73/ 24.00
Information Technology	28.00/ 23.00/ 25.00
International Relations	26.00/ 28.00/ 28.00
Language Studies	24.72/ 24.34/ 25.47
Law and Ethics	21.26/ 26.77/ 22.83
Literature and Linguistics	26.32/ 25.51/ 23.48
Logical Reasoning	20.72/ 21.12/ 25.10
Materials Science	25.00/ 19.00/ 18.00
Media and Communication	18.00/ 19.00/ 19.00
Music and Performing Arts	25.00/ 27.00/ 27.00
Physics	22.60/ 24.40/ 24.20
Politics and Governance	29.05/ 31.12/ 30.29
Psychology	32.00/ 28.00/ 27.00
Public Administration	26.00/ 21.00/ 19.00
Religion and Spirituality	29.00/ 31.00/ 33.00
Social Welfare and Development	21.00/ 23.00/ 21.00
Sociology	19.82/ 22.52/ 23.42
Sports and Recreation	32.05/ 30.13/ 30.77
Technology and Innovation	21.00/ 24.00/ 29.00
Transportation and Logistics	23.00/ 29.00/ 27.00

Table 11: Detailed subject-wise evaluation for ABHINAND/TAMIL-LLAMA-7B-INSTRUCT-V0.2 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	hi
Agriculture	26.56/ 26.96/ 29.98
Anthropology	30.30/ 26.26/ 30.30
Architecture and Design	28.00/ 31.00/ 26.00
Arts and Culture	25.45/ 24.65/ 27.45
Astronomy and Astrophysics	24.22/ 29.69/ 29.69
Biology	25.05/ 32.06/ 32.06
Business and Management	27.87/ 29.51/ 26.70
Chemistry	29.66/ 31.46/ 31.26
Computer Science	28.60/ 30.60/ 30.60
Defense and Security	28.91/ 22.27/ 24.64
Earth Sciences	26.05/ 27.45/ 26.25
Economics	26.81/ 25.20/ 27.02
Education	27.22/ 25.15/ 26.04
Energy and Power	31.53/ 29.83/ 31.53
Engineering	24.00/ 25.60/ 26.40
Environmental Science	23.65/ 25.85/ 25.65
Ethics and Human Rights	33.00/ 25.00/ 29.00
Finance and Investment	26.60/ 28.20/ 30.20
Food Science	24.24/ 31.31/ 33.33
Geography	26.31/ 27.11/ 26.71
Health and Medicine	29.06/ 32.87/ 32.67
History	27.02/ 28.63/ 27.62
Information Technology	32.96/ 34.64/ 32.96
International Relations	31.91/ 31.91/ 31.52
Language Studies	25.05/ 28.86/ 25.25
Law and Ethics	29.20/ 29.40/ 29.20
Literature and Linguistics	32.87/ 31.86/ 30.06
Logical Reasoning	26.45/ 27.45/ 27.25
Materials Science	24.67/ 28.60/ 25.55
Media and Communication	23.81/ 25.85/ 24.49
Music and Performing Arts	25.17/ 24.50/ 25.83
Physics	26.20/ 25.20/ 23.40
Politics and Governance	28.71/ 27.71/ 26.91
Psychology	19.38/ 23.26/ 23.26
Public Administration	39.39/ 41.41/ 40.40
Religion and Spirituality	27.82/ 29.03/ 29.44
Social Welfare and Development	22.28/ 23.32/ 27.98
Sociology	23.40/ 23.20/ 24.20
Sports and Recreation	29.40/ 30.20/ 31.60
Technology and Innovation	26.20/ 28.60/ 30.60
Transportation and Logistics	25.87/ 26.50/ 27.13

Table 12: Detailed subject-wise evaluation for AI4BHARAT/AIRAVATA on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	hi
Agriculture	27.97/ 27.36/ 28.97
Anthropology	27.27/ 26.26/ 20.20
Architecture and Design	23.00/ 20.00/ 25.00
Arts and Culture	28.06/ 28.26/ 24.25
Astronomy and Astrophysics	27.34/ 30.47/ 35.94
Biology	31.06/ 30.06/ 30.46
Business and Management	26.23/ 24.12/ 29.27
Chemistry	28.26/ 25.65/ 30.06
Computer Science	25.80/ 28.80/ 29.80
Defense and Security	23.22/ 30.81/ 28.91
Earth Sciences	23.65/ 26.45/ 29.46
Economics	27.02/ 27.82/ 29.44
Education	26.33/ 28.99/ 31.07
Energy and Power	27.80/ 26.44/ 29.83
Engineering	20.40/ 20.20/ 24.00
Environmental Science	25.45/ 26.25/ 30.06
Ethics and Human Rights	27.00/ 24.00/ 24.00
Finance and Investment	23.60/ 25.00/ 29.20
Food Science	32.32/ 37.37/ 28.28
Geography	26.71/ 25.90/ 28.71
Health and Medicine	30.46/ 30.46/ 30.66
History	26.01/ 28.02/ 29.03
Information Technology	27.93/ 31.28/ 36.87
International Relations	25.68/ 25.29/ 24.12
Language Studies	23.05/ 24.85/ 26.65
Law and Ethics	26.60/ 26.60/ 28.80
Literature and Linguistics	27.45/ 27.25/ 29.26
Logical Reasoning	22.24/ 27.05/ 26.05
Materials Science	26.86/ 29.26/ 29.48
Media and Communication	23.81/ 26.53/ 32.65
Music and Performing Arts	25.17/ 22.52/ 28.48
Physics	24.20/ 25.40/ 27.80
Politics and Governance	23.09/ 24.70/ 27.11
Psychology	26.36/ 25.58/ 25.58
Public Administration	36.36/ 34.34/ 32.32
Religion and Spirituality	25.81/ 22.98/ 29.03
Social Welfare and Development	25.91/ 24.87/ 25.91
Sociology	26.20/ 25.60/ 24.80
Sports and Recreation	26.00/ 26.40/ 26.60
Technology and Innovation	28.80/ 28.20/ 30.40
Transportation and Logistics	25.24/ 23.97/ 24.29

Table 13: Detailed subject-wise evaluation for BHABHAAI/GAJENDRA-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	kn
Agriculture	25.00/ 28.00/ 23.00
Anthropology	25.00/ 26.00/ 26.00
Architecture and Design	25.00/ 31.00/ 28.00
Arts and Culture	26.99/ 29.45/ 24.54
Astronomy and Astrophysics	28.00/ 30.00/ 30.00
Biology	29.10/ 26.23/ 28.28
Business and Management	36.00/ 25.00/ 26.00
Chemistry	31.37/ 25.82/ 28.43
Computer Science	28.40/ 33.14/ 33.14
Defense and Security	27.00/ 31.00/ 33.00
Earth Sciences	29.00/ 31.00/ 33.00
Economics	29.00/ 32.00/ 37.00
Education	28.00/ 30.00/ 29.00
Energy and Power	26.00/ 22.00/ 19.00
Engineering	24.60/ 27.20/ 28.20
Environmental Science	21.88/ 23.44/ 29.69
Ethics and Human Rights	29.00/ 19.00/ 26.00
Finance and Investment	29.63/ 28.40/ 31.48
Food Science	23.00/ 30.00/ 38.00
Geography	27.85/ 27.40/ 28.77
Health and Medicine	23.00/ 29.00/ 24.00
History	24.54/ 28.57/ 29.67
Information Technology	26.00/ 35.00/ 31.00
International Relations	32.00/ 36.00/ 26.00
Language Studies	28.68/ 19.38/ 27.91
Law and Ethics	25.48/ 22.29/ 26.75
Literature and Linguistics	18.33/ 23.33/ 23.33
Logical Reasoning	21.99/ 21.58/ 25.73
Materials Science	28.00/ 24.00/ 23.00
Media and Communication	32.00/ 22.00/ 28.00
Music and Performing Arts	32.00/ 29.00/ 30.00
Physics	26.40/ 26.60/ 26.00
Politics and Governance	27.53/ 23.89/ 25.51
Psychology	23.00/ 34.00/ 31.00
Public Administration	36.36/ 30.30/ 32.32
Religion and Spirituality	27.00/ 27.00/ 31.00
Social Welfare and Development	28.00/ 26.00/ 32.00
Sociology	24.00/ 33.00/ 30.00
Sports and Recreation	22.47/ 29.78/ 27.53
Technology and Innovation	25.00/ 29.00/ 32.00
Transportation and Logistics	25.25/ 26.26/ 29.29

Table 14: Detailed subject-wise evaluation for COGNITIVE-LAB/AMBARI-7B-BASE-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	kn
Agriculture	31.00/ 26.00/ 23.00
Anthropology	29.00/ 23.00/ 22.00
Architecture and Design	29.00/ 21.00/ 23.00
Arts and Culture	25.15/ 28.83/ 30.06
Astronomy and Astrophysics	29.00/ 24.00/ 33.00
Biology	27.87/ 25.82/ 29.10
Business and Management	38.00/ 31.00/ 28.00
Chemistry	31.37/ 30.07/ 31.05
Computer Science	24.26/ 25.44/ 30.77
Defense and Security	27.00/ 22.00/ 19.00
Earth Sciences	26.00/ 28.00/ 27.00
Economics	31.00/ 32.00/ 29.00
Education	32.00/ 30.00/ 32.00
Energy and Power	26.00/ 21.00/ 26.00
Engineering	23.20/ 26.20/ 28.20
Environmental Science	25.78/ 28.12/ 26.56
Ethics and Human Rights	29.00/ 25.00/ 28.00
Finance and Investment	23.46/ 25.31/ 27.78
Food Science	22.00/ 21.00/ 33.00
Geography	31.51/ 24.20/ 31.05
Health and Medicine	17.00/ 22.00/ 21.00
History	27.11/ 28.57/ 30.04
Information Technology	33.00/ 32.00/ 35.00
International Relations	22.00/ 27.00/ 26.00
Language Studies	25.58/ 30.23/ 23.26
Law and Ethics	28.03/ 23.57/ 21.66
Literature and Linguistics	23.33/ 22.50/ 24.17
Logical Reasoning	21.16/ 21.16/ 24.07
Materials Science	22.00/ 29.00/ 26.00
Media and Communication	30.00/ 25.00/ 26.00
Music and Performing Arts	29.00/ 27.00/ 25.00
Physics	26.40/ 23.60/ 24.20
Politics and Governance	22.67/ 29.15/ 25.51
Psychology	17.00/ 18.00/ 25.00
Public Administration	24.24/ 23.23/ 28.28
Religion and Spirituality	27.00/ 26.00/ 29.00
Social Welfare and Development	29.00/ 25.00/ 27.00
Sociology	18.00/ 19.00/ 22.00
Sports and Recreation	19.10/ 24.16/ 24.16
Technology and Innovation	32.00/ 23.00/ 25.00
Transportation and Logistics	32.32/ 31.31/ 27.27

Table 15: Detailed subject-wise evaluation for COGNITIVE-LAB/AMBARI-7B-INSTRUCT-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	hi
Agriculture	29.38/ 30.18/ 36.62
Anthropology	37.37/ 32.32/ 40.40
Architecture and Design	26.00/ 28.00/ 35.00
Arts and Culture	26.25/ 29.46/ 40.48
Astronomy and Astrophysics	37.50/ 42.19/ 46.88
Biology	35.27/ 35.07/ 50.30
Business and Management	31.15/ 32.32/ 38.64
Chemistry	28.86/ 38.68/ 48.30
Computer Science	32.60/ 34.00/ 36.60
Defense and Security	28.91/ 36.02/ 35.55
Earth Sciences	32.06/ 28.66/ 40.48
Economics	27.42/ 32.06/ 41.53
Education	23.96/ 27.51/ 30.77
Energy and Power	26.10/ 31.19/ 43.39
Engineering	25.80/ 26.60/ 33.20
Environmental Science	29.66/ 30.26/ 44.89
Ethics and Human Rights	30.00/ 31.00/ 41.00
Finance and Investment	27.40/ 34.20/ 38.00
Food Science	24.24/ 33.33/ 47.47
Geography	31.33/ 30.72/ 42.57
Health and Medicine	34.27/ 34.67/ 48.70
History	27.02/ 32.86/ 36.90
Information Technology	34.64/ 37.99/ 50.84
International Relations	31.52/ 31.13/ 42.41
Language Studies	22.65/ 27.05/ 29.26
Law and Ethics	28.00/ 29.80/ 38.60
Literature and Linguistics	24.85/ 28.86/ 33.47
Logical Reasoning	25.45/ 27.25/ 29.66
Materials Science	31.88/ 30.35/ 35.59
Media and Communication	25.85/ 26.53/ 40.82
Music and Performing Arts	21.19/ 30.46/ 33.77
Physics	32.60/ 32.80/ 40.00
Politics and Governance	25.50/ 27.71/ 41.97
Psychology	23.26/ 22.48/ 42.64
Public Administration	30.30/ 27.27/ 36.36
Religion and Spirituality	28.23/ 33.87/ 42.34
Social Welfare and Development	26.94/ 27.98/ 35.75
Sociology	23.80/ 26.60/ 28.00
Sports and Recreation	29.00/ 34.80/ 43.80
Technology and Innovation	31.80/ 30.80/ 41.60
Transportation and Logistics	23.97/ 23.97/ 30.28

Table 16: Detailed subject-wise evaluation for GENVRADMIN/ARYABHATTA-GEMMAGENZ-VIKAS-MERGED on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	hi
Agriculture	29.38/ 28.97/ 27.16
Anthropology	36.36/ 36.36/ 28.28
Architecture and Design	34.00/ 31.00/ 33.00
Arts and Culture	34.07/ 33.67/ 34.87
Astronomy and Astrophysics	42.97/ 41.41/ 37.50
Biology	36.67/ 38.08/ 36.07
Business and Management	32.55/ 31.15/ 29.04
Chemistry	37.47/ 34.67/ 31.66
Computer Science	37.40/ 34.00/ 35.80
Defense and Security	30.33/ 28.44/ 27.49
Earth Sciences	31.86/ 31.46/ 33.07
Economics	34.68/ 33.06/ 30.44
Education	31.36/ 29.88/ 30.77
Energy and Power	34.58/ 32.20/ 32.54
Engineering	28.60/ 27.60/ 27.80
Environmental Science	31.86/ 31.26/ 30.26
Ethics and Human Rights	36.00/ 35.00/ 36.00
Finance and Investment	29.80/ 30.20/ 28.40
Food Science	36.36/ 36.36/ 28.28
Geography	33.33/ 32.73/ 30.12
Health and Medicine	44.09/ 40.28/ 38.88
History	39.92/ 38.10/ 36.49
Information Technology	47.49/ 47.49/ 43.02
International Relations	40.08/ 40.47/ 34.63
Language Studies	29.86/ 29.06/ 27.45
Law and Ethics	41.60/ 39.20/ 40.00
Literature and Linguistics	29.66/ 30.26/ 29.86
Logical Reasoning	30.46/ 28.46/ 29.06
Materials Science	29.26/ 31.00/ 30.35
Media and Communication	36.73/ 34.01/ 31.29
Music and Performing Arts	36.42/ 32.45/ 34.44
Physics	33.60/ 29.80/ 29.00
Politics and Governance	36.95/ 36.75/ 32.33
Psychology	41.86/ 37.21/ 33.33
Public Administration	32.32/ 33.33/ 34.34
Religion and Spirituality	29.84/ 31.45/ 29.44
Social Welfare and Development	30.05/ 30.57/ 30.57
Sociology	30.00/ 30.00/ 27.40
Sports and Recreation	32.00/ 30.80/ 28.60
Technology and Innovation	36.60/ 36.80/ 34.00
Transportation and Logistics	30.60/ 29.65/ 28.71

Table 17: Detailed subject-wise evaluation for MANISHIITG/OPEN-ADITI-V6-LLAMA3 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	hi
Agriculture	22.74/ 22.74/ 21.93
Anthropology	25.25/ 24.24/ 24.24
Architecture and Design	32.00/ 28.00/ 29.00
Arts and Culture	24.85/ 26.05/ 26.25
Astronomy and Astrophysics	35.94/ 35.94/ 32.03
Biology	24.85/ 24.45/ 24.05
Business and Management	25.29/ 25.06/ 25.76
Chemistry	25.85/ 25.25/ 24.45
Computer Science	26.40/ 27.00/ 25.20
Defense and Security	27.01/ 24.64/ 25.12
Earth Sciences	25.45/ 24.05/ 23.85
Economics	28.02/ 27.42/ 27.42
Education	26.04/ 24.85/ 24.85
Energy and Power	23.73/ 24.07/ 25.08
Engineering	23.60/ 23.40/ 24.00
Environmental Science	28.86/ 28.26/ 27.86
Ethics and Human Rights	30.00/ 31.00/ 22.00
Finance and Investment	24.00/ 25.00/ 25.40
Food Science	27.27/ 29.29/ 27.27
Geography	27.11/ 25.90/ 26.51
Health and Medicine	25.45/ 26.05/ 23.85
History	24.80/ 25.81/ 26.21
Information Technology	26.26/ 27.37/ 25.14
International Relations	26.85/ 28.40/ 29.96
Language Studies	21.44/ 21.44/ 22.24
Law and Ethics	24.80/ 24.80/ 22.80
Literature and Linguistics	23.65/ 23.85/ 22.24
Logical Reasoning	27.05/ 27.86/ 27.45
Materials Science	28.17/ 29.04/ 28.60
Media and Communication	28.57/ 31.29/ 26.53
Music and Performing Arts	21.19/ 22.52/ 23.84
Physics	24.40/ 24.00/ 21.60
Politics and Governance	26.71/ 25.30/ 22.29
Psychology	27.91/ 27.91/ 26.36
Public Administration	26.26/ 27.27/ 28.28
Religion and Spirituality	26.21/ 29.44/ 27.82
Social Welfare and Development	26.94/ 27.98/ 26.42
Sociology	27.00/ 28.00/ 26.40
Sports and Recreation	27.80/ 27.40/ 26.60
Technology and Innovation	23.80/ 25.20/ 26.40
Transportation and Logistics	23.97/ 26.81/ 24.92

Table 18: Detailed subject-wise evaluation for NICKMALHOTRA/PROJECTINDUS on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	hi
Agriculture	24.35/ 30.18/ 29.98
Anthropology	20.20/ 25.25/ 26.26
Architecture and Design	23.00/ 32.00/ 27.00
Arts and Culture	27.66/ 27.05/ 28.06
Astronomy and Astrophysics	24.22/ 31.25/ 33.59
Biology	27.45/ 26.65/ 28.86
Business and Management	26.70/ 25.53/ 29.74
Chemistry	26.65/ 26.25/ 28.86
Computer Science	26.80/ 29.20/ 25.80
Defense and Security	27.96/ 24.64/ 27.96
Earth Sciences	22.04/ 24.65/ 26.85
Economics	26.21/ 30.85/ 33.27
Education	24.26/ 26.92/ 25.15
Energy and Power	30.17/ 31.19/ 32.54
Engineering	22.20/ 21.40/ 25.40
Environmental Science	25.05/ 25.65/ 27.05
Ethics and Human Rights	31.00/ 33.00/ 36.00
Finance and Investment	27.00/ 27.80/ 26.20
Food Science	20.20/ 24.24/ 30.30
Geography	24.30/ 27.91/ 27.11
Health and Medicine	33.07/ 28.66/ 30.86
History	29.44/ 30.24/ 32.66
Information Technology	25.70/ 30.17/ 35.20
International Relations	26.46/ 28.40/ 32.30
Language Studies	25.45/ 24.85/ 28.46
Law and Ethics	28.60/ 28.80/ 34.20
Literature and Linguistics	26.25/ 25.85/ 27.45
Logical Reasoning	28.86/ 23.85/ 28.26
Materials Science	22.93/ 24.45/ 25.76
Media and Communication	29.93/ 27.89/ 24.49
Music and Performing Arts	27.15/ 24.50/ 24.50
Physics	26.00/ 27.20/ 29.60
Politics and Governance	23.49/ 24.90/ 27.91
Psychology	24.81/ 23.26/ 26.36
Public Administration	29.29/ 33.33/ 32.32
Religion and Spirituality	28.63/ 26.21/ 30.24
Social Welfare and Development	27.46/ 25.91/ 26.42
Sociology	28.80/ 28.40/ 25.80
Sports and Recreation	28.80/ 29.00/ 31.40
Technology and Innovation	29.00/ 29.60/ 30.40
Transportation and Logistics	29.02/ 24.29/ 27.44

Table 19: Detailed subject-wise evaluation for SARVAMAI/OPENHATHI-7B-HI-V0.1-BASE on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	kn
Agriculture	23.00/ 29.00/ 28.00
Anthropology	30.00/ 41.00/ 38.00
Architecture and Design	26.00/ 27.00/ 25.00
Arts and Culture	26.38/ 26.99/ 29.45
Astronomy and Astrophysics	26.00/ 30.00/ 21.00
Biology	23.77/ 29.51/ 28.69
Business and Management	30.00/ 28.00/ 28.00
Chemistry	28.10/ 27.12/ 25.82
Computer Science	29.59/ 36.09/ 36.69
Defense and Security	25.00/ 26.00/ 28.00
Earth Sciences	24.00/ 24.00/ 24.00
Economics	30.00/ 32.00/ 29.00
Education	28.00/ 38.00/ 36.00
Energy and Power	28.00/ 22.00/ 25.00
Engineering	23.00/ 23.60/ 24.80
Environmental Science	23.44/ 32.81/ 30.47
Ethics and Human Rights	24.00/ 28.00/ 30.00
Finance and Investment	26.54/ 31.48/ 29.63
Food Science	27.00/ 29.00/ 31.00
Geography	32.88/ 31.96/ 26.94
Health and Medicine	17.00/ 27.00/ 30.00
History	27.11/ 26.01/ 31.14
Information Technology	33.00/ 37.00/ 35.00
International Relations	25.00/ 25.00/ 29.00
Language Studies	26.36/ 19.38/ 24.03
Law and Ethics	26.11/ 26.75/ 24.84
Literature and Linguistics	22.50/ 25.00/ 21.67
Logical Reasoning	22.41/ 23.24/ 22.82
Materials Science	25.00/ 26.00/ 24.00
Media and Communication	28.00/ 32.00/ 32.00
Music and Performing Arts	30.00/ 28.00/ 29.00
Physics	26.20/ 28.20/ 26.60
Politics and Governance	26.72/ 29.55/ 30.36
Psychology	29.00/ 29.00/ 25.00
Public Administration	31.31/ 29.29/ 31.31
Religion and Spirituality	26.00/ 27.00/ 35.00
Social Welfare and Development	32.00/ 34.00/ 31.00
Sociology	29.00/ 28.00/ 30.00
Sports and Recreation	26.40/ 30.90/ 29.21
Technology and Innovation	29.00/ 30.00/ 34.00
Transportation and Logistics	36.36/ 31.31/ 32.32

Table 20: Detailed subject-wise evaluation for TENSOIC/KAN-LLAMA-7B-BASE on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	ml
Agriculture	22.22/ 18.18/ 25.25
Anthropology	24.00/ 20.00/ 21.00
Architecture and Design	22.00/ 21.00/ 23.00
Arts and Culture	25.00/ 20.00/ 23.00
Astronomy and Astrophysics	27.00/ 28.00/ 26.00
Biology	24.24/ 27.27/ 26.26
Business and Management	26.00/ 24.00/ 28.00
Chemistry	25.00/ 26.00/ 26.00
Computer Science	24.00/ 28.00/ 27.00
Defense and Security	18.00/ 18.00/ 18.00
Earth Sciences	30.00/ 31.00/ 34.00
Economics	23.00/ 26.00/ 27.00
Education	23.00/ 21.00/ 22.00
Energy and Power	32.00/ 28.00/ 30.00
Engineering	27.35/ 27.78/ 25.64
Environmental Science	30.00/ 31.00/ 28.00
Ethics and Human Rights	29.00/ 19.00/ 24.00
Finance and Investment	35.00/ 31.00/ 35.00
Food Science	27.00/ 22.00/ 20.00
Geography	30.30/ 27.27/ 24.24
Health and Medicine	23.00/ 23.00/ 20.00
History	30.00/ 22.00/ 25.00
Information Technology	35.00/ 37.00/ 37.00
International Relations	23.00/ 25.00/ 24.00
Language Studies	21.00/ 17.00/ 20.00
Law and Ethics	27.00/ 30.00/ 27.00
Literature and Linguistics	28.00/ 24.00/ 25.00
Logical Reasoning	29.00/ 26.00/ 27.00
Materials Science	28.00/ 25.00/ 20.00
Media and Communication	24.00/ 24.00/ 22.00
Music and Performing Arts	25.00/ 21.00/ 24.00
Physics	25.79/ 27.89/ 30.53
Politics and Governance	30.00/ 28.00/ 31.00
Psychology	26.00/ 25.00/ 24.00
Public Administration	23.00/ 25.00/ 20.00
Religion and Spirituality	27.00/ 24.00/ 26.00
Social Welfare and Development	34.00/ 33.00/ 30.00
Sociology	20.00/ 21.00/ 20.00
Sports and Recreation	31.00/ 27.00/ 25.00
Technology and Innovation	25.00/ 26.00/ 26.00
Transportation and Logistics	24.00/ 25.00/ 32.00

Table 21: Detailed subject-wise evaluation for ABHINAND/MALAYALAM-LLAMA-7B-INSTRUCT-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	te
Agriculture	25.00/ 23.00/ 21.00
Anthropology	33.00/ 30.00/ 33.00
Architecture and Design	24.00/ 20.00/ 20.00
Arts and Culture	28.64/ 29.65/ 28.14
Astronomy and Astrophysics	25.00/ 25.00/ 25.00
Biology	24.53/ 23.27/ 22.01
Business and Management	36.79/ 36.79/ 39.62
Chemistry	25.69/ 27.43/ 27.93
Computer Science	25.56/ 30.04/ 28.70
Defense and Security	26.00/ 25.00/ 25.00
Earth Sciences	27.00/ 29.00/ 29.00
Economics	24.55/ 28.74/ 29.94
Education	22.00/ 22.00/ 28.00
Energy and Power	26.00/ 26.00/ 24.00
Engineering	30.40/ 29.20/ 28.00
Environmental Science	21.23/ 23.46/ 23.46
Ethics and Human Rights	21.00/ 22.00/ 20.00
Finance and Investment	26.79/ 24.64/ 26.79
Food Science	28.00/ 25.00/ 21.00
Geography	28.65/ 28.12/ 23.44
Health and Medicine	28.04/ 30.84/ 28.04
History	31.71/ 32.48/ 29.67
Information Technology	30.00/ 38.00/ 33.00
International Relations	27.00/ 29.00/ 31.00
Language Studies	27.01/ 28.16/ 27.59
Law and Ethics	27.50/ 24.37/ 26.87
Literature and Linguistics	25.35/ 21.83/ 18.31
Logical Reasoning	24.75/ 24.50/ 25.25
Materials Science	26.00/ 24.00/ 27.00
Media and Communication	23.00/ 23.00/ 25.00
Music and Performing Arts	28.00/ 23.00/ 30.00
Physics	22.80/ 25.20/ 28.20
Politics and Governance	27.59/ 27.59/ 29.48
Psychology	28.00/ 22.00/ 22.00
Public Administration	35.00/ 31.00/ 28.00
Religion and Spirituality	22.00/ 24.00/ 24.00
Social Welfare and Development	30.00/ 27.00/ 29.00
Sociology	21.18/ 20.59/ 18.82
Sports and Recreation	24.76/ 25.71/ 26.67
Technology and Innovation	29.00/ 24.00/ 22.00
Transportation and Logistics	26.00/ 22.00/ 26.00

Table 22: Detailed subject-wise evaluation for ABHINAND/TELUGU-LLAMA-7B-INSTRUCT-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	mr
Agriculture	20.00/ 20.00/ 17.00
Anthropology	15.00/ 20.00/ 20.00
Architecture and Design	18.00/ 19.00/ 19.00
Arts and Culture	20.22/ 20.77/ 22.95
Astronomy and Astrophysics	20.00/ 19.00/ 19.00
Biology	24.18/ 23.58/ 23.28
Business and Management	20.00/ 21.00/ 22.00
Chemistry	23.75/ 25.33/ 25.59
Computer Science	26.63/ 25.44/ 27.22
Defense and Security	24.00/ 22.00/ 19.00
Earth Sciences	23.00/ 29.00/ 28.00
Economics	28.17/ 26.76/ 28.17
Education	24.20/ 23.57/ 22.29
Energy and Power	27.00/ 25.00/ 22.00
Engineering	21.60/ 21.60/ 22.20
Environmental Science	23.16/ 23.16/ 26.84
Ethics and Human Rights	20.00/ 18.00/ 17.00
Finance and Investment	23.16/ 24.29/ 24.29
Food Science	25.00/ 27.00/ 21.00
Geography	21.30/ 20.71/ 18.93
Health and Medicine	20.54/ 27.68/ 23.21
History	24.91/ 24.54/ 25.65
Information Technology	25.00/ 24.00/ 23.00
International Relations	24.00/ 27.00/ 23.00
Language Studies	21.48/ 23.49/ 23.49
Law and Ethics	21.12/ 21.55/ 23.71
Literature and Linguistics	20.35/ 23.01/ 22.12
Logical Reasoning	26.18/ 24.79/ 26.74
Materials Science	16.00/ 19.00/ 16.00
Media and Communication	20.00/ 17.00/ 17.00
Music and Performing Arts	27.00/ 27.00/ 31.00
Physics	25.00/ 24.40/ 23.40
Politics and Governance	21.07/ 20.22/ 22.75
Psychology	24.00/ 25.00/ 24.00
Public Administration	24.24/ 25.25/ 18.18
Religion and Spirituality	19.00/ 22.00/ 24.00
Social Welfare and Development	18.00/ 19.00/ 22.00
Sociology	18.24/ 19.50/ 21.38
Sports and Recreation	25.99/ 22.03/ 22.03
Technology and Innovation	18.00/ 20.00/ 18.00
Transportation and Logistics	21.43/ 26.53/ 24.49

Table 23: Detailed subject-wise evaluation for SMALLSTEPAI/MISAL-7B-INSTRUCT-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	mr
Agriculture	34.00/ 28.00/ 26.00
Anthropology	27.00/ 35.00/ 37.00
Architecture and Design	21.00/ 18.00/ 19.00
Arts and Culture	25.68/ 25.14/ 26.23
Astronomy and Astrophysics	25.00/ 23.00/ 24.00
Biology	23.88/ 25.97/ 25.07
Business and Management	19.00/ 27.00/ 26.00
Chemistry	29.82/ 27.70/ 29.02
Computer Science	20.12/ 24.26/ 23.08
Defense and Security	23.00/ 26.00/ 22.00
Earth Sciences	30.00/ 28.00/ 25.00
Economics	27.46/ 29.58/ 30.99
Education	38.22/ 28.66/ 28.66
Energy and Power	25.00/ 24.00/ 28.00
Engineering	21.60/ 28.00/ 26.80
Environmental Science	26.32/ 30.53/ 32.11
Ethics and Human Rights	20.00/ 22.00/ 25.00
Finance and Investment	25.99/ 22.03/ 24.29
Food Science	19.00/ 17.00/ 21.00
Geography	24.85/ 21.30/ 23.67
Health and Medicine	29.46/ 33.93/ 26.79
History	26.77/ 27.14/ 28.25
Information Technology	19.00/ 17.00/ 21.00
International Relations	30.00/ 30.00/ 25.00
Language Studies	22.15/ 22.15/ 20.81
Law and Ethics	25.00/ 29.31/ 32.33
Literature and Linguistics	22.12/ 27.43/ 29.20
Logical Reasoning	30.64/ 30.64/ 30.36
Materials Science	37.00/ 27.00/ 24.00
Media and Communication	24.00/ 31.00/ 31.00
Music and Performing Arts	23.00/ 23.00/ 30.00
Physics	26.00/ 24.00/ 23.00
Politics and Governance	30.62/ 25.84/ 26.12
Psychology	30.00/ 22.00/ 21.00
Public Administration	29.29/ 25.25/ 25.25
Religion and Spirituality	18.00/ 26.00/ 28.00
Social Welfare and Development	25.00/ 18.00/ 23.00
Sociology	25.79/ 19.50/ 16.35
Sports and Recreation	26.55/ 24.29/ 29.38
Technology and Innovation	28.00/ 30.00/ 31.00
Transportation and Logistics	27.55/ 25.51/ 28.57

Table 24: Detailed subject-wise evaluation for SMALLSTEPAI/MISAL-7B-BASE-V0.1 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Subject	te
Agriculture	25.00/ 21.00/ 29.00
Anthropology	22.00/ 19.00/ 21.00
Architecture and Design	19.00/ 14.00/ 17.00
Arts and Culture	24.62/ 21.11/ 22.11
Astronomy and Astrophysics	31.00/ 27.00/ 28.00
Biology	19.81/ 19.50/ 22.33
Business and Management	17.92/ 16.98/ 18.87
Chemistry	27.93/ 27.43/ 29.43
Computer Science	27.80/ 27.35/ 30.49
Defense and Security	35.00/ 29.00/ 32.00
Earth Sciences	31.00/ 32.00/ 31.00
Economics	26.95/ 23.35/ 23.35
Education	30.00/ 24.00/ 23.00
Energy and Power	24.00/ 20.00/ 26.00
Engineering	26.80/ 27.00/ 26.60
Environmental Science	26.82/ 23.46/ 24.02
Ethics and Human Rights	17.00/ 18.00/ 19.00
Finance and Investment	25.00/ 22.50/ 24.29
Food Science	22.00/ 24.00/ 25.00
Geography	26.56/ 26.56/ 27.08
Health and Medicine	25.23/ 28.04/ 29.91
History	29.16/ 26.85/ 26.09
Information Technology	32.00/ 30.00/ 33.00
International Relations	16.00/ 21.00/ 21.00
Language Studies	28.74/ 27.59/ 29.31
Law and Ethics	27.50/ 20.62/ 23.75
Literature and Linguistics	28.17/ 23.94/ 25.35
Logical Reasoning	26.49/ 26.73/ 26.24
Materials Science	24.00/ 27.00/ 31.00
Media and Communication	33.00/ 36.00/ 34.00
Music and Performing Arts	23.00/ 17.00/ 25.00
Physics	26.80/ 25.40/ 27.40
Politics and Governance	29.25/ 28.77/ 27.36
Psychology	26.00/ 26.00/ 23.00
Public Administration	27.00/ 32.00/ 34.00
Religion and Spirituality	20.00/ 27.00/ 26.00
Social Welfare and Development	32.00/ 27.00/ 30.00
Sociology	25.29/ 24.71/ 25.88
Sports and Recreation	21.43/ 22.38/ 21.90
Technology and Innovation	22.00/ 21.00/ 21.00
Transportation and Logistics	22.00/ 24.00/ 27.00

Table 25: Detailed subject-wise evaluation for TELUGU-LLM-LABS/TELUGU-LLAMA2-7B-V0-BASE on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	70	79.5	79.8	75.7	74	63.6	67	63	74	65	72
Anthropology	79	80	69	76.8	73	73	83	72	72	76	80
Architecture and Design	77	80	80	77	70	75	72	74	69	66	81
Arts and Culture	86.1	85.8	83	82	86.5	80	83.6	78	76	74.5	82.9
Astronomy and Astrophysics	88	95	92	88.3	88	85	84	87	85	88	88
Biology	87.3	92.4	83	90	82	82.8	79.1	82	89	76.8	81.4
Business and Management	67	73.4	62	73.8	64	61	67	65	67	63	59.4
Chemistry	78.6	91.2	76.7	89.2	75.5	79	78.1	78.8	82	73.6	78.6
Computer Science	56.8	65.6	63	61.2	65.7	62	55.6	61	71	57	56.1
Defense and Security	88	87	82	89.1	92	80	90	82	81	76	86
Earth Sciences	81	90	78	83.8	85	79	78	63.6	74	84	80
Economics	79.1	81.3	75	81	77	63	70.4	71	80	77.8	70.7
Education	70	72.6	58	66.6	65	58	65	50	57	59	62
Energy and Power	81	86.5	81	80.7	78	78	81	77	83	65	76
Engineering	73	72.6	75.4	66	76.4	73.9	72.6	64.3	56	60.2	74.6
Environmental Science	81	86.6	80	82	84.4	74	80	71.7	74	78	77.7
Ethics and Human Rights	69	82	71	71	68	66	64	67	62	69	72
Finance and Investment	60.9	66	62	65.2	66	53	59.3	62	56	52.5	51.1
Food Science	81	88	80	86.9	80	76	80	74	82	82	86
Geography	84.4	85.4	80	82.9	86.8	74.7	74	81	77	71.4	77.1
Health and Medicine	82	89.8	82	90	86	83	71.4	79	87	81	81.3
History	90.7	80.7	84	85.1	86.4	72	76.6	73	80	74.9	75.4
Information Technology	90	92	86	87.2	90	91	92	89	86	81	85
International Relations	85	91.7	80	84	82	82	80	77	76	84	80
Language Studies	57	92	57	61.5	58.9	56	59.7	58	60	52.1	56.3
Law and Ethics	91.9	86.6	76	86.6	80.3	74	75	73	84	81.1	78.1
Literature and Linguistics	84.7	83.2	82	79.2	83.3	80	82.3	69	78	51	70.4
Logical Reasoning	66.1	62.7	62.3	58.3	66.4	55	61.3	58	64	61.4	57.9
Materials Science	84	84.8	73	78.6	71	75	71	65	67	58	73
Media and Communication	78	84	79	82.3	84	77	75	78	78	81	81
Music and Performing Arts	76	86	80	85.4	79	78	80	72	81	71	84
Physics	72.8	82.2	65.2	75.8	70	68.4	71.4	62.2	64	62.4	65
Politics and Governance	90.2	79.4	78	84.7	87	78	70.5	75	71	80.1	74.3
Psychology	77	91	80	81.4	72	76	74	77	75	66	77
Public Administration	60	73.7	65.7	76.8	61.6	62	51.5	59	66.7	66	66
Religion and Spirituality	87	93	85	89.5	88	82	85	78	80	80	81
Social Welfare and Development	73	77.7	75	80.8	69	68	78	69	78	64	72
Sociology	65.2	68.6	55	63.6	61	56	72.3	49	53	59.5	56.5
Sports and Recreation	86.6	90.4	83	86.4	86.5	85	84.2	81	83	80.8	85.7
Technology and Innovation	84	86.8	83	82.8	86	84	84	75	82	73	74
Transportation and Logistics	69	72.3	66	69.4	66.7	58	61.2	60.6	67	64	56

Table 26: Detailed subject-wise evaluation for GPT-4o on MILU across different languages. The results reported are for 0-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	60	64	60.6	61	56	43.4	45	43	51	51	53
Anthropology	55	67	49	63.6	48	49	60	39	58	52	60
Architecture and Design	55	60	52	53	48	50	53	39	51	51	56
Arts and Culture	65.5	76	54	65.9	57.7	57	57.9	45	62	53.9	55.8
Astronomy and Astrophysics	76	92	78	76.6	73	68	68	62	67	67	71
Biology	75	86.8	68	79.6	57.8	59.6	57.9	49	68	53.8	58.8
Business and Management	62	63.8	56	60.2	58	46	54	44	57	49	44.3
Chemistry	62.1	84.6	64.7	75.2	58.2	62	58.6	53.8	70	46.9	54.4
Computer Science	47.9	56.8	51	50.6	54.4	51	45.6	44	58	43.6	45.7
Defense and Security	68	81	56	66.8	58	55	61	52	54	52	55
Earth Sciences	61	77.5	61	62.9	56	60	65	46.5	56	60	57
Economics	59.7	70.9	54	62.3	65	51	47.2	48	62	57.1	54.5
Education	49	58.3	47	51.8	55	47	55.4	36	41	48	54
Energy and Power	62	77	68	65.8	59	60	59	49	56	47	59
Engineering	57.2	57.6	59	52.6	54.2	53.8	57.4	47.3	47	42.8	54.4
Environmental Science	66	73.5	64	67.5	68.8	53	64.2	49.5	58	55	58.7
Ethics and Human Rights	64	71	61	62	61	62	60	55	57	50	63
Finance and Investment	54	56.8	45	53	51.9	50	48.6	40	51	44.7	36.8
Food Science	65	77	63	66.7	63	64	71	60	64	60	67
Geography	60.9	69.6	59	67.5	61.2	49.5	53.3	49	55	50.3	59.4
Health and Medicine	78	83.2	68	81.8	73	73	58.9	62	80	61.2	62.6
History	69.9	65.7	63	64.1	61.9	56	52	49	71	55.3	47.1
Information Technology	88	88	81	87.7	80	88	80	80	76	70	84
International Relations	66	84.2	66	69.6	65	60	60	53	65	65	62
Language Studies	37	87.2	48	50.9	44.2	41	47	46	53	41.2	38.5
Law and Ethics	71.6	71	53	68	56.1	50	54.7	35	46	48	48.1
Literature and Linguistics	63.1	69.7	55	61.1	50	55	56.6	38	56	40.1	51.4
Logical Reasoning	47.9	47.7	40.3	41.9	40.2	38	45.5	39	47	39	43.8
Materials Science	60	73	46	59.4	51	55	53	38	59	38	55
Media and Communication	66	82	71	73.5	69	60	65	58	66	64	64
Music and Performing Arts	57	69	56	61.6	60	50	56	34	49	45	61
Physics	54.6	76	52.4	60.4	49.4	52.1	55.6	39.2	52	46	46.8
Politics and Governance	69.8	63.5	49	66.3	64.8	57	50.6	44	50	58.5	53.3
Psychology	59	87	65	66.7	58	57	57	52	67	57	68
Public Administration	49	66.7	53.5	62.6	43.4	44	41.4	53	58.6	47	50
Religion and Spirituality	65	86	68	74.2	67	68	61	52	57	57	71
Social Welfare and Development	60	64.3	59	64.2	50	41	61	39	57	52	48
Sociology	47.8	56.8	42	49.2	53	40	50.6	41	51	44.1	47.1
Sports and Recreation	64.9	75	58	66.2	62.4	53	65	45	55	55.8	58.6
Technology and Innovation	60	76	69	67.6	67	64	65	60	63	58	58
Transportation and Logistics	45	60	56	49.8	42.4	45	44.9	46.5	52	55	38

Table 27: Detailed subject-wise evaluation for GPT-4O-MINI on MILU across different languages. The results reported are for 0-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	60	76.7	66.7	72.2	66	51.5	53	52	59	50	60
Anthropology	67	73	57	68.7	60	49	69	50	58	60	69
Architecture and Design	59	69	46	65	48	53	67	51	52	57	57
Arts and Culture	58.8	77.8	58	69.9	61.3	51	62.3	51	56	56.4	58.3
Astronomy and Astrophysics	81	94	89	82.8	82	73	81	81	74	78	74
Biology	80.3	90.2	73	85.8	83.2	77.8	71.9	70	81	74.8	70.1
Business and Management	61	71.8	56	67.9	61	62	61	44	62	60	51.9
Chemistry	80.5	90.8	75	88.2	81	76	76.5	69.2	83	76	78.1
Computer Science	55.2	58.8	61	53.4	59.8	56	49.1	55	61	52.9	49.3
Defense and Security	66	83	59	73	63	55	71	63	64	56	65
Earth Sciences	67	87.5	67	75.6	73	71	70	50.5	59	68	66
Economics	73.4	75.3	64	74.2	75	60	60.6	57	61	77.8	59.3
Education	54	67.8	54	60.4	55	56	55.4	44	48	59	56
Energy and Power	68	82.4	71	70.5	69	76	67	59	62	60	64
Engineering	65.8	69.2	68.8	60.4	70.4	64.5	69.6	53.9	54	59.6	67.8
Environmental Science	73	80	67	78.8	73.4	66	63.7	56.6	69	67	70.9
Ethics and Human Rights	58	72	60	61	56	60	56	58	59	61	61
Finance and Investment	70.3	73	73	67.4	75.3	58	67.8	59	52	76	61.4
Food Science	73	85	73	77.8	69	76	78	67	79	77	78
Geography	64.8	78	67	71.5	74.9	65.7	63.9	62	68	66.3	63
Health and Medicine	76	87.4	66	85.8	84	79	68.8	72	83	75	74.8
History	72	75.1	69	74.2	75.5	54	60.6	61	68	67.6	54.5
Information Technology	79	88	78	77.7	79	78	80	70	71	74	70
International Relations	63	81.7	68	70	70	71	72	55	68	70	70
Language Studies	52	89	49	57.1	55.8	43	58.4	47	62	49.1	42
Law and Ethics	70.9	76.8	60	71.4	72.6	56	62.5	45	71	69.3	55.6
Literature and Linguistics	63.1	74.3	52	65.1	58.3	59	59.3	50	57	36.4	55.6
Logical Reasoning	60.9	57.8	63	55.9	60.6	51	54.9	55	50	52.2	55.2
Materials Science	80	77.9	71	69.2	65	68	68	56	69	61	77
Media and Communication	70	81	68	74.1	70	59	65	51	68	69	65
Music and Performing Arts	51	69	60	65.6	58	59	65	46	58	56	60
Physics	78	84.2	75.5	74	79.8	74.2	80.4	64.6	70	73.8	72.4
Politics and Governance	71	69.7	51	71.1	66.4	60	52.5	53	60	70.1	54.7
Psychology	72	92	71	69	66	63	69	66	73	62	73
Public Administration	51	62.6	51.5	62.6	52.5	44	38.4	44	50.5	53	50
Religion and Spirituality	68	85	65	75.4	73	68	68	61	61	72	68
Social Welfare and Development	59	69.6	65	73.1	56	48	65	43	68	59	57
Sociology	60	66.4	49	57.2	50	41	55.3	47	51	46.8	53.5
Sports and Recreation	66.3	80	58	72.4	67.4	71	72.3	52	73	60.3	61.4
Technology and Innovation	72	79.2	73	73.4	69	72	75	63	64	66	64
Transportation and Logistics	43	66.9	66	59	48.5	47	45.9	47.5	57	51	43

Table 28: Detailed subject-wise evaluation for GEMINI-1.5-PRO on MILU across different languages. The results reported are for 0-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	64	65.6	58.6	62.6	60	55.6	57	59	57	52	63
Anthropology	58	66	49	57.6	58	54	62	59	65	54	56
Architecture and Design	58	60	57	60	51	52	64	46	58	52	51
Arts and Culture	58.2	71	60	63.9	62.6	54	59.6	46	54	60	54.8
Astronomy and Astrophysics	79	91	84	83.6	78	71	83	80	76	80	71
Biology	75.4	87	72	84.6	78.6	72.7	69.3	66	82	69.7	73
Business and Management	50	64.7	46	59.3	56	53	51	56	62	52	46.2
Chemistry	71.4	87.6	74.1	84.4	70.6	73	66.5	73.1	77	66.2	70.8
Computer Science	49.5	56.6	57	51	55	59	48.5	53	53	54.1	46.2
Defense and Security	53	76	52	56.9	56	49	53	53	48	51	50
Earth Sciences	66	79.3	65	70.9	81	69	66	54.5	65	66	68
Economics	61.2	71.1	67	65.9	66	51	54.2	63	70	62.7	55.7
Education	50	57	48	52.4	51	49	54.1	37	48	48	47
Energy and Power	62	75	74	66.1	70	71	71	62	67	63	69
Engineering	59.6	59.6	66.2	59.6	62.4	65.4	59.2	61.3	57	55	62.8
Environmental Science	63	74.5	68	70.9	75	60	67.9	66.7	73	61	62
Ethics and Human Rights	61	65	66	60	63	63	59	58	65	61	59
Finance and Investment	49	57.4	46	51.2	55.6	47	49.2	44	57	50.3	44.3
Food Science	65	84	71	81.8	77	73	75	71	72	71	79
Geography	62.6	69.2	59	71.5	68.5	57.6	60.9	63	64	60	61.5
Health and Medicine	74	82.4	72	81	81	80	63.4	71	80	70.7	76.6
History	64.5	62.9	58	66.5	71.1	64	55.8	52	70	56.7	51.7
Information Technology	83	83	78	81.6	76	82	75	79	74	67	75
International Relations	58	72.5	67	68.1	65	62	63	62	59	70	58
Language Studies	44	85.2	54	51.9	51.9	44	51.7	48	61	44.6	45.4
Law and Ethics	70.9	69	57	66.2	59.9	57	59.5	46	62	54.3	53.1
Literature and Linguistics	52.3	65.1	51	59.1	55.8	50	50.4	44	62	43.7	52.8
Logical Reasoning	54.8	53	56.5	53.9	58.9	55	52.4	51	51	51.8	52.7
Materials Science	72	73.4	58	65.3	66	59	68	54	63	55	58
Media and Communication	68	76	67	72.8	71	61	59	60	62	65	63
Music and Performing Arts	40	59	53	57.6	53	48	53	46	56	43	44
Physics	68	75.8	66	69.4	68.2	70	65.8	58.7	62	66	61.8
Politics and Governance	70.6	63.1	52	63.9	64	52	51.1	44	46	61	51.7
Psychology	70	88	63	65.1	67	64	65	70	67	60	68
Public Administration	48	55.6	51.5	51.5	44.4	48	40.4	55	54.5	47	45
Religion and Spirituality	68	79	70	69	69	66	66	54	61	65	65
Social Welfare and Development	57	65.2	59	58	50	41	51	48	63	54	58
Sociology	55.7	54	45	55.6	56	47	50.3	42	52	46.8	48.2
Sports and Recreation	58.9	66.6	53	58.4	62.9	55	61.6	53	56	54.5	58.6
Technology and Innovation	66	73.2	68	67.4	61	73	64	64	67	55	57
Transportation and Logistics	44	56.9	54	55.5	52.5	48	38.8	44.4	53	48	51

Table 29: Detailed subject-wise evaluation for GEMINI-1.5-FLASH on MILU across different languages. The results reported are for 0-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	51	45.3	52	42.7	49	36.7	38	38	37.9	42	37
Anthropology	47	42	43	43.4	41	40.2	44	38	40	36	35
Architecture and Design	38	50	41.9	43	45	42	49.5	32.3	34	34	43.4
Arts and Culture	55.8	57.4	54	54.8	52.8	52	46.1	37	44	48.8	47.7
Astronomy and Astrophysics	53	53	57	55.9	53	50	57.6	43	37	51	45
Biology	57.4	63.7	61	57.6	60.7	45.5	47.6	40	48	48.1	46.1
Business and Management	40	41.2	46	43.7	43	38	42.9	35	41	36	29.2
Chemistry	32.4	49.7	40	49.2	34.3	40	34.9	30.8	44	34.5	34.8
Computer Science	26	35.1	24	31.1	36.1	34	28	34	29	24.4	23.8
Defense and Security	52	57	45	43.9	39	36	49	41	42	36	43
Earth Sciences	48	50.5	45	45.7	53	49	42	33.3	30	49	46
Economics	46	47.5	49	45.1	54	38	45.5	33	39	43.7	38.9
Education	29	36.3	32	32	32	33	34.7	27	30	31.2	37
Energy and Power	57	54.7	47.9	47.8	54	54.5	47.2	48	52	41	48
Engineering	41.8	36.6	39.4	33.4	41.8	40.3	39	34.3	32	38.2	38
Environmental Science	49	52.9	48.3	50.3	51.2	43	40.6	40.8	48	42.1	47.5
Ethics and Human Rights	43	48.5	44.7	53	40	39.8	38	41.5	34	39	37.4
Finance and Investment	37.6	33	32	31.8	40.7	38	35	25	35	32.6	26.4
Food Science	46	62	49	54.5	52	52	52	41	48	52	56
Geography	54.2	46	60	52.8	58.9	45.5	48.8	41	43.8	46.5	44.5
Health and Medicine	63	62	56	62.3	67	57	48.4	46	53	50.9	46.7
History	61.3	43.8	59.6	56.1	62.5	46	45.5	43	50	50.6	41
Information Technology	63	64	54.9	61.2	60	61	56	55	46	47	62
International Relations	48	51.7	46	46.3	46	45	43	36.8	41	53	47
Language Studies	19	56.4	34	38.7	27.9	32	26.8	27	37	28.1	28.2
Law and Ethics	46.6	39.2	38.4	46.2	42.7	41	37.7	33	35	41.7	32.9
Literature and Linguistics	51.4	45.9	42	50.3	50	39	44.2	39	38	41.1	47.2
Logical Reasoning	27	24.7	26.1	27.4	29	21	27.1	22	24	24	28.1
Materials Science	39	45.6	38	41	49	44	39	33	45	31	32
Media and Communication	48	52	48	45.6	51	50.5	48	35	49	45	55
Music and Performing Arts	39	45	39	43.7	48	40	52	43	42	36	46
Physics	34.2	46.2	32.5	42.1	37.4	32.1	34.9	30.3	35	32.6	32.9
Politics and Governance	59.2	42.9	48	48.2	55.9	44.4	38.9	35	45	51.9	39.8
Psychology	57	61	54	53.5	44	47.9	44.4	42	47	52	53
Public Administration	40	46.5	41.4	45.7	34.3	32.3	30.2	41.4	37.8	34	40
Religion and Spirituality	62	62	64	61.5	66	59	61	51	57	55	59
Social Welfare and Development	38	47.7	48	38.1	41	34	46.5	33	41	41	40.8
Sociology	35.7	38	35	35.8	34	32	34.2	34	34	37.3	34.3
Sports and Recreation	54.5	50.6	42	51.8	51.1	40	49.4	35	47	45.8	45.7
Technology and Innovation	48	50.8	38	42.2	56	49	48	45	38	44	38
Transportation and Logistics	28	36.2	41	35.1	31.3	33	35.8	33.3	33	40	28

Table 30: Detailed subject-wise evaluation for KRUTRIM-SPECTRE-V2 on MILU across different languages. The results reported are for 0-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	32	35.4	38.4	31	30	33.3	27	31	30	28	28
Anthropology	41	40	27	38.4	31	27	31	29	34	26	35
Architecture and Design	31	35	32	30	30	20	19	30	31	30	34
Arts and Culture	28.5	37.2	29	31.3	32.5	24	31.1	27	23	29.7	30.1
Astronomy and Astrophysics	46	45	42	42.2	39	24	41	31	37	40	37
Biology	40.5	46.9	36	38.7	38.9	30.3	31.3	30	39	31.2	34.9
Business and Management	31	37.8	30	34	27	27	29	25	28	34	25.5
Chemistry	34.1	41.3	33.6	40.9	33.3	38	33.8	26	30	27.9	34.9
Computer Science	29.2	31.6	42	31.2	36.1	34	30.2	28	33	30.2	34.5
Defense and Security	29	24	35	31.8	32	29	26	31	37	23	31
Earth Sciences	41	44.1	31	35.7	37	37	42	32.3	36	33	36
Economics	37.4	36.1	39	31.1	43	30	37.3	34	30	36.5	32.9
Education	31	23.9	19	25.1	26	29	35	32	31	28	38
Energy and Power	30	28.4	40	35.9	42	27	36	28	30	36	27
Engineering	30.4	32.6	32.2	29.8	29.4	30.8	30.4	30.4	24	28.4	30.2
Environmental Science	43	39.9	26	38.3	35.2	29	36.3	36.4	45	26	36.3
Ethics and Human Rights	34	30	41	34	29	32	29	27	29	34	24
Finance and Investment	34.2	32.6	38	31.4	31.5	32	28.2	27	29	30.2	28.2
Food Science	39	46	42	45.5	36	31	43	37	40	39	51
Geography	40.8	35.2	34	34.3	37	31.3	31.9	28	33	34.9	30.2
Health and Medicine	33	49.3	43	43.7	45	34	39.3	40	53	39.7	35.5
History	28.5	36.5	42	34.9	36.3	31	37.2	29	41	34.5	32.7
Information Technology	50	42	46	43.6	45	45	46	45	36	38	42
International Relations	36	42.5	34	38.5	34	36	43	40	40	26	40
Language Studies	35	33.8	31	24.6	26.4	25	27.5	32	39	27.3	22.4
Law and Ethics	30.4	39	38	36.6	33.8	37	28.4	30	37	32.3	29.4
Literature and Linguistics	38.7	31.1	33	28.5	31.7	35	25.7	30	30	24.7	34.5
Logical Reasoning	29.2	22.2	19.5	25.1	23.6	28	25.1	28	27	25.1	26.7
Materials Science	32	42.2	26	34.3	36	34	29	24	26	25	19
Media and Communication	39	35	37	27.9	42	36	40	37	35	30	44
Music and Performing Arts	34	39	32	20.5	32	28	24	27	24	25	24
Physics	27.4	40.4	32.3	34.6	27.2	25.8	32.4	31.2	37	28.8	31.4
Politics and Governance	36.5	31.9	32	31.3	34.8	26	29.5	30	19	33.6	34
Psychology	34	36	41	29.5	39	37	31	39	36	31	37
Public Administration	25	30.3	28.3	30.3	35.4	27	21.2	28	36.4	31	32
Religion and Spirituality	34	41	40	33.9	36	47	29	39	29	39	37
Social Welfare and Development	25	31.2	32	26.9	35	31	33	23	34	20	24
Sociology	32.2	28.2	32	30	31	30	24.5	31	23	27	34.1
Sports and Recreation	36.6	33.8	29	30.2	35.4	21	35	31	28	33.3	29
Technology and Innovation	38	37	30	35.6	35	26	43	28	38	35	27
Transportation and Logistics	25	26.9	30	26.5	27.3	32	17.3	28.3	27	22	22

Table 31: Detailed subject-wise evaluation for SARVAMAI/SARVAM-1 on MILU across different languages. The results reported are for 5-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	28.0/ 34.0/ 29.0	24.3/ 29.0/ 29.8	29.3/ 30.3/ 23.2	22.1/ 26.8/ 26.0	21.0/ 27.0/ 26.0	38.4/ 31.3/ 31.3	27.0/ 26.0/ 27.0	29.0/ 26.0/ 30.0	19.0/ 26.0/ 25.0	26.0/ 29.0/ 26.0	26.0/ 25.0/ 26.0
Anthropology	21.0/ 15.0/ 25.0	24.0/ 37.0/ 39.0	21.0/ 23.0/ 28.0	21.2/ 24.2/ 28.3	31.0/ 27.0/ 28.0	25.0/ 25.0/ 29.0	23.0/ 25.0/ 21.0	28.0/ 29.0/ 28.0	19.0/ 20.0/ 23.0	27.0/ 23.0/ 23.0	19.0/ 24.0/ 19.0
Architecture and Design	25.0/ 19.0/ 25.0	27.0/ 23.0/ 39.0	30.0/ 25.0/ 20.0	22.0/ 29.0/ 38.0	28.0/ 24.0/ 22.0	28.0/ 27.0/ 25.0	23.0/ 24.0/ 25.0	17.0/ 23.0/ 22.0	25.0/ 23.0/ 27.0	29.0/ 32.0/ 29.0	28.0/ 26.0/ 24.0
Arts and Culture	24.2/ 24.9/ 29.7	31.0/ 32.8/ 36.8	30.0/ 24.0/ 22.0	23.2/ 23.6/ 27.1	23.3/ 20.9/ 24.5	31.0/ 27.0/ 27.0	21.9/ 24.6/ 24.0	26.0/ 24.0/ 21.0	29.0/ 25.0/ 27.0	29.7/ 26.1/ 23.6	27.1/ 28.1/ 23.1
Astronomy and Astrophysics	22.0/ 24.0/ 20.0	27.0/ 37.0/ 35.0	30.0/ 34.0/ 33.0	24.2/ 28.9/ 28.9	24.0/ 22.0/ 23.0	24.0/ 24.0/ 23.0	24.0/ 27.0/ 35.0	25.0/ 30.0/ 29.0	24.0/ 23.0/ 24.0	21.0/ 20.0/ 16.0	20.0/ 31.0/ 19.0
Biology	23.6/ 24.3/ 21.5	33.7/ 40.7/ 44.7	21.0/ 17.0/ 21.0	26.1/ 28.3/ 30.7	23.8/ 28.3/ 25.8	24.2/ 22.2/ 24.2	34.3/ 29.5/ 31.9	26.0/ 20.0/ 27.0	21.0/ 24.0/ 18.0	25.2/ 25.5/ 25.2	20.8/ 24.2/ 23.6
Business and Management	28.0/ 22.0/ 21.0	27.2/ 36.2/ 35.0	20.0/ 17.0/ 20.0	26.2/ 27.9/ 29.3	25.0/ 24.0/ 24.0	30.0/ 22.0/ 26.0	20.0/ 22.0/ 16.0	16.0/ 21.0/ 18.0	27.0/ 26.0/ 24.0	24.0/ 21.0/ 23.0	17.0/ 21.7/ 20.8
Chemistry	23.6/ 27.2/ 27.2	32.9/ 34.1/ 35.9	26.7/ 24.1/ 24.1	26.9/ 30.3/ 26.9	26.1/ 31.4/ 29.1	27.0/ 22.0/ 26.0	31.1/ 28.2/ 29.5	26.9/ 26.0/ 23.1	19.0/ 22.0/ 19.0	27.6/ 28.8/ 28.2	25.9/ 28.9/ 28.2
Computer Science	24.5/ 27.1/ 24.0	27.8/ 30.2/ 29.2	28.0/ 21.0/ 27.0	27.2/ 26.8/ 24.8	28.4/ 29.0/ 27.8	23.0/ 32.0/ 33.0	28.4/ 33.1/ 31.9	28.0/ 29.0/ 26.0	25.0/ 20.0/ 21.0	23.3/ 26.7/ 26.2	29.1/ 31.4/ 30.9
Defense and Security	25.0/ 25.0/ 19.0	27.0/ 36.0/ 37.0	23.0/ 21.0/ 22.0	24.6/ 24.2/ 22.3	23.0/ 22.0/ 20.0	22.0/ 24.0/ 24.0	21.0/ 22.0/ 20.0	28.0/ 23.0/ 18.0	33.0/ 27.0/ 26.0	29.0/ 24.0/ 29.0	26.0/ 31.0/ 24.0
Earth Sciences	28.0/ 20.0/ 26.0	29.3/ 36.1/ 39.8	30.0/ 27.0/ 33.0	26.9/ 29.3/ 28.3	29.0/ 27.0/ 26.0	18.0/ 17.0/ 21.0	24.0/ 24.0/ 27.0	29.3/ 26.3/ 24.2	26.0/ 27.0/ 23.0	28.0/ 27.0/ 27.0	26.0/ 28.0/ 32.0
Economics	28.8/ 24.5/ 30.2	29.1/ 33.5/ 32.1	21.0/ 18.0/ 18.0	29.8/ 29.8/ 28.8	22.0/ 15.0/ 25.0	29.0/ 28.0/ 33.0	29.6/ 31.0/ 32.4	24.0/ 27.0/ 27.0	20.0/ 25.0/ 24.0	28.6/ 24.6/ 24.6	26.4/ 28.7/ 23.9
Education	24.0/ 30.0/ 28.0	25.8/ 29.6/ 25.5	22.0/ 24.0/ 27.0	29.6/ 24.3/ 24.6	24.0/ 20.0/ 23.0	20.0/ 24.0/ 25.0	32.5/ 28.0/ 28.0	22.0/ 23.0/ 28.0	24.0/ 27.0/ 24.0	24.0/ 26.0/ 29.0	24.0/ 27.0/ 32.0
Energy and Power	28.0/ 22.0/ 21.0	23.6/ 31.1/ 32.4	33.0/ 26.0/ 28.0	29.1/ 27.8/ 28.1	26.0/ 22.0/ 33.0	27.0/ 26.0/ 30.0	26.0/ 31.0/ 28.0	21.0/ 27.0/ 25.0	27.0/ 29.0/ 31.0	33.0/ 23.0/ 25.0	36.0/ 37.0/ 35.0
Engineering	23.2/ 24.8/ 25.2	30.0/ 29.2/ 30.8	20.6/ 24.0/ 22.2	26.8/ 26.6/ 26.4	22.8/ 21.4/ 23.4	24.8/ 23.9/ 25.6	25.0/ 24.2/ 21.4	25.3/ 26.8/ 26.5	23.0/ 27.0/ 29.0	23.0/ 27.0/ 24.8	27.2/ 23.4/ 23.4
Environmental Science	26.0/ 30.0/ 31.0	32.9/ 36.3/ 39.7	31.0/ 24.0/ 23.0	25.2/ 27.3/ 29.3	21.1/ 25.0/ 25.8	23.0/ 29.0/ 24.0	22.1/ 24.2/ 23.7	22.2/ 24.2/ 25.2	31.0/ 32.0/ 27.0	25.0/ 27.0/ 27.0	27.9/ 28.5/ 26.8
Ethics and Human Rights	24.0/ 21.0/ 31.0	30.0/ 33.0/ 34.0	23.0/ 24.0/ 25.0	25.0/ 30.0/ 22.0	26.0/ 30.0/ 31.0	19.0/ 18.0/ 18.0	27.0/ 29.0/ 29.0	25.0/ 23.0/ 18.0	23.0/ 20.0/ 20.0	21.0/ 21.0/ 19.0	26.0/ 31.0/ 29.0
Finance and Investment	26.7/ 28.2/ 30.2	29.8/ 30.6/ 32.0	22.0/ 24.0/ 23.0	27.2/ 28.0/ 26.0	25.9/ 25.9/ 28.4	25.0/ 28.0/ 21.0	25.4/ 25.4/ 27.1	22.0/ 30.0/ 23.0	18.0/ 28.0/ 21.0	24.0/ 25.1/ 22.9	25.0/ 28.2/ 27.5
Food Science	21.0/ 25.0/ 28.0	34.0/ 39.0/ 45.0	26.0/ 26.0/ 27.0	21.2/ 23.2/ 25.2	19.0/ 23.0/ 30.0	22.0/ 23.0/ 21.0	22.0/ 21.0/ 25.0	20.0/ 23.0/ 26.0	22.0/ 20.0/ 25.0	20.0/ 18.0/ 26.0	20.0/ 26.0/ 26.0
Geography	25.1/ 23.5/ 27.9	26.6/ 32.2/ 33.8	24.0/ 31.0/ 35.0	28.7/ 29.7/ 27.5	27.9/ 21.0/ 22.4	31.3/ 24.2/ 22.2	31.9/ 31.9/ 31.9	32.0/ 31.0/ 27.0	21.0/ 23.0/ 30.0	26.3/ 26.9/ 29.1	27.6/ 22.9/ 25.5
Health and Medicine	20.0/ 20.0/ 23.0	34.7/ 40.3/ 43.5	27.0/ 26.0/ 23.0	25.2/ 29.9/ 30.1	15.0/ 22.0/ 26.0	34.0/ 31.0/ 30.0	21.4/ 24.1/ 29.5	26.0/ 22.0/ 19.0	24.0/ 24.0/ 22.0	27.6/ 23.3/ 25.9	29.9/ 27.1/ 27.1
History	24.3/ 22.4/ 21.9	29.7/ 31.1/ 33.5	26.0/ 26.0/ 27.0	27.2/ 28.0/ 26.2	26.0/ 21.2/ 20.5	24.0/ 24.0/ 25.0	23.4/ 24.9/ 24.2	32.0/ 25.0/ 28.0	29.0/ 29.0/ 27.0	23.6/ 24.7/ 26.6	27.4/ 25.1/ 29.7
Information Technology	26.0/ 28.0/ 30.0	26.0/ 32.0/ 36.0	32.0/ 22.0/ 32.0	33.5/ 25.1/ 29.0	27.0/ 27.0/ 29.0	28.0/ 23.0/ 30.0	29.0/ 34.0/ 36.0	30.0/ 27.0/ 20.0	31.0/ 31.0/ 26.0	22.0/ 22.0/ 27.0	32.0/ 30.0/ 30.0
International Relations	25.0/ 24.0/ 27.0	33.3/ 39.2/ 36.7	28.0/ 20.0/ 32.0	29.2/ 31.1/ 31.1	25.0/ 33.0/ 27.0	31.0/ 28.0/ 30.0	25.0/ 29.0/ 31.0	25.0/ 25.0/ 28.0	28.0/ 25.0/ 26.0	37.0/ 33.0/ 32.0	16.0/ 19.0/ 16.0
Language Studies	30.0/ 21.0/ 29.0	30.2/ 31.6/ 31.2	31.0/ 25.0/ 32.0	24.9/ 25.9/ 27.3	27.9/ 24.8/ 21.7	23.0/ 27.0/ 30.0	28.9/ 30.9/ 30.9	23.0/ 24.0/ 21.0	26.0/ 28.0/ 29.0	26.6/ 23.2/ 27.7	26.4/ 27.6/ 22.4
Law and Ethics	31.8/ 31.1/ 24.3	34.2/ 35.2/ 35.0	28.0/ 26.0/ 26.0	29.6/ 26.0/ 25.4	19.1/ 17.2/ 22.3	28.0/ 22.0/ 18.0	26.3/ 24.1/ 23.3	25.0/ 26.0/ 22.0	27.0/ 30.0/ 34.0	22.1/ 23.6/ 19.7	26.9/ 26.2/ 26.9
Literature and Linguistics	27.9/ 31.5/ 31.5	28.3/ 33.1/ 31.7	22.0/ 21.0/ 24.0	23.6/ 25.2/ 25.4	27.5/ 25.0/ 27.5	27.0/ 29.0/ 24.0	23.9/ 25.7/ 28.3	19.0/ 26.0/ 26.0	33.0/ 27.0/ 27.0	30.4/ 30.0/ 27.5	26.8/ 28.9/ 26.1
Logical Reasoning	25.3/ 27.0/ 25.1	27.6/ 26.8/ 26.6	26.0/ 26.6/ 25.3	24.9/ 25.7/ 24.6	29.9/ 26.6/ 29.5	29.0/ 27.0/ 27.0	23.7/ 25.9/ 24.5	24.0/ 26.0/ 33.0	26.0/ 32.0/ 20.0	27.1/ 25.9/ 26.3	26.0/ 23.0/ 23.5
Materials Science	26.0/ 28.0/ 33.0	28.1/ 33.1/ 33.1	21.0/ 24.0/ 24.0	23.6/ 24.7/ 26.9	24.0/ 23.0/ 25.0	25.0/ 31.0/ 29.0	27.0/ 29.0/ 31.0	26.0/ 27.0/ 29.0	22.0/ 26.0/ 26.0	31.0/ 25.0/ 26.0	20.0/ 19.0/ 26.0
Media and Communication	32.0/ 28.0/ 25.0	31.0/ 41.0/ 39.0	26.0/ 25.0/ 23.0	25.9/ 23.8/ 23.8	34.0/ 30.0/ 24.0	24.0/ 23.0/ 23.0	27.0/ 30.0/ 21.0	35.0/ 31.0/ 25.0	18.0/ 18.0/ 25.0	25.0/ 26.0/ 31.0	29.0/ 25.0/ 27.0
Music and Performing Arts	30.0/ 34.0/ 26.0	33.0/ 41.0/ 41.0	27.0/ 20.0/ 20.0	28.5/ 29.1/ 31.1	26.0/ 29.0/ 25.0	35.0/ 29.0/ 26.0	26.0/ 26.0/ 28.0	25.0/ 23.0/ 29.0	25.0/ 28.0/ 28.0	30.0/ 30.0/ 27.0	32.0/ 30.0/ 27.0
Physics	25.4/ 25.6/ 23.6	27.2/ 29.8/ 32.8	23.4/ 30.6/ 26.5	23.2/ 25.6/ 24.2	26.4/ 25.0/ 27.0	28.4/ 29.5/ 23.2	24.4/ 26.2/ 24.8	23.6/ 24.0/ 28.8	28.0/ 27.0/ 23.0	25.4/ 21.4/ 21.4	23.4/ 24.6/ 23.4
Politics and Governance	18.8/ 24.7/ 25.1	30.5/ 34.1/ 31.9	37.0/ 35.0/ 28.0	27.5/ 27.1/ 28.1	22.3/ 26.3/ 25.5	22.0/ 25.0/ 22.0	23.3/ 23.6/ 24.4	30.0/ 21.0/ 25.0	22.0/ 29.0/ 28.0	27.0/ 27.4/ 27.8	22.2/ 28.8/ 28.5
Psychology	20.0/ 19.0/ 19.0	36.0/ 32.0/ 39.0	20.0/ 21.0/ 23.0	24.0/ 30.2/ 27.9	26.0/ 14.0/ 23.0	27.0/ 27.0/ 31.0	26.0/ 27.0/ 30.0	21.0/ 23.0/ 20.0	23.0/ 25.0/ 23.0	22.0/ 20.0/ 22.0	29.0/ 16.0/ 21.0
Public Administration	26.0/ 20.0/ 21.0	33.3/ 44.4/ 44.4	25.2/ 21.2/ 28.3	31.3/ 30.3/ 28.3	26.3/ 29.3/ 36.4	23.0/ 22.0/ 29.0	27.3/ 26.3/ 28.3	31.0/ 34.0/ 28.0	29.3/ 26.3/ 26.3	26.0/ 26.0/ 32.0	19.0/ 28.0/ 31.0
Religion and Spirituality	24.0/ 23.0/ 18.0	27.0/ 35.0/ 37.0	32.0/ 28.0/ 30.0	25.4/ 27.0/ 21.8	25.0/ 27.0/ 24.0	26.0/ 27.0/ 19.0	31.0/ 31.0/ 31.0	24.0/ 20.0/ 19.0	26.0/ 24.0/ 24.0	24.0/ 25.0/ 27.0	24.0/ 27.0/ 25.0
Social Welfare and Development	21.0/ 21.0/ 19.0	24.1/ 29.5/ 34.8	23.0/ 23.0/ 26.0	25.9/ 26.9/ 27.5	20.0/ 26.0/ 22.0	28.0/ 26.0/ 22.0	26.0/ 36.0/ 30.0	27.0/ 24.0/ 29.0	27.0/ 24.0/ 25.0	24.0/ 23.0/ 21.0	27.0/ 22.0/ 25.0
Sociology	24.3/ 28.7/ 27.0	31.8/ 33.2/ 32.4	32.0/ 23.0/ 33.0	24.6/ 27.4/ 27.6	22.0/ 20.0/ 21.0	28.0/ 30.0/ 22.0	25.8/ 26.4/ 23.9	29.0/ 29.0/ 27.0	23.0/ 25.0/ 15.0	25.2/ 22.5/ 24.3	21.8/ 22.9/ 25.9
Sports and Recreation	24.8/ 25.2/ 29.2	29.6/ 35.0/ 35.4	25.0/ 17.0/ 21.0	23.4/ 26.2/ 28.2	21.9/ 23.0/ 23.0	29.0/ 26.0/ 27.0	26.0/ 23.7/ 24.3	20.0/ 18.0/ 17.0	27.0/ 26.0/ 25.0	26.3/ 30.1/ 24.4	29.5/ 26.2/ 23.8
Technology and Innovation	21.0/ 20.0/ 21.0	30.2/ 33.8/ 34.0	26.0/ 26.0/ 25.0	27.2/ 27.0/ 26.8	26.0/ 27.0/ 25.0	30.0/ 23.0/ 19.0	24.0/ 25.0/ 24.0	26.0/ 21.0/ 27.0	23.0/ 23.0/ 20.0	23.0/ 30.0/ 29.0	27.0/ 24.0/ 26.0
Transportation and Logistics	24.0/ 19.0/ 28.0	20.8/ 23.1/ 30.0	20.0/ 24.0/ 24.0	25.2/ 23.3/ 23.3	23.2/ 27.3/ 26.3	27.0/ 25.0/ 22.0	24.5/ 21.4/ 21.4	23.2/ 23.2/ 22.2	19.0/ 20.0/ 18.0	24.0/ 22.0/ 20.0	14.0/ 20.0/ 25.0

Table 32: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.2-1B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	31.0/ 25.0/ 24.0	26.8/ 26.8/ 25.6	31.3/ 29.3/ 24.2	27.4/ 24.8/ 23.3	31.0/ 24.0/ 25.0	30.3/ 24.2/ 23.2	29.0/ 23.0/ 28.0	27.0/ 30.0/ 24.0	27.0/ 22.0/ 23.0	22.0/ 23.0/ 26.0	18.0/ 22.0/ 26.0
Anthropology	26.0/ 33.0/ 30.0	26.0/ 27.0/ 23.0	29.0/ 30.0/ 25.0	33.3/ 37.4/ 25.2	29.0/ 26.0/ 31.0	24.0/ 25.0/ 21.0	29.0/ 27.0/ 24.0	31.0/ 29.0/ 20.0	21.0/ 21.0/ 18.0	21.0/ 25.0/ 30.0	27.0/ 27.0/ 26.0
Architecture and Design	19.0/ 18.0/ 19.0	23.0/ 23.0/ 19.0	23.0/ 28.0/ 20.0	36.0/ 25.0/ 22.0	26.0/ 23.0/ 21.0	28.0/ 30.0/ 28.0	32.0/ 28.0/ 21.0	26.0/ 27.0/ 25.0	32.0/ 27.0/ 27.0	31.0/ 25.0/ 27.0	21.0/ 26.0/ 22.0
Arts and Culture	32.7/ 26.7/ 31.5	32.4/ 25.8/ 27.4	22.0/ 25.0/ 25.0	25.4/ 28.1/ 28.1	23.9/ 22.7/ 21.5	29.0/ 26.0/ 26.0	22.4/ 23.5/ 25.1	23.0/ 27.0/ 23.0	27.0/ 28.0/ 29.0	29.7/ 27.9/ 27.9	22.6/ 25.6/ 29.1
Astronomy and Astrophysics	26.0/ 30.0/ 32.0	34.0/ 28.0/ 29.0	27.0/ 28.0/ 30.0	21.1/ 27.3/ 25.8	26.0/ 29.0/ 30.0	26.0/ 22.0/ 23.0	25.0/ 34.0/ 35.0	25.0/ 18.0/ 24.0	19.0/ 25.0/ 26.0	24.0/ 18.0/ 28.0	27.0/ 22.0/ 26.0
Biology	29.2/ 23.6/ 21.5	30.1/ 26.1/ 23.4	26.0/ 24.0/ 28.0	28.7/ 28.7/ 28.5	27.9/ 26.6/ 26.2	24.2/ 23.2/ 26.3	25.1/ 27.8/ 21.8	28.0/ 26.0/ 26.0	23.0/ 23.0/ 24.0	25.5/ 27.7/ 27.1	23.9/ 23.3/ 20.1
Business and Management	23.0/ 20.0/ 22.0	29.4/ 29.4/ 27.9	26.0/ 31.0/ 20.0	28.8/ 26.5/ 27.9	27.0/ 25.0/ 24.0	22.0/ 29.0/ 31.0	19.0/ 26.0/ 24.0	28.0/ 29.0/ 23.0	28.0/ 24.0/ 26.0	22.0/ 24.0/ 24.0	22.6/ 21.7/ 27.4
Chemistry	28.3/ 29.4/ 27.2	30.9/ 29.5/ 26.1	22.4/ 26.7/ 22.4	33.1/ 27.9/ 27.9	25.5/ 25.5/ 25.2	30.0/ 21.0/ 21.0	28.2/ 26.7/ 25.9	26.9/ 24.0/ 23.1	29.0/ 33.0/ 31.0	27.0/ 24.9/ 27.6	27.9/ 25.4/ 27.2
Computer Science	24.5/ 24.5/ 27.6	29.8/ 27.4/ 26.0	16.0/ 22.0/ 28.0	25.6/ 27.0/ 22.6	24.9/ 22.5/ 27.2	34.0/ 26.0/ 21.0	30.8/ 32.5/ 27.8	23.0/ 32.0/ 25.0	32.0/ 28.0/ 30.0	23.8/ 29.6/ 27.9	27.4/ 26.0/ 26.0
Defense and Security	27.0/ 23.0/ 27.0	29.0/ 25.0/ 23.0	30.0/ 25.0/ 31.0	27.5/ 23.7/ 27.0	30.0/ 35.0/ 35.0	21.0/ 21.0/ 27.0	34.0/ 31.0/ 22.0	25.0/ 20.0/ 24.0	30.0/ 24.0/ 31.0	33.0/ 31.0/ 24.0	34.0/ 37.0/ 36.0
Earth Sciences	39.0/ 28.0/ 23.0	29.1/ 26.4/ 28.0	28.0/ 23.0/ 24.0	26.7/ 25.2/ 25.2	22.0/ 28.0/ 24.0	25.0/ 21.0/ 29.0	27.0/ 26.0/ 27.0	23.2/ 22.2/ 24.2	20.0/ 22.0/ 25.0	29.0/ 21.0/ 24.0	28.0/ 22.0/ 28.0
Economics	27.3/ 25.2/ 27.3	25.1/ 22.7/ 22.7	25.0/ 24.0/ 25.0	26.0/ 28.0/ 25.4	20.0/ 21.0/ 26.0	28.0/ 23.0/ 25.0	27.5/ 24.6/ 21.1	33.0/ 24.0/ 27.0	19.0/ 23.0/ 20.0	24.6/ 21.4/ 25.4	22.2/ 25.8/ 25.1
Education	33.0/ 29.0/ 35.0	28.3/ 23.6/ 23.9	30.0/ 30.0/ 27.0	28.1/ 28.1/ 27.2	33.0/ 33.0/ 33.0	23.0/ 29.0/ 22.0	27.4/ 19.8/ 24.2	23.0/ 26.0/ 26.0	20.0/ 20.0/ 21.0	34.0/ 34.0/ 29.0	31.0/ 27.0/ 36.0
Energy and Power	24.0/ 23.0/ 23.0	31.1/ 29.0/ 24.3	32.0/ 30.0/ 28.0	26.8/ 25.4/ 25.1	26.0/ 25.0/ 17.0	20.0/ 25.0/ 24.0	22.0/ 28.0/ 25.0	24.0/ 25.0/ 25.0	31.0/ 29.0/ 26.0	26.0/ 30.0/ 30.0	24.0/ 20.0/ 23.0
Engineering	25.4/ 22.0/ 23.6	21.8/ 25.6/ 26.6	24.0/ 22.6/ 22.6	26.6/ 25.6/ 24.0	21.8/ 22.8/ 25.2	22.7/ 20.5/ 25.2	25.0/ 25.0/ 23.6	23.2/ 21.7/ 25.6	27.0/ 20.0/ 21.0	28.2/ 25.2/ 27.6	25.4/ 26.6/ 24.8
Environmental Science	28.0/ 31.0/ 29.0	29.5/ 24.2/ 25.7	28.0/ 29.0/ 33.0	24.6/ 27.5/ 29.3	27.3/ 27.3/ 28.1	23.0/ 22.0/ 24.0	28.4/ 23.2/ 27.9	16.2/ 21.2/ 21.2	25.0/ 25.0/ 22.0	28.0/ 30.0/ 29.0	30.7/ 24.6/ 26.8
Ethics and Human Rights	27.0/ 27.0/ 27.0	31.0/ 32.0/ 26.0	29.0/ 27.0/ 26.0	21.0/ 23.0/ 23.0	29.0/ 26.0/ 24.0	25.0/ 16.0/ 21.0	25.0/ 27.0/ 29.0	29.0/ 25.0/ 23.0	30.0/ 25.0/ 21.0	16.0/ 16.0/ 25.0	22.0/ 25.0/ 24.0
Finance and Investment	22.3/ 21.8/ 21.3	28.2/ 27.4/ 26.4	26.0/ 25.0/ 25.0	25.8/ 26.0/ 22.4	22.8/ 24.1/ 27.8	26.0/ 29.0/ 25.0	21.5/ 20.3/ 20.9	15.0/ 13.0/ 22.0	19.0/ 19.0/ 21.0	30.2/ 26.8/ 26.8	26.8/ 25.7/ 23.2
Food Science	26.0/ 21.0/ 19.0	23.0/ 25.0/ 27.0	34.0/ 28.0/ 27.0	27.3/ 26.3/ 22.2	27.0/ 23.0/ 24.0	35.0/ 25.0/ 30.0	29.0/ 30.0/ 26.0	21.0/ 25.0/ 26.0	24.0/ 32.0/ 20.0	21.0/ 17.0/ 19.0	20.0/ 32.0/ 28.0
Geography	27.4/ 21.8/ 21.8	30.0/ 27.4/ 26.4	33.0/ 26.0/ 26.0	28.3/ 26.3/ 26.1	25.1/ 21.0/ 22.8	24.2/ 22.2/ 27.3	27.2/ 23.7/ 23.7	30.0/ 29.0/ 26.0	28.0/ 28.0/ 28.0	33.1/ 33.7/ 32.0	29.2/ 27.6/ 27.6
Health and Medicine	33.0/ 29.0/ 24.0	41.1/ 33.7/ 30.7	24.0/ 26.0/ 30.0	31.7/ 27.9/ 25.9	37.0/ 26.0/ 22.0	26.0/ 22.0/ 18.0	25.9/ 31.2/ 29.5	22.0/ 26.0/ 27.0	29.0/ 27.0/ 29.0	25.9/ 30.2/ 28.4	26.2/ 30.8/ 31.8
History	26.7/ 25.6/ 28.5	31.7/ 28.3/ 27.9	31.0/ 23.0/ 24.0	26.6/ 27.0/ 24.2	28.6/ 24.2/ 27.1	34.0/ 27.0/ 33.0	23.1/ 23.1/ 24.5	34.0/ 29.0/ 29.0	35.0/ 34.0/ 25.0	26.2/ 24.0/ 24.7	28.4/ 25.1/ 26.6
Information Technology	38.0/ 36.0/ 30.0	26.0/ 26.0/ 26.0	30.0/ 24.0/ 24.0	36.3/ 30.7/ 33.0	25.0/ 24.0/ 29.0	27.0/ 24.0/ 23.0	33.0/ 29.0/ 28.0	24.0/ 24.0/ 26.0	33.0/ 25.0/ 27.0	34.0/ 26.0/ 30.0	28.0/ 33.0/ 39.0
International Relations	28.0/ 33.0/ 25.0	24.2/ 30.0/ 23.3	32.0/ 27.0/ 28.0	31.5/ 26.1/ 22.2	29.0/ 35.0/ 25.0	30.0/ 32.0/ 23.0	33.0/ 29.0/ 27.0	30.0/ 26.0/ 27.0	32.0/ 33.0/ 28.0	33.0/ 29.0/ 22.0	22.0/ 17.0/ 19.0
Language Studies	26.0/ 32.0/ 27.0	30.8/ 27.8/ 26.8	33.0/ 33.0/ 30.0	27.7/ 26.2/ 27.9	27.1/ 25.6/ 33.3	21.0/ 23.0/ 22.0	25.5/ 31.5/ 28.9	25.0/ 24.0/ 22.0	32.0/ 30.0/ 27.0	28.5/ 22.9/ 24.3	29.3/ 29.3/ 26.4
Law and Ethics	24.3/ 18.9/ 22.3	30.4/ 27.8/ 28.4	15.0/ 15.0/ 16.0	27.8/ 24.8/ 25.2	25.5/ 23.6/ 22.3	23.0/ 16.0/ 19.0	32.3/ 31.5/ 30.6	24.0/ 23.0/ 23.0	24.0/ 18.0/ 14.0	25.2/ 31.5/ 33.9	20.6/ 22.5/ 22.5
Literature and Linguistics	27.9/ 23.4/ 25.2	25.9/ 27.9/ 24.2	23.0/ 26.0/ 21.0	31.1/ 30.7/ 28.9	26.7/ 29.2/ 29.2	23.0/ 25.0/ 32.0	30.1/ 31.9/ 29.2	22.0/ 23.0/ 16.0	21.0/ 22.0/ 24.0	28.3/ 24.7/ 25.5	28.9/ 28.9/ 27.5
Logical Reasoning	26.8/ 25.8/ 27.5	24.2/ 24.6/ 24.4	27.9/ 29.9/ 27.9	26.5/ 25.2/ 23.8	25.7/ 28.2/ 28.2	26.0/ 24.0/ 24.0	21.4/ 22.6/ 25.4	23.0/ 15.0/ 19.0	36.0/ 28.0/ 30.0	30.3/ 29.5/ 29.9	25.0/ 25.7/ 25.7
Materials Science	33.0/ 27.0/ 33.0	32.7/ 25.9/ 23.2	23.0/ 26.0/ 28.0	26.2/ 22.7/ 25.3	29.0/ 25.0/ 21.0	26.0/ 27.0/ 27.0	38.0/ 29.0/ 24.0	25.0/ 27.0/ 25.0	26.0/ 33.0/ 21.0	26.0/ 22.0/ 23.0	22.0/ 28.0/ 25.0
Media and Communication	28.0/ 26.0/ 32.0	32.0/ 21.0/ 26.0	30.0/ 28.0/ 26.0	30.6/ 23.8/ 20.4	26.0/ 27.0/ 28.0	29.0/ 31.0/ 27.0	34.0/ 30.0/ 32.0	23.0/ 30.0/ 27.0	26.0/ 29.0/ 25.0	23.0/ 23.0/ 21.0	33.0/ 37.0/ 33.0
Music and Performing Arts	27.0/ 20.0/ 20.0	33.0/ 31.0/ 31.0	26.0/ 25.0/ 26.0	24.5/ 22.5/ 24.5	25.0/ 28.0/ 26.0	30.0/ 24.0/ 22.0	29.0/ 33.0/ 26.0	34.0/ 35.0/ 38.0	25.0/ 21.0/ 23.0	29.0/ 19.0/ 24.0	30.0/ 21.0/ 25.0
Physics	26.4/ 25.8/ 25.8	32.2/ 28.2/ 27.8	23.4/ 22.8/ 21.7	26.2/ 25.0/ 24.2	25.4/ 27.6/ 25.2	29.5/ 27.9/ 26.8	29.4/ 28.2/ 25.4	22.9/ 22.9/ 24.0	24.0/ 26.0/ 32.0	22.8/ 23.0/ 29.0	26.8/ 24.2/ 26.8
Politics and Governance	27.8/ 27.1/ 25.9	30.9/ 28.9/ 25.9	24.0/ 24.0/ 23.0	25.9/ 25.9/ 24.9	27.9/ 26.3/ 23.5	22.0/ 19.0/ 27.0	27.5/ 24.2/ 25.8	33.0/ 26.0/ 30.0	28.0/ 30.0/ 30.0	26.6/ 25.3/ 23.2	27.1/ 24.5/ 26.9
Psychology	22.0/ 22.0/ 23.0	28.0/ 22.0/ 23.0	30.0/ 27.0/ 31.0	31.8/ 35.7/ 20.2	26.0/ 26.0/ 30.0	25.0/ 27.0/ 21.0	28.0/ 34.0/ 31.0	25.0/ 22.0/ 25.0	31.0/ 27.0/ 32.0	26.0/ 24.0/ 30.0	28.0/ 23.0/ 33.0
Public Administration	21.0/ 23.0/ 16.0	24.2/ 20.2/ 24.2	26.3/ 24.2/ 27.3	33.3/ 35.4/ 30.3	27.3/ 21.2/ 27.3	24.0/ 23.0/ 22.0	29.3/ 27.3/ 19.2	27.0/ 23.0/ 20.0	29.3/ 24.2/ 22.2	22.0/ 13.0/ 24.0	33.0/ 29.0/ 32.0
Religion and Spirituality	26.0/ 24.0/ 25.0	27.0/ 30.0/ 26.0	34.0/ 23.0/ 27.0	25.4/ 18.9/ 22.6	26.0/ 18.0/ 18.0	30.0/ 25.0/ 26.0	35.0/ 30.0/ 30.0	24.0/ 28.0/ 24.0	28.0/ 31.0/ 32.0	25.0/ 26.0/ 24.0	30.0/ 22.0/ 25.0
Social Welfare and Development	22.0/ 21.0/ 20.0	24.1/ 29.5/ 23.2	32.0/ 30.0/ 29.0	23.8/ 26.9/ 25.9	27.0/ 24.0/ 27.0	23.0/ 19.0/ 22.0	32.0/ 26.0/ 25.0	28.0/ 22.0/ 24.0	24.0/ 23.0/ 14.0	28.0/ 25.0/ 28.0	24.0/ 24.0/ 27.0
Sociology	20.9/ 23.5/ 21.7	22.4/ 23.4/ 24.4	27.0/ 28.0/ 29.0	27.0/ 23.4/ 22.8	20.0/ 30.0/ 26.0	18.0/ 19.0/ 24.0	25.8/ 27.7/ 32.1	24.0/ 21.0/ 16.0	25.0/ 21.0/ 25.0	27.0/ 26.1/ 27.9	21.8/ 22.9/ 24.7
Sports and Recreation	24.3/ 21.8/ 16.8	27.8/ 26.0/ 24.4	18.0/ 17.0/ 18.0	21.8/ 21.2/ 18.8	25.8/ 28.6/ 27.0	24.0/ 28.0/ 27.0	26.6/ 26.6/ 23.7	21.0/ 24.0/ 27.0	25.0/ 22.0/ 23.0	26.3/ 25.6/ 25.6	24.8/ 23.3/ 20.5
Technology and Innovation	30.0/ 26.0/ 24.0	31.2/ 26.2/ 24.4	26.0/ 26.0/ 27.0	27.8/ 24.0/ 26.0	25.0/ 33.0/ 31.0	28.0/ 27.0/ 30.0	33.0/ 27.0/ 29.0	26.0/ 23.0/ 18.0	31.0/ 31.0/ 26.0	27.0/ 29.0/ 25.0	20.0/ 14.0/ 20.0
Transportation and Logistics	20.0/ 28.0/ 25.0	27.7/ 31.5/ 24.6	20.0/ 20.0/ 20.0	24.9/ 25.9/ 24.6	24.2/ 21.2/ 25.2	24.0/ 22.0/ 28.0	22.4/ 26.5/ 27.6	23.2/ 26.3/ 28.3	18.0/ 22.0/ 15.0	26.0/ 20.0/ 26.0	23.0/ 23.0/ 24.0

Table 33: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.2-1B-INSTRUCT on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	bn	en	gu	hi	kn	ml	mr	or	pa	ta	te
Agriculture	25.0/ 21.0/ 20.0	42.9/ 48.1/ 49.3	27.3/ 26.3/ 27.3	26.2/ 28.0/ 27.4	16.0/ 20.0/ 25.0	24.2/ 28.3/ 25.2	20.0/ 27.0/ 24.0	25.0/ 32.0/ 27.0	23.0/ 29.0/ 34.0	22.0/ 30.0/ 26.0	28.0/ 24.0/ 28.0
Anthropology	33.0/ 33.0/ 36.0	48.0/ 49.0/ 58.0	31.0/ 34.0/ 36.0	29.3/ 32.3/ 38.4	27.0/ 34.0/ 34.0	30.0/ 38.0/ 39.0	32.0/ 41.0/ 38.0	23.0/ 23.0/ 20.0	23.0/ 28.0/ 27.0	33.0/ 29.0/ 27.0	28.0/ 31.0/ 28.0
Architecture and Design	27.0/ 18.0/ 21.0	37.0/ 47.0/ 46.0	28.0/ 23.0/ 19.0	31.0/ 30.0/ 27.0	33.0/ 31.0/ 24.0	29.0/ 24.0/ 24.0	30.0/ 20.0/ 20.0	26.0/ 22.0/ 22.0	36.0/ 35.0/ 36.0	26.0/ 21.0/ 24.0	24.0/ 25.0/ 22.0
Arts and Culture	20.6/ 25.4/ 21.8	48.4/ 55.8/ 57.8	29.0/ 23.0/ 27.0	28.3/ 28.5/ 29.1	26.4/ 27.0/ 28.2	22.0/ 31.0/ 28.0	25.7/ 26.8/ 23.5	19.0/ 20.0/ 21.0	25.0/ 27.0/ 33.0	32.1/ 32.7/ 33.3	22.1/ 26.6/ 24.6
Astronomy and Astrophysics	30.0/ 23.0/ 27.0	54.0/ 64.0/ 68.0	24.0/ 35.0/ 30.0	35.2/ 39.8/ 37.5	31.0/ 24.0/ 34.0	33.0/ 29.0/ 30.0	35.0/ 41.0/ 37.0	29.0/ 19.0/ 27.0	28.0/ 21.0/ 27.0	27.0/ 32.0/ 29.0	28.0/ 26.0/ 31.0
Biology	28.9/ 28.9/ 31.7	62.3/ 67.3/ 69.9	20.0/ 22.0/ 24.0	30.9/ 37.7/ 36.9	29.1/ 34.8/ 31.6	24.2/ 28.3/ 31.3	28.4/ 30.4/ 32.5	25.0/ 24.0/ 25.0	27.0/ 25.0/ 29.0	26.8/ 29.0/ 29.3	25.5/ 31.8/ 30.2
Business and Management	27.0/ 32.0/ 25.0	41.5/ 48.9/ 47.7	30.0/ 28.0/ 26.0	34.4/ 34.2/ 36.3	25.0/ 27.0/ 24.0	20.0/ 30.0/ 28.0	27.0/ 30.0/ 33.0	22.0/ 27.0/ 28.0	24.0/ 28.0/ 34.0	36.0/ 31.0/ 34.0	32.1/ 30.2/ 23.6
Chemistry	29.9/ 36.0/ 38.2	52.3/ 59.1/ 58.3	25.0/ 26.7/ 26.7	33.3/ 37.7/ 36.3	29.7/ 34.6/ 33.3	32.0/ 33.0/ 32.0	31.7/ 32.7/ 34.0	30.8/ 27.9/ 33.7	26.0/ 36.0/ 36.0	31.4/ 33.8/ 36.2	29.2/ 30.7/ 33.2
Computer Science	28.1/ 30.7/ 33.3	33.0/ 35.8/ 37.6	30.0/ 30.0/ 34.0	29.2/ 31.8/ 32.8	29.0/ 32.5/ 33.1	27.0/ 22.0/ 30.0	28.4/ 38.5/ 34.9	23.0/ 24.0/ 20.0	22.0/ 30.0/ 37.0	29.6/ 33.1/ 37.2	31.8/ 33.2/ 32.7
Defense and Security	33.0/ 26.0/ 27.0	39.0/ 58.0/ 60.0	23.0/ 27.0/ 24.0	24.2/ 25.6/ 24.2	30.0/ 29.0/ 24.0	30.0/ 24.0/ 23.0	29.0/ 35.0/ 35.0	25.0/ 27.0/ 28.0	36.0/ 32.0/ 31.0	37.0/ 31.0/ 30.0	24.0/ 28.0/ 23.0
Earth Sciences	33.0/ 39.0/ 43.0	47.9/ 53.2/ 57.0	31.0/ 37.0/ 32.0	29.3/ 32.1/ 33.5	27.0/ 25.0/ 30.0	23.0/ 29.0/ 33.0	32.0/ 35.0/ 35.0	22.2/ 22.2/ 21.2	32.0/ 29.0/ 28.0	26.0/ 36.0/ 29.0	33.0/ 30.0/ 28.0
Economics	30.9/ 25.9/ 36.0	45.4/ 51.0/ 51.2	24.0/ 29.0/ 31.0	31.6/ 32.9/ 35.3	37.0/ 37.0/ 36.0	36.0/ 30.0/ 28.0	24.6/ 29.6/ 29.6	27.0/ 27.0/ 24.0	30.0/ 30.0/ 30.0	28.6/ 28.6/ 30.2	32.3/ 26.4/ 28.7
Education	23.0/ 26.0/ 25.0	36.9/ 40.5/ 43.0	23.0/ 23.0/ 23.0	29.3/ 28.7/ 29.6	32.0/ 33.0/ 32.0	25.0/ 28.0/ 28.0	25.5/ 31.2/ 31.2	21.0/ 26.0/ 29.0	32.0/ 35.0/ 35.0	31.0/ 29.0/ 28.0	30.0/ 32.0/ 23.0
Energy and Power	34.0/ 30.0/ 28.0	40.5/ 51.3/ 53.4	29.0/ 33.0/ 28.0	33.2/ 39.0/ 40.0	25.0/ 28.0/ 31.0	25.0/ 30.0/ 30.0	27.0/ 38.0/ 40.0	24.0/ 23.0/ 26.0	29.0/ 29.0/ 33.0	28.0/ 32.0/ 38.0	27.0/ 28.0/ 26.0
Engineering	25.2/ 28.4/ 27.0	34.4/ 39.4/ 39.8	26.0/ 25.6/ 27.0	27.0/ 31.6/ 32.0	26.6/ 27.2/ 29.0	24.8/ 24.8/ 26.9	32.6/ 28.0/ 29.4	20.5/ 22.6/ 23.2	32.0/ 30.0/ 29.0	29.8/ 29.0/ 30.8	24.4/ 27.2/ 27.6
Environmental Science	22.0/ 23.0/ 25.0	50.3/ 56.3/ 55.7	30.0/ 28.0/ 29.0	29.9/ 35.7/ 37.5	21.9/ 24.2/ 26.6	30.0/ 38.0/ 35.0	27.4/ 33.7/ 32.6	29.3/ 25.2/ 31.3	27.0/ 30.0/ 42.0	30.0/ 31.0/ 34.0	29.0/ 25.7/ 28.5
Ethics and Human Rights	25.0/ 20.0/ 33.0	32.0/ 51.0/ 52.0	34.0/ 29.0/ 35.0	28.0/ 32.0/ 34.0	32.0/ 36.0/ 36.0	28.0/ 28.0/ 32.0	28.0/ 28.0/ 26.0	15.0/ 14.0/ 22.0	31.0/ 33.0/ 27.0	24.0/ 29.0/ 32.0	30.0/ 28.0/ 34.0
Finance and Investment	26.2/ 28.7/ 35.1	41.4/ 43.6/ 46.4	32.0/ 32.0/ 34.0	30.8/ 30.4/ 30.2	28.4/ 32.7/ 31.5	26.0/ 26.0/ 32.0	24.3/ 27.1/ 33.3	20.0/ 23.0/ 22.0	31.0/ 25.0/ 28.0	33.5/ 34.6/ 33.5	22.5/ 24.6/ 25.4
Food Science	19.0/ 25.0/ 27.0	46.0/ 54.0/ 56.0	27.0/ 30.0/ 28.0	31.3/ 28.3/ 36.4	29.0/ 23.0/ 30.0	20.0/ 26.0/ 27.0	21.0/ 29.0/ 23.0	30.0/ 28.0/ 32.0	20.0/ 19.0/ 22.0	25.0/ 20.0/ 26.0	23.0/ 27.0/ 29.0
Geography	22.9/ 29.6/ 26.3	42.0/ 48.4/ 51.2	30.0/ 35.0/ 36.0	31.1/ 28.7/ 29.9	26.5/ 27.9/ 28.8	22.2/ 29.3/ 25.2	31.4/ 30.8/ 30.8	24.0/ 23.0/ 28.0	24.0/ 22.0/ 23.0	29.1/ 27.4/ 28.6	29.2/ 25.5/ 24.0
Health and Medicine	28.0/ 28.0/ 31.0	59.1/ 66.7/ 68.3	23.0/ 23.0/ 21.0	31.9/ 34.1/ 38.3	21.0/ 24.0/ 29.0	22.0/ 22.0/ 28.0	21.4/ 41.1/ 40.2	22.0/ 25.0/ 29.0	30.0/ 38.0/ 31.0	25.9/ 27.6/ 29.3	28.0/ 31.8/ 35.5
History	24.8/ 25.6/ 28.0	40.6/ 44.2/ 45.2	22.0/ 29.0/ 28.0	29.4/ 30.4/ 32.7	28.2/ 27.8/ 29.3	24.0/ 34.0/ 31.0	26.8/ 27.1/ 29.0	26.0/ 28.0/ 29.0	31.0/ 35.0/ 30.0	26.9/ 29.8/ 28.0	29.2/ 30.2/ 29.7
Information Technology	41.0/ 44.0/ 42.0	53.0/ 62.0/ 63.0	46.0/ 35.0/ 36.0	35.8/ 39.1/ 41.3	38.0/ 36.0/ 39.0	24.0/ 35.0/ 36.0	31.0/ 41.0/ 35.0	28.0/ 30.0/ 26.0	33.0/ 36.0/ 35.0	35.0/ 43.0/ 41.0	40.0/ 39.0/ 40.0
International Relations	28.0/ 23.0/ 25.0	58.3/ 60.0/ 59.2	34.0/ 43.0/ 33.0	31.9/ 33.9/ 37.4	21.0/ 27.0/ 31.0	35.0/ 35.0/ 32.0	33.0/ 34.0/ 34.0	30.0/ 24.0/ 16.0	27.0/ 31.0/ 36.0	26.0/ 33.0/ 34.0	27.0/ 26.0/ 26.0
Language Studies	24.0/ 29.0/ 33.0	45.0/ 52.8/ 54.4	35.0/ 25.0/ 24.0	23.2/ 29.5/ 29.3	22.5/ 24.0/ 27.1	22.0/ 26.0/ 30.0	24.8/ 29.5/ 32.2	22.0/ 22.0/ 18.0	33.0/ 31.0/ 29.0	31.5/ 24.7/ 25.1	24.1/ 23.6/ 25.3
Law and Ethics	31.8/ 35.8/ 40.5	41.4/ 52.8/ 53.0	33.0/ 31.0/ 28.0	31.0/ 33.0/ 35.4	21.0/ 22.3/ 26.8	31.0/ 28.0/ 34.0	25.0/ 33.6/ 33.6	25.0/ 29.0/ 27.0	24.0/ 28.0/ 21.0	33.1/ 29.9/ 33.1	30.0/ 24.4/ 25.0
Literature and Linguistics	26.1/ 27.0/ 25.2	39.9/ 40.3/ 42.9	22.0/ 28.0/ 26.0	24.1/ 26.7/ 27.7	30.8/ 26.7/ 25.8	25.0/ 21.0/ 22.0	26.6/ 23.9/ 22.1	21.0/ 23.0/ 23.0	24.0/ 33.0/ 30.0	30.4/ 32.8/ 30.4	26.1/ 26.1/ 30.3
Logical Reasoning	29.5/ 28.0/ 27.0	28.6/ 32.4/ 30.8	27.3/ 26.6/ 26.6	26.7/ 28.7/ 27.1	29.9/ 27.0/ 28.2	23.0/ 24.0/ 25.0	26.7/ 29.0/ 27.3	30.0/ 21.0/ 17.0	17.0/ 24.0/ 25.0	25.9/ 26.7/ 26.7	27.5/ 29.9/ 31.4
Materials Science	27.0/ 31.0/ 27.0	50.2/ 53.2/ 56.3	26.0/ 25.0/ 25.0	27.3/ 31.2/ 30.8	22.0/ 23.0/ 28.0	24.0/ 31.0/ 33.0	33.0/ 33.0/ 33.0	29.0/ 24.0/ 23.0	21.0/ 30.0/ 33.0	32.0/ 26.0/ 33.0	24.0/ 25.0/ 30.0
Media and Communication	30.0/ 26.0/ 30.0	51.0/ 61.0/ 60.0	32.0/ 38.0/ 42.0	29.9/ 36.0/ 29.2	30.0/ 32.0/ 30.0	41.0/ 31.0/ 33.0	30.0/ 41.0/ 36.0	32.0/ 33.0/ 33.0	28.0/ 26.0/ 33.0	26.0/ 29.0/ 27.0	27.0/ 31.0/ 33.0
Music and Performing Arts	22.0/ 25.0/ 24.0	43.0/ 44.0/ 45.0	31.0/ 25.0/ 27.0	27.8/ 27.8/ 23.2	28.0/ 25.0/ 27.0	26.0/ 27.0/ 22.0	30.0/ 27.0/ 31.0	27.0/ 25.0/ 24.0	23.0/ 25.0/ 27.0	24.0/ 29.0/ 26.0	31.0/ 28.0/ 31.0
Physics	30.8/ 33.0/ 32.6	50.8/ 51.4/ 54.2	29.8/ 33.4/ 30.6	29.8/ 31.2/ 32.2	26.0/ 32.0/ 31.0	25.3/ 23.7/ 28.4	30.8/ 30.6/ 32.6	23.6/ 24.3/ 25.7	34.0/ 38.0/ 35.0	28.2/ 26.0/ 29.8	28.0/ 29.0/ 30.4
Politics and Governance	27.1/ 26.3/ 29.4	42.7/ 48.1/ 46.9	27.0/ 28.0/ 29.0	27.7/ 32.5/ 31.9	23.1/ 22.3/ 22.7	23.0/ 25.0/ 24.0	24.4/ 27.5/ 29.5	22.0/ 35.0/ 30.0	20.0/ 26.0/ 25.0	25.3/ 29.5/ 29.9	27.8/ 26.2/ 25.7
Psychology	24.0/ 27.0/ 23.0	53.0/ 60.0/ 58.0	33.0/ 24.0/ 34.0	29.5/ 37.2/ 38.0	25.0/ 27.0/ 32.0	25.0/ 29.0/ 28.0	27.0/ 26.0/ 27.0	34.0/ 40.0/ 34.0	26.0/ 35.0/ 37.0	21.0/ 28.0/ 26.0	29.0/ 35.0/ 41.0
Public Administration	25.0/ 26.0/ 26.0	36.4/ 39.4/ 40.4	34.3/ 33.3/ 34.3	23.2/ 29.3/ 28.3	24.2/ 26.3/ 28.3	30.0/ 27.0/ 30.0	22.2/ 32.3/ 26.3	24.0/ 24.0/ 25.0	33.3/ 28.3/ 32.3	29.0/ 27.0/ 32.0	36.0/ 26.0/ 24.0
Religion and Spirituality	29.0/ 27.0/ 26.0	47.0/ 55.0/ 56.0	28.0/ 30.0/ 25.0	24.6/ 25.4/ 25.8	26.0/ 29.0/ 29.0	30.0/ 24.0/ 28.0	31.0/ 30.0/ 34.0	27.0/ 24.0/ 24.0	32.0/ 33.0/ 24.0	27.0/ 20.0/ 18.0	28.0/ 28.0/ 28.0
Social Welfare and Development	30.0/ 30.0/ 38.0	42.9/ 46.4/ 45.5	32.0/ 33.0/ 36.0	31.1/ 36.8/ 39.4	30.0/ 28.0/ 33.0	22.0/ 26.0/ 28.0	24.0/ 33.0/ 30.0	24.0/ 25.0/ 26.0	32.0/ 26.0/ 37.0	22.0/ 28.0/ 28.0	29.0/ 26.0/ 25.0
Sociology	26.1/ 24.3/ 20.9	36.2/ 38.4/ 36.6	17.0/ 22.0/ 29.0	24.4/ 28.6/ 28.2	26.0/ 30.0/ 29.0	19.0/ 22.0/ 23.0	26.4/ 22.6/ 27.0	27.0/ 25.0/ 28.0	22.0/ 25.0/ 26.0	21.6/ 30.6/ 23.4	20.0/ 25.9/ 24.1
Sports and Recreation	22.3/ 27.7/ 25.2	53.0/ 52.8/ 53.6	32.0/ 29.0/ 25.0	30.4/ 28.4/ 29.4	25.8/ 25.3/ 29.2	27.0/ 25.0/ 25.0	22.6/ 28.8/ 29.9	27.0/ 28.0/ 29.0	19.0/ 32.0/ 30.0	27.6/ 28.2/ 27.6	24.8/ 25.2/ 24.8
Technology and Innovation	28.0/ 31.0/ 34.0	47.4/ 52.6/ 54.2	28.0/ 27.0/ 27.0	30.4/ 32.4/ 34.2	33.0/ 34.0/ 41.0	22.0/ 29.0/ 26.0	29.0/ 35.0/ 37.0	20.0/ 19.0/ 19.0	28.0/ 30.0/ 32.0	29.0/ 32.0/ 35.0	29.0/ 26.0/ 30.0
Transportation and Logistics	31.0/ 29.0/ 29.0	36.9/ 40.8/ 41.5	23.0/ 26.0/ 22.0	23.3/ 26.8/ 27.1	24.2/ 23.2/ 27.3	30.0/ 23.0/ 26.0	25.5/ 21.4/ 25.5	22.2/ 24.2/ 24.2	27.0/ 30.0/ 27.0	32.0/ 33.0/ 30.0	33.0/ 27.0/ 27.0

Table 34: Detailed subject-wise evaluation for GOOGLE/GEMMA-2-2B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	bn	en	gu	hi	kn	ml	mr	or	pa	ta	te
Agriculture	34.0/ 31.0/ 32.0	41.6/ 47.9/ 51.1	29.3/ 35.4/ 34.3	27.4/ 32.2/ 33.6	17.0/ 19.0/ 16.0	19.2/ 19.2/ 22.2	25.0/ 24.0/ 26.0	26.0/ 27.0/ 30.0	25.0/ 28.0/ 26.0	27.0/ 28.0/ 29.0	28.0/ 31.0/ 28.0
Anthropology	28.0/ 28.0/ 29.0	47.0/ 55.0/ 53.0	31.0/ 30.0/ 32.0	44.4/ 39.4/ 42.4	27.0/ 26.0/ 28.0	33.0/ 29.0/ 40.0	41.0/ 39.0/ 38.0	27.0/ 26.0/ 29.0	24.0/ 29.0/ 27.0	33.0/ 27.0/ 28.0	30.0/ 29.0/ 29.0
Architecture and Design	23.0/ 27.0/ 27.0	40.0/ 51.0/ 48.0	32.0/ 31.0/ 28.0	37.0/ 30.0/ 34.0	19.0/ 24.0/ 27.0	31.0/ 24.0/ 21.0	40.0/ 35.0/ 36.0	20.0/ 25.0/ 20.0	26.0/ 28.0/ 31.0	24.0/ 27.0/ 23.0	23.0/ 27.0/ 25.0
Arts and Culture	26.1/ 27.3/ 29.1	44.6/ 54.4/ 57.0	17.0/ 21.0/ 20.0	29.9/ 32.9/ 32.5	21.5/ 22.7/ 19.6	30.0/ 33.0/ 29.0	30.0/ 29.5/ 26.8	23.0/ 20.0/ 21.0	31.0/ 28.0/ 32.0	24.9/ 24.2/ 30.9	21.6/ 20.6/ 19.6
Astronomy and Astrophysics	31.0/ 28.0/ 32.0	66.0/ 71.0/ 76.0	23.0/ 25.0/ 26.0	35.2/ 38.3/ 42.2	29.0/ 27.0/ 28.0	30.0/ 33.0/ 33.0	37.0/ 37.0/ 38.0	25.0/ 20.0/ 22.0	24.0/ 23.0/ 28.0	33.0/ 32.0/ 31.0	32.0/ 29.0/ 31.0
Biology	31.0/ 29.9/ 26.8	60.1/ 71.3/ 71.5	25.0/ 22.0/ 21.0	34.9/ 40.5/ 40.5	25.8/ 26.2/ 27.9	29.3/ 34.3/ 34.3	31.0/ 31.3/ 32.5	26.0/ 27.0/ 31.0	36.0/ 30.0/ 33.0	32.8/ 31.5/ 31.2	30.5/ 33.0/ 30.5
Business and Management	32.0/ 30.0/ 32.0	41.5/ 52.0/ 48.6	15.0/ 21.0/ 27.0	35.6/ 38.9/ 38.4	27.0/ 33.0/ 29.0	33.0/ 32.0/ 34.0	24.0/ 31.0/ 29.0	36.0/ 33.0/ 30.0	38.0/ 29.0/ 35.0	30.0/ 29.0/ 39.0	27.4/ 28.3/ 30.2
Chemistry	35.2/ 33.2/ 36.3	52.5/ 58.5/ 60.3	28.4/ 31.9/ 33.6	38.5/ 39.3/ 39.3	30.1/ 30.7/ 29.1	32.0/ 30.0/ 32.0	33.5/ 35.6/ 36.7	33.7/ 32.7/ 29.8	30.0/ 30.0/ 29.0	36.5/ 33.5/ 37.7	29.4/ 31.4/ 35.7
Computer Science	33.3/ 36.5/ 33.3	41.0/ 40.4/ 39.0	29.0/ 29.0/ 31.0	29.2/ 36.4/ 35.2	31.9/ 36.7/ 34.9	35.0/ 41.0/ 37.0	35.5/ 32.5/ 34.9	31.0/ 30.0/ 29.0	35.0/ 36.0/ 37.0	34.9/ 33.1/ 37.2	34.1/ 39.0/ 37.2
Defense and Security	29.0/ 30.0/ 27.0	47.0/ 60.0/ 58.0	18.0/ 26.0/ 22.0	26.1/ 24.6/ 26.5	21.0/ 24.0/ 23.0	23.0/ 22.0/ 19.0	24.0/ 34.0/ 32.0	26.0/ 29.0/ 30.0	31.0/ 27.0/ 32.0	27.0/ 31.0/ 34.0	30.0/ 37.0/ 33.0
Earth Sciences	32.0/ 28.0/ 32.0	48.9/ 57.7/ 58.0	39.0/ 41.0/ 39.0	31.5/ 34.1/ 36.3	27.0/ 35.0/ 33.0	34.0/ 37.0/ 35.0	30.0/ 33.0/ 32.0	24.2/ 17.2/ 22.2	32.0/ 28.0/ 31.0	30.0/ 31.0/ 27.0	31.0/ 30.0/ 31.0
Economics	30.2/ 30.2/ 30.9	45.8/ 51.6/ 53.4	24.0/ 25.0/ 27.0	33.5/ 39.1/ 35.5	32.0/ 34.0/ 38.0	29.0/ 33.0/ 31.0	31.0/ 34.5/ 33.1	30.0/ 25.0/ 28.0	31.0/ 29.0/ 27.0	29.4/ 30.9/ 33.3	26.4/ 29.9/ 26.4
Education	31.0/ 26.0/ 27.0	36.3/ 39.8/ 42.4	18.0/ 17.0/ 25.0	31.7/ 30.8/ 33.7	27.0/ 29.0/ 33.0	28.0/ 37.0/ 30.0	32.5/ 39.5/ 36.3	29.0/ 31.0/ 29.0	26.0/ 23.0/ 23.0	23.0/ 28.0/ 25.0	28.0/ 27.0/ 29.0
Energy and Power	31.0/ 28.0/ 25.0	47.3/ 52.0/ 51.3	26.0/ 28.0/ 28.0	31.2/ 36.6/ 35.2	24.0/ 26.0/ 31.0	26.0/ 25.0/ 23.0	27.0/ 31.0/ 32.0	19.0/ 18.0/ 29.0	24.0/ 31.0/ 24.0	26.0/ 26.0/ 25.0	30.0/ 29.0/ 29.0
Engineering	25.2/ 25.4/ 28.8	36.4/ 37.4/ 43.4	28.8/ 27.8/ 29.4	28.8/ 29.6/ 29.0	27.2/ 26.2/ 27.6	26.5/ 27.8/ 26.5	27.8/ 29.0/ 33.0	22.3/ 25.3/ 27.7	29.0/ 22.0/ 25.0	28.8/ 29.0/ 30.6	25.4/ 28.4/ 28.4
Environmental Science	31.0/ 34.0/ 35.0	47.7/ 57.1/ 57.9	32.0/ 33.0/ 32.0	35.1/ 38.5/ 39.5	26.6/ 28.1/ 27.3	30.0/ 29.0/ 29.0	37.9/ 33.2/ 32.1	32.3/ 30.3/ 29.3	30.0/ 32.0/ 34.0	31.0/ 26.0/ 23.0	29.6/ 29.0/ 29.6
Ethics and Human Rights	31.0/ 29.0/ 29.0	43.0/ 47.0/ 49.0	35.0/ 31.0/ 32.0	34.0/ 33.0/ 31.0	29.0/ 33.0/ 27.0	32.0/ 34.0/ 35.0	24.0/ 25.0/ 27.0	21.0/ 21.0/ 23.0	26.0/ 26.0/ 26.0	28.0/ 34.0/ 33.0	38.0/ 36.0/ 32.0
Finance and Investment	27.7/ 30.7/ 33.2	42.4/ 47.6/ 47.6	29.0/ 31.0/ 26.0	31.0/ 31.2/ 35.0	25.3/ 27.8/ 30.2	17.0/ 24.0/ 19.0	24.3/ 29.9/ 30.5	12.0/ 19.0/ 18.0	27.0/ 27.0/ 22.0	25.1/ 30.7/ 33.0	21.1/ 24.3/ 27.9
Food Science	21.0/ 19.0/ 24.0	54.0/ 63.0/ 62.0	31.0/ 27.0/ 26.0	34.3/ 34.3/ 38.4	26.0/ 27.0/ 24.0	22.0/ 29.0/ 31.0	39.0/ 33.0/ 40.0	24.0/ 29.0/ 30.0	23.0/ 25.0/ 24.0	31.0/ 27.0/ 26.0	31.0/ 32.0/ 29.0
Geography	27.9/ 24.6/ 24.6	44.0/ 49.6/ 50.4	31.0/ 33.0/ 32.0	34.7/ 32.1/ 29.7	24.7/ 25.6/ 26.0	22.2/ 25.2/ 28.3	37.3/ 36.1/ 35.5	28.0/ 30.0/ 30.0	25.0/ 20.0/ 21.0	33.1/ 30.9/ 31.4	22.9/ 26.0/ 29.7
Health and Medicine	32.0/ 33.0/ 36.0	57.9/ 67.5/ 68.9	25.0/ 22.0/ 22.0	35.3/ 41.1/ 39.3	36.0/ 33.0/ 30.0	32.0/ 26.0/ 25.0	28.6/ 29.5/ 28.6	23.0/ 27.0/ 25.0	24.0/ 24.0/ 28.0	30.2/ 26.7/ 31.9	34.6/ 36.4/ 31.8
History	25.1/ 25.3/ 25.1	38.1/ 43.8/ 44.0	28.0/ 33.0/ 33.0	29.2/ 33.9/ 33.9	24.9/ 25.6/ 27.8	21.0/ 33.0/ 32.0	27.5/ 24.2/ 26.4	27.0/ 23.0/ 26.0	25.0/ 28.0/ 29.0	28.4/ 30.2/ 29.1	28.4/ 28.4/ 27.6
Information Technology	35.0/ 32.0/ 39.0	56.0/ 65.0/ 64.0	40.0/ 43.0/ 43.0	36.9/ 44.1/ 48.0	31.0/ 36.0/ 41.0	32.0/ 37.0/ 30.0	36.0/ 43.0/ 40.0	31.0/ 31.0/ 24.0	33.0/ 36.0/ 38.0	39.0/ 38.0/ 42.0	43.0/ 46.0/ 43.0
International Relations	28.0/ 22.0/ 28.0	51.7/ 62.5/ 65.0	28.0/ 34.0/ 34.0	32.7/ 36.6/ 39.3	33.0/ 26.0/ 28.0	25.0/ 34.0/ 38.0	41.0/ 41.0/ 41.0	35.0/ 32.0/ 26.0	20.0/ 33.0/ 26.0	27.0/ 29.0/ 29.0	27.0/ 25.0/ 25.0
Language Studies	29.0/ 31.0/ 36.0	51.4/ 59.0/ 62.0	37.0/ 38.0/ 33.0	27.5/ 31.5/ 30.3	22.5/ 28.7/ 27.1	27.0/ 31.0/ 25.0	33.6/ 36.9/ 31.5	27.0/ 27.0/ 25.0	37.0/ 34.0/ 32.0	34.1/ 28.8/ 25.5	27.0/ 31.0/ 27.6
Law and Ethics	34.5/ 37.2/ 35.1	43.8/ 48.0/ 51.6	42.0/ 29.0/ 33.0	33.4/ 34.6/ 35.2	29.9/ 32.5/ 23.6	23.0/ 32.0/ 31.0	23.7/ 24.6/ 25.0	18.0/ 16.0/ 22.0	24.0/ 24.0/ 29.0	27.6/ 25.2/ 25.2	22.5/ 24.4/ 21.9
Literature and Linguistics	27.0/ 27.0/ 26.1	41.5/ 50.3/ 48.1	29.0/ 29.0/ 30.0	25.1/ 29.9/ 29.5	28.3/ 20.0/ 25.8	19.0/ 16.0/ 19.0	30.1/ 25.7/ 29.2	33.0/ 25.0/ 25.0	28.0/ 32.0/ 34.0	23.5/ 25.1/ 28.7	26.1/ 26.1/ 26.1
Logical Reasoning	29.7/ 30.2/ 31.2	29.4/ 31.6/ 33.0	21.4/ 24.7/ 25.3	28.9/ 33.1/ 31.1	26.1/ 25.3/ 23.6	28.0/ 31.0/ 32.0	27.6/ 29.2/ 28.4	23.0/ 26.0/ 25.0	24.0/ 33.0/ 29.0	26.3/ 30.7/ 28.7	24.3/ 28.2/ 28.0
Materials Science	29.0/ 32.0/ 33.0	41.4/ 50.2/ 53.2	30.0/ 30.0/ 36.0	28.0/ 29.9/ 32.5	34.0/ 26.0/ 32.0	32.0/ 41.0/ 34.0	18.0/ 32.0/ 33.0	21.0/ 23.0/ 23.0	23.0/ 27.0/ 27.0	28.0/ 29.0/ 34.0	28.0/ 29.0/ 30.0
Media and Communication	36.0/ 32.0/ 35.0	52.0/ 65.0/ 62.0	30.0/ 31.0/ 32.0	25.2/ 32.0/ 34.7	35.0/ 40.0/ 39.0	37.0/ 36.0/ 38.0	23.0/ 29.0/ 33.0	28.0/ 33.0/ 35.0	28.0/ 28.0/ 23.0	26.0/ 30.0/ 29.0	31.0/ 39.0/ 39.0
Music and Performing Arts	25.0/ 23.0/ 31.0	41.0/ 35.0/ 43.0	18.0/ 30.0/ 25.0	25.8/ 22.5/ 23.8	22.0/ 26.0/ 27.0	22.0/ 24.0/ 23.0	28.0/ 27.0/ 24.0	26.0/ 26.0/ 26.0	27.0/ 25.0/ 23.0	29.0/ 31.0/ 32.0	31.0/ 22.0/ 24.0
Physics	29.2/ 32.6/ 36.6	49.6/ 53.0/ 57.4	28.1/ 31.2/ 33.1	35.0/ 37.6/ 36.4	26.8/ 30.8/ 30.6	24.7/ 26.8/ 26.8	30.8/ 33.0/ 35.8	22.9/ 25.7/ 24.6	30.0/ 32.0/ 34.0	23.6/ 29.4/ 31.8	28.2/ 31.0/ 33.0
Politics and Governance	29.4/ 27.1/ 27.1	40.5/ 48.9/ 47.3	32.0/ 24.0/ 25.0	28.5/ 32.5/ 32.1	31.2/ 26.7/ 29.1	30.0/ 28.0/ 34.0	26.4/ 28.6/ 28.6	26.0/ 18.0/ 23.0	27.0/ 26.0/ 25.0	26.1/ 22.8/ 26.1	27.6/ 27.8/ 26.7
Psychology	31.0/ 30.0/ 25.0	48.0/ 63.0/ 68.0	29.0/ 26.0/ 25.0	34.1/ 41.1/ 41.1	26.0/ 28.0/ 31.0	33.0/ 40.0/ 36.0	34.0/ 35.0/ 32.0	33.0/ 37.0/ 42.0	31.0/ 30.0/ 30.0	32.0/ 27.0/ 31.0	32.0/ 38.0/ 38.0
Public Administration	33.0/ 29.0/ 32.0	42.4/ 48.5/ 45.5	33.3/ 30.3/ 33.3	31.3/ 37.4/ 31.3	23.2/ 24.2/ 26.3	32.0/ 26.0/ 20.0	23.2/ 25.2/ 27.3	15.0/ 20.0/ 15.0	29.3/ 33.3/ 34.3	35.0/ 26.0/ 28.0	29.0/ 29.0/ 32.0
Religion and Spirituality	26.0/ 27.0/ 29.0	52.0/ 56.0/ 59.0	24.0/ 33.0/ 29.0	29.8/ 29.0/ 28.6	21.0/ 25.0/ 24.0	20.0/ 32.0/ 27.0	26.0/ 33.0/ 34.0	26.0/ 27.0/ 26.0	35.0/ 32.0/ 28.0	28.0/ 28.0/ 25.0	30.0/ 32.0/ 29.0
Social Welfare and Development	31.0/ 26.0/ 31.0	34.8/ 48.2/ 46.4	30.0/ 26.0/ 24.0	32.1/ 34.7/ 37.3	30.0/ 30.0/ 31.0	16.0/ 23.0/ 21.0	35.0/ 44.0/ 44.0	30.0/ 26.0/ 27.0	22.0/ 28.0/ 34.0	16.0/ 19.0/ 21.0	31.0/ 34.0/ 32.0
Sociology	20.0/ 21.7/ 24.3	36.0/ 36.0/ 38.6	23.0/ 27.0/ 27.0	24.8/ 28.0/ 27.6	24.0/ 29.0/ 27.0	22.0/ 27.0/ 24.0	30.2/ 29.6/ 28.9	20.0/ 23.0/ 24.0	31.0/ 33.0/ 31.0	25.2/ 25.2/ 24.3	20.6/ 20.6/ 18.8
Sports and Recreation	29.2/ 28.2/ 29.2	48.6/ 55.4/ 55.0	29.0/ 27.0/ 30.0	28.8/ 28.4/ 30.2	33.1/ 31.5/ 31.5	32.0/ 26.0/ 24.0	35.0/ 31.6/ 28.8	26.0/ 29.0/ 23.0	28.0/ 28.0/ 30.0	28.2/ 28.2/ 29.5	24.8/ 26.2/ 26.2
Technology and Innovation	31.0/ 34.0/ 34.0	47.2/ 53.8/ 54.8	26.0/ 26.0/ 22.0	33.0/ 38.8/ 38.8	34.0/ 37.0/ 39.0	38.0/ 33.0/ 33.0	31.0/ 38.0/ 41.0	24.0/ 19.0/ 30.0	30.0/ 27.0/ 33.0	31.0/ 30.0/ 34.0	28.0/ 28.0/ 30.0
Transportation and Logistics	28.0/ 30.0/ 32.0	45.4/ 47.7/ 46.2	25.0/ 26.0/ 24.0	28.1/ 28.7/ 27.8	21.2/ 25.2/ 22.2	24.0/ 29.0/ 24.0	28.6/ 25.5/ 28.6	29.3/ 29.3/ 24.2	26.0/ 27.0/ 25.0	27.0/ 25.0/ 28.0	25.0/ 24.0/ 23.0

Table 35: Detailed subject-wise evaluation for GOOGLE/GEMMA-2-2B-IT on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	23.0/ 23.0/ 28.0	27.2/ 33.6/ 33.2	31.3/ 27.3/ 24.2	28.4/ 28.2/ 32.8	19.0/ 26.0/ 29.0	25.2/ 28.3/ 32.3	27.0/ 22.0/ 27.0	31.0/ 26.0/ 25.0	22.0/ 27.0/ 26.0	31.0/ 23.0/ 29.0	32.0/ 24.0/ 27.0
Anthropology	25.0/ 19.0/ 27.0	32.0/ 29.0/ 34.0	27.0/ 31.0/ 35.0	25.2/ 30.3/ 34.3	28.0/ 20.0/ 29.0	24.0/ 13.0/ 24.0	22.0/ 22.0/ 24.0	29.0/ 33.0/ 30.0	21.0/ 30.0/ 27.0	22.0/ 19.0/ 25.0	25.0/ 30.0/ 28.0
Architecture and Design	24.0/ 20.0/ 26.0	31.0/ 34.0/ 30.0	26.0/ 30.0/ 26.0	31.0/ 27.0/ 30.0	30.0/ 29.0/ 26.0	24.0/ 18.0/ 19.0	27.0/ 28.0/ 25.0	29.0/ 27.0/ 22.0	29.0/ 33.0/ 30.0	21.0/ 25.0/ 31.0	22.0/ 28.0/ 31.0
Arts and Culture	26.7/ 24.9/ 29.1	29.2/ 36.8/ 37.0	30.0/ 26.0/ 30.0	22.7/ 24.6/ 27.5	21.5/ 28.8/ 28.2	22.0/ 23.0/ 24.0	25.1/ 25.7/ 25.1	21.0/ 39.0/ 32.0	28.0/ 31.0/ 30.0	29.1/ 25.4/ 26.1	19.6/ 22.6/ 24.1
Astronomy and Astrophysics	26.0/ 33.0/ 34.0	27.0/ 39.0/ 48.0	33.0/ 32.0/ 39.0	35.2/ 31.2/ 43.8	37.0/ 32.0/ 35.0	32.0/ 21.0/ 25.0	30.0/ 27.0/ 32.0	25.0/ 32.0/ 29.0	26.0/ 35.0/ 31.0	29.0/ 24.0/ 29.0	27.0/ 40.0/ 34.0
Biology	22.9/ 26.8/ 30.3	33.1/ 40.1/ 47.3	27.0/ 24.0/ 25.0	28.7/ 27.1/ 34.3	24.2/ 26.2/ 32.0	31.3/ 28.3/ 31.3	23.6/ 26.0/ 30.4	25.0/ 25.0/ 23.0	32.0/ 35.0/ 32.0	29.9/ 26.1/ 28.7	28.9/ 26.7/ 32.1
Business and Management	32.0/ 23.0/ 22.0	24.5/ 26.0/ 29.7	22.0/ 18.0/ 17.0	25.8/ 29.0/ 31.6	27.0/ 26.0/ 28.0	28.0/ 25.0/ 21.0	21.0/ 24.0/ 24.0	26.0/ 22.0/ 22.0	27.0/ 26.0/ 26.0	25.0/ 30.0/ 31.0	19.8/ 20.8/ 19.8
Chemistry	29.7/ 32.7/ 34.9	30.3/ 32.9/ 36.3	20.7/ 37.1/ 38.8	29.3/ 29.1/ 37.7	27.1/ 32.7/ 31.4	29.0/ 39.0/ 38.0	30.6/ 31.4/ 34.3	32.7/ 24.0/ 27.9	34.0/ 32.0/ 34.0	26.4/ 28.8/ 31.4	29.9/ 32.4/ 34.4
Computer Science	28.1/ 25.0/ 26.6	28.0/ 31.8/ 35.6	32.0/ 29.0/ 23.0	27.2/ 31.2/ 31.8	33.1/ 29.0/ 30.8	26.0/ 36.0/ 33.0	32.5/ 32.5/ 27.8	30.0/ 27.0/ 29.0	23.0/ 31.0/ 33.0	33.1/ 31.4/ 27.3	28.7/ 30.5/ 28.2
Defense and Security	23.0/ 21.0/ 22.0	35.0/ 31.0/ 31.0	24.0/ 23.0/ 23.0	21.8/ 22.3/ 28.0	27.0/ 28.0/ 30.0	30.0/ 31.0/ 31.0	29.0/ 23.0/ 26.0	25.0/ 29.0/ 26.0	23.0/ 24.0/ 25.0	23.0/ 24.0/ 22.0	20.0/ 23.0/ 25.0
Earth Sciences	34.0/ 26.0/ 41.0	26.8/ 37.5/ 42.5	22.0/ 27.0/ 32.0	23.4/ 30.5/ 37.1	30.0/ 21.0/ 31.0	21.0/ 26.0/ 29.0	24.0/ 32.0/ 39.0	27.3/ 29.3/ 30.3	30.0/ 31.0/ 30.0	28.0/ 28.0/ 27.0	29.0/ 31.0/ 37.0
Economics	36.7/ 38.1/ 41.0	31.3/ 35.5/ 38.4	33.0/ 36.0/ 33.0	27.8/ 30.2/ 30.6	25.0/ 28.0/ 33.0	23.0/ 29.0/ 30.0	26.8/ 26.1/ 30.3	31.0/ 32.0/ 29.0	32.0/ 32.0/ 34.0	19.1/ 28.6/ 30.9	28.1/ 28.1/ 27.5
Education	25.0/ 26.0/ 28.0	22.9/ 26.1/ 28.0	25.0/ 25.0/ 29.0	25.4/ 24.0/ 27.5	30.0/ 26.0/ 24.0	16.0/ 26.0/ 19.0	27.4/ 22.3/ 31.2	29.0/ 27.0/ 30.0	18.0/ 30.0/ 35.0	23.0/ 26.0/ 34.0	25.0/ 18.0/ 24.0
Energy and Power	22.0/ 24.0/ 30.0	31.8/ 40.5/ 40.5	26.0/ 27.0/ 35.0	30.5/ 32.9/ 34.6	36.0/ 30.0/ 36.0	24.0/ 22.0/ 25.0	21.0/ 28.0/ 27.0	22.0/ 21.0/ 24.0	24.0/ 27.0/ 30.0	27.0/ 27.0/ 32.0	29.0/ 25.0/ 29.0
Engineering	27.2/ 24.2/ 27.8	26.0/ 27.8/ 30.4	24.6/ 28.2/ 28.0	26.0/ 25.8/ 26.4	24.6/ 23.2/ 27.6	24.4/ 26.1/ 25.6	25.4/ 26.8/ 28.8	25.6/ 26.2/ 26.8	20.0/ 24.0/ 20.0	28.2/ 24.8/ 25.2	26.8/ 23.6/ 29.6
Environmental Science	26.0/ 26.0/ 33.0	29.5/ 31.9/ 36.7	21.0/ 27.0/ 39.0	24.1/ 29.7/ 35.1	23.4/ 21.9/ 19.5	24.0/ 27.0/ 25.0	27.4/ 24.2/ 24.7	20.2/ 22.2/ 35.4	24.0/ 35.0/ 41.0	34.0/ 32.0/ 28.0	27.9/ 26.3/ 26.8
Ethics and Human Rights	21.0/ 25.0/ 28.0	33.0/ 33.0/ 33.0	21.0/ 29.0/ 37.0	28.0/ 26.0/ 31.0	24.0/ 31.0/ 32.0	23.0/ 29.0/ 35.0	23.0/ 27.0/ 35.0	25.0/ 29.0/ 26.0	33.0/ 31.0/ 33.0	27.0/ 28.0/ 28.0	23.0/ 24.0/ 23.0
Finance and Investment	27.7/ 30.2/ 32.7	30.6/ 30.4/ 29.6	27.0/ 36.0/ 36.0	27.8/ 29.8/ 31.6	29.0/ 30.9/ 32.7	35.0/ 27.0/ 32.0	28.8/ 29.4/ 27.7	24.0/ 21.0/ 30.0	27.0/ 32.0/ 36.0	26.3/ 25.7/ 25.7	25.4/ 27.5/ 29.3
Food Science	26.0/ 30.0/ 33.0	26.0/ 38.0/ 42.0	39.0/ 28.0/ 27.0	27.3/ 26.3/ 33.3	29.0/ 26.0/ 27.0	26.0/ 24.0/ 26.0	27.0/ 24.0/ 28.0	22.0/ 24.0/ 25.0	18.0/ 30.0/ 25.0	27.0/ 25.0/ 27.0	28.0/ 26.0/ 30.0
Geography	24.6/ 24.0/ 27.4	24.2/ 32.0/ 34.2	28.0/ 33.0/ 31.0	21.7/ 26.7/ 28.7	24.2/ 29.2/ 32.9	21.2/ 23.2/ 27.3	23.7/ 23.7/ 26.0	24.0/ 24.0/ 30.0	24.0/ 33.0/ 30.0	22.9/ 20.6/ 27.4	23.4/ 29.2/ 25.5
Health and Medicine	23.0/ 28.0/ 38.0	26.1/ 40.7/ 42.1	27.0/ 28.0/ 28.0	30.7/ 29.5/ 38.1	24.0/ 37.0/ 38.0	27.0/ 32.0/ 31.0	28.6/ 24.1/ 25.9	27.0/ 21.0/ 30.0	32.0/ 42.0/ 44.0	25.9/ 28.4/ 28.4	34.6/ 32.7/ 34.6
History	24.0/ 27.7/ 30.7	28.3/ 33.7/ 35.7	31.0/ 28.0/ 33.0	28.0/ 27.0/ 30.2	25.3/ 26.7/ 30.8	21.0/ 24.0/ 25.0	25.3/ 30.9/ 29.0	34.0/ 26.0/ 29.0	35.0/ 29.0/ 32.0	27.6/ 24.7/ 29.1	27.6/ 28.4/ 28.1
Information Technology	32.0/ 34.0/ 45.0	26.0/ 30.0/ 45.0	32.0/ 39.0/ 40.0	29.6/ 30.2/ 33.5	28.0/ 33.0/ 38.0	29.0/ 36.0/ 34.0	33.0/ 36.0/ 42.0	34.0/ 29.0/ 31.0	31.0/ 33.0/ 34.0	35.0/ 38.0/ 37.0	31.0/ 26.0/ 34.0
International Relations	29.0/ 32.0/ 36.0	25.0/ 33.3/ 36.7	21.0/ 32.0/ 28.0	26.5/ 30.0/ 33.1	31.0/ 29.0/ 30.0	34.0/ 23.0/ 22.0	24.0/ 32.0/ 40.0	27.0/ 30.0/ 22.0	33.0/ 28.0/ 29.0	28.0/ 24.0/ 33.0	28.0/ 34.0/ 30.0
Language Studies	39.0/ 29.0/ 34.0	30.8/ 31.2/ 31.4	28.0/ 30.0/ 37.0	24.2/ 26.1/ 26.1	26.4/ 24.8/ 26.4	22.0/ 24.0/ 26.0	28.2/ 30.9/ 26.9	37.0/ 34.0/ 35.0	29.0/ 36.0/ 36.0	27.7/ 26.2/ 26.6	31.0/ 32.2/ 25.9
Law and Ethics	27.7/ 33.1/ 33.8	27.0/ 30.8/ 33.2	30.0/ 26.0/ 27.0	28.0/ 30.6/ 34.4	27.4/ 21.0/ 18.5	22.0/ 24.0/ 23.0	26.3/ 23.7/ 28.0	29.0/ 30.0/ 31.0	23.0/ 23.0/ 26.0	31.5/ 29.9/ 31.5	25.6/ 25.0/ 23.8
Literature and Linguistics	26.1/ 27.0/ 25.2	27.1/ 27.1/ 29.3	27.0/ 27.0/ 27.0	25.1/ 23.6/ 26.9	25.0/ 30.8/ 32.5	20.0/ 19.0/ 21.0	21.2/ 22.1/ 24.8	19.0/ 20.0/ 23.0	15.0/ 28.0/ 29.0	30.0/ 29.1/ 32.0	31.7/ 25.4/ 29.6
Logical Reasoning	31.4/ 32.2/ 29.5	25.6/ 24.4/ 29.0	29.2/ 25.3/ 25.3	26.7/ 25.9/ 28.9	29.5/ 22.4/ 21.2	29.0/ 28.0/ 24.0	31.5/ 26.2/ 27.6	27.0/ 24.0/ 19.0	32.0/ 29.0/ 22.0	29.1/ 28.3/ 25.5	26.2/ 22.3/ 24.0
Materials Science	22.0/ 25.0/ 31.0	26.6/ 31.2/ 36.1	16.0/ 24.0/ 22.0	24.9/ 27.1/ 31.4	25.0/ 30.0/ 30.0	24.0/ 29.0/ 30.0	23.0/ 23.0/ 22.0	15.0/ 25.0/ 24.0	29.0/ 22.0/ 26.0	19.0/ 15.0/ 24.0	29.0/ 29.0/ 30.0
Media and Communication	39.0/ 26.0/ 33.0	28.0/ 35.0/ 31.0	22.0/ 34.0/ 32.0	21.8/ 23.1/ 28.6	37.0/ 30.0/ 36.0	25.0/ 15.0/ 19.0	23.0/ 33.0/ 34.0	32.0/ 31.0/ 31.0	22.0/ 20.0/ 22.0	34.0/ 28.0/ 32.0	32.0/ 32.0/ 37.0
Music and Performing Arts	31.0/ 23.0/ 23.0	30.0/ 27.0/ 31.0	25.0/ 29.0/ 33.0	21.9/ 28.5/ 25.8	22.0/ 23.0/ 24.0	29.0/ 29.0/ 34.0	22.0/ 30.0/ 32.0	20.0/ 19.0/ 22.0	24.0/ 21.0/ 18.0	21.0/ 21.0/ 17.0	27.0/ 26.0/ 22.0
Physics	21.4/ 23.4/ 28.4	28.6/ 33.4/ 35.6	22.8/ 25.6/ 28.4	26.0/ 28.0/ 32.6	23.2/ 27.8/ 27.6	24.7/ 24.2/ 19.5	27.8/ 29.0/ 33.2	23.3/ 26.4/ 24.3	25.0/ 26.0/ 23.0	24.6/ 24.6/ 26.4	25.4/ 28.8/ 27.0
Politics and Governance	23.1/ 25.9/ 31.8	26.9/ 34.5/ 36.1	27.0/ 28.0/ 26.0	28.1/ 29.5/ 29.3	27.1/ 25.1/ 25.9	20.0/ 19.0/ 15.0	26.7/ 26.7/ 25.0	27.0/ 26.0/ 22.0	23.0/ 21.0/ 17.0	22.0/ 25.3/ 27.0	25.5/ 27.4/ 28.8
Psychology	17.0/ 16.0/ 29.0	26.0/ 25.0/ 27.0	32.0/ 22.0/ 27.0	22.5/ 25.6/ 26.4	32.0/ 33.0/ 35.0	23.0/ 18.0/ 21.0	22.0/ 17.0/ 27.0	28.0/ 29.0/ 28.0	24.0/ 18.0/ 21.0	18.0/ 15.0/ 19.0	26.0/ 33.0/ 34.0
Public Administration	19.0/ 22.0/ 25.0	19.2/ 33.3/ 32.3	22.2/ 31.3/ 24.2	28.3/ 27.3/ 40.4	29.3/ 29.3/ 28.3	18.0/ 21.0/ 19.0	22.2/ 18.2/ 31.3	21.0/ 16.0/ 29.0	24.2/ 30.3/ 33.3	24.0/ 22.0/ 32.0	20.0/ 21.0/ 29.0
Religion and Spirituality	24.0/ 32.0/ 31.0	20.0/ 28.0/ 34.0	29.0/ 32.0/ 33.0	24.6/ 26.2/ 27.4	14.0/ 22.0/ 23.0	22.0/ 24.0/ 25.0	22.0/ 24.0/ 25.0	17.0/ 24.0/ 36.0	22.0/ 27.0/ 21.0	23.0/ 28.0/ 23.0	21.0/ 25.0/ 27.0
Social Welfare and Development	21.0/ 20.0/ 25.0	24.1/ 30.4/ 33.0	20.0/ 30.0/ 30.0	21.8/ 25.4/ 35.2	24.0/ 28.0/ 30.0	25.0/ 26.0/ 26.0	30.0/ 27.0/ 33.0	20.0/ 24.0/ 24.0	22.0/ 34.0/ 37.0	23.0/ 26.0/ 32.0	37.0/ 36.0/ 33.0
Sociology	21.7/ 20.0/ 21.7	28.0/ 27.8/ 30.8	20.0/ 24.0/ 34.0	22.2/ 24.6/ 26.6	19.0/ 27.0/ 26.0	16.0/ 23.0/ 26.0	24.5/ 23.3/ 20.8	22.0/ 23.0/ 24.0	23.0/ 22.0/ 19.0	15.3/ 18.9/ 23.4	27.7/ 33.5/ 27.7
Sports and Recreation	22.8/ 24.8/ 21.3	23.8/ 31.2/ 34.2	23.0/ 23.0/ 30.0	22.6/ 23.6/ 29.4	25.3/ 26.4/ 27.5	29.0/ 19.0/ 25.0	23.2/ 23.2/ 27.1	23.0/ 22.0/ 28.0	25.0/ 33.0/ 32.0	25.6/ 25.0/ 26.3	18.1/ 26.2/ 30.0
Technology and Innovation	27.0/ 20.0/ 31.0	26.0/ 31.0/ 34.2	22.0/ 25.0/ 23.0	26.0/ 27.8/ 34.2	30.0/ 31.0/ 36.0	23.0/ 26.0/ 26.0	33.0/ 37.0/ 35.0	26.0/ 26.0/ 26.0	20.0/ 31.0/ 38.0	25.0/ 25.0/ 30.0	28.0/ 27.0/ 32.0
Transportation and Logistics	27.0/ 28.0/ 20.0	26.2/ 23.1/ 23.1	22.0/ 28.0/ 23.0	25.2/ 27.4/ 25.2	27.3/ 21.2/ 28.3	23.0/ 24.0/ 26.0	22.4/ 23.5/ 21.4	30.3/ 27.3/ 26.3	28.0/ 21.0/ 24.0	24.0/ 28.0/ 23.0	23.0/ 29.0/ 28.0

Table 36: Detailed subject-wise evaluation for SARVAMAI/SARVAM-2B-V0.5 on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	31.0/ 32.0/ 35.0	34.0/ 46.1/ 47.5	20.2/ 27.3/ 29.3	25.4/ 29.4/ 35.2	27.0/ 33.0/ 29.0	28.3/ 27.3/ 27.3	31.0/ 34.0/ 34.0	33.0/ 33.0/ 29.0	29.0/ 25.0/ 22.0	28.0/ 28.0/ 28.0	18.0/ 23.0/ 26.0
Anthropology	34.0/ 34.0/ 38.0	37.0/ 57.0/ 58.0	36.0/ 41.0/ 33.0	33.3/ 31.3/ 25.2	25.0/ 24.0/ 27.0	23.0/ 27.0/ 24.0	27.0/ 32.0/ 34.0	26.0/ 38.0/ 32.0	30.0/ 27.0/ 27.0	30.0/ 27.0/ 34.0	26.0/ 25.0/ 19.0
Architecture and Design	27.0/ 32.0/ 34.0	38.0/ 60.0/ 62.0	27.0/ 21.0/ 24.0	27.0/ 35.0/ 34.0	32.0/ 38.0/ 34.0	24.0/ 25.0/ 28.0	30.0/ 38.0/ 36.0	24.0/ 24.0/ 27.0	25.0/ 27.0/ 31.0	34.0/ 30.0/ 26.0	23.0/ 30.0/ 28.0
Arts and Culture	27.9/ 29.1/ 32.7	40.0/ 59.6/ 59.6	30.0/ 31.0/ 30.0	25.9/ 31.1/ 33.7	27.6/ 28.8/ 32.5	26.0/ 34.0/ 31.0	22.9/ 26.2/ 31.1	15.0/ 15.0/ 21.0	24.0/ 26.0/ 29.0	27.3/ 27.9/ 26.7	26.1/ 23.6/ 28.1
Astronomy and Astrophysics	29.0/ 28.0/ 34.0	43.0/ 56.0/ 69.0	29.0/ 39.0/ 35.0	31.2/ 33.6/ 34.4	37.0/ 39.0/ 35.0	31.0/ 34.0/ 37.0	33.0/ 34.0/ 38.0	24.0/ 28.0/ 22.0	30.0/ 34.0/ 31.0	27.0/ 35.0/ 37.0	34.0/ 38.0/ 34.0
Biology	25.7/ 28.5/ 33.1	45.5/ 62.5/ 66.7	25.0/ 28.0/ 27.0	32.7/ 34.9/ 39.5	33.2/ 34.4/ 33.6	28.3/ 28.3/ 25.2	30.8/ 32.8/ 35.8	28.0/ 35.0/ 33.0	29.0/ 34.0/ 35.0	28.3/ 33.4/ 29.9	27.4/ 31.4/ 32.4
Business and Management	29.0/ 31.0/ 31.0	36.2/ 48.9/ 50.8	31.0/ 30.0/ 38.0	30.9/ 34.4/ 37.7	33.0/ 32.0/ 29.0	30.0/ 27.0/ 27.0	31.0/ 35.0/ 38.0	24.0/ 18.0/ 23.0	32.0/ 30.0/ 36.0	34.0/ 33.0/ 30.0	30.2/ 31.1/ 27.4
Chemistry	28.3/ 29.4/ 33.5	39.1/ 54.9/ 56.7	29.3/ 33.6/ 30.2	33.3/ 35.3/ 38.1	29.4/ 25.5/ 30.1	33.0/ 25.0/ 31.0	30.1/ 31.4/ 33.5	25.0/ 30.8/ 29.8	27.0/ 24.0/ 34.0	26.4/ 27.9/ 29.4	32.7/ 30.4/ 33.2
Computer Science	29.2/ 30.2/ 34.9	34.4/ 38.8/ 39.4	28.0/ 33.0/ 37.0	32.6/ 30.6/ 32.8	30.2/ 32.5/ 32.5	26.0/ 29.0/ 30.0	27.2/ 33.1/ 34.9	27.0/ 25.0/ 24.0	30.0/ 31.0/ 22.0	28.5/ 30.2/ 36.6	30.9/ 31.8/ 36.3
Defense and Security	23.0/ 24.0/ 30.0	30.0/ 52.0/ 50.0	24.0/ 23.0/ 24.0	25.6/ 24.6/ 25.6	25.0/ 24.0/ 25.0	21.0/ 20.0/ 24.0	16.0/ 20.0/ 25.0	27.0/ 26.0/ 29.0	33.0/ 32.0/ 34.0	28.0/ 33.0/ 32.0	26.0/ 26.0/ 24.0
Earth Sciences	32.0/ 36.0/ 32.0	36.6/ 55.9/ 59.6	30.0/ 36.0/ 34.0	30.5/ 27.9/ 34.7	30.0/ 30.0/ 25.0	28.0/ 28.0/ 31.0	21.0/ 31.0/ 28.0	25.2/ 32.3/ 28.3	30.0/ 39.0/ 30.0	24.0/ 26.0/ 28.0	25.0/ 21.0/ 25.0
Economics	30.2/ 34.5/ 35.2	36.4/ 48.6/ 51.4	35.0/ 29.0/ 32.0	28.6/ 34.1/ 38.1	28.0/ 29.0/ 32.0	29.0/ 30.0/ 29.0	31.0/ 30.3/ 28.2	25.0/ 31.0/ 29.0	29.0/ 25.0/ 27.0	39.7/ 36.5/ 33.3	32.3/ 38.9/ 35.3
Education	32.0/ 26.0/ 31.0	31.5/ 39.5/ 38.2	25.0/ 29.0/ 28.0	28.7/ 33.1/ 33.7	33.0/ 36.0/ 37.0	23.0/ 26.0/ 32.0	36.3/ 30.6/ 38.2	25.0/ 22.0/ 27.0	31.0/ 29.0/ 25.0	27.0/ 30.0/ 32.0	24.0/ 29.0/ 27.0
Energy and Power	28.0/ 29.0/ 25.0	37.8/ 54.7/ 54.7	29.0/ 37.0/ 41.0	26.8/ 30.9/ 40.7	31.0/ 32.0/ 27.0	32.0/ 27.0/ 18.0	19.0/ 20.0/ 30.0	28.0/ 28.0/ 30.0	27.0/ 30.0/ 29.0	33.0/ 29.0/ 41.0	33.0/ 33.0/ 32.0
Engineering	30.0/ 25.6/ 32.4	31.2/ 36.8/ 40.8	25.8/ 26.2/ 28.2	25.6/ 27.4/ 28.2	25.2/ 29.0/ 29.8	29.1/ 28.2/ 30.3	27.6/ 25.0/ 28.0	29.5/ 29.5/ 28.6	28.0/ 34.0/ 32.0	27.8/ 29.4/ 30.6	24.0/ 27.0/ 26.2
Environmental Science	37.0/ 38.0/ 31.0	40.7/ 54.9/ 55.5	27.0/ 34.0/ 31.0	28.7/ 33.1/ 36.5	27.3/ 27.3/ 32.8	35.0/ 29.0/ 32.0	25.3/ 32.1/ 34.2	30.3/ 29.3/ 27.3	43.0/ 38.0/ 34.0	34.0/ 28.0/ 31.0	31.8/ 29.6/ 25.1
Ethics and Human Rights	35.0/ 30.0/ 41.0	40.0/ 52.0/ 49.0	32.0/ 30.0/ 31.0	35.0/ 44.0/ 39.0	38.0/ 33.0/ 32.0	24.0/ 29.0/ 28.0	33.0/ 36.0/ 39.0	29.0/ 28.0/ 32.0	24.0/ 26.0/ 25.0	29.0/ 37.0/ 35.0	36.0/ 35.0/ 36.0
Finance and Investment	27.7/ 29.2/ 30.7	36.8/ 42.2/ 43.2	35.0/ 37.0/ 35.0	33.8/ 33.0/ 34.0	30.9/ 37.0/ 38.9	18.0/ 24.0/ 29.0	27.1/ 33.3/ 33.9	20.0/ 26.0/ 23.0	36.0/ 35.0/ 31.0	32.4/ 32.4/ 34.1	31.4/ 31.4/ 32.9
Food Science	21.0/ 29.0/ 31.0	50.0/ 60.0/ 65.0	30.0/ 32.0/ 32.0	27.3/ 33.3/ 39.4	31.0/ 33.0/ 40.0	28.0/ 36.0/ 32.0	21.0/ 27.0/ 32.0	32.0/ 27.0/ 24.0	26.0/ 28.0/ 26.0	31.0/ 32.0/ 28.0	33.0/ 28.0/ 29.0
Geography	31.3/ 35.2/ 35.2	36.6/ 46.6/ 50.0	37.0/ 31.0/ 29.0	26.5/ 28.3/ 31.3	28.3/ 34.7/ 36.5	31.3/ 41.4/ 45.5	28.4/ 30.8/ 34.3	22.0/ 29.0/ 30.0	30.0/ 30.0/ 28.0	30.9/ 32.6/ 30.9	25.0/ 27.1/ 28.1
Health and Medicine	31.0/ 34.0/ 31.0	43.3/ 65.7/ 68.7	30.0/ 30.0/ 35.0	33.5/ 37.3/ 42.9	25.0/ 28.0/ 28.0	31.0/ 33.0/ 31.0	35.7/ 33.0/ 32.1	23.0/ 25.0/ 23.0	35.0/ 37.0/ 42.0	32.8/ 32.8/ 29.3	29.9/ 29.9/ 32.7
History	30.9/ 28.8/ 31.2	38.8/ 44.8/ 44.6	35.0/ 35.0/ 37.0	32.7/ 32.9/ 32.9	31.9/ 30.0/ 33.3	34.0/ 27.0/ 30.0	32.7/ 30.5/ 29.4	36.0/ 32.0/ 30.0	29.0/ 30.0/ 33.0	31.6/ 33.8/ 32.7	32.2/ 31.7/ 29.9
Information Technology	40.0/ 40.0/ 42.0	43.0/ 58.0/ 57.0	37.0/ 35.0/ 37.0	36.9/ 41.3/ 45.2	38.0/ 37.0/ 41.0	31.0/ 33.0/ 35.0	43.0/ 46.0/ 44.0	29.0/ 29.0/ 31.0	43.0/ 48.0/ 40.0	30.0/ 29.0/ 31.0	43.0/ 51.0/ 46.0
International Relations	23.0/ 25.0/ 27.0	43.3/ 53.3/ 57.5	31.0/ 31.0/ 32.0	31.9/ 34.2/ 38.5	21.0/ 30.0/ 30.0	30.0/ 31.0/ 30.0	30.0/ 33.0/ 43.0	27.0/ 35.0/ 32.0	35.0/ 35.0/ 35.0	39.0/ 37.0/ 39.0	24.0/ 22.0/ 28.0
Language Studies	28.0/ 29.0/ 33.0	38.2/ 46.8/ 49.6	25.0/ 26.0/ 31.0	25.1/ 28.1/ 28.5	30.2/ 24.0/ 24.8	25.0/ 21.0/ 22.0	29.5/ 33.6/ 34.9	32.0/ 31.0/ 30.0	38.0/ 38.0/ 37.0	26.6/ 30.0/ 28.8	24.7/ 29.3/ 29.3
Law and Ethics	40.5/ 39.2/ 43.2	44.6/ 52.8/ 51.2	29.0/ 32.0/ 30.0	35.2/ 37.2/ 38.8	29.3/ 25.5/ 26.8	32.0/ 32.0/ 31.0	29.3/ 26.7/ 29.7	32.0/ 32.0/ 26.0	28.0/ 37.0/ 34.0	26.0/ 28.3/ 28.3	29.4/ 25.6/ 26.9
Literature and Linguistics	28.8/ 30.6/ 36.0	36.9/ 46.7/ 44.9	29.0/ 21.0/ 26.0	24.1/ 25.7/ 31.7	27.5/ 30.8/ 30.0	24.0/ 27.0/ 28.0	22.1/ 29.2/ 26.6	21.0/ 19.0/ 25.0	28.0/ 33.0/ 32.0	30.4/ 32.4/ 29.1	22.5/ 23.9/ 28.9
Logical Reasoning	28.5/ 34.9/ 35.9	31.0/ 34.2/ 31.2	26.6/ 29.2/ 29.2	27.1/ 26.2/ 29.3	31.1/ 25.7/ 31.9	31.0/ 26.0/ 25.0	22.8/ 26.2/ 27.0	19.0/ 23.0/ 26.0	34.0/ 37.0/ 27.0	28.7/ 29.9/ 29.5	25.7/ 28.0/ 27.5
Materials Science	25.0/ 28.0/ 30.0	33.8/ 48.7/ 51.7	26.0/ 23.0/ 22.0	26.4/ 31.2/ 32.3	27.0/ 26.0/ 28.0	32.0/ 32.0/ 34.0	29.0/ 30.0/ 26.0	27.0/ 19.0/ 21.0	26.0/ 28.0/ 27.0	27.0/ 26.0/ 28.0	30.0/ 26.0/ 24.0
Media and Communication	33.0/ 35.0/ 33.0	41.0/ 57.0/ 48.0	35.0/ 36.0/ 32.0	34.0/ 32.0/ 35.4	38.0/ 41.0/ 32.0	36.0/ 31.0/ 34.0	26.0/ 29.0/ 28.0	37.0/ 32.0/ 34.0	26.0/ 26.0/ 28.0	30.0/ 25.0/ 30.0	35.0/ 35.0/ 34.0
Music and Performing Arts	32.0/ 31.0/ 28.0	38.0/ 47.0/ 46.0	27.0/ 34.0/ 32.0	27.8/ 33.1/ 30.5	27.0/ 27.0/ 29.0	23.0/ 23.0/ 20.0	28.0/ 32.0/ 35.0	21.0/ 32.0/ 30.0	23.0/ 32.0/ 34.0	27.0/ 30.0/ 29.0	31.0/ 32.0/ 28.0
Physics	28.8/ 31.8/ 31.8	37.8/ 49.4/ 52.6	32.0/ 29.5/ 28.1	27.6/ 30.6/ 32.8	29.4/ 32.0/ 33.0	26.3/ 25.3/ 27.9	27.4/ 30.8/ 35.4	28.8/ 28.1/ 28.5	27.0/ 27.0/ 27.0	29.0/ 28.8/ 29.2	28.8/ 33.2/ 32.0
Politics and Governance	27.8/ 31.8/ 30.6	37.9/ 47.1/ 45.3	25.0/ 31.0/ 30.0	27.5/ 31.5/ 34.9	23.9/ 25.1/ 26.7	26.0/ 33.0/ 33.0	27.5/ 27.8/ 29.5	23.0/ 23.0/ 21.0	28.0/ 28.0/ 31.0	36.1/ 31.9/ 33.2	30.2/ 25.2/ 29.7
Psychology	28.0/ 30.0/ 34.0	37.0/ 57.0/ 59.0	28.0/ 35.0/ 31.0	28.7/ 34.9/ 34.9	34.0/ 40.0/ 43.0	25.0/ 27.0/ 32.0	29.0/ 28.0/ 31.0	37.0/ 37.0/ 44.0	29.0/ 35.0/ 27.0	37.0/ 39.0/ 34.0	29.0/ 34.0/ 29.0
Public Administration	25.0/ 22.0/ 27.0	35.4/ 39.4/ 43.4	29.3/ 30.3/ 34.3	32.3/ 35.4/ 37.4	27.3/ 25.2/ 26.3	23.0/ 24.0/ 23.0	26.3/ 27.3/ 28.3	21.0/ 27.0/ 29.0	27.3/ 30.3/ 29.3	30.0/ 29.0/ 30.0	27.0/ 24.0/ 30.0
Religion and Spirituality	17.0/ 27.0/ 28.0	33.0/ 55.0/ 59.0	37.0/ 41.0/ 43.0	22.6/ 31.4/ 31.4	35.0/ 38.0/ 34.0	34.0/ 35.0/ 39.0	28.0/ 36.0/ 30.0	27.0/ 30.0/ 28.0	27.0/ 27.0/ 31.0	18.0/ 25.0/ 25.0	29.0/ 39.0/ 37.0
Social Welfare and Development	29.0/ 34.0/ 37.0	31.2/ 50.0/ 48.2	38.0/ 35.0/ 37.0	30.0/ 29.0/ 31.1	28.0/ 34.0/ 36.0	18.0/ 20.0/ 18.0	32.0/ 35.0/ 44.0	34.0/ 29.0/ 29.0	26.0/ 29.0/ 29.0	28.0/ 37.0/ 28.0	31.0/ 30.0/ 24.0
Sociology	21.7/ 21.7/ 24.3	36.4/ 38.2/ 38.6	36.0/ 34.0/ 33.0	28.2/ 30.2/ 30.2	31.0/ 27.0/ 29.0	18.0/ 23.0/ 26.0	27.7/ 25.2/ 26.4	20.0/ 29.0/ 26.0	28.0/ 28.0/ 30.0	21.6/ 20.7/ 25.2	24.7/ 20.0/ 22.4
Sports and Recreation	27.7/ 26.7/ 29.7	40.0/ 53.2/ 52.2	31.0/ 25.0/ 36.0	30.4/ 32.2/ 37.8	28.1/ 28.6/ 32.0	30.0/ 27.0/ 26.0	27.7/ 32.2/ 38.4	26.0/ 32.0/ 28.0	28.0/ 34.0/ 33.0	32.0/ 35.3/ 30.8	26.7/ 29.0/ 29.0
Technology and Innovation	42.0/ 42.0/ 34.0	38.6/ 49.8/ 50.6	24.0/ 30.0/ 33.0	31.6/ 33.4/ 39.4	31.0/ 35.0/ 36.0	27.0/ 32.0/ 28.0	29.0/ 33.0/ 31.0	28.0/ 36.0/ 32.0	25.0/ 31.0/ 29.0	29.0/ 36.0/ 39.0	30.0/ 27.0/ 29.0
Transportation and Logistics	27.0/ 28.0/ 32.0	33.9/ 39.2/ 43.1	29.0/ 33.0/ 29.0	28.4/ 27.4/ 27.8	31.3/ 34.3/ 33.3	19.0/ 29.0/ 27.0	25.5/ 29.6/ 29.6	25.2/ 22.2/ 26.3	21.0/ 23.0/ 24.0	31.0/ 29.0/ 27.0	23.0/ 24.0/ 34.0

Table 37: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.2-3B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	27.0/ 22.0/ 32.0	36.4/ 33.8/ 32.6	32.3/ 29.3/ 27.3	31.4/ 27.4/ 25.6	32.0/ 21.0/ 32.0	25.2/ 29.3/ 33.3	35.0/ 30.0/ 30.0	27.0/ 25.0/ 33.0	31.0/ 25.0/ 27.0	31.0/ 29.0/ 33.0	32.0/ 33.0/ 35.0
Anthropology	31.0/ 30.0/ 38.0	32.0/ 43.0/ 30.0	33.0/ 33.0/ 33.0	35.4/ 37.4/ 36.4	29.0/ 37.0/ 40.0	24.0/ 27.0/ 30.0	31.0/ 24.0/ 31.0	31.0/ 32.0/ 26.0	34.0/ 33.0/ 37.0	34.0/ 31.0/ 40.0	26.0/ 28.0/ 39.0
Architecture and Design	31.0/ 30.0/ 40.0	40.0/ 33.0/ 32.0	30.0/ 33.0/ 27.0	33.0/ 29.0/ 35.0	34.0/ 30.0/ 42.0	36.0/ 34.0/ 31.0	39.0/ 33.0/ 31.0	31.0/ 25.0/ 26.0	32.0/ 30.0/ 33.0	29.0/ 39.0/ 35.0	33.0/ 33.0/ 31.0
Arts and Culture	27.9/ 29.1/ 29.1	39.2/ 36.6/ 28.8	27.0/ 22.0/ 27.0	29.7/ 30.3/ 27.5	27.0/ 23.3/ 22.7	32.0/ 32.0/ 36.0	25.1/ 32.2/ 24.0	24.0/ 20.0/ 23.0	27.0/ 33.0/ 28.0	30.3/ 25.4/ 27.9	25.6/ 29.1/ 27.6
Astronomy and Astrophysics	36.0/ 32.0/ 36.0	57.0/ 51.0/ 36.0	39.0/ 40.0/ 41.0	39.1/ 39.8/ 39.1	39.0/ 35.0/ 38.0	35.0/ 37.0/ 36.0	39.0/ 35.0/ 36.0	30.0/ 33.0/ 32.0	38.0/ 29.0/ 41.0	31.0/ 34.0/ 44.0	32.0/ 37.0/ 31.0
Biology	32.4/ 27.5/ 29.2	43.5/ 40.1/ 33.7	31.0/ 25.0/ 27.0	36.7/ 37.1/ 33.9	30.7/ 32.8/ 30.7	30.3/ 34.3/ 22.2	29.5/ 29.0/ 25.7	24.0/ 25.0/ 23.0	31.0/ 32.0/ 38.0	27.1/ 22.9/ 27.7	30.5/ 34.0/ 32.4
Business and Management	25.0/ 30.0/ 40.0	37.8/ 37.5/ 31.3	34.0/ 35.0/ 34.0	23.9/ 24.6/ 29.0	27.0/ 26.0/ 26.0	32.0/ 25.0/ 32.0	28.0/ 24.0/ 31.0	28.0/ 25.0/ 23.0	27.0/ 23.0/ 27.0	29.0/ 33.0/ 36.0	26.4/ 26.4/ 31.1
Chemistry	33.0/ 27.5/ 38.2	48.1/ 40.3/ 34.9	27.6/ 29.3/ 39.7	32.9/ 33.7/ 34.9	26.5/ 28.1/ 27.5	23.0/ 21.0/ 30.0	26.9/ 30.9/ 33.0	26.0/ 28.8/ 26.9	26.0/ 31.0/ 35.0	27.0/ 30.0/ 30.6	27.4/ 31.2/ 32.9
Computer Science	27.1/ 27.1/ 27.6	32.6/ 32.6/ 32.6	31.0/ 26.0/ 27.0	26.2/ 28.4/ 29.6	29.0/ 22.5/ 36.1	36.0/ 27.0/ 38.0	24.9/ 24.3/ 27.8	31.0/ 25.0/ 21.0	22.0/ 21.0/ 21.0	28.5/ 27.9/ 33.7	32.3/ 30.0/ 35.4
Defense and Security	29.0/ 28.0/ 30.0	45.0/ 40.0/ 33.0	36.0/ 29.0/ 26.0	28.4/ 24.6/ 27.5	29.0/ 36.0/ 40.0	25.0/ 35.0/ 30.0	33.0/ 27.0/ 23.0	26.0/ 27.0/ 22.0	28.0/ 35.0/ 30.0	28.0/ 29.0/ 36.0	33.0/ 41.0/ 32.0
Earth Sciences	35.0/ 38.0/ 46.0	43.2/ 40.7/ 35.9	34.0/ 33.0/ 35.0	32.1/ 29.3/ 28.9	28.0/ 38.0/ 46.0	29.0/ 28.0/ 30.0	33.0/ 33.0/ 24.0	23.2/ 22.2/ 23.2	29.0/ 34.0/ 30.0	29.0/ 31.0/ 32.0	33.0/ 33.0/ 35.0
Economics	33.8/ 36.7/ 33.1	38.1/ 37.5/ 29.9	22.0/ 30.0/ 28.0	29.2/ 32.1/ 28.4	33.0/ 27.0/ 41.0	32.0/ 26.0/ 35.0	33.1/ 27.5/ 28.2	28.0/ 31.0/ 26.0	31.0/ 29.0/ 34.0	32.5/ 30.9/ 28.6	25.1/ 32.3/ 28.1
Education	33.0/ 34.0/ 38.0	33.4/ 28.7/ 27.4	28.0/ 33.0/ 31.0	33.7/ 36.1/ 30.8	39.0/ 32.0/ 34.0	22.0/ 24.0/ 30.0	37.6/ 33.8/ 36.3	29.0/ 31.0/ 33.0	23.0/ 28.0/ 32.0	31.0/ 37.0/ 23.0	41.0/ 39.0/ 39.0
Energy and Power	27.0/ 25.0/ 35.0	43.9/ 37.2/ 29.7	27.0/ 32.0/ 36.0	31.9/ 32.9/ 29.1	33.0/ 30.0/ 30.0	24.0/ 25.0/ 29.0	31.0/ 37.0/ 31.0	27.0/ 26.0/ 31.0	35.0/ 39.0/ 40.0	30.0/ 33.0/ 30.0	27.0/ 32.0/ 25.0
Engineering	25.8/ 28.4/ 25.8	32.8/ 26.4/ 26.0	23.4/ 29.2/ 30.4	27.4/ 27.4/ 28.6	30.0/ 26.4/ 30.0	31.2/ 29.9/ 29.5	26.0/ 23.2/ 25.4	26.5/ 27.1/ 28.3	29.0/ 25.0/ 30.0	26.2/ 26.4/ 29.2	29.2/ 28.4/ 31.8
Environmental Science	32.0/ 25.0/ 36.0	43.9/ 40.5/ 31.5	30.0/ 31.0/ 38.0	33.1/ 32.1/ 31.7	30.5/ 25.8/ 37.5	29.0/ 30.0/ 34.0	28.4/ 29.5/ 32.6	20.2/ 20.2/ 24.2	35.0/ 34.0/ 39.0	33.0/ 32.0/ 33.0	34.1/ 30.7/ 34.6
Ethics and Human Rights	32.0/ 29.0/ 28.0	38.0/ 47.0/ 33.0	30.0/ 33.0/ 28.0	27.0/ 27.0/ 27.0	27.0/ 25.0/ 25.0	28.0/ 27.0/ 31.0	35.0/ 35.0/ 35.0	19.0/ 20.0/ 18.0	24.0/ 28.0/ 23.0	27.0/ 35.0/ 38.0	31.0/ 29.0/ 27.0
Finance and Investment	32.2/ 31.2/ 33.7	30.8/ 30.4/ 29.2	30.0/ 29.0/ 33.0	26.4/ 26.6/ 26.8	34.0/ 34.0/ 30.2	24.0/ 41.0/ 26.0	28.2/ 29.4/ 30.5	27.0/ 33.0/ 25.0	26.0/ 27.0/ 35.0	25.7/ 32.4/ 31.8	28.2/ 28.6/ 31.4
Food Science	28.0/ 23.0/ 35.0	38.0/ 37.0/ 24.0	31.0/ 28.0/ 37.0	27.3/ 33.3/ 26.3	31.0/ 21.0/ 29.0	31.0/ 30.0/ 26.0	35.0/ 36.0/ 37.0	26.0/ 27.0/ 31.0	25.0/ 30.0/ 28.0	21.0/ 28.0/ 28.0	30.0/ 35.0/ 33.0
Geography	28.5/ 29.0/ 30.2	40.6/ 36.2/ 32.0	31.0/ 38.0/ 35.0	30.3/ 31.1/ 31.1	31.5/ 26.0/ 33.3	32.3/ 29.3/ 31.3	32.5/ 28.4/ 28.4	27.0/ 24.0/ 27.0	26.0/ 28.0/ 23.0	30.3/ 24.6/ 31.4	28.1/ 28.1/ 32.8
Health and Medicine	25.0/ 21.0/ 33.0	54.1/ 49.3/ 38.7	27.0/ 35.0/ 29.0	37.3/ 37.1/ 34.7	37.0/ 27.0/ 43.0	27.0/ 20.0/ 24.0	28.6/ 21.4/ 25.0	26.0/ 30.0/ 25.0	29.0/ 36.0/ 36.0	26.7/ 29.3/ 39.7	31.8/ 29.9/ 33.6
History	33.1/ 32.0/ 36.3	37.5/ 35.3/ 29.9	42.0/ 36.0/ 39.0	31.9/ 34.7/ 33.1	34.8/ 28.9/ 36.3	28.0/ 35.0/ 29.0	29.7/ 29.7/ 32.0	27.0/ 28.0/ 26.0	29.0/ 32.0/ 35.0	34.2/ 34.5/ 34.5	28.1/ 29.9/ 34.0
Information Technology	38.0/ 30.0/ 41.0	34.0/ 45.0/ 35.0	45.0/ 42.0/ 44.0	41.3/ 43.0/ 38.6	36.0/ 30.0/ 39.0	39.0/ 34.0/ 37.0	42.0/ 38.0/ 39.0	30.0/ 31.0/ 31.0	36.0/ 42.0/ 41.0	35.0/ 26.0/ 30.0	41.0/ 44.0/ 43.0
International Relations	29.0/ 24.0/ 35.0	46.7/ 45.8/ 35.0	36.0/ 38.0/ 40.0	33.5/ 32.7/ 28.8	33.0/ 32.0/ 30.0	43.0/ 35.0/ 40.0	38.0/ 28.0/ 30.0	28.0/ 23.0/ 26.0	36.0/ 41.0/ 42.0	34.0/ 37.0/ 40.0	26.0/ 28.0/ 25.0
Language Studies	30.0/ 31.0/ 31.0	42.6/ 38.6/ 35.8	29.0/ 26.0/ 29.0	29.9/ 27.5/ 30.7	32.6/ 26.4/ 30.2	23.0/ 23.0/ 25.0	30.2/ 28.2/ 26.9	28.0/ 20.0/ 21.0	32.0/ 30.0/ 33.0	25.5/ 22.1/ 24.7	27.0/ 21.3/ 24.7
Law and Ethics	45.3/ 45.3/ 47.3	41.0/ 41.6/ 39.0	26.0/ 29.0/ 33.0	36.4/ 37.2/ 36.4	29.9/ 28.0/ 27.4	24.0/ 28.0/ 31.0	27.2/ 28.9/ 28.0	25.0/ 30.0/ 29.0	31.0/ 29.0/ 30.0	29.1/ 26.8/ 35.4	23.8/ 23.8/ 28.7
Literature and Linguistics	27.0/ 25.2/ 25.2	35.1/ 35.7/ 27.5	32.0/ 31.0/ 30.0	32.3/ 30.7/ 30.3	31.7/ 27.5/ 29.2	34.0/ 28.0/ 27.0	25.7/ 27.4/ 24.8	20.0/ 23.0/ 26.0	25.0/ 29.0/ 30.0	27.9/ 27.1/ 28.7	26.8/ 23.9/ 31.0
Logical Reasoning	25.8/ 29.5/ 32.9	24.0/ 26.6/ 27.8	23.4/ 24.0/ 29.2	29.3/ 32.1/ 28.9	23.2/ 27.8/ 27.4	26.0/ 27.0/ 32.0	24.2/ 24.2/ 27.0	17.0/ 17.0/ 21.0	30.0/ 30.0/ 32.0	27.1/ 23.9/ 25.9	26.7/ 24.8/ 27.5
Materials Science	31.0/ 25.0/ 32.0	39.5/ 36.1/ 33.8	27.0/ 33.0/ 34.0	30.1/ 29.0/ 30.3	27.0/ 34.0/ 31.0	32.0/ 31.0/ 40.0	33.0/ 36.0/ 36.0	26.0/ 29.0/ 24.0	35.0/ 29.0/ 30.0	29.0/ 32.0/ 33.0	25.0/ 23.0/ 25.0
Media and Communication	36.0/ 35.0/ 39.0	46.0/ 39.0/ 36.0	32.0/ 34.0/ 33.0	32.6/ 28.6/ 32.6	41.0/ 33.0/ 39.0	32.0/ 27.0/ 30.0	30.0/ 35.0/ 34.0	34.0/ 31.0/ 38.0	27.0/ 30.0/ 32.0	32.0/ 30.0/ 36.0	38.0/ 39.0/ 39.0
Music and Performing Arts	28.0/ 30.0/ 36.0	37.0/ 34.0/ 30.0	37.0/ 28.0/ 32.0	27.2/ 26.5/ 24.5	30.0/ 26.0/ 31.0	31.0/ 30.0/ 28.0	25.0/ 28.0/ 28.0	21.0/ 24.0/ 25.0	32.0/ 31.0/ 40.0	23.0/ 23.0/ 27.0	35.0/ 33.0/ 34.0
Physics	25.8/ 24.8/ 31.6	39.8/ 36.4/ 35.8	27.9/ 28.1/ 31.8	29.4/ 28.0/ 33.0	28.2/ 25.4/ 29.6	25.8/ 26.8/ 30.5	27.6/ 27.8/ 27.6	25.7/ 26.0/ 28.1	26.0/ 33.0/ 35.0	32.4/ 28.0/ 30.6	27.8/ 30.6/ 31.4
Politics and Governance	29.4/ 29.8/ 35.7	40.1/ 36.3/ 32.5	27.0/ 38.0/ 38.0	33.7/ 34.3/ 31.5	33.2/ 26.3/ 34.0	28.0/ 26.0/ 28.0	24.4/ 24.2/ 24.2	26.0/ 21.0/ 30.0	33.0/ 32.0/ 29.0	24.9/ 29.0/ 29.9	29.2/ 30.4/ 30.2
Psychology	40.0/ 27.0/ 44.0	43.0/ 47.0/ 30.0	33.0/ 30.0/ 30.0	38.8/ 34.9/ 34.1	34.0/ 30.0/ 42.0	32.0/ 28.0/ 27.0	29.0/ 27.0/ 25.0	25.0/ 28.0/ 29.0	27.0/ 21.0/ 27.0	36.0/ 36.0/ 40.0	32.0/ 30.0/ 32.0
Public Administration	24.0/ 28.0/ 20.0	33.3/ 30.3/ 29.3	22.2/ 28.3/ 31.3	28.3/ 28.3/ 27.3	28.3/ 25.2/ 27.3	27.0/ 30.0/ 30.0	28.3/ 26.3/ 25.2	21.0/ 22.0/ 25.0	31.3/ 30.3/ 32.3	23.0/ 28.0/ 32.0	27.0/ 24.0/ 29.0
Religion and Spirituality	26.0/ 25.0/ 25.0	46.0/ 43.0/ 31.0	37.0/ 42.0/ 41.0	30.2/ 31.1/ 25.8	31.0/ 23.0/ 25.0	32.0/ 33.0/ 26.0	39.0/ 33.0/ 25.0	24.0/ 27.0/ 26.0	31.0/ 34.0/ 37.0	32.0/ 31.0/ 28.0	30.0/ 27.0/ 29.0
Social Welfare and Development	22.0/ 28.0/ 32.0	36.6/ 35.7/ 33.9	38.0/ 39.0/ 40.0	28.5/ 32.1/ 32.6	23.0/ 33.0/ 33.0	29.0/ 30.0/ 32.0	35.0/ 31.0/ 34.0	25.0/ 24.0/ 20.0	28.0/ 36.0/ 31.0	37.0/ 35.0/ 32.0	32.0/ 34.0/ 37.0
Sociology	26.1/ 22.6/ 24.3	30.0/ 29.0/ 26.0	35.0/ 35.0/ 40.0	25.0/ 24.0/ 26.4	31.0/ 26.0/ 27.0	18.0/ 20.0/ 22.0	25.8/ 30.2/ 28.3	14.0/ 15.0/ 22.0	29.0/ 28.0/ 32.0	34.2/ 35.1/ 25.2	30.0/ 25.3/ 25.9
Sports and Recreation	24.8/ 23.8/ 25.7	44.0/ 37.2/ 29.8	27.0/ 25.0/ 23.0	33.8/ 31.0/ 29.2	26.4/ 27.5/ 24.2	21.0/ 31.0/ 27.0	32.2/ 29.4/ 23.7	21.0/ 26.0/ 25.0	25.0/ 28.0/ 27.0	33.3/ 29.5/ 25.6	26.7/ 27.6/ 25.2
Technology and Innovation	30.0/ 34.0/ 40.0	41.6/ 37.0/ 32.0	30.0/ 36.0/ 43.0	32.6/ 34.0/ 32.2	35.0/ 28.0/ 32.0	24.0/ 25.0/ 31.0	32.0/ 29.0/ 29.0	32.0/ 30.0/ 33.0	37.0/ 35.0/ 36.0	28.0/ 28.0/ 35.0	29.0/ 29.0/ 27.0
Transportation and Logistics	26.0/ 25.0/ 25.0	37.7/ 38.5/ 26.2	25.0/ 32.0/ 28.0	30.9/ 29.6/ 29.6	26.3/ 24.2/ 28.3	24.0/ 20.0/ 29.0	23.5/ 17.3/ 17.3	20.2/ 28.3/ 25.2	20.0/ 22.0/ 28.0	27.0/ 24.0/ 27.0	22.0/ 26.0/ 26.0

Table 38: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.2-3B-INSTRUCT on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	25	46.7	22.2	52.5	36	31.3	29	20	27	32	28
Anthropology	27	48	22	51.5	26	23	37	24	25	24	19
Architecture and Design	34	54	28	50	26	25	50	22	30	27	26
Arts and Culture	26.1	58.6	28	60.1	28.8	26	48.6	24	37	26.1	24.1
Astronomy and Astrophysics	31	72	26	58.6	24	25	56	23	27	30	29
Biology	33.1	66.3	22	65.9	30.3	28.3	36.4	32	28	27.7	31.4
Business and Management	33	43.3	25	47.8	29	28	40	31	21	37	28.3
Chemistry	33.8	57.7	23.3	57.3	24.8	28	40.1	31.7	30	26.7	26.9
Computer Science	36.5	39.8	21	41.4	30.8	28	36.1	26	27	37.2	40.4
Defense and Security	37	54	23	51.2	29	22	43	36	30	33	24
Earth Sciences	39	59.6	29	55.5	25	33	43	24.2	23	25	37
Economics	33.8	54.4	24	51.8	28	23	41.5	32	24	25.4	31.1
Education	34	46.2	20	38.8	31	29	40.8	31	24	28	28
Energy and Power	28	56.8	29	50.2	28	28	46	30	29	31	31
Engineering	25	35.6	26.6	36	25.2	25.2	34.6	25.9	21	28	25.8
Environmental Science	32	51.3	23	58.1	28.1	24	45.3	16.2	26	31	25.7
Ethics and Human Rights	34	43	23	54	25	26	46	19	25	30	28
Finance and Investment	26.7	39.6	27	45	30.9	37	34.5	24	32	31.8	29.3
Food Science	32	61	24	52.5	29	27	53	27	25	29	20
Geography	24	52	31	53.4	26.9	32.3	39.1	30	26	23.4	26.6
Health and Medicine	39	63.1	25	69.3	21	22	40.2	34	29	33.6	29
History	32.3	51	25	60.7	27.5	29	45.7	33	22	27.6	23.3
Information Technology	44	60	32	56.4	37	33	53	32	31	45	39
International Relations	30	57.5	30	54.5	22	31	54	25	31	25	33
Language Studies	32	59.8	28	40.3	23.3	29	29.5	21	29	29.2	27
Law and Ethics	43.2	56.8	27	60.8	29.9	23	36.2	27	22	23.6	26.2
Literature and Linguistics	29.7	48.3	15	49.5	34.2	22	39.8	23	23	27.5	23.9
Logical Reasoning	27	34.6	26	30.7	27	21	29.8	22	30	28.7	28
Materials Science	33	48.3	28	44.8	26	29	38	23	20	29	27
Media and Communication	42	59	21	53.7	43	30	47	37	25	37	30
Music and Performing Arts	28	47	25	51	26	24	43	33	28	20	29
Physics	32.2	55.2	29.5	45.8	26.2	24.7	36	24.6	28	25	29
Politics and Governance	25.9	50.9	28	54.4	20.6	28	37.9	26	24	28.2	25.2
Psychology	30	69	21	56.6	28	23	33	33	36	31	38
Public Administration	32	48.5	23.2	50.5	29.3	33	23.2	27	25.2	24	24
Religion and Spirituality	43	54	31	58.5	27	25	51	26	27	26	33
Social Welfare and Development	36	45.5	27	53.9	40	24	45	24	24	25	32
Sociology	30.4	42.2	26	40.8	22	28	29.6	35	29	17.1	30
Sports and Recreation	26.7	57	26	58.6	27	23	40.1	18	22	17.9	32.9
Technology and Innovation	35	51.2	29	51.8	21	32	53	27	27	22	27
Transportation and Logistics	32	40.8	28	40.4	22.2	29	29.6	29.3	16	26	26

Table 39: Detailed subject-wise evaluation for NVIDIA/NEMOTRON-4-MINI-HINDI-4B-BASE on MILU across different languages. The results reported are for 5-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	29.0/ 22.0/ 30.0	35.8/ 46.1/ 46.5	21.2/ 23.2/ 22.2	26.2/ 27.4/ 29.0	19.0/ 26.0/ 23.0	28.3/ 27.3/ 28.3	28.0/ 26.0/ 27.0	27.0/ 23.0/ 24.0	26.0/ 30.0/ 32.0	29.0/ 21.0/ 22.0	26.0/ 25.0/ 25.0
Anthropology	33.0/ 21.0/ 25.0	46.0/ 50.0/ 48.0	28.0/ 30.0/ 26.0	25.2/ 24.2/ 27.3	23.0/ 22.0/ 23.0	20.0/ 20.0/ 21.0	28.0/ 32.0/ 25.0	24.0/ 26.0/ 23.0	28.0/ 24.0/ 23.0	25.0/ 25.0/ 27.0	24.0/ 27.0/ 24.0
Architecture and Design	34.0/ 30.0/ 26.0	37.0/ 53.0/ 44.0	27.0/ 23.0/ 25.0	27.0/ 34.0/ 35.0	19.0/ 14.0/ 17.0	27.0/ 25.0/ 27.0	26.0/ 24.0/ 24.0	27.0/ 26.0/ 19.0	21.0/ 24.0/ 24.0	19.0/ 23.0/ 26.0	19.0/ 16.0/ 16.0
Arts and Culture	26.7/ 24.9/ 21.8	48.4/ 58.8/ 57.4	24.0/ 27.0/ 28.0	25.7/ 27.7/ 28.1	23.9/ 27.0/ 27.0	24.0/ 28.0/ 29.0	22.9/ 22.4/ 22.4	15.0/ 14.0/ 23.0	19.0/ 26.0/ 23.0	26.1/ 25.4/ 26.1	23.1/ 24.1/ 20.6
Astronomy and Astrophysics	16.0/ 28.0/ 21.0	53.0/ 69.0/ 66.0	24.0/ 26.0/ 23.0	25.0/ 28.1/ 30.5	30.0/ 23.0/ 27.0	22.0/ 25.0/ 19.0	32.0/ 28.0/ 29.0	17.0/ 17.0/ 17.0	32.0/ 24.0/ 26.0	27.0/ 26.0/ 28.0	21.0/ 28.0/ 25.0
Biology	23.6/ 23.9/ 22.9	49.5/ 62.7/ 62.5	24.0/ 25.0/ 22.0	27.5/ 26.7/ 24.4	27.9/ 23.8/ 23.8	29.3/ 29.3/ 31.3	26.3/ 28.7/ 29.5	24.0/ 24.0/ 32.0	31.0/ 22.0/ 27.0	27.4/ 28.0/ 25.5	18.9/ 20.1/ 22.6
Business and Management	25.0/ 23.0/ 29.0	37.1/ 48.6/ 45.5	28.0/ 25.0/ 22.0	28.3/ 26.7/ 30.4	27.0/ 19.0/ 18.0	21.0/ 21.0/ 24.0	20.0/ 24.0/ 27.0	24.0/ 27.0/ 25.0	26.0/ 31.0/ 29.0	30.0/ 28.0/ 29.0	21.7/ 20.8/ 17.0
Chemistry	28.8/ 31.0/ 31.3	44.9/ 51.1/ 51.1	26.7/ 22.4/ 23.3	28.3/ 27.3/ 29.1	30.1/ 30.7/ 30.4	24.0/ 26.0/ 25.0	29.8/ 30.9/ 29.5	22.1/ 24.0/ 26.9	24.0/ 21.0/ 20.0	25.5/ 26.4/ 29.1	28.9/ 28.7/ 29.7
Computer Science	30.2/ 35.9/ 33.9	37.0/ 41.2/ 40.4	36.0/ 37.0/ 33.0	28.2/ 32.0/ 28.6	31.4/ 40.8/ 40.2	31.0/ 28.0/ 29.0	33.1/ 32.5/ 31.4	31.0/ 29.0/ 25.0	30.0/ 21.0/ 24.0	36.0/ 36.0/ 36.6	33.2/ 32.3/ 31.8
Defense and Security	25.0/ 28.0/ 25.0	49.0/ 57.0/ 58.0	25.0/ 29.0/ 21.0	27.5/ 24.6/ 23.7	26.0/ 27.0/ 22.0	30.0/ 27.0/ 30.0	26.0/ 27.0/ 27.0	27.0/ 35.0/ 29.0	27.0/ 30.0/ 27.0	34.0/ 22.0/ 23.0	29.0/ 30.0/ 24.0
Earth Sciences	35.0/ 27.0/ 43.0	43.6/ 54.1/ 54.3	22.0/ 19.0/ 25.0	25.7/ 27.1/ 27.1	24.0/ 21.0/ 19.0	19.0/ 21.0/ 23.0	24.0/ 26.0/ 24.0	18.2/ 24.2/ 28.3	25.0/ 21.0/ 26.0	23.0/ 26.0/ 29.0	32.0/ 32.0/ 24.0
Economics	25.2/ 28.1/ 26.6	41.4/ 51.0/ 50.2	23.0/ 28.0/ 29.0	29.4/ 31.9/ 33.7	30.0/ 21.0/ 22.0	27.0/ 29.0/ 27.0	29.6/ 29.6/ 30.3	26.0/ 26.0/ 28.0	31.0/ 22.0/ 26.0	24.6/ 28.6/ 28.6	18.6/ 19.8/ 24.6
Education	22.0/ 20.0/ 20.0	35.7/ 44.3/ 41.7	25.0/ 25.0/ 24.0	26.9/ 26.3/ 26.3	35.0/ 24.0/ 27.0	27.0/ 25.0/ 24.0	33.8/ 28.7/ 37.6	29.0/ 30.0/ 28.0	22.0/ 25.0/ 24.0	29.0/ 30.0/ 24.0	34.0/ 20.0/ 27.0
Energy and Power	28.0/ 28.0/ 27.0	33.1/ 51.3/ 53.4	29.0/ 35.0/ 36.0	23.7/ 26.1/ 30.5	30.0/ 24.0/ 28.0	37.0/ 29.0/ 30.0	27.0/ 24.0/ 35.0	25.0/ 23.0/ 25.0	32.0/ 31.0/ 26.0	26.0/ 14.0/ 23.0	23.0/ 21.0/ 25.0
Engineering	26.8/ 28.4/ 27.8	35.2/ 35.8/ 35.8	24.2/ 25.0/ 22.2	25.2/ 24.8/ 28.4	26.0/ 29.4/ 29.2	26.5/ 26.5/ 28.6	25.0/ 27.0/ 28.8	26.2/ 25.0/ 25.0	25.0/ 25.0/ 27.0	26.8/ 29.4/ 27.6	25.8/ 26.0/ 25.8
Environmental Science	24.0/ 27.0/ 34.0	41.3/ 53.5/ 51.3	29.0/ 28.0/ 22.0	27.7/ 26.9/ 30.9	21.9/ 23.4/ 19.5	26.0/ 26.0/ 29.0	18.9/ 25.3/ 26.3	32.3/ 28.3/ 32.3	22.0/ 21.0/ 22.0	22.0/ 23.0/ 31.0	25.1/ 22.9/ 28.5
Ethics and Human Rights	18.0/ 23.0/ 27.0	35.0/ 47.0/ 45.0	23.0/ 27.0/ 26.0	29.0/ 26.0/ 23.0	29.0/ 27.0/ 28.0	21.0/ 21.0/ 21.0	29.0/ 28.0/ 35.0	21.0/ 21.0/ 16.0	23.0/ 27.0/ 27.0	30.0/ 32.0/ 30.0	20.0/ 22.0/ 22.0
Finance and Investment	20.8/ 25.2/ 28.2	36.0/ 42.0/ 38.0	26.0/ 24.0/ 22.0	28.6/ 27.0/ 28.8	24.7/ 25.9/ 21.0	30.0/ 24.0/ 31.0	24.3/ 24.9/ 25.4	22.0/ 23.0/ 21.0	32.0/ 27.0/ 23.0	24.0/ 27.4/ 29.0	25.4/ 23.2/ 21.1
Food Science	37.0/ 25.0/ 29.0	53.0/ 51.0/ 57.0	28.0/ 32.0/ 30.0	21.2/ 28.3/ 27.3	27.0/ 34.0/ 26.0	26.0/ 28.0/ 26.0	26.0/ 30.0/ 24.0	31.0/ 25.0/ 25.0	28.0/ 22.0/ 15.0	21.0/ 26.0/ 27.0	22.0/ 26.0/ 26.0
Geography	25.7/ 30.7/ 27.9	42.0/ 52.2/ 49.4	29.0/ 19.0/ 21.0	27.5/ 28.3/ 26.9	23.7/ 28.8/ 31.5	22.2/ 23.2/ 18.2	24.9/ 30.2/ 29.6	28.0/ 27.0/ 19.0	31.0/ 29.0/ 30.0	23.4/ 30.3/ 28.6	26.6/ 30.2/ 27.1
Health and Medicine	34.0/ 28.0/ 33.0	51.3/ 64.1/ 60.7	25.0/ 29.0/ 29.0	29.5/ 26.7/ 28.7	26.0/ 26.0/ 29.0	21.0/ 25.0/ 15.0	26.8/ 31.2/ 30.4	30.0/ 27.0/ 18.0	26.0/ 28.0/ 24.0	21.6/ 20.7/ 17.2	19.6/ 25.2/ 29.0
History	24.3/ 24.8/ 26.7	37.4/ 45.0/ 42.6	26.0/ 20.0/ 23.0	29.0/ 32.5/ 33.1	27.5/ 24.2/ 22.7	27.0/ 26.0/ 31.0	27.5/ 23.1/ 26.4	31.0/ 31.0/ 29.0	20.0/ 26.0/ 21.0	24.0/ 27.3/ 29.4	28.1/ 24.0/ 26.1
Information Technology	27.0/ 36.0/ 35.0	44.0/ 57.0/ 55.0	29.0/ 30.0/ 27.0	27.9/ 25.7/ 27.9	26.0/ 34.0/ 33.0	26.0/ 29.0/ 30.0	31.0/ 36.0/ 30.0	30.0/ 32.0/ 34.0	33.0/ 37.0/ 36.0	19.0/ 26.0/ 26.0	27.0/ 34.0/ 31.0
International Relations	30.0/ 33.0/ 34.0	58.3/ 61.7/ 61.7	29.0/ 27.0/ 20.0	26.9/ 29.6/ 28.8	28.0/ 22.0/ 23.0	24.0/ 28.0/ 24.0	27.0/ 26.0/ 32.0	27.0/ 26.0/ 28.0	30.0/ 27.0/ 25.0	28.0/ 25.0/ 27.0	18.0/ 16.0/ 14.0
Language Studies	28.0/ 29.0/ 26.0	42.4/ 52.0/ 52.8	16.0/ 28.0/ 28.0	21.6/ 25.2/ 25.2	31.0/ 23.3/ 20.9	26.0/ 21.0/ 23.0	22.8/ 25.5/ 30.9	21.0/ 26.0/ 18.0	39.0/ 33.0/ 38.0	24.3/ 24.3/ 25.1	27.6/ 29.3/ 28.7
Law and Ethics	29.7/ 27.0/ 25.0	45.2/ 55.4/ 53.6	21.0/ 23.0/ 22.0	33.8/ 31.4/ 34.0	22.3/ 19.8/ 21.7	24.0/ 30.0/ 30.0	26.3/ 24.1/ 30.2	22.0/ 25.0/ 25.0	25.0/ 25.0/ 21.0	23.6/ 27.6/ 26.8	26.2/ 18.1/ 20.0
Literature and Linguistics	27.0/ 27.9/ 25.2	40.9/ 49.5/ 45.7	16.0/ 25.0/ 20.0	27.1/ 25.7/ 26.5	26.7/ 25.8/ 25.0	28.0/ 27.0/ 25.0	22.1/ 22.1/ 23.9	33.0/ 38.0/ 35.0	26.0/ 24.0/ 21.0	24.7/ 27.5/ 27.5	23.9/ 27.5/ 21.8
Logical Reasoning	22.6/ 24.6/ 22.9	33.0/ 32.6/ 32.8	19.5/ 27.9/ 26.6	24.1/ 25.7/ 25.9	24.5/ 25.3/ 29.9	32.0/ 21.0/ 25.0	23.7/ 27.6/ 24.2	25.0/ 27.0/ 23.0	15.0/ 28.0/ 22.0	24.7/ 27.5/ 25.9	30.7/ 29.5/ 25.5
Materials Science	24.0/ 17.0/ 22.0	39.9/ 47.9/ 43.7	26.0/ 22.0/ 18.0	28.0/ 24.4/ 26.9	37.0/ 33.0/ 31.0	41.0/ 27.0/ 31.0	31.0/ 29.0/ 29.0	25.0/ 21.0/ 19.0	23.0/ 20.0/ 18.0	29.0/ 22.0/ 27.0	29.0/ 29.0/ 26.0
Media and Communication	24.0/ 23.0/ 24.0	37.0/ 56.0/ 55.0	26.0/ 23.0/ 27.0	26.5/ 29.9/ 33.3	27.0/ 24.0/ 23.0	23.0/ 25.0/ 23.0	24.0/ 28.0/ 24.0	29.0/ 28.0/ 23.0	24.0/ 20.0/ 22.0	26.0/ 29.0/ 29.0	30.0/ 31.0/ 34.0
Music and Performing Arts	28.0/ 13.0/ 22.0	43.0/ 50.0/ 51.0	28.0/ 20.0/ 25.0	27.2/ 29.1/ 28.5	21.0/ 23.0/ 24.0	22.0/ 22.0/ 18.0	20.0/ 25.0/ 21.0	31.0/ 28.0/ 33.0	32.0/ 27.0/ 22.0	23.0/ 27.0/ 23.0	20.0/ 24.0/ 26.0
Physics	21.2/ 25.6/ 24.6	36.2/ 47.0/ 44.4	22.0/ 28.1/ 27.3	25.6/ 25.2/ 26.8	22.8/ 27.0/ 24.4	22.6/ 24.7/ 23.2	23.6/ 26.0/ 24.0	20.1/ 25.0/ 27.4	22.0/ 33.0/ 33.0	20.6/ 23.4/ 22.2	25.6/ 27.2/ 24.2
Politics and Governance	25.9/ 27.5/ 27.1	39.1/ 48.9/ 44.3	26.0/ 26.0/ 21.0	25.5/ 27.5/ 26.1	26.7/ 19.0/ 23.1	19.0/ 23.0/ 23.0	21.6/ 20.8/ 23.9	24.0/ 19.0/ 20.0	27.0/ 31.0/ 30.0	31.1/ 22.8/ 26.1	25.2/ 25.5/ 23.1
Psychology	27.0/ 26.0/ 24.0	47.0/ 52.0/ 56.0	27.0/ 25.0/ 26.0	28.7/ 23.3/ 21.7	26.0/ 30.0/ 28.0	26.0/ 23.0/ 28.0	35.0/ 23.0/ 26.0	33.0/ 28.0/ 24.0	29.0/ 30.0/ 34.0	25.0/ 21.0/ 21.0	28.0/ 26.0/ 25.0
Public Administration	31.0/ 25.0/ 21.0	38.4/ 48.5/ 48.5	25.2/ 26.3/ 23.2	33.3/ 33.3/ 28.3	37.4/ 30.3/ 26.3	29.0/ 27.0/ 29.0	28.3/ 19.2/ 24.2	29.0/ 25.0/ 31.0	23.2/ 26.3/ 29.3	28.0/ 25.0/ 29.0	32.0/ 28.0/ 31.0
Religion and Spirituality	15.0/ 24.0/ 29.0	44.0/ 63.0/ 57.0	34.0/ 30.0/ 32.0	23.8/ 28.6/ 25.0	17.0/ 33.0/ 23.0	28.0/ 27.0/ 26.0	16.0/ 22.0/ 26.0	28.0/ 27.0/ 29.0	28.0/ 30.0/ 25.0	22.0/ 24.0/ 26.0	18.0/ 24.0/ 31.0
Social Welfare and Development	33.0/ 29.0/ 28.0	38.4/ 47.3/ 42.9	26.0/ 20.0/ 26.0	25.9/ 23.3/ 29.0	22.0/ 32.0/ 28.0	23.0/ 25.0/ 26.0	34.0/ 30.0/ 33.0	25.0/ 26.0/ 24.0	27.0/ 34.0/ 29.0	25.0/ 33.0/ 28.0	32.0/ 30.0/ 32.0
Sociology	21.7/ 21.7/ 20.0	33.6/ 39.6/ 37.2	29.0/ 20.0/ 23.0	25.2/ 28.8/ 29.0	16.0/ 22.0/ 22.0	26.0/ 29.0/ 26.0	30.2/ 25.8/ 25.8	19.0/ 23.0/ 28.0	19.0/ 24.0/ 30.0	22.5/ 28.8/ 26.1	24.7/ 27.7/ 30.0
Sports and Recreation	23.8/ 29.2/ 27.2	43.0/ 58.8/ 56.8	20.0/ 20.0/ 17.0	26.0/ 28.4/ 26.6	25.3/ 20.8/ 24.7	23.0/ 29.0/ 29.0	23.7/ 29.9/ 26.0	18.0/ 25.0/ 21.0	21.0/ 26.0/ 25.0	25.6/ 26.9/ 28.2	23.8/ 22.9/ 22.4
Technology and Innovation	32.0/ 26.0/ 21.0	43.2/ 52.2/ 52.0	27.0/ 26.0/ 26.0	29.4/ 25.4/ 29.6	30.0/ 27.0/ 31.0	25.0/ 29.0/ 32.0	30.0/ 27.0/ 28.0	24.0/ 23.0/ 34.0	25.0/ 24.0/ 27.0	32.0/ 28.0/ 26.0	19.0/ 19.0/ 17.0
Transportation and Logistics	22.0/ 30.0/ 20.0	32.3/ 38.5/ 34.6	24.0/ 17.0/ 25.0	25.9/ 26.5/ 25.2	24.2/ 21.2/ 24.2	30.0/ 29.0/ 27.0	23.5/ 21.4/ 20.4	25.2/ 20.2/ 29.3	23.0/ 29.0/ 27.0	23.0/ 27.0/ 31.0	24.0/ 25.0/ 29.0

Table 40: Detailed subject-wise evaluation for META-LLAMA/LLAMA-2-7B-HF on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	21.0/ 32.0/ 23.0	28.0/ 28.0/ 27.4	29.3/ 31.3/ 26.3	28.2/ 27.6/ 26.8	24.0/ 27.0/ 21.0	28.3/ 32.3/ 36.4	31.0/ 35.0/ 27.0	24.0/ 26.0/ 29.0	20.0/ 29.0/ 25.0	28.0/ 29.0/ 25.0	24.0/ 25.0/ 23.0
Anthropology	29.0/ 29.0/ 29.0	28.0/ 30.0/ 27.0	25.0/ 30.0/ 17.0	34.3/ 24.2/ 25.2	31.0/ 27.0/ 26.0	26.0/ 22.0/ 24.0	23.0/ 24.0/ 28.0	20.0/ 27.0/ 31.0	23.0/ 21.0/ 29.0	31.0/ 26.0/ 33.0	30.0/ 18.0/ 20.0
Architecture and Design	33.0/ 31.0/ 26.0	38.0/ 24.0/ 25.0	27.0/ 28.0/ 18.0	31.0/ 25.0/ 24.0	22.0/ 24.0/ 22.0	31.0/ 33.0/ 28.0	29.0/ 23.0/ 19.0	24.0/ 27.0/ 17.0	24.0/ 18.0/ 26.0	20.0/ 19.0/ 22.0	22.0/ 20.0/ 19.0
Arts and Culture	29.1/ 24.9/ 24.9	31.8/ 30.6/ 30.0	20.0/ 24.0/ 24.0	22.9/ 21.6/ 23.1	25.8/ 25.1/ 19.0	13.0/ 21.0/ 22.0	25.7/ 23.5/ 20.2	26.0/ 22.0/ 22.0	24.0/ 24.0/ 24.0	26.7/ 24.9/ 29.1	21.1/ 23.1/ 29.1
Astronomy and Astrophysics	29.0/ 24.0/ 25.0	37.0/ 30.0/ 35.0	25.0/ 24.0/ 30.0	21.1/ 28.1/ 26.6	22.0/ 25.0/ 27.0	27.0/ 24.0/ 32.0	26.0/ 34.0/ 29.0	20.0/ 21.0/ 24.0	21.0/ 30.0/ 27.0	28.0/ 29.0/ 34.0	28.0/ 31.0/ 25.0
Biology	21.1/ 24.3/ 20.1	33.5/ 26.7/ 29.5	26.0/ 26.0/ 24.0	23.1/ 25.4/ 22.9	27.9/ 27.9/ 28.3	28.3/ 24.2/ 26.3	25.4/ 23.3/ 26.3	26.0/ 30.0/ 21.0	25.0/ 19.0/ 26.0	29.3/ 28.3/ 29.6	20.4/ 23.3/ 24.8
Business and Management	22.0/ 28.0/ 19.0	28.5/ 27.6/ 27.9	26.0/ 24.0/ 23.0	23.6/ 25.8/ 23.6	25.0/ 23.0/ 25.0	24.0/ 21.0/ 28.0	24.0/ 27.0/ 27.0	33.0/ 30.0/ 27.0	29.0/ 26.0/ 22.0	34.0/ 30.0/ 26.0	25.5/ 15.1/ 23.6
Chemistry	26.9/ 28.8/ 26.7	36.9/ 30.5/ 29.1	23.3/ 22.4/ 24.1	25.7/ 26.1/ 23.2	23.2/ 23.2/ 26.5	21.0/ 19.0/ 16.0	23.5/ 24.3/ 23.0	20.2/ 25.0/ 24.0	20.0/ 20.0/ 17.0	22.9/ 21.1/ 21.4	26.9/ 27.7/ 25.9
Computer Science	26.0/ 25.0/ 29.2	27.2/ 28.2/ 24.8	28.0/ 26.0/ 31.0	27.6/ 25.4/ 28.2	26.0/ 29.6/ 29.6	29.0/ 23.0/ 23.0	29.0/ 25.4/ 34.3	24.0/ 32.0/ 26.0	24.0/ 25.0/ 25.0	27.3/ 29.1/ 31.4	26.0/ 28.2/ 30.0
Defense and Security	24.0/ 24.0/ 19.0	34.0/ 30.0/ 25.0	33.0/ 34.0/ 25.0	27.5/ 28.9/ 29.9	29.0/ 25.0/ 29.0	28.0/ 26.0/ 22.0	26.0/ 27.0/ 32.0	17.0/ 20.0/ 19.0	37.0/ 26.0/ 34.0	24.0/ 28.0/ 20.0	28.0/ 29.0/ 23.0
Earth Sciences	37.0/ 31.0/ 33.0	29.8/ 29.1/ 25.9	25.0/ 26.0/ 23.0	29.5/ 28.1/ 25.9	33.0/ 24.0/ 22.0	25.0/ 22.0/ 22.0	32.0/ 34.0/ 30.0	30.3/ 28.3/ 25.2	27.0/ 26.0/ 31.0	23.0/ 25.0/ 29.0	24.0/ 24.0/ 16.0
Economics	18.0/ 23.7/ 28.1	32.5/ 29.5/ 30.5	24.0/ 21.0/ 19.0	24.2/ 24.6/ 24.0	30.0/ 26.0/ 27.0	27.0/ 23.0/ 22.0	29.6/ 23.9/ 24.6	25.0/ 18.0/ 33.0	19.0/ 20.0/ 18.0	20.6/ 19.1/ 22.2	25.1/ 27.5/ 26.4
Education	27.0/ 20.0/ 24.0	26.8/ 29.3/ 29.9	25.0/ 21.0/ 28.0	28.7/ 28.4/ 25.7	39.0/ 33.0/ 27.0	25.0/ 20.0/ 26.0	22.9/ 26.1/ 23.6	27.0/ 26.0/ 23.0	24.0/ 24.0/ 30.0	25.0/ 22.0/ 26.0	29.0/ 33.0/ 26.0
Energy and Power	33.0/ 26.0/ 29.0	20.9/ 30.4/ 20.3	27.0/ 28.0/ 33.0	24.1/ 28.8/ 24.8	24.0/ 20.0/ 18.0	29.0/ 24.0/ 23.0	28.0/ 19.0/ 21.0	22.0/ 23.0/ 24.0	24.0/ 23.0/ 25.0	27.0/ 24.0/ 16.0	26.0/ 28.0/ 25.0
Engineering	26.4/ 26.0/ 25.8	28.2/ 27.6/ 28.2	22.4/ 25.0/ 21.4	23.2/ 25.2/ 24.6	27.2/ 26.8/ 26.6	29.1/ 26.5/ 27.8	27.8/ 26.4/ 29.0	27.1/ 22.0/ 22.3	26.0/ 27.0/ 23.0	26.2/ 23.8/ 28.8	26.4/ 26.4/ 28.2
Environmental Science	18.0/ 22.0/ 22.0	30.1/ 29.5/ 28.5	29.0/ 27.0/ 24.0	25.4/ 23.6/ 24.9	25.0/ 24.2/ 28.1	24.0/ 23.0/ 21.0	19.5/ 25.8/ 24.2	21.2/ 18.2/ 25.2	31.0/ 27.0/ 35.0	28.0/ 28.0/ 33.0	24.0/ 25.1/ 26.8
Ethics and Human Rights	26.0/ 28.0/ 32.0	34.0/ 31.0/ 31.0	15.0/ 24.0/ 15.0	21.0/ 27.0/ 29.0	23.0/ 27.0/ 19.0	20.0/ 18.0/ 25.0	33.0/ 28.0/ 29.0	27.0/ 27.0/ 24.0	18.0/ 26.0/ 24.0	20.0/ 18.0/ 27.0	19.0/ 16.0/ 9.0
Finance and Investment	24.8/ 23.8/ 26.7	31.2/ 30.0/ 30.2	28.0/ 23.0/ 30.0	23.6/ 26.0/ 24.4	29.0/ 26.5/ 23.5	27.0/ 32.0/ 26.0	26.6/ 24.3/ 22.0	23.0/ 28.0/ 33.0	34.0/ 27.0/ 27.0	27.9/ 32.4/ 28.5	25.0/ 26.8/ 26.1
Food Science	32.0/ 29.0/ 31.0	31.0/ 32.0/ 34.0	23.0/ 24.0/ 23.0	27.3/ 29.3/ 26.3	22.0/ 19.0/ 17.0	28.0/ 22.0/ 21.0	22.0/ 29.0/ 28.0	24.0/ 20.0/ 21.0	22.0/ 22.0/ 24.0	24.0/ 18.0/ 19.0	20.0/ 30.0/ 21.0
Geography	29.0/ 25.7/ 24.6	29.0/ 29.6/ 26.0	35.0/ 33.0/ 37.0	24.5/ 28.3/ 23.3	18.7/ 22.4/ 24.2	16.2/ 16.2/ 27.3	22.5/ 24.9/ 23.1	24.0/ 21.0/ 23.0	30.0/ 21.0/ 18.0	28.0/ 24.6/ 27.4	24.5/ 24.5/ 25.0
Health and Medicine	24.0/ 28.0/ 22.0	35.1/ 29.9/ 30.5	22.0/ 19.0/ 21.0	23.6/ 24.6/ 24.1	21.0/ 24.0/ 28.0	31.0/ 30.0/ 24.0	22.3/ 24.1/ 17.9	26.0/ 19.0/ 20.0	23.0/ 27.0/ 21.0	20.7/ 19.8/ 23.3	27.1/ 29.9/ 30.8
History	30.7/ 30.1/ 24.5	31.5/ 30.9/ 33.1	33.0/ 27.0/ 18.0	27.0/ 24.8/ 26.8	24.5/ 20.9/ 27.5	20.0/ 18.0/ 22.0	26.0/ 20.8/ 23.4	31.0/ 32.0/ 32.0	25.0/ 22.0/ 24.0	25.8/ 28.0/ 21.1	26.1/ 25.6/ 23.8
Information Technology	30.0/ 25.0/ 33.0	29.0/ 32.0/ 34.0	31.0/ 27.0/ 26.0	26.3/ 28.5/ 31.8	30.0/ 28.0/ 29.0	20.0/ 23.0/ 20.0	17.0/ 19.0/ 23.0	26.0/ 28.0/ 24.0	27.0/ 24.0/ 30.0	18.0/ 23.0/ 19.0	29.0/ 28.0/ 25.0
International Relations	22.0/ 25.0/ 25.0	38.3/ 30.8/ 38.3	32.0/ 27.0/ 30.0	26.1/ 23.4/ 23.0	18.0/ 24.0/ 20.0	22.0/ 23.0/ 22.0	20.0/ 23.0/ 25.0	22.0/ 30.0/ 28.0	33.0/ 30.0/ 31.0	24.0/ 22.0/ 23.0	20.0/ 22.0/ 18.0
Language Studies	28.0/ 33.0/ 33.0	33.0/ 32.8/ 33.4	21.0/ 24.0/ 21.0	23.4/ 26.9/ 24.1	27.1/ 31.8/ 32.6	25.0/ 21.0/ 32.0	26.9/ 24.8/ 24.8	26.0/ 26.0/ 26.0	30.0/ 36.0/ 32.0	23.6/ 26.6/ 25.5	28.7/ 29.9/ 31.0
Law and Ethics	24.3/ 29.7/ 20.9	37.0/ 36.2/ 33.4	24.0/ 25.0/ 28.0	29.6/ 27.6/ 27.0	24.8/ 23.6/ 22.3	25.0/ 29.0/ 26.0	28.4/ 23.7/ 29.3	17.0/ 19.0/ 21.0	21.0/ 20.0/ 20.0	30.7/ 29.1/ 31.5	26.9/ 23.8/ 28.7
Literature and Linguistics	26.1/ 27.0/ 24.3	28.7/ 25.7/ 26.5	26.0/ 23.0/ 29.0	27.1/ 29.3/ 28.1	22.5/ 27.5/ 22.5	26.0/ 27.0/ 28.0	20.3/ 23.0/ 27.4	23.0/ 30.0/ 29.0	28.0/ 24.0/ 21.0	28.3/ 23.5/ 22.3	26.1/ 23.9/ 31.0
Logical Reasoning	25.3/ 24.1/ 23.3	26.6/ 24.8/ 26.6	26.0/ 29.9/ 31.8	28.3/ 27.9/ 26.7	29.5/ 27.4/ 29.5	30.0/ 32.0/ 31.0	25.1/ 25.9/ 26.7	30.0/ 31.0/ 28.0	25.0/ 23.0/ 23.0	27.5/ 27.5/ 25.5	29.5/ 29.0/ 29.5
Materials Science	29.0/ 18.0/ 16.0	29.3/ 28.9/ 28.5	26.0/ 24.0/ 27.0	24.2/ 26.9/ 25.3	33.0/ 25.0/ 29.0	31.0/ 26.0/ 38.0	29.0/ 22.0/ 22.0	25.0/ 31.0/ 35.0	28.0/ 25.0/ 24.0	22.0/ 19.0/ 18.0	27.0/ 27.0/ 28.0
Media and Communication	27.0/ 28.0/ 22.0	28.0/ 28.0/ 31.0	31.0/ 23.0/ 28.0	31.3/ 29.9/ 28.6	24.0/ 24.0/ 20.0	31.0/ 29.0/ 24.0	27.0/ 25.0/ 27.0	29.0/ 30.0/ 37.0	27.0/ 25.0/ 19.0	17.0/ 18.0/ 14.0	29.0/ 38.0/ 26.0
Music and Performing Arts	27.0/ 19.0/ 22.0	36.0/ 36.0/ 36.0	29.0/ 21.0/ 25.0	21.2/ 21.9/ 27.2	20.0/ 27.0/ 22.0	25.0/ 24.0/ 30.0	23.0/ 24.0/ 28.0	33.0/ 30.0/ 27.0	25.0/ 31.0/ 29.0	27.0/ 28.0/ 27.0	25.0/ 19.0/ 28.0
Physics	23.6/ 25.8/ 23.4	30.4/ 29.4/ 28.0	24.2/ 23.1/ 21.2	23.0/ 23.8/ 25.6	26.6/ 26.4/ 24.0	26.8/ 24.2/ 25.8	25.2/ 27.4/ 24.8	29.2/ 26.0/ 29.2	27.0/ 33.0/ 23.0	23.4/ 24.0/ 20.2	26.4/ 25.6/ 24.0
Politics and Governance	27.5/ 28.6/ 27.8	34.1/ 35.7/ 31.7	24.0/ 24.0/ 21.0	24.1/ 25.3/ 25.7	24.7/ 23.1/ 25.1	20.0/ 22.0/ 28.0	23.0/ 26.1/ 24.4	30.0/ 31.0/ 33.0	23.0/ 28.0/ 28.0	28.2/ 28.6/ 29.0	25.9/ 25.5/ 22.4
Psychology	28.0/ 34.0/ 28.0	35.0/ 27.0/ 30.0	23.0/ 31.0/ 25.0	26.4/ 20.2/ 18.6	29.0/ 22.0/ 27.0	30.0/ 23.0/ 26.0	25.0/ 21.0/ 23.0	27.0/ 29.0/ 25.0	25.0/ 22.0/ 29.0	32.0/ 23.0/ 30.0	26.0/ 21.0/ 22.0
Public Administration	25.0/ 24.0/ 24.0	31.3/ 29.3/ 31.3	17.2/ 21.2/ 23.2	31.3/ 29.3/ 30.3	29.3/ 32.3/ 28.3	24.0/ 25.0/ 30.0	25.2/ 20.2/ 19.2	27.0/ 28.0/ 30.0	29.3/ 28.3/ 30.3	24.0/ 20.0/ 23.0	26.0/ 26.0/ 31.0
Religion and Spirituality	16.0/ 13.0/ 20.0	35.0/ 32.0/ 28.0	31.0/ 24.0/ 24.0	21.0/ 23.8/ 25.8	20.0/ 22.0/ 20.0	24.0/ 18.0/ 16.0	19.0/ 26.0/ 27.0	33.0/ 27.0/ 29.0	29.0/ 29.0/ 24.0	20.0/ 24.0/ 25.0	21.0/ 20.0/ 27.0
Social Welfare and Development	16.0/ 22.0/ 16.0	33.0/ 32.1/ 33.9	27.0/ 29.0/ 23.0	25.4/ 22.3/ 24.9	25.0/ 26.0/ 25.0	24.0/ 26.0/ 26.0	23.0/ 28.0/ 27.0	22.0/ 22.0/ 23.0	23.0/ 26.0/ 27.0	26.0/ 23.0/ 28.0	33.0/ 30.0/ 27.0
Sociology	20.0/ 27.0/ 23.5	31.0/ 26.0/ 28.8	26.0/ 27.0/ 28.0	25.4/ 20.8/ 22.4	23.0/ 28.0/ 25.0	22.0/ 23.0/ 22.0	22.0/ 18.9/ 20.8	24.0/ 22.0/ 19.0	28.0/ 21.0/ 25.0	23.4/ 24.3/ 32.4	28.8/ 32.9/ 23.5
Sports and Recreation	23.8/ 28.7/ 28.2	33.2/ 27.2/ 27.0	28.0/ 27.0/ 25.0	25.2/ 23.4/ 25.4	23.0/ 20.2/ 22.5	38.0/ 25.0/ 26.0	26.0/ 27.7/ 26.6	17.0/ 16.0/ 17.0	21.0/ 28.0/ 26.0	24.4/ 22.4/ 19.2	18.6/ 22.4/ 24.8
Technology and Innovation	34.0/ 31.0/ 25.0	30.0/ 30.2/ 31.6	18.0/ 19.0/ 23.0	26.2/ 27.6/ 24.4	22.0/ 21.0/ 21.0	23.0/ 29.0/ 23.0	24.0/ 30.0/ 27.0	24.0/ 24.0/ 26.0	26.0/ 25.0/ 26.0	24.0/ 36.0/ 24.0	27.0/ 25.0/ 30.0
Transportation and Logistics	20.0/ 24.0/ 26.0	30.0/ 26.9/ 26.9	28.0/ 23.0/ 24.0	23.0/ 27.8/ 24.6	25.2/ 23.2/ 27.3	23.0/ 19.0/ 18.0	30.6/ 26.5/ 26.5	25.2/ 24.2/ 23.2	29.0/ 28.0/ 24.0	29.0/ 25.0/ 22.0	23.0/ 20.0/ 22.0

Table 41: Detailed subject-wise evaluation for META-LLAMA/LLAMA-2-7B-CHAT-HF on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	32	48.5	37.4	37.6	35	33.3	31	35	39	32	32
Anthropology	43	50	35	36.4	26	32	31	27	39	30	27
Architecture and Design	48	52	36	34	30	39	32	25	31	27	35
Arts and Culture	36.4	55.2	33	37.1	33.7	25	34.4	22	36	35.8	30.6
Astronomy and Astrophysics	56	82	46	46.1	36	38	44	39	38	42	42
Biology	46.5	71.3	44	55.1	36.5	36.4	35.5	26	39	28	42.1
Business and Management	33	52.6	40	43.1	37	38	35	31	37	39	20.8
Chemistry	46.4	69.3	45.7	53.7	33	34	40.4	35.6	39	30.9	40.9
Computer Science	39.1	45.6	44	37.6	42	33	42.6	35	38	29.6	34.5
Defense and Security	46	65	32	37.4	36	31	29	29	39	25	37
Earth Sciences	45	67.7	33	41.1	30	38	39	25.2	30	32	41
Economics	49.6	58.8	33	40.9	36	32	39.4	28	32	29.4	39.5
Education	43	45.5	36	39.1	33	36	43.3	40	36	29	36
Energy and Power	48	56.1	42	43.4	28	34	41	34	35	26	40
Engineering	41.2	49.6	35.8	37.4	34.4	30.8	38.4	32.4	33	32.8	40.6
Environmental Science	43	61.7	40	43.5	35.9	34	35.8	38.4	42	34	37.4
Ethics and Human Rights	48	59	38	49	37	38	42	31	39	43	45
Finance and Investment	44.5	50.2	39	41.6	40.7	30	37.9	32	39	41.9	36.8
Food Science	48	76	46	53.5	36	41	49	35	48	39	53
Geography	36.9	50.4	40	41.2	32.4	23.2	34.3	26	30	30.3	28.6
Health and Medicine	49	71.7	39	50.5	41	41	42.9	35	47	35.3	44.9
History	38.7	46.2	40	41.9	37	35	32.3	35	34	26.9	32.7
Information Technology	71	64	53	64.2	45	60	58	47	44	45	54
International Relations	45	60.8	42	47.9	36	47	44	47	43	48	31
Language Studies	33	71.8	38	36.1	33.3	34	30.9	38	37	30.7	35.1
Law and Ethics	46.6	53.4	34	43.2	24.8	24	31	37	38	33.1	30.6
Literature and Linguistics	34.2	51.7	27	37.7	30.8	32	30.1	37	36	30	19
Logical Reasoning	38.3	40.8	22.1	35.1	26.1	27	27	25	34	30.7	31.9
Materials Science	40	58.6	26	43	34	36	45	28	31	36	46
Media and Communication	55	73	44	49	36	41	35	42	41	43	50
Music and Performing Arts	31	47	28	32.5	32	29	36	29	32	20	29
Physics	43.6	63.2	36.2	45.4	35.2	37.4	40	33.7	39	34.4	42.6
Politics and Governance	39.6	52.3	30	37	35.6	22	32	28	35	28.6	31.8
Psychology	45	71	42	46.5	35	40	42	32	34	38	51
Public Administration	31	54.5	32.3	36.4	27.3	31	30.3	26	40.4	40	27
Religion and Spirituality	39	53	45	41.1	33	29	43	25	33	25	36
Social Welfare and Development	42	53.6	34	40.9	35	33	46	31	32	29	28
Sociology	38.3	41.2	28	36.4	23	20	23.3	24	32	41.4	28.2
Sports and Recreation	30.7	57.4	25	41.2	28.1	25	35.6	31	30	26.9	28.1
Technology and Innovation	50	60	43	49.6	40	43	48	38	39	35	45
Transportation and Logistics	28	49.2	37	37.2	24.2	26	29.6	20.2	24	28	34

Table 42: Detailed subject-wise evaluation for NEULAB/PANGEA-7B on MILU across different languages. The results reported are for 5-shot experiments.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	26.0/ 24.0/ 28.0	41.0/ 43.5/ 41.4	19.2/ 22.2/ 23.2	37.0/ 36.2/ 36.4	31.0/ 32.0/ 34.0	34.3/ 30.3/ 32.3	23.0/ 27.0/ 27.0	29.0/ 30.0/ 32.0	25.0/ 29.0/ 32.0	29.0/ 29.0/ 33.0	28.0/ 27.0/ 27.0
Anthropology	37.0/ 32.0/ 34.0	49.0/ 45.0/ 43.0	27.0/ 32.0/ 37.0	36.4/ 41.4/ 40.4	39.0/ 30.0/ 34.0	32.0/ 31.0/ 33.0	25.0/ 31.0/ 30.0	31.0/ 31.0/ 33.0	27.0/ 38.0/ 32.0	29.0/ 27.0/ 29.0	30.0/ 29.0/ 24.0
Architecture and Design	31.0/ 30.0/ 35.0	46.0/ 52.0/ 50.0	26.0/ 23.0/ 21.0	33.0/ 29.0/ 31.0	27.0/ 23.0/ 32.0	25.0/ 25.0/ 27.0	33.0/ 31.0/ 30.0	19.0/ 28.0/ 29.0	29.0/ 24.0/ 26.0	26.0/ 28.0/ 31.0	20.0/ 21.0/ 19.0
Arts and Culture	25.4/ 24.2/ 27.3	46.2/ 48.4/ 46.2	28.0/ 28.0/ 27.0	39.3/ 39.5/ 37.9	28.8/ 28.8/ 26.4	26.0/ 19.0/ 20.0	27.9/ 29.0/ 27.9	27.0/ 30.0/ 34.0	31.0/ 36.0/ 36.0	30.3/ 31.5/ 30.3	24.1/ 27.1/ 29.1
Astronomy and Astrophysics	29.0/ 25.0/ 27.0	63.0/ 62.0/ 62.0	25.0/ 32.0/ 25.0	43.0/ 41.4/ 43.8	27.0/ 26.0/ 28.0	27.0/ 25.0/ 27.0	33.0/ 43.0/ 41.0	33.0/ 28.0/ 30.0	28.0/ 29.0/ 31.0	32.0/ 39.0/ 37.0	28.0/ 26.0/ 25.0
Biology	25.0/ 25.0/ 25.7	52.3/ 53.9/ 54.1	30.0/ 28.0/ 28.0	38.1/ 41.3/ 43.9	27.1/ 24.6/ 25.4	26.3/ 30.3/ 23.2	34.9/ 34.6/ 33.4	24.0/ 18.0/ 20.0	28.0/ 36.0/ 33.0	26.4/ 25.5/ 23.2	24.2/ 24.2/ 24.2
Business and Management	27.0/ 31.0/ 27.0	41.8/ 43.6/ 44.0	25.0/ 34.0/ 29.0	34.0/ 38.2/ 38.2	26.0/ 26.0/ 30.0	37.0/ 30.0/ 30.0	24.0/ 30.0/ 28.0	27.0/ 31.0/ 25.0	27.0/ 23.0/ 18.0	36.0/ 30.0/ 30.0	18.9/ 16.0/ 15.1
Chemistry	25.6/ 25.6/ 23.1	50.3/ 48.9/ 51.5	28.4/ 24.1/ 22.4	36.3/ 39.3/ 38.7	33.7/ 29.7/ 29.4	28.0/ 23.0/ 27.0	28.5/ 29.5/ 31.7	31.7/ 30.8/ 30.8	23.0/ 28.0/ 25.0	25.5/ 27.0/ 24.3	27.4/ 27.2/ 28.7
Computer Science	28.6/ 31.8/ 31.2	40.2/ 36.8/ 37.6	25.0/ 27.0/ 26.0	30.6/ 33.0/ 29.8	26.0/ 28.4/ 24.3	23.0/ 27.0/ 30.0	26.6/ 31.9/ 29.0	34.0/ 31.0/ 32.0	18.0/ 19.0/ 22.0	30.2/ 28.5/ 27.3	25.6/ 28.7/ 24.7
Defense and Security	27.0/ 26.0/ 27.0	47.0/ 40.0/ 39.0	30.0/ 30.0/ 27.0	31.8/ 30.3/ 31.3	27.0/ 31.0/ 31.0	25.0/ 27.0/ 32.0	29.0/ 31.0/ 30.0	26.0/ 34.0/ 31.0	33.0/ 33.0/ 29.0	24.0/ 30.0/ 29.0	25.0/ 24.0/ 16.0
Earth Sciences	35.0/ 32.0/ 32.0	49.5/ 51.1/ 50.0	30.0/ 21.0/ 25.0	34.9/ 38.3/ 38.9	24.0/ 23.0/ 23.0	29.0/ 30.0/ 29.0	27.0/ 33.0/ 33.0	22.2/ 22.2/ 21.2	23.0/ 18.0/ 20.0	23.0/ 26.0/ 25.0	33.0/ 30.0/ 27.0
Economics	30.9/ 25.9/ 23.0	44.4/ 43.0/ 43.4	35.0/ 39.0/ 34.0	36.1/ 35.3/ 38.7	33.0/ 25.0/ 28.0	27.0/ 24.0/ 24.0	29.6/ 38.0/ 31.0	24.0/ 26.0/ 22.0	27.0/ 25.0/ 26.0	30.9/ 31.8/ 29.4	27.0/ 28.7/ 31.7
Education	28.0/ 29.0/ 28.0	39.2/ 36.6/ 35.4	23.0/ 18.0/ 22.0	31.1/ 31.4/ 32.8	27.0/ 25.0/ 24.0	31.0/ 25.0/ 23.0	30.6/ 32.5/ 32.5	17.0/ 22.0/ 29.0	23.0/ 27.0/ 19.0	31.0/ 27.0/ 32.0	32.0/ 29.0/ 33.0
Energy and Power	33.0/ 31.0/ 35.0	49.3/ 48.6/ 50.0	29.0/ 34.0/ 37.0	34.6/ 36.3/ 35.9	22.0/ 26.0/ 27.0	21.0/ 23.0/ 24.0	28.0/ 33.0/ 36.0	27.0/ 32.0/ 31.0	29.0/ 27.0/ 24.0	24.0/ 25.0/ 29.0	26.0/ 26.0/ 26.0
Engineering	27.2/ 27.4/ 28.4	39.4/ 40.4/ 40.6	28.8/ 29.6/ 28.0	34.8/ 34.8/ 33.0	27.6/ 30.6/ 30.8	24.8/ 23.9/ 26.5	26.4/ 28.6/ 29.6	27.7/ 24.4/ 27.7	29.0/ 26.0/ 24.0	27.4/ 26.0/ 26.8	27.4/ 27.8/ 29.4
Environmental Science	32.0/ 32.0/ 29.0	48.5/ 49.7/ 50.3	25.0/ 22.0/ 21.0	36.3/ 36.7/ 39.5	28.1/ 30.5/ 27.3	31.0/ 26.0/ 29.0	28.4/ 31.1/ 32.1	29.3/ 30.3/ 29.3	26.0/ 22.0/ 26.0	28.0/ 26.0/ 29.0	26.8/ 27.4/ 27.9
Ethics and Human Rights	31.0/ 31.0/ 25.0	46.0/ 51.0/ 52.0	24.0/ 21.0/ 20.0	40.0/ 38.0/ 39.0	20.0/ 29.0/ 22.0	24.0/ 33.0/ 33.0	35.0/ 32.0/ 34.0	22.0/ 22.0/ 27.0	25.0/ 23.0/ 25.0	32.0/ 26.0/ 26.0	30.0/ 26.0/ 26.0
Finance and Investment	27.2/ 30.2/ 31.7	37.0/ 40.2/ 39.4	27.0/ 34.0/ 39.0	28.8/ 32.6/ 32.0	38.3/ 34.6/ 40.7	24.0/ 25.0/ 25.0	29.9/ 26.6/ 29.9	24.0/ 17.0/ 22.0	34.0/ 30.0/ 32.0	32.4/ 31.8/ 34.6	26.4/ 27.1/ 27.9
Food Science	33.0/ 35.0/ 39.0	50.0/ 57.0/ 60.0	24.0/ 31.0/ 31.0	43.4/ 45.5/ 47.5	29.0/ 36.0/ 34.0	26.0/ 33.0/ 34.0	37.0/ 30.0/ 30.0	30.0/ 29.0/ 36.0	34.0/ 45.0/ 44.0	32.0/ 27.0/ 28.0	22.0/ 22.0/ 22.0
Geography	21.2/ 23.5/ 26.3	40.8/ 44.8/ 43.8	24.0/ 30.0/ 29.0	32.1/ 30.9/ 32.7	27.4/ 25.1/ 26.0	33.3/ 31.3/ 33.3	30.8/ 26.6/ 26.6	25.0/ 26.0/ 30.0	18.0/ 27.0/ 28.0	29.1/ 32.6/ 25.1	30.2/ 27.6/ 25.5
Health and Medicine	32.0/ 35.0/ 39.0	55.5/ 58.9/ 57.1	22.0/ 27.0/ 29.0	44.9/ 45.5/ 46.9	28.0/ 27.0/ 29.0	31.0/ 32.0/ 32.0	25.9/ 28.6/ 25.9	19.0/ 22.0/ 27.0	30.0/ 31.0/ 32.0	30.2/ 29.3/ 27.6	24.3/ 27.1/ 25.2
History	25.3/ 23.2/ 25.6	36.5/ 40.8/ 39.0	24.0/ 23.0/ 23.0	36.3/ 34.3/ 34.5	24.2/ 23.8/ 24.9	29.0/ 31.0/ 31.0	30.5/ 26.4/ 29.7	27.0/ 25.0/ 26.0	31.0/ 28.0/ 30.0	29.8/ 34.2/ 34.5	24.6/ 27.4/ 25.8
Information Technology	37.0/ 33.0/ 41.0	47.0/ 50.0/ 49.0	30.0/ 29.0/ 25.0	44.7/ 51.4/ 45.8	25.0/ 29.0/ 32.0	35.0/ 36.0/ 31.0	41.0/ 44.0/ 46.0	29.0/ 27.0/ 27.0	24.0/ 27.0/ 27.0	31.0/ 34.0/ 32.0	39.0/ 35.0/ 34.0
International Relations	30.0/ 34.0/ 27.0	56.7/ 58.3/ 55.8	21.0/ 22.0/ 21.0	44.0/ 39.7/ 37.7	24.0/ 22.0/ 27.0	27.0/ 27.0/ 26.0	35.0/ 28.0/ 32.0	27.0/ 32.0/ 25.0	27.0/ 29.0/ 26.0	29.0/ 29.0/ 26.0	25.0/ 30.0/ 26.0
Language Studies	25.0/ 30.0/ 26.0	51.8/ 53.8/ 55.2	24.0/ 25.0/ 24.0	36.7/ 31.9/ 31.7	28.7/ 26.4/ 27.9	29.0/ 27.0/ 28.0	29.5/ 26.2/ 26.9	31.0/ 33.0/ 32.0	25.0/ 26.0/ 29.0	25.8/ 29.2/ 26.2	27.6/ 28.2/ 26.4
Law and Ethics	16.9/ 21.6/ 20.9	43.6/ 46.0/ 47.2	23.0/ 26.0/ 27.0	39.2/ 37.4/ 38.4	22.9/ 26.8/ 26.8	23.0/ 21.0/ 21.0	32.8/ 34.1/ 29.3	29.0/ 28.0/ 25.0	27.0/ 30.0/ 30.0	28.3/ 26.0/ 29.9	30.0/ 26.9/ 28.7
Literature and Linguistics	27.0/ 31.5/ 31.5	42.1/ 41.7/ 41.1	30.0/ 30.0/ 33.0	35.1/ 33.7/ 34.7	32.5/ 25.0/ 25.8	26.0/ 25.0/ 27.0	36.3/ 37.2/ 36.3	30.0/ 33.0/ 29.0	14.0/ 21.0/ 23.0	31.6/ 32.4/ 34.0	25.4/ 23.9/ 21.8
Logical Reasoning	29.0/ 31.0/ 29.0	34.0/ 34.2/ 34.0	26.0/ 25.3/ 28.6	28.1/ 25.7/ 26.2	27.8/ 24.5/ 25.3	31.0/ 31.0/ 32.0	25.6/ 27.9/ 27.9	21.0/ 24.0/ 20.0	25.0/ 30.0/ 23.0	31.5/ 31.1/ 30.7	28.5/ 25.7/ 27.5
Materials Science	29.0/ 30.0/ 26.0	46.0/ 44.9/ 46.4	19.0/ 25.0/ 23.0	34.1/ 34.7/ 35.4	22.0/ 26.0/ 24.0	33.0/ 41.0/ 35.0	32.0/ 34.0/ 33.0	28.0/ 30.0/ 32.0	36.0/ 27.0/ 29.0	28.0/ 32.0/ 33.0	29.0/ 29.0/ 27.0
Media and Communication	27.0/ 31.0/ 30.0	58.0/ 57.0/ 63.0	28.0/ 31.0/ 29.0	40.1/ 39.5/ 40.8	30.0/ 26.0/ 23.0	42.0/ 33.0/ 34.0	36.0/ 29.0/ 26.0	23.0/ 24.0/ 21.0	23.0/ 24.0/ 24.0	26.0/ 26.0/ 24.0	30.0/ 29.0/ 28.0
Music and Performing Arts	27.0/ 19.0/ 24.0	45.0/ 43.0/ 41.0	20.0/ 23.0/ 19.0	28.5/ 31.1/ 27.2	30.0/ 26.0/ 26.0	21.0/ 26.0/ 27.0	34.0/ 30.0/ 28.0	31.0/ 32.0/ 31.0	20.0/ 23.0/ 30.0	32.0/ 27.0/ 26.0	29.0/ 30.0/ 30.0
Physics	28.0/ 25.8/ 26.6	43.8/ 48.4/ 50.2	26.2/ 26.7/ 25.9	34.2/ 35.4/ 35.2	23.4/ 25.4/ 25.2	26.3/ 27.9/ 26.3	28.8/ 28.8/ 30.0	30.6/ 26.0/ 27.1	15.0/ 30.0/ 23.0	26.6/ 28.4/ 29.6	25.6/ 26.2/ 26.6
Politics and Governance	29.0/ 28.6/ 25.9	44.3/ 44.5/ 44.9	28.0/ 31.0/ 28.0	36.5/ 34.5/ 34.7	30.4/ 27.5/ 28.3	36.0/ 41.0/ 33.0	29.8/ 29.8/ 30.1	26.0/ 26.0/ 22.0	22.0/ 28.0/ 24.0	30.3/ 33.2/ 30.3	27.6/ 26.4/ 27.1
Psychology	27.0/ 34.0/ 36.0	60.0/ 58.0/ 62.0	30.0/ 27.0/ 25.0	44.2/ 37.2/ 43.4	30.0/ 26.0/ 19.0	30.0/ 33.0/ 30.0	25.0/ 28.0/ 36.0	33.0/ 29.0/ 30.0	25.0/ 23.0/ 25.0	32.0/ 33.0/ 32.0	27.0/ 29.0/ 28.0
Public Administration	23.0/ 26.0/ 25.0	45.5/ 44.4/ 41.4	28.3/ 26.3/ 22.2	34.3/ 39.4/ 39.4	21.2/ 33.3/ 26.3	32.0/ 28.0/ 29.0	27.3/ 26.3/ 25.2	31.0/ 28.0/ 30.0	18.2/ 19.2/ 22.2	36.0/ 31.0/ 33.0	34.0/ 31.0/ 27.0
Religion and Spirituality	24.0/ 24.0/ 27.0	42.0/ 45.0/ 38.0	17.0/ 20.0/ 24.0	31.9/ 33.5/ 32.7	25.0/ 24.0/ 27.0	35.0/ 34.0/ 33.0	39.0/ 30.0/ 32.0	30.0/ 27.0/ 22.0	35.0/ 24.0/ 30.0	28.0/ 25.0/ 30.0	37.0/ 25.0/ 31.0
Social Welfare and Development	25.0/ 22.0/ 28.0	42.0/ 43.8/ 44.6	28.0/ 28.0/ 28.0	32.6/ 38.3/ 38.9	26.0/ 35.0/ 32.0	28.0/ 31.0/ 28.0	35.0/ 37.0/ 36.0	25.0/ 25.0/ 23.0	25.0/ 18.0/ 20.0	30.0/ 35.0/ 32.0	26.0/ 25.0/ 16.0
Sociology	23.5/ 21.7/ 20.9	34.6/ 36.2/ 36.2	32.0/ 35.0/ 36.0	32.8/ 33.2/ 32.4	27.0/ 33.0/ 27.0	25.0/ 30.0/ 30.0	27.7/ 32.1/ 29.6	23.0/ 20.0/ 24.0	24.0/ 27.0/ 28.0	28.8/ 27.0/ 27.0	24.1/ 19.4/ 23.5
Sports and Recreation	23.8/ 25.7/ 25.7	46.2/ 47.6/ 46.2	22.0/ 23.0/ 20.0	37.2/ 36.8/ 38.2	27.5/ 23.6/ 23.0	28.0/ 25.0/ 27.0	30.5/ 33.9/ 36.7	25.0/ 28.0/ 26.0	21.0/ 32.0/ 32.0	27.6/ 23.7/ 24.4	21.4/ 20.9/ 21.9
Technology and Innovation	30.0/ 32.0/ 27.0	48.6/ 52.8/ 51.8	30.0/ 29.0/ 33.0	37.4/ 39.0/ 39.4	25.0/ 28.0/ 25.0	30.0/ 19.0/ 21.0	34.0/ 37.0/ 39.0	33.0/ 33.0/ 33.0	23.0/ 20.0/ 22.0	29.0/ 30.0/ 30.0	23.0/ 27.0/ 19.0
Transportation and Logistics	27.0/ 25.0/ 25.0	43.9/ 43.9/ 44.6	32.0/ 27.0/ 27.0	29.6/ 34.1/ 33.8	35.4/ 30.3/ 29.3	31.0/ 25.0/ 25.0	31.6/ 29.6/ 31.6	29.3/ 26.3/ 24.2	26.0/ 26.0/ 25.0	31.0/ 34.0/ 31.0	30.0/ 25.0/ 27.0

Table 43: Detailed subject-wise evaluation for COHEREFORAI/AYA-23-8B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	37.0/ 35.0/ 32.0	38.8/ 56.5/ 53.5	32.3/ 34.3/ 30.3	27.6/ 35.8/ 38.0	33.0/ 38.0/ 41.0	31.3/ 31.3/ 35.4	29.0/ 34.0/ 38.0	35.0/ 37.0/ 38.0	28.0/ 29.0/ 30.0	25.0/ 35.0/ 36.0	36.0/ 32.0/ 34.0
Anthropology	37.0/ 37.0/ 38.0	46.0/ 66.0/ 62.0	35.0/ 38.0/ 40.0	34.3/ 42.4/ 39.4	32.0/ 34.0/ 34.0	28.0/ 35.0/ 31.0	37.0/ 42.0/ 39.0	36.0/ 39.0/ 38.0	33.0/ 35.0/ 35.0	29.0/ 36.0/ 32.0	30.0/ 34.0/ 32.0
Architecture and Design	28.0/ 30.0/ 35.0	52.0/ 57.0/ 60.0	33.0/ 35.0/ 38.0	31.0/ 41.0/ 42.0	39.0/ 40.0/ 35.0	31.0/ 35.0/ 33.0	31.0/ 39.0/ 41.0	29.0/ 29.0/ 27.0	28.0/ 38.0/ 37.0	29.0/ 34.0/ 35.0	38.0/ 30.0/ 32.0
Arts and Culture	29.1/ 32.1/ 31.5	48.4/ 64.6/ 63.8	26.0/ 45.0/ 42.0	27.7/ 39.1/ 41.3	30.1/ 33.7/ 36.8	38.0/ 44.0/ 39.0	22.9/ 36.1/ 40.4	17.0/ 18.0/ 20.0	46.0/ 40.0/ 36.0	30.3/ 43.6/ 37.0	28.6/ 35.7/ 35.2
Astronomy and Astrophysics	28.0/ 39.0/ 48.0	60.0/ 80.0/ 79.0	35.0/ 49.0/ 48.0	40.6/ 44.5/ 47.7	35.0/ 43.0/ 48.0	40.0/ 46.0/ 45.0	33.0/ 56.0/ 50.0	36.0/ 38.0/ 38.0	40.0/ 46.0/ 53.0	41.0/ 42.0/ 39.0	33.0/ 39.0/ 42.0
Biology	29.2/ 37.0/ 33.8	53.3/ 73.4/ 74.2	29.0/ 33.0/ 32.0	34.9/ 46.7/ 50.1	35.2/ 41.4/ 40.6	36.4/ 32.3/ 32.3	31.6/ 36.1/ 38.8	26.0/ 32.0/ 31.0	42.0/ 47.0/ 51.0	34.1/ 41.1/ 36.0	30.5/ 36.5/ 34.3
Business and Management	32.0/ 35.0/ 29.0	39.3/ 50.5/ 51.4	26.0/ 39.0/ 38.0	37.9/ 38.6/ 42.4	34.0/ 35.0/ 37.0	43.0/ 33.0/ 34.0	30.0/ 36.0/ 36.0	31.0/ 27.0/ 27.0	25.0/ 34.0/ 33.0	29.0/ 33.0/ 34.0	28.3/ 26.4/ 22.6
Chemistry	34.1/ 40.1/ 40.9	54.1/ 67.5/ 69.7	35.3/ 38.8/ 43.1	38.5/ 48.3/ 49.5	33.0/ 35.3/ 37.9	32.0/ 43.0/ 42.0	36.1/ 42.2/ 43.0	25.0/ 33.7/ 28.8	32.0/ 35.0/ 38.0	32.3/ 34.1/ 35.0	37.9/ 40.6/ 38.9
Computer Science	27.6/ 26.0/ 30.2	39.4/ 44.8/ 43.6	37.0/ 37.0/ 41.0	33.2/ 41.2/ 39.4	36.1/ 43.8/ 43.8	19.0/ 33.0/ 31.0	30.2/ 33.7/ 34.3	26.0/ 33.0/ 35.0	33.0/ 32.0/ 39.0	32.6/ 38.4/ 38.4	29.1/ 37.7/ 39.5
Defense and Security	31.0/ 34.0/ 38.0	45.0/ 66.0/ 64.0	32.0/ 35.0/ 29.0	33.7/ 36.0/ 40.3	33.0/ 38.0/ 37.0	34.0/ 33.0/ 37.0	30.0/ 41.0/ 37.0	31.0/ 33.0/ 29.0	32.0/ 40.0/ 42.0	32.0/ 32.0/ 30.0	30.0/ 33.0/ 37.0
Earth Sciences	42.0/ 41.0/ 40.0	49.1/ 62.1/ 65.9	34.0/ 33.0/ 30.0	31.1/ 41.1/ 42.7	38.0/ 37.0/ 38.0	31.0/ 38.0/ 38.0	32.0/ 39.0/ 40.0	34.3/ 34.3/ 33.3	33.0/ 36.0/ 36.0	38.0/ 35.0/ 38.0	32.0/ 35.0/ 34.0
Economics	36.0/ 36.7/ 41.7	48.4/ 58.4/ 60.6	31.0/ 36.0/ 36.0	30.0/ 40.7/ 48.0	36.0/ 42.0/ 40.0	31.0/ 27.0/ 23.0	29.6/ 33.8/ 34.5	29.0/ 28.0/ 27.0	25.0/ 32.0/ 36.0	34.1/ 38.9/ 36.5	28.7/ 34.7/ 37.7
Education	38.0/ 38.0/ 40.0	38.5/ 49.7/ 51.6	28.0/ 36.0/ 38.0	33.1/ 33.7/ 39.1	28.0/ 33.0/ 35.0	23.0/ 27.0/ 30.0	31.9/ 38.2/ 38.2	28.0/ 36.0/ 33.0	25.0/ 30.0/ 24.0	37.0/ 43.0/ 39.0	25.0/ 31.0/ 29.0
Energy and Power	35.0/ 44.0/ 47.0	46.0/ 63.5/ 66.2	39.0/ 41.0/ 39.0	34.6/ 44.1/ 52.2	28.0/ 34.0/ 35.0	34.0/ 42.0/ 39.0	25.0/ 44.0/ 42.0	25.0/ 34.0/ 34.0	33.0/ 36.0/ 38.0	39.0/ 38.0/ 36.0	37.0/ 38.0/ 38.0
Engineering	26.4/ 32.8/ 33.0	38.2/ 48.2/ 49.4	27.6/ 29.0/ 30.6	31.0/ 36.4/ 34.2	29.4/ 28.8/ 29.4	27.8/ 30.3/ 31.6	25.8/ 29.2/ 31.8	29.2/ 25.0/ 29.8	29.0/ 33.0/ 33.0	29.2/ 36.0/ 34.0	28.2/ 31.0/ 33.8
Environmental Science	26.0/ 34.0/ 36.0	45.7/ 65.3/ 61.3	33.0/ 35.0/ 35.0	31.1/ 43.9/ 45.9	36.7/ 32.8/ 34.4	30.0/ 37.0/ 39.0	25.8/ 37.4/ 41.6	39.4/ 38.4/ 43.4	40.0/ 43.0/ 37.0	32.0/ 42.0/ 42.0	25.7/ 31.8/ 32.4
Ethics and Human Rights	33.0/ 35.0/ 45.0	44.0/ 62.0/ 64.0	36.0/ 41.0/ 38.0	41.0/ 42.0/ 49.0	32.0/ 41.0/ 37.0	33.0/ 38.0/ 41.0	36.0/ 43.0/ 40.0	36.0/ 33.0/ 38.0	30.0/ 35.0/ 28.0	30.0/ 38.0/ 38.0	30.0/ 34.0/ 39.0
Finance and Investment	32.2/ 39.1/ 37.1	42.6/ 49.6/ 50.0	39.0/ 40.0/ 36.0	33.4/ 43.2/ 42.4	37.0/ 42.6/ 37.0	36.0/ 32.0/ 31.0	29.9/ 41.8/ 36.7	26.0/ 27.0/ 26.0	35.0/ 35.0/ 35.0	39.7/ 42.5/ 37.4	33.9/ 35.0/ 35.0
Food Science	29.0/ 32.0/ 37.0	53.0/ 74.0/ 74.0	32.0/ 34.0/ 33.0	28.3/ 48.5/ 53.5	29.0/ 34.0/ 40.0	36.0/ 38.0/ 37.0	22.0/ 35.0/ 41.0	29.0/ 36.0/ 38.0	32.0/ 36.0/ 38.0	30.0/ 31.0/ 36.0	33.0/ 42.0/ 48.0
Geography	36.3/ 38.6/ 41.3	43.0/ 58.0/ 59.2	44.0/ 42.0/ 49.0	29.1/ 38.8/ 41.0	32.4/ 40.6/ 36.1	38.4/ 43.4/ 44.4	30.2/ 34.9/ 36.7	41.0/ 43.0/ 41.0	30.0/ 42.0/ 37.0	29.1/ 30.9/ 26.3	30.7/ 33.9/ 36.5
Health and Medicine	31.0/ 38.0/ 46.0	50.7/ 75.0/ 74.2	31.0/ 47.0/ 41.0	37.1/ 50.3/ 55.9	41.0/ 47.0/ 45.0	37.0/ 41.0/ 42.0	34.8/ 33.9/ 35.7	33.0/ 44.0/ 42.0	49.0/ 46.0/ 47.0	35.3/ 37.9/ 36.2	32.7/ 40.2/ 42.1
History	34.9/ 41.6/ 42.1	42.2/ 59.0/ 59.6	43.0/ 51.0/ 47.0	33.5/ 43.8/ 46.0	38.8/ 42.5/ 45.1	31.0/ 37.0/ 33.0	34.9/ 39.4/ 37.5	35.0/ 38.0/ 36.0	32.0/ 36.0/ 32.0	35.6/ 41.1/ 41.1	37.3/ 35.5/ 38.6
Information Technology	35.0/ 41.0/ 48.0	53.0/ 68.0/ 71.0	43.0/ 41.0/ 41.0	43.0/ 57.5/ 54.8	36.0/ 43.0/ 47.0	41.0/ 45.0/ 44.0	43.0/ 49.0/ 53.0	29.0/ 38.0/ 44.0	39.0/ 47.0/ 44.0	39.0/ 50.0/ 45.0	42.0/ 51.0/ 49.0
International Relations	30.0/ 37.0/ 42.0	55.0/ 70.0/ 69.2	34.0/ 40.0/ 39.0	33.9/ 40.9/ 45.1	29.0/ 33.0/ 39.0	37.0/ 34.0/ 35.0	30.0/ 40.0/ 40.0	36.0/ 40.0/ 47.0	49.0/ 47.0/ 48.0	38.0/ 48.0/ 50.0	30.0/ 31.0/ 32.0
Language Studies	29.0/ 26.0/ 25.0	53.0/ 69.6/ 70.0	27.0/ 27.0/ 25.0	30.1/ 35.1/ 33.9	30.2/ 33.3/ 33.3	26.0/ 29.0/ 24.0	26.2/ 30.2/ 30.9	28.0/ 33.0/ 34.0	38.0/ 42.0/ 37.0	32.2/ 30.0/ 30.0	20.7/ 29.9/ 29.3
Law and Ethics	38.5/ 48.6/ 49.3	51.0/ 65.4/ 64.8	22.0/ 37.0/ 35.0	41.8/ 51.8/ 52.4	32.5/ 38.2/ 40.8	25.0/ 29.0/ 31.0	36.6/ 40.1/ 39.2	32.0/ 29.0/ 34.0	42.0/ 43.0/ 42.0	34.6/ 37.0/ 35.4	26.9/ 30.0/ 33.1
Literature and Linguistics	29.7/ 37.8/ 41.4	45.5/ 62.9/ 60.1	29.0/ 32.0/ 31.0	26.7/ 32.5/ 35.9	32.5/ 37.5/ 37.5	33.0/ 32.0/ 33.0	30.1/ 31.9/ 31.9	28.0/ 30.0/ 31.0	27.0/ 33.0/ 31.0	31.6/ 31.6/ 36.0	30.3/ 33.8/ 33.1
Logical Reasoning	31.2/ 34.6/ 34.6	27.6/ 32.4/ 31.8	22.7/ 29.2/ 29.9	28.3/ 31.5/ 32.5	30.3/ 31.9/ 29.5	32.0/ 22.0/ 29.0	25.6/ 25.1/ 26.2	27.0/ 27.0/ 24.0	27.0/ 23.0/ 26.0	32.7/ 31.1/ 31.9	29.5/ 29.5/ 27.0
Materials Science	31.0/ 32.0/ 40.0	43.4/ 56.3/ 59.3	25.0/ 19.0/ 18.0	31.4/ 34.3/ 34.5	29.0/ 28.0/ 30.0	41.0/ 43.0/ 35.0	33.0/ 39.0/ 44.0	20.0/ 25.0/ 19.0	31.0/ 36.0/ 37.0	33.0/ 37.0/ 37.0	22.0/ 33.0/ 38.0
Media and Communication	34.0/ 35.0/ 33.0	50.0/ 71.0/ 64.0	41.0/ 39.0/ 37.0	32.0/ 43.5/ 44.9	41.0/ 36.0/ 40.0	44.0/ 39.0/ 34.0	29.0/ 39.0/ 40.0	33.0/ 31.0/ 38.0	40.0/ 39.0/ 35.0	38.0/ 41.0/ 41.0	39.0/ 39.0/ 44.0
Music and Performing Arts	31.0/ 37.0/ 38.0	46.0/ 63.0/ 59.0	29.0/ 30.0/ 33.0	29.8/ 37.8/ 36.4	36.0/ 35.0/ 36.0	29.0/ 30.0/ 32.0	34.0/ 42.0/ 38.0	29.0/ 34.0/ 38.0	37.0/ 47.0/ 44.0	30.0/ 30.0/ 35.0	37.0/ 39.0/ 39.0
Physics	35.2/ 36.2/ 39.4	50.6/ 61.4/ 60.8	34.5/ 33.7/ 32.0	31.2/ 38.0/ 39.0	35.2/ 35.6/ 37.6	26.8/ 32.1/ 36.3	30.8/ 40.6/ 40.4	30.9/ 28.8/ 32.3	34.0/ 40.0/ 39.0	30.2/ 35.8/ 34.8	33.2/ 35.4/ 36.8
Politics and Governance	33.7/ 40.4/ 44.3	42.7/ 57.3/ 57.1	29.0/ 35.0/ 33.0	36.8/ 44.6/ 43.2	30.4/ 33.6/ 36.0	37.0/ 29.0/ 32.0	31.5/ 36.8/ 37.6	31.0/ 28.0/ 33.0	28.0/ 30.0/ 30.0	41.9/ 42.7/ 40.2	30.9/ 34.0/ 33.7
Psychology	36.0/ 40.0/ 37.0	57.0/ 71.0/ 64.0	33.0/ 36.0/ 36.0	29.5/ 41.1/ 46.5	42.0/ 37.0/ 41.0	31.0/ 33.0/ 39.0	30.0/ 35.0/ 36.0	34.0/ 45.0/ 41.0	35.0/ 33.0/ 34.0	35.0/ 41.0/ 38.0	34.0/ 37.0/ 40.0
Public Administration	33.0/ 30.0/ 31.0	37.4/ 56.6/ 61.6	31.3/ 35.4/ 28.3	32.3/ 36.4/ 30.3	31.3/ 27.3/ 34.3	25.0/ 33.0/ 33.0	37.4/ 31.3/ 33.3	26.0/ 30.0/ 33.0	37.4/ 38.4/ 32.3	40.0/ 37.0/ 33.0	28.0/ 33.0/ 34.0
Religion and Spirituality	30.0/ 39.0/ 42.0	49.0/ 72.0/ 69.0	39.0/ 45.0/ 45.0	37.9/ 41.9/ 46.0	43.0/ 33.0/ 46.0	39.0/ 36.0/ 44.0	40.0/ 38.0/ 45.0	34.0/ 36.0/ 35.0	38.0/ 33.0/ 30.0	37.0/ 37.0/ 40.0	35.0/ 41.0/ 34.0
Social Welfare and Development	35.0/ 38.0/ 38.0	42.9/ 56.2/ 58.9	35.0/ 38.0/ 40.0	34.7/ 38.3/ 42.0	26.0/ 37.0/ 35.0	25.0/ 31.0/ 30.0	44.0/ 40.0/ 34.0	37.0/ 29.0/ 23.0	30.0/ 28.0/ 33.0	34.0/ 35.0/ 35.0	30.0/ 39.0/ 36.0
Sociology	27.0/ 28.7/ 30.4	36.6/ 44.0/ 45.2	31.0/ 29.0/ 34.0	31.8/ 38.4/ 39.0	28.0/ 31.0/ 32.0	28.0/ 29.0/ 29.0	29.6/ 32.1/ 32.7	25.0/ 29.0/ 25.0	27.0/ 34.0/ 27.0	37.8/ 32.4/ 37.8	28.8/ 24.7/ 23.5
Sports and Recreation	27.2/ 31.7/ 30.7	46.6/ 68.0/ 67.0	25.0/ 35.0/ 34.0	29.0/ 39.4/ 43.6	29.8/ 29.8/ 36.5	37.0/ 30.0/ 31.0	28.8/ 42.4/ 41.8	23.0/ 26.0/ 24.0	29.0/ 33.0/ 34.0	35.3/ 36.5/ 30.1	29.5/ 28.6/ 31.9
Technology and Innovation	41.0/ 44.0/ 49.0	49.4/ 64.2/ 63.8	32.0/ 35.0/ 42.0	38.2/ 43.8/ 45.4	43.0/ 45.0/ 41.0	38.0/ 37.0/ 38.0	36.0/ 40.0/ 40.0	30.0/ 39.0/ 43.0	28.0/ 38.0/ 40.0	35.0/ 38.0/ 42.0	36.0/ 29.0/ 31.0
Transportation and Logistics	24.0/ 31.0/ 29.0	36.9/ 53.8/ 56.1	20.0/ 27.0/ 27.0	32.5/ 35.6/ 35.0	29.3/ 24.2/ 22.2	30.0/ 30.0/ 28.0	22.4/ 25.5/ 31.6	24.2/ 28.3/ 24.2	34.0/ 33.0/ 31.0	34.0/ 31.0/ 32.0	29.0/ 30.0/ 27.0

Table 44: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.1-8B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	30.0/ 24.0/ 31.0	39.4/ 39.4/ 40.6	30.3/ 31.3/ 28.3	33.6/ 34.2/ 31.2	23.0/ 31.0/ 30.0	31.3/ 30.3/ 25.2	35.0/ 33.0/ 27.0	36.0/ 29.0/ 31.0	34.0/ 38.0/ 30.0	20.0/ 28.0/ 31.0	36.0/ 38.0/ 25.0
Anthropology	46.0/ 41.0/ 36.0	50.0/ 52.0/ 44.0	38.0/ 37.0/ 37.0	45.5/ 31.3/ 32.3	42.0/ 40.0/ 41.0	33.0/ 33.0/ 36.0	39.0/ 36.0/ 37.0	38.0/ 40.0/ 35.0	46.0/ 47.0/ 37.0	43.0/ 38.0/ 41.0	36.0/ 35.0/ 36.0
Architecture and Design	37.0/ 31.0/ 25.0	46.0/ 45.0/ 48.0	38.0/ 42.0/ 36.0	38.0/ 30.0/ 37.0	37.0/ 30.0/ 35.0	39.0/ 39.0/ 36.0	35.0/ 29.0/ 32.0	26.0/ 28.0/ 27.0	27.0/ 27.0/ 24.0	30.0/ 31.0/ 29.0	35.0/ 33.0/ 31.0
Arts and Culture	37.6/ 34.5/ 33.9	46.0/ 41.8/ 46.6	32.0/ 42.0/ 30.0	35.5/ 32.5/ 37.1	40.5/ 35.6/ 35.6	34.0/ 35.0/ 31.0	27.9/ 29.0/ 30.0	27.0/ 27.0/ 20.0	40.0/ 37.0/ 33.0	40.0/ 36.4/ 37.6	34.7/ 36.7/ 35.2
Astronomy and Astrophysics	47.0/ 44.0/ 41.0	66.0/ 60.0/ 67.0	52.0/ 41.0/ 42.0	49.2/ 39.1/ 42.2	52.0/ 49.0/ 40.0	53.0/ 47.0/ 39.0	48.0/ 40.0/ 44.0	41.0/ 40.0/ 41.0	43.0/ 37.0/ 37.0	49.0/ 42.0/ 41.0	53.0/ 50.0/ 45.0
Biology	43.7/ 33.8/ 33.8	47.5/ 49.3/ 53.3	34.0/ 31.0/ 33.0	41.9/ 36.5/ 38.7	39.8/ 43.4/ 35.2	35.4/ 32.3/ 35.4	36.1/ 33.4/ 30.8	35.0/ 31.0/ 30.0	37.0/ 36.0/ 37.0	37.3/ 34.4/ 31.9	34.9/ 38.4/ 29.2
Business and Management	32.0/ 33.0/ 27.0	45.8/ 39.6/ 42.4	43.0/ 51.0/ 37.0	37.2/ 31.4/ 32.6	37.0/ 33.0/ 39.0	38.0/ 33.0/ 38.0	38.0/ 40.0/ 37.0	35.0/ 29.0/ 29.0	39.0/ 32.0/ 30.0	36.0/ 31.0/ 32.0	30.2/ 26.4/ 23.6
Chemistry	38.7/ 34.9/ 35.7	50.5/ 49.3/ 53.7	29.3/ 37.1/ 42.2	39.9/ 37.5/ 37.7	35.0/ 33.0/ 35.9	41.0/ 33.0/ 41.0	34.6/ 34.3/ 31.7	34.6/ 34.6/ 30.8	35.0/ 22.0/ 29.0	35.6/ 32.0/ 32.3	37.4/ 37.4/ 37.4
Computer Science	26.0/ 22.9/ 26.6	33.8/ 34.4/ 37.6	36.0/ 38.0/ 41.0	30.4/ 28.0/ 31.4	35.5/ 32.5/ 36.7	42.0/ 40.0/ 41.0	26.0/ 28.4/ 29.6	27.0/ 36.0/ 33.0	32.0/ 28.0/ 31.0	33.1/ 39.5/ 38.4	31.4/ 34.1/ 31.8
Defense and Security	32.0/ 31.0/ 37.0	61.0/ 49.0/ 62.0	40.0/ 32.0/ 35.0	37.9/ 35.1/ 37.9	31.0/ 39.0/ 41.0	37.0/ 32.0/ 34.0	43.0/ 38.0/ 37.0	39.0/ 31.0/ 28.0	38.0/ 45.0/ 37.0	27.0/ 32.0/ 35.0	39.0/ 34.0/ 43.0
Earth Sciences	53.0/ 38.0/ 42.0	50.2/ 50.9/ 52.3	38.0/ 39.0/ 36.0	35.9/ 34.1/ 32.3	37.0/ 40.0/ 45.0	39.0/ 38.0/ 41.0	38.0/ 39.0/ 37.0	28.3/ 31.3/ 30.3	38.0/ 38.0/ 35.0	39.0/ 36.0/ 37.0	35.0/ 34.0/ 32.0
Economics	43.2/ 36.0/ 38.9	48.4/ 41.8/ 42.6	34.0/ 40.0/ 32.0	34.5/ 34.1/ 36.3	43.0/ 40.0/ 42.0	37.0/ 26.0/ 29.0	37.3/ 35.9/ 38.0	28.0/ 35.0/ 28.0	39.0/ 35.0/ 35.0	36.5/ 28.6/ 30.2	37.1/ 35.3/ 33.5
Education	38.0/ 41.0/ 37.0	45.9/ 36.3/ 43.6	33.0/ 36.0/ 28.0	34.3/ 32.8/ 34.6	34.0/ 33.0/ 33.0	31.0/ 32.0/ 29.0	39.5/ 42.0/ 42.0	27.0/ 28.0/ 31.0	32.0/ 28.0/ 27.0	43.0/ 34.0/ 27.0	44.0/ 39.0/ 34.0
Energy and Power	41.0/ 34.0/ 37.0	53.4/ 47.3/ 48.6	43.0/ 42.0/ 38.0	32.2/ 35.2/ 35.6	38.0/ 37.0/ 36.0	35.0/ 32.0/ 33.0	36.0/ 35.0/ 37.0	44.0/ 39.0/ 39.0	43.0/ 37.0/ 41.0	32.0/ 27.0/ 34.0	32.0/ 29.0/ 30.0
Engineering	33.2/ 27.0/ 26.2	35.6/ 34.0/ 36.8	30.8/ 30.6/ 29.0	30.4/ 28.6/ 29.4	36.8/ 32.4/ 32.2	33.3/ 33.3/ 34.2	29.0/ 27.0/ 31.0	31.9/ 28.0/ 30.9	23.0/ 26.0/ 30.0	29.8/ 28.4/ 28.8	35.0/ 29.4/ 31.6
Environmental Science	36.0/ 25.0/ 27.0	48.9/ 42.3/ 49.7	46.0/ 42.0/ 33.0	41.5/ 40.3/ 41.9	48.4/ 34.4/ 38.3	36.0/ 39.0/ 36.0	42.1/ 33.7/ 37.4	29.3/ 26.3/ 30.3	38.0/ 38.0/ 36.0	44.0/ 36.0/ 40.0	41.9/ 40.8/ 30.7
Ethics and Human Rights	40.0/ 41.0/ 38.0	53.0/ 47.0/ 47.0	38.0/ 37.0/ 32.0	42.0/ 37.0/ 42.0	40.0/ 38.0/ 31.0	37.0/ 32.0/ 34.0	40.0/ 43.0/ 40.0	27.0/ 26.0/ 22.0	33.0/ 37.0/ 31.0	40.0/ 38.0/ 34.0	45.0/ 42.0/ 37.0
Finance and Investment	30.7/ 28.7/ 34.2	38.2/ 36.2/ 39.6	31.0/ 34.0/ 33.0	31.4/ 29.0/ 31.2	38.3/ 36.4/ 37.0	35.0/ 30.0/ 35.0	30.5/ 28.8/ 28.8	24.0/ 29.0/ 23.0	22.0/ 25.0/ 27.0	33.0/ 38.0/ 35.2	32.9/ 35.0/ 31.4
Food Science	42.0/ 34.0/ 37.0	46.0/ 48.0/ 54.0	37.0/ 32.0/ 33.0	44.4/ 41.4/ 42.4	33.0/ 34.0/ 32.0	42.0/ 46.0/ 44.0	37.0/ 32.0/ 28.0	30.0/ 34.0/ 33.0	41.0/ 32.0/ 33.0	45.0/ 34.0/ 34.0	41.0/ 35.0/ 38.0
Geography	39.1/ 25.7/ 31.8	47.8/ 42.0/ 47.2	45.0/ 45.0/ 46.0	39.0/ 35.5/ 33.1	41.1/ 39.3/ 34.7	39.4/ 37.4/ 35.4	31.9/ 30.2/ 30.8	31.0/ 27.0/ 32.0	36.0/ 30.0/ 36.0	36.6/ 33.7/ 33.7	42.2/ 37.0/ 38.5
Health and Medicine	40.0/ 48.0/ 39.0	60.3/ 54.9/ 58.5	39.0/ 42.0/ 36.0	48.5/ 43.9/ 43.9	46.0/ 50.0/ 47.0	47.0/ 39.0/ 43.0	33.0/ 34.8/ 31.2	44.0/ 37.0/ 35.0	44.0/ 44.0/ 39.0	41.4/ 39.7/ 34.5	43.0/ 37.4/ 42.1
History	43.5/ 43.5/ 41.9	45.6/ 42.6/ 46.2	46.0/ 50.0/ 37.0	44.4/ 42.7/ 43.4	45.1/ 46.9/ 39.9	36.0/ 37.0/ 40.0	36.1/ 33.5/ 32.7	34.0/ 31.0/ 32.0	46.0/ 42.0/ 39.0	36.7/ 36.7/ 38.6	38.4/ 37.9/ 38.9
Information Technology	49.0/ 42.0/ 42.0	54.0/ 51.0/ 50.0	47.0/ 37.0/ 35.0	50.3/ 45.8/ 46.4	47.0/ 37.0/ 45.0	52.0/ 48.0/ 49.0	44.0/ 40.0/ 41.0	42.0/ 42.0/ 38.0	45.0/ 36.0/ 40.0	47.0/ 40.0/ 43.0	56.0/ 48.0/ 50.0
International Relations	36.0/ 39.0/ 37.0	57.5/ 48.3/ 60.0	45.0/ 43.0/ 40.0	40.1/ 40.9/ 38.9	35.0/ 37.0/ 30.0	35.0/ 34.0/ 32.0	49.0/ 45.0/ 44.0	31.0/ 34.0/ 35.0	44.0/ 42.0/ 41.0	43.0/ 38.0/ 40.0	31.0/ 32.0/ 28.0
Language Studies	29.0/ 29.0/ 37.0	51.4/ 43.8/ 52.2	33.0/ 28.0/ 39.0	31.1/ 30.7/ 31.1	27.9/ 28.7/ 31.0	34.0/ 31.0/ 34.0	32.2/ 28.9/ 27.5	38.0/ 30.0/ 34.0	41.0/ 40.0/ 36.0	30.0/ 26.6/ 25.8	32.8/ 26.4/ 25.9
Law and Ethics	53.4/ 38.5/ 48.6	49.2/ 49.4/ 49.8	31.0/ 37.0/ 29.0	45.2/ 42.4/ 43.6	46.5/ 41.4/ 36.9	28.0/ 32.0/ 33.0	37.9/ 39.7/ 36.2	29.0/ 29.0/ 33.0	46.0/ 40.0/ 42.0	40.2/ 34.6/ 34.6	38.8/ 35.6/ 35.6
Literature and Linguistics	36.0/ 30.6/ 36.0	43.7/ 39.9/ 44.1	33.0/ 39.0/ 35.0	35.1/ 34.1/ 34.9	33.3/ 33.3/ 35.0	32.0/ 29.0/ 31.0	31.9/ 26.6/ 31.9	32.0/ 27.0/ 29.0	30.0/ 34.0/ 33.0	37.6/ 29.1/ 31.2	26.1/ 31.0/ 29.6
Logical Reasoning	36.1/ 33.7/ 34.6	27.4/ 24.0/ 26.6	26.0/ 26.0/ 27.9	30.1/ 28.9/ 28.9	30.7/ 30.3/ 31.5	23.0/ 19.0/ 21.0	29.5/ 27.6/ 25.4	21.0/ 22.0/ 22.0	25.0/ 25.0/ 21.0	33.1/ 30.7/ 30.3	27.5/ 28.2/ 30.2
Materials Science	34.0/ 29.0/ 32.0	45.6/ 41.1/ 46.8	23.0/ 19.0/ 28.0	37.5/ 33.0/ 33.2	44.0/ 42.0/ 36.0	37.0/ 36.0/ 44.0	40.0/ 34.0/ 43.0	31.0/ 36.0/ 34.0	31.0/ 34.0/ 26.0	40.0/ 40.0/ 41.0	34.0/ 42.0/ 37.0
Media and Communication	46.0/ 35.0/ 38.0	54.0/ 52.0/ 51.0	37.0/ 36.0/ 40.0	38.8/ 33.3/ 42.2	50.0/ 45.0/ 42.0	47.0/ 41.0/ 38.0	42.0/ 30.0/ 38.0	36.0/ 44.0/ 40.0	40.0/ 32.0/ 36.0	40.0/ 44.0/ 39.0	43.0/ 46.0/ 39.0
Music and Performing Arts	36.0/ 30.0/ 26.0	51.0/ 42.0/ 49.0	35.0/ 33.0/ 37.0	38.4/ 31.1/ 27.2	38.0/ 38.0/ 39.0	29.0/ 25.0/ 20.0	38.0/ 32.0/ 29.0	23.0/ 30.0/ 33.0	32.0/ 33.0/ 38.0	37.0/ 31.0/ 30.0	40.0/ 33.0/ 42.0
Physics	34.6/ 32.2/ 32.2	42.6/ 43.0/ 48.8	30.1/ 31.8/ 31.5	35.6/ 33.4/ 35.6	33.0/ 33.2/ 35.0	33.7/ 37.4/ 33.2	32.0/ 29.2/ 31.8	29.2/ 31.2/ 28.5	32.0/ 36.0/ 39.0	36.4/ 35.6/ 35.8	34.6/ 33.4/ 35.2
Politics and Governance	42.4/ 35.3/ 40.8	48.1/ 44.5/ 47.7	34.0/ 33.0/ 31.0	37.0/ 35.9/ 38.0	37.2/ 37.2/ 33.2	27.0/ 28.0/ 32.0	38.8/ 34.8/ 32.0	30.0/ 28.0/ 29.0	33.0/ 25.0/ 29.0	38.2/ 40.2/ 34.8	33.0/ 33.2/ 30.9
Psychology	41.0/ 28.0/ 35.0	52.0/ 43.0/ 45.0	33.0/ 34.0/ 29.0	36.4/ 34.1/ 35.7	42.0/ 44.0/ 47.0	31.0/ 30.0/ 31.0	37.0/ 41.0/ 38.0	30.0/ 41.0/ 36.0	32.0/ 32.0/ 30.0	40.0/ 42.0/ 37.0	41.0/ 46.0/ 38.0
Public Administration	33.0/ 32.0/ 35.0	39.4/ 35.4/ 36.4	25.2/ 28.3/ 32.3	36.4/ 42.4/ 40.4	24.2/ 26.3/ 34.3	26.0/ 30.0/ 30.0	24.2/ 29.3/ 36.4	24.0/ 27.0/ 25.0	33.3/ 31.3/ 29.3	28.0/ 30.0/ 30.0	31.0/ 27.0/ 33.0
Religion and Spirituality	41.0/ 32.0/ 32.0	49.0/ 44.0/ 51.0	40.0/ 39.0/ 40.0	37.9/ 36.3/ 34.3	53.0/ 49.0/ 44.0	47.0/ 34.0/ 36.0	45.0/ 38.0/ 35.0	31.0/ 22.0/ 25.0	39.0/ 38.0/ 35.0	37.0/ 35.0/ 30.0	42.0/ 35.0/ 29.0
Social Welfare and Development	41.0/ 34.0/ 36.0	44.6/ 47.3/ 50.9	37.0/ 44.0/ 40.0	39.9/ 37.3/ 36.3	37.0/ 34.0/ 41.0	37.0/ 31.0/ 29.0	36.0/ 33.0/ 37.0	31.0/ 29.0/ 29.0	40.0/ 34.0/ 28.0	38.0/ 39.0/ 41.0	37.0/ 34.0/ 36.0
Sociology	37.4/ 30.4/ 30.4	32.6/ 34.8/ 35.0	37.0/ 30.0/ 29.0	30.0/ 30.8/ 32.4	35.0/ 29.0/ 34.0	30.0/ 25.0/ 26.0	37.7/ 34.6/ 32.1	20.0/ 21.0/ 19.0	38.0/ 37.0/ 34.0	30.6/ 35.1/ 34.2	31.2/ 32.4/ 30.0
Sports and Recreation	36.1/ 31.7/ 31.7	53.8/ 46.2/ 50.4	30.0/ 31.0/ 30.0	37.0/ 31.4/ 34.2	30.9/ 33.1/ 29.2	34.0/ 37.0/ 32.0	39.6/ 36.2/ 36.7	26.0/ 28.0/ 28.0	30.0/ 26.0/ 27.0	41.0/ 36.5/ 33.3	35.7/ 35.2/ 30.0
Technology and Innovation	38.0/ 40.0/ 40.0	50.4/ 45.6/ 48.8	33.0/ 39.0/ 39.0	34.4/ 33.8/ 37.4	40.0/ 37.0/ 43.0	35.0/ 38.0/ 40.0	38.0/ 32.0/ 39.0	35.0/ 34.0/ 27.0	40.0/ 43.0/ 41.0	36.0/ 39.0/ 36.0	35.0/ 35.0/ 30.0
Transportation and Logistics	35.0/ 31.0/ 30.0	37.7/ 36.1/ 40.0	24.0/ 31.0/ 24.0	33.4/ 30.9/ 31.9	32.3/ 29.3/ 24.2	33.0/ 24.0/ 29.0	32.6/ 25.5/ 27.6	27.3/ 25.2/ 18.2	26.0/ 25.0/ 27.0	32.0/ 35.0/ 31.0	31.0/ 25.0/ 24.0

Table 45: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.1-8B-INSTRUCT on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	42.0/ 44.0/ 48.0	63.8/ 67.4/ 65.8	37.4/ 51.5/ 49.5	49.5/ 52.9/ 58.1	47.0/ 43.0/ 47.0	38.4/ 38.4/ 39.4	31.0/ 38.0/ 43.0	24.0/ 27.0/ 32.0	43.0/ 43.0/ 46.0	34.0/ 32.0/ 35.0	40.0/ 47.0/ 45.0
Anthropology	48.0/ 46.0/ 47.0	60.0/ 64.0/ 62.0	39.0/ 45.0/ 47.0	50.5/ 46.5/ 48.5	44.0/ 41.0/ 47.0	42.0/ 44.0/ 49.0	48.0/ 44.0/ 40.0	42.0/ 36.0/ 40.0	44.0/ 47.0/ 46.0	42.0/ 42.0/ 37.0	40.0/ 45.0/ 46.0
Architecture and Design	43.0/ 46.0/ 42.0	68.0/ 66.0/ 68.0	44.0/ 46.0/ 44.0	46.0/ 47.0/ 47.0	38.0/ 41.0/ 47.0	43.0/ 46.0/ 47.0	44.0/ 51.0/ 47.0	43.0/ 50.0/ 43.0	44.0/ 45.0/ 48.0	39.0/ 37.0/ 42.0	43.0/ 47.0/ 44.0
Arts and Culture	45.5/ 43.0/ 43.6	69.8/ 71.2/ 71.4	50.0/ 50.0/ 51.0	50.9/ 53.1/ 52.7	39.3/ 42.3/ 47.2	40.0/ 49.0/ 45.0	45.4/ 49.7/ 50.8	24.0/ 28.0/ 26.0	41.0/ 44.0/ 41.0	35.1/ 35.8/ 41.2	34.2/ 36.2/ 38.7
Astronomy and Astrophysics	61.0/ 72.0/ 73.0	86.0/ 87.0/ 86.0	63.0/ 72.0/ 73.0	59.4/ 68.0/ 64.8	64.0/ 70.0/ 73.0	62.0/ 67.0/ 65.0	63.0/ 65.0/ 64.0	49.0/ 43.0/ 47.0	58.0/ 57.0/ 59.0	60.0/ 72.0/ 74.0	61.0/ 59.0/ 66.0
Biology	51.1/ 59.9/ 64.4	79.6/ 83.0/ 83.6	57.0/ 59.0/ 58.0	69.3/ 73.4/ 74.6	53.7/ 59.0/ 61.5	57.6/ 59.6/ 59.6	47.8/ 55.2/ 56.7	41.0/ 44.0/ 41.0	69.0/ 68.0/ 70.0	53.8/ 62.1/ 59.2	52.5/ 56.9/ 56.0
Business and Management	38.0/ 43.0/ 43.0	57.3/ 64.7/ 67.2	38.0/ 55.0/ 51.0	50.1/ 51.0/ 52.9	39.0/ 47.0/ 52.0	38.0/ 40.0/ 45.0	43.0/ 43.0/ 41.0	42.0/ 41.0/ 41.0	50.0/ 49.0/ 46.0	42.0/ 40.0/ 46.0	34.0/ 38.7/ 34.0
Chemistry	58.0/ 64.6/ 64.8	82.0/ 83.6/ 85.0	56.0/ 55.2/ 56.9	63.7/ 69.7/ 70.7	49.7/ 54.9/ 57.2	55.0/ 59.0/ 56.0	50.1/ 59.1/ 62.3	43.3/ 48.1/ 46.2	61.0/ 69.0/ 69.0	49.9/ 53.1/ 57.9	54.6/ 57.6/ 64.3
Computer Science	38.0/ 41.1/ 40.6	49.6/ 52.0/ 52.4	51.0/ 53.0/ 51.0	42.8/ 47.4/ 51.0	52.7/ 58.0/ 61.5	39.0/ 48.0/ 43.0	43.2/ 49.7/ 43.2	41.0/ 47.0/ 48.0	51.0/ 54.0/ 52.0	42.4/ 47.1/ 45.4	39.9/ 45.7/ 40.8
Defense and Security	40.0/ 48.0/ 52.0	63.0/ 79.0/ 78.0	41.0/ 40.0/ 37.0	45.5/ 48.8/ 52.1	41.0/ 45.0/ 55.0	36.0/ 42.0/ 41.0	39.0/ 45.0/ 49.0	35.0/ 41.0/ 43.0	42.0/ 46.0/ 41.0	38.0/ 44.0/ 44.0	43.0/ 54.0/ 58.0
Earth Sciences	46.0/ 51.0/ 54.0	73.6/ 77.0/ 77.0	42.0/ 47.0/ 50.0	51.1/ 56.3/ 57.7	55.0/ 57.0/ 65.0	45.0/ 58.0/ 56.0	49.0/ 55.0/ 57.0	36.4/ 32.3/ 35.4	48.0/ 48.0/ 49.0	53.0/ 56.0/ 59.0	44.0/ 49.0/ 45.0
Economics	57.6/ 56.1/ 59.0	64.9/ 72.5/ 73.5	47.0/ 48.0/ 53.0	47.0/ 55.6/ 56.9	57.0/ 58.0/ 62.0	43.0/ 50.0/ 49.0	45.8/ 54.2/ 49.3	36.0/ 33.0/ 34.0	48.0/ 49.0/ 52.0	46.8/ 52.4/ 54.8	47.9/ 45.5/ 47.9
Education	41.0/ 44.0/ 45.0	53.8/ 58.3/ 58.9	33.0/ 35.0/ 39.0	42.3/ 45.6/ 46.5	44.0/ 42.0/ 43.0	42.0/ 43.0/ 45.0	40.8/ 49.0/ 48.4	36.0/ 40.0/ 35.0	36.0/ 40.0/ 40.0	40.0/ 42.0/ 43.0	41.0/ 37.0/ 37.0
Energy and Power	59.0/ 60.0/ 59.0	69.6/ 79.0/ 79.0	56.0/ 66.0/ 66.0	52.9/ 58.6/ 60.0	52.0/ 60.0/ 60.0	46.0/ 60.0/ 58.0	47.0/ 58.0/ 56.0	45.0/ 41.0/ 52.0	51.0/ 52.0/ 54.0	43.0/ 46.0/ 51.0	57.0/ 54.0/ 60.0
Engineering	41.8/ 51.2/ 48.6	54.8/ 54.6/ 59.2	43.0/ 48.0/ 49.6	41.8/ 44.4/ 46.8	41.2/ 46.6/ 48.2	43.6/ 47.0/ 48.7	40.2/ 47.2/ 47.6	33.6/ 30.9/ 34.5	41.0/ 45.0/ 43.0	35.8/ 39.6/ 43.8	40.4/ 46.2/ 49.6
Environmental Science	45.0/ 52.0/ 55.0	69.1/ 74.8/ 74.0	46.0/ 56.0/ 59.0	55.5/ 59.9/ 61.3	51.6/ 53.9/ 54.7	46.0/ 53.0/ 57.0	46.8/ 52.1/ 52.1	40.4/ 44.4/ 40.4	46.0/ 51.0/ 53.0	47.0/ 50.0/ 52.0	40.8/ 44.7/ 48.0
Ethics and Human Rights	49.0/ 57.0/ 56.0	58.0/ 67.0/ 66.0	46.0/ 57.0/ 54.0	51.0/ 56.0/ 51.0	43.0/ 52.0/ 55.0	40.0/ 50.0/ 51.0	37.0/ 46.0/ 49.0	29.0/ 36.0/ 37.0	44.0/ 52.0/ 55.0	45.0/ 51.0/ 53.0	41.0/ 46.0/ 45.0
Finance and Investment	51.0/ 49.0/ 52.0	56.6/ 60.8/ 62.2	47.0/ 48.0/ 56.0	45.6/ 48.4/ 51.4	49.4/ 55.6/ 59.9	47.0/ 52.0/ 53.0	42.9/ 46.3/ 50.8	41.0/ 36.0/ 40.0	44.0/ 45.0/ 41.0	45.8/ 48.6/ 53.1	37.1/ 39.3/ 45.4
Food Science	58.0/ 52.0/ 62.0	72.0/ 79.0/ 85.0	53.0/ 60.0/ 62.0	68.7/ 75.8/ 78.8	49.0/ 58.0/ 59.0	52.0/ 58.0/ 61.0	57.0/ 66.0/ 59.0	52.0/ 51.0/ 51.0	53.0/ 64.0/ 65.0	48.0/ 55.0/ 56.0	52.0/ 62.0/ 63.0
Geography	42.5/ 48.0/ 49.7	64.4/ 70.6/ 72.4	49.0/ 52.0/ 50.0	48.0/ 53.0/ 53.4	48.4/ 52.5/ 59.4	40.4/ 43.4/ 48.5	42.0/ 47.9/ 52.1	43.0/ 42.0/ 43.0	39.0/ 45.0/ 42.0	46.9/ 52.0/ 51.4	42.2/ 42.2/ 49.0
Health and Medicine	59.0/ 60.0/ 69.0	79.0/ 83.0/ 84.0	51.0/ 58.0/ 55.0	67.7/ 73.6/ 73.0	63.0/ 66.0/ 71.0	70.0/ 71.0/ 71.0	49.1/ 53.6/ 60.7	54.0/ 50.0/ 53.0	65.0/ 72.0/ 71.0	56.9/ 61.2/ 59.5	58.9/ 59.8/ 56.1
History	46.1/ 50.1/ 53.3	60.2/ 66.9/ 67.5	53.0/ 56.0/ 53.0	51.4/ 57.3/ 58.7	49.1/ 52.0/ 54.9	36.0/ 42.0/ 48.0	40.5/ 44.2/ 43.9	38.0/ 42.0/ 40.0	44.0/ 47.0/ 46.0	43.6/ 46.6/ 49.5	39.1/ 44.2/ 44.5
Information Technology	62.0/ 64.0/ 70.0	74.0/ 78.0/ 80.0	61.0/ 65.0/ 70.0	62.6/ 67.0/ 72.1	61.0/ 72.0/ 70.0	65.0/ 69.0/ 69.0	61.0/ 68.0/ 73.0	62.0/ 55.0/ 58.0	55.0/ 69.0/ 66.0	69.0/ 60.0/ 64.0	55.0/ 71.0/ 72.0
International Relations	57.0/ 55.0/ 62.0	72.5/ 77.5/ 79.2	48.0/ 57.0/ 61.0	52.5/ 54.9/ 58.0	39.0/ 52.0/ 55.0	52.0/ 50.0/ 58.0	57.0/ 58.0/ 58.0	42.0/ 50.0/ 53.0	52.0/ 50.0/ 51.0	54.0/ 59.0/ 53.0	41.0/ 57.0/ 60.0
Language Studies	33.0/ 34.0/ 28.0	75.0/ 81.0/ 81.8	36.0/ 41.0/ 51.0	38.5/ 43.5/ 42.9	34.9/ 39.5/ 41.9	37.0/ 45.0/ 37.0	36.2/ 39.6/ 41.6	38.0/ 38.0/ 44.0	40.0/ 51.0/ 47.0	35.2/ 37.8/ 41.2	39.7/ 43.1/ 40.2
Law and Ethics	64.9/ 67.6/ 71.0	67.8/ 75.2/ 75.2	56.0/ 60.0/ 57.0	58.2/ 65.2/ 66.4	55.4/ 58.6/ 63.1	47.0/ 52.0/ 53.0	44.0/ 49.6/ 49.1	40.0/ 41.0/ 38.0	48.0/ 56.0/ 54.0	51.2/ 59.8/ 56.7	38.1/ 41.9/ 41.9
Literature and Linguistics	36.0/ 40.5/ 40.5	60.7/ 64.1/ 65.9	34.0/ 39.0/ 41.0	36.5/ 39.7/ 45.3	38.3/ 38.3/ 40.0	27.0/ 30.0/ 33.0	33.6/ 34.5/ 35.4	36.0/ 37.0/ 40.0	37.0/ 45.0/ 43.0	35.6/ 35.6/ 34.8	31.7/ 35.9/ 30.3
Logical Reasoning	33.9/ 38.6/ 39.1	40.8/ 40.4/ 39.8	27.3/ 37.7/ 40.3	34.9/ 41.9/ 40.9	30.3/ 31.9/ 38.6	26.0/ 28.0/ 32.0	31.2/ 31.5/ 32.0	21.0/ 22.0/ 17.0	28.0/ 35.0/ 32.0	33.5/ 33.5/ 35.5	33.9/ 34.9/ 36.6
Materials Science	55.0/ 56.0/ 56.0	70.3/ 71.1/ 73.8	46.0/ 46.0/ 48.0	49.8/ 55.0/ 55.5	45.0/ 45.0/ 46.0	45.0/ 54.0/ 54.0	51.0/ 57.0/ 52.0	34.0/ 40.0/ 40.0	50.0/ 50.0/ 55.0	37.0/ 52.0/ 53.0	50.0/ 55.0/ 59.0
Media and Communication	54.0/ 56.0/ 57.0	75.0/ 78.0/ 78.0	54.0/ 54.0/ 56.0	51.7/ 55.8/ 61.2	52.0/ 53.0/ 60.0	50.0/ 52.0/ 59.0	44.0/ 53.0/ 55.0	47.0/ 52.0/ 52.0	50.0/ 42.0/ 38.0	46.0/ 50.0/ 54.0	55.0/ 60.0/ 58.0
Music and Performing Arts	32.0/ 31.0/ 35.0	63.0/ 62.0/ 60.0	39.0/ 40.0/ 41.0	33.1/ 42.4/ 37.1	30.0/ 35.0/ 35.0	28.0/ 35.0/ 39.0	42.0/ 35.0/ 38.0	35.0/ 35.0/ 35.0	37.0/ 38.0/ 42.0	32.0/ 23.0/ 28.0	37.0/ 41.0/ 39.0
Physics	56.0/ 58.6/ 61.4	70.4/ 72.2/ 75.0	50.1/ 50.4/ 54.0	55.2/ 56.8/ 63.6	46.8/ 58.0/ 60.4	48.9/ 57.4/ 57.9	55.4/ 58.2/ 59.8	35.1/ 32.6/ 44.1	53.0/ 52.0/ 52.0	50.8/ 56.6/ 60.2	55.0/ 54.2/ 58.4
Politics and Governance	50.6/ 52.9/ 58.8	60.1/ 63.9/ 65.3	41.0/ 52.0/ 47.0	49.0/ 52.6/ 54.0	47.4/ 50.6/ 52.6	49.0/ 44.0/ 51.0	38.5/ 43.8/ 45.5	40.0/ 36.0/ 40.0	49.0/ 54.0/ 52.0	44.4/ 47.7/ 45.2	39.1/ 39.9/ 42.2
Psychology	53.0/ 57.0/ 62.0	75.0/ 82.0/ 82.0	37.0/ 40.0/ 46.0	51.9/ 55.0/ 62.0	44.0/ 57.0/ 53.0	50.0/ 54.0/ 49.0	50.0/ 52.0/ 52.0	44.0/ 48.0/ 46.0	54.0/ 57.0/ 56.0	41.0/ 45.0/ 48.0	55.0/ 62.0/ 62.0
Public Administration	39.0/ 42.0/ 43.0	60.6/ 62.6/ 63.6	47.5/ 48.5/ 46.5	46.5/ 50.5/ 48.5	42.4/ 40.4/ 42.4	29.0/ 37.0/ 34.0	36.4/ 39.4/ 37.4	36.0/ 45.0/ 39.0	44.4/ 51.5/ 45.5	36.0/ 48.0/ 55.0	36.0/ 54.0/ 50.0
Religion and Spirituality	48.0/ 55.0/ 59.0	73.0/ 81.0/ 80.0	47.0/ 61.0/ 59.0	60.5/ 62.9/ 60.5	49.0/ 62.0/ 67.0	46.0/ 60.0/ 63.0	47.0/ 58.0/ 54.0	37.0/ 34.0/ 31.0	51.0/ 55.0/ 50.0	48.0/ 50.0/ 48.0	48.0/ 58.0/ 62.0
Social Welfare and Development	43.0/ 51.0/ 53.0	62.5/ 71.4/ 69.6	44.0/ 51.0/ 54.0	52.8/ 57.5/ 59.1	36.0/ 42.0/ 46.0	37.0/ 40.0/ 39.0	43.0/ 45.0/ 46.0	32.0/ 36.0/ 29.0	44.0/ 50.0/ 51.0	38.0/ 43.0/ 44.0	42.0/ 46.0/ 50.0
Sociology	28.7/ 34.8/ 34.8	45.8/ 50.4/ 50.8	30.0/ 29.0/ 31.0	36.8/ 42.0/ 43.4	31.0/ 42.0/ 37.0	33.0/ 32.0/ 34.0	32.1/ 40.9/ 39.6	29.0/ 35.0/ 37.0	30.0/ 37.0/ 33.0	40.5/ 42.3/ 46.9	40.6/ 40.6/ 40.0
Sports and Recreation	40.1/ 48.0/ 50.5	71.2/ 74.2/ 75.2	40.0/ 37.0/ 42.0	47.0/ 53.2/ 57.6	38.8/ 48.9/ 55.6	41.0/ 43.0/ 42.0	49.7/ 54.8/ 55.9	30.0/ 30.0/ 38.0	46.0/ 55.0/ 51.0	48.1/ 48.1/ 51.3	41.4/ 43.8/ 45.7
Technology and Innovation	51.0/ 56.0/ 59.0	69.2/ 72.2/ 73.0	47.0/ 47.0/ 52.0	56.6/ 56.2/ 59.4	49.0/ 52.0/ 61.0	50.0/ 48.0/ 52.0	49.0/ 57.0/ 58.0	42.0/ 40.0/ 49.0	49.0/ 49.0/ 51.0	44.0/ 45.0/ 49.0	36.0/ 45.0/ 51.0
Transportation and Logistics	29.0/ 33.0/ 41.0	56.9/ 64.6/ 61.5	39.0/ 40.0/ 41.0	38.8/ 42.6/ 44.5	34.3/ 38.4/ 39.4	33.0/ 33.0/ 33.0	25.5/ 29.6/ 28.6	24.2/ 26.3/ 31.3	33.0/ 37.0/ 39.0	36.0/ 37.0/ 43.0	28.0/ 39.0/ 32.0

Table 46: Detailed subject-wise evaluation for GOOGLE/GEMMA-2-9B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	25.0/ 35.0/ 37.0	51.3/ 60.0/ 59.8	30.3/ 42.4/ 31.3	37.6/ 43.9/ 42.0	38.0/ 44.0/ 38.0	30.3/ 35.4/ 36.4	35.0/ 36.0/ 38.0	28.0/ 28.0/ 25.0	35.0/ 36.0/ 32.0	35.0/ 34.0/ 29.0	34.0/ 37.0/ 29.0
Anthropology	37.0/ 41.0/ 44.0	51.0/ 60.0/ 60.0	39.0/ 44.0/ 32.0	41.4/ 46.5/ 39.4	36.0/ 33.0/ 31.0	38.0/ 39.0/ 35.0	37.0/ 41.0/ 27.0	30.0/ 31.0/ 29.0	39.0/ 39.0/ 30.0	32.0/ 36.0/ 35.0	34.0/ 36.0/ 30.0
Architecture and Design	38.0/ 32.0/ 32.0	49.0/ 58.0/ 56.0	31.0/ 35.0/ 31.0	35.0/ 35.0/ 41.0	36.0/ 38.0/ 29.0	39.0/ 40.0/ 41.0	31.0/ 37.0/ 25.0	29.0/ 27.0/ 26.0	40.0/ 34.0/ 31.0	33.0/ 37.0/ 27.0	27.0/ 31.0/ 32.0
Arts and Culture	35.1/ 38.8/ 39.4	55.6/ 65.8/ 67.0	33.0/ 35.0/ 40.0	37.3/ 42.1/ 40.9	36.2/ 38.0/ 35.6	36.0/ 35.0/ 36.0	30.6/ 30.6/ 35.5	31.0/ 32.0/ 22.0	34.0/ 43.0/ 43.0	39.4/ 33.9/ 37.0	32.7/ 31.2/ 34.7
Astronomy and Astrophysics	47.0/ 56.0/ 58.0	70.0/ 83.0/ 78.0	59.0/ 67.0/ 60.0	46.1/ 52.3/ 51.6	59.0/ 63.0/ 65.0	47.0/ 57.0/ 51.0	44.0/ 45.0/ 44.0	36.0/ 44.0/ 44.0	40.0/ 48.0/ 55.0	46.0/ 55.0/ 54.0	53.0/ 54.0/ 54.0
Biology	40.1/ 47.9/ 53.5	65.9/ 78.2/ 80.4	32.0/ 42.0/ 43.0	48.1/ 54.7/ 58.1	41.8/ 43.0/ 42.6	48.5/ 49.5/ 53.5	34.9/ 37.9/ 36.7	35.0/ 37.0/ 28.0	48.0/ 49.0/ 51.0	38.9/ 40.8/ 39.2	40.6/ 39.6/ 43.1
Business and Management	31.0/ 33.0/ 36.0	51.4/ 57.0/ 58.8	31.0/ 32.0/ 36.0	41.9/ 40.8/ 42.6	32.0/ 34.0/ 25.0	42.0/ 41.0/ 40.0	41.0/ 41.0/ 43.0	33.0/ 29.0/ 34.0	40.0/ 42.0/ 41.0	44.0/ 40.0/ 36.0	32.1/ 41.5/ 29.2
Chemistry	43.7/ 52.2/ 56.3	63.5/ 74.6/ 79.4	44.0/ 42.2/ 43.1	49.9/ 54.9/ 53.3	41.2/ 42.2/ 49.0	44.0/ 51.0/ 53.0	37.7/ 42.7/ 42.2	39.4/ 34.6/ 43.3	47.0/ 53.0/ 51.0	44.8/ 48.4/ 45.7	42.1/ 46.6/ 47.6
Computer Science	34.4/ 39.1/ 39.1	43.4/ 48.0/ 48.2	44.0/ 50.0/ 48.0	35.2/ 42.2/ 42.2	39.1/ 48.5/ 48.5	37.0/ 43.0/ 41.0	33.7/ 40.8/ 39.6	41.0/ 43.0/ 44.0	49.0/ 47.0/ 49.0	41.9/ 42.4/ 37.8	40.4/ 37.7/ 36.8
Defense and Security	38.0/ 36.0/ 40.0	53.0/ 63.0/ 64.0	39.0/ 39.0/ 32.0	37.0/ 36.5/ 38.4	46.0/ 42.0/ 44.0	28.0/ 34.0/ 36.0	34.0/ 28.0/ 31.0	39.0/ 30.0/ 37.0	36.0/ 44.0/ 38.0	41.0/ 42.0/ 34.0	49.0/ 45.0/ 36.0
Earth Sciences	43.0/ 43.0/ 44.0	53.6/ 71.6/ 73.2	32.0/ 40.0/ 34.0	42.5/ 46.9/ 47.1	48.0/ 38.0/ 52.0	37.0/ 43.0/ 44.0	44.0/ 43.0/ 42.0	31.3/ 30.3/ 27.3	34.0/ 31.0/ 29.0	40.0/ 52.0/ 43.0	44.0/ 44.0/ 39.0
Economics	41.0/ 46.8/ 48.2	48.8/ 61.0/ 63.0	34.0/ 42.0/ 36.0	42.3/ 47.4/ 48.4	49.0/ 52.0/ 58.0	33.0/ 37.0/ 40.0	33.8/ 39.4/ 43.7	37.0/ 33.0/ 40.0	39.0/ 45.0/ 48.0	45.2/ 49.2/ 38.1	35.9/ 39.5/ 37.1
Education	39.0/ 40.0/ 49.0	43.0/ 51.6/ 53.5	31.0/ 41.0/ 37.0	35.5/ 36.7/ 40.8	44.0/ 38.0/ 38.0	38.0/ 38.0/ 40.0	41.4/ 41.4/ 40.1	27.0/ 27.0/ 30.0	28.0/ 37.0/ 33.0	43.0/ 40.0/ 34.0	38.0/ 40.0/ 42.0
Energy and Power	49.0/ 51.0/ 56.0	54.7/ 69.6/ 69.6	39.0/ 43.0/ 42.0	35.9/ 45.1/ 50.5	36.0/ 47.0/ 48.0	41.0/ 44.0/ 40.0	41.0/ 47.0/ 45.0	27.0/ 32.0/ 37.0	37.0/ 45.0/ 43.0	37.0/ 35.0/ 43.0	42.0/ 36.0/ 39.0
Engineering	30.4/ 38.0/ 41.0	43.6/ 50.0/ 52.8	35.8/ 36.4/ 34.0	33.2/ 32.0/ 38.8	35.0/ 35.8/ 39.2	40.6/ 37.6/ 38.9	34.4/ 33.2/ 33.6	28.9/ 27.1/ 31.6	38.0/ 34.0/ 39.0	32.8/ 32.0/ 36.0	34.2/ 37.4/ 37.8
Environmental Science	34.0/ 37.0/ 42.0	55.7/ 65.3/ 68.1	37.0/ 42.0/ 37.0	46.9/ 53.3/ 50.5	39.1/ 36.7/ 43.8	45.0/ 44.0/ 41.0	41.0/ 41.0/ 43.2	34.3/ 43.4/ 41.4	46.0/ 45.0/ 47.0	42.0/ 45.0/ 48.0	44.7/ 46.9/ 43.0
Ethics and Human Rights	38.0/ 41.0/ 44.0	44.0/ 54.0/ 50.0	44.0/ 47.0/ 47.0	51.0/ 48.0/ 47.0	44.0/ 38.0/ 46.0	39.0/ 39.0/ 40.0	39.0/ 41.0/ 41.0	33.0/ 30.0/ 34.0	40.0/ 39.0/ 43.0	39.0/ 40.0/ 41.0	46.0/ 44.0/ 44.0
Finance and Investment	47.0/ 49.5/ 50.0	49.2/ 57.0/ 58.8	42.0/ 46.0/ 47.0	37.2/ 42.0/ 45.8	45.1/ 49.4/ 52.5	42.0/ 45.0/ 48.0	37.3/ 41.8/ 45.2	35.0/ 29.0/ 36.0	29.0/ 29.0/ 37.0	36.9/ 45.2/ 46.4	35.4/ 37.5/ 36.8
Food Science	36.0/ 49.0/ 51.0	65.0/ 78.0/ 82.0	37.0/ 43.0/ 38.0	37.4/ 54.5/ 54.5	44.0/ 45.0/ 42.0	44.0/ 44.0/ 41.0	36.0/ 39.0/ 42.0	35.0/ 34.0/ 36.0	37.0/ 51.0/ 41.0	40.0/ 45.0/ 37.0	43.0/ 49.0/ 45.0
Geography	34.6/ 48.0/ 45.2	50.4/ 62.6/ 63.6	35.0/ 43.0/ 40.0	38.6/ 43.2/ 44.4	40.6/ 45.2/ 41.1	36.4/ 38.4/ 35.4	30.2/ 31.4/ 34.9	29.0/ 29.0/ 30.0	38.0/ 32.0/ 32.0	36.6/ 43.4/ 36.0	37.5/ 40.1/ 37.5
Health and Medicine	43.0/ 56.0/ 57.0	64.9/ 76.3/ 77.0	41.0/ 52.0/ 48.0	50.5/ 59.5/ 59.3	47.0/ 50.0/ 55.0	48.0/ 54.0/ 51.0	33.0/ 35.7/ 30.4	42.0/ 48.0/ 45.0	47.0/ 55.0/ 50.0	48.3/ 39.7/ 39.7	49.5/ 46.7/ 49.5
History	44.0/ 41.9/ 44.0	52.2/ 57.8/ 59.4	43.0/ 37.0/ 40.0	39.1/ 42.9/ 45.2	45.8/ 44.3/ 47.2	41.0/ 37.0/ 32.0	33.8/ 33.8/ 34.6	33.0/ 28.0/ 30.0	37.0/ 42.0/ 40.0	43.6/ 40.7/ 37.1	37.3/ 35.3/ 39.1
Information Technology	44.0/ 52.0/ 57.0	58.0/ 73.0/ 75.0	46.0/ 48.0/ 49.0	53.1/ 59.2/ 62.6	54.0/ 52.0/ 57.0	55.0/ 58.0/ 55.0	48.0/ 56.0/ 52.0	45.0/ 45.0/ 47.0	47.0/ 50.0/ 53.0	49.0/ 51.0/ 42.0	39.0/ 49.0/ 45.0
International Relations	48.0/ 53.0/ 52.0	57.5/ 72.5/ 74.2	46.0/ 48.0/ 47.0	46.3/ 54.5/ 49.0	35.0/ 47.0/ 47.0	39.0/ 48.0/ 36.0	47.0/ 51.0/ 51.0	41.0/ 44.0/ 45.0	44.0/ 46.0/ 39.0	52.0/ 45.0/ 45.0	39.0/ 37.0/ 30.0
Language Studies	33.0/ 39.0/ 46.0	59.8/ 69.6/ 71.8	34.0/ 37.0/ 38.0	32.3/ 37.9/ 37.7	27.9/ 32.6/ 36.4	36.0/ 36.0/ 33.0	32.9/ 26.9/ 27.5	43.0/ 39.0/ 39.0	36.0/ 45.0/ 45.0	28.5/ 28.8/ 29.6	28.7/ 29.9/ 33.9
Law and Ethics	45.3/ 51.3/ 54.7	57.2/ 63.4/ 64.8	46.0/ 48.0/ 42.0	49.8/ 52.4/ 56.0	43.3/ 47.8/ 54.1	46.0/ 44.0/ 45.0	39.2/ 44.8/ 40.1	27.0/ 28.0/ 31.0	48.0/ 55.0/ 43.0	43.3/ 52.0/ 49.6	34.4/ 35.0/ 33.8
Literature and Linguistics	29.7/ 33.3/ 37.8	50.1/ 59.9/ 61.7	27.0/ 37.0/ 36.0	32.5/ 33.3/ 36.1	37.5/ 35.8/ 40.8	37.0/ 31.0/ 39.0	31.9/ 33.6/ 27.4	36.0/ 34.0/ 40.0	30.0/ 40.0/ 37.0	29.5/ 29.5/ 28.3	33.1/ 40.1/ 33.8
Logical Reasoning	36.1/ 42.8/ 39.3	32.0/ 36.8/ 40.2	33.8/ 34.4/ 36.4	37.1/ 39.7/ 38.1	36.5/ 36.5/ 40.2	42.0/ 36.0/ 40.0	32.9/ 37.3/ 33.4	23.0/ 28.0/ 25.0	42.0/ 32.0/ 35.0	35.9/ 41.4/ 40.6	36.4/ 37.4/ 37.9
Materials Science	39.0/ 38.0/ 46.0	52.5/ 62.4/ 65.4	35.0/ 34.0/ 31.0	41.0/ 43.0/ 42.4	45.0/ 41.0/ 37.0	35.0/ 45.0/ 46.0	43.0/ 39.0/ 39.0	29.0/ 25.0/ 32.0	37.0/ 46.0/ 44.0	36.0/ 37.0/ 34.0	42.0/ 34.0/ 39.0
Media and Communication	34.0/ 44.0/ 50.0	61.0/ 71.0/ 72.0	42.0/ 43.0/ 46.0	44.9/ 45.6/ 41.5	40.0/ 43.0/ 41.0	44.0/ 46.0/ 44.0	35.0/ 39.0/ 38.0	34.0/ 37.0/ 38.0	40.0/ 44.0/ 51.0	37.0/ 37.0/ 35.0	50.0/ 51.0/ 46.0
Music and Performing Arts	38.0/ 23.0/ 25.0	40.0/ 52.0/ 55.0	36.0/ 38.0/ 29.0	29.8/ 36.4/ 30.5	24.0/ 22.0/ 20.0	33.0/ 26.0/ 32.0	29.0/ 34.0/ 31.0	28.0/ 25.0/ 22.0	39.0/ 43.0/ 34.0	36.0/ 41.0/ 31.0	45.0/ 35.0/ 31.0
Physics	41.6/ 47.6/ 52.6	54.8/ 67.2/ 68.8	40.9/ 42.6/ 43.2	38.2/ 42.0/ 45.4	39.4/ 42.8/ 48.4	41.0/ 43.2/ 48.9	41.4/ 41.4/ 44.8	26.7/ 30.6/ 37.9	43.0/ 38.0/ 44.0	40.8/ 45.0/ 47.0	42.6/ 44.4/ 48.4
Politics and Governance	41.6/ 53.7/ 52.5	49.5/ 60.5/ 60.3	34.0/ 37.0/ 32.0	34.9/ 46.6/ 42.8	33.6/ 33.6/ 40.1	36.0/ 45.0/ 40.0	30.6/ 39.0/ 33.7	28.0/ 37.0/ 34.0	40.0/ 43.0/ 40.0	44.8/ 45.6/ 45.2	39.4/ 41.5/ 38.0
Psychology	34.0/ 44.0/ 47.0	61.0/ 73.0/ 75.0	38.0/ 39.0/ 32.0	36.4/ 36.4/ 41.1	38.0/ 47.0/ 57.0	36.0/ 36.0/ 37.0	33.0/ 31.0/ 37.0	42.0/ 36.0/ 42.0	42.0/ 45.0/ 43.0	33.0/ 38.0/ 34.0	46.0/ 43.0/ 49.0
Public Administration	26.0/ 34.0/ 36.0	49.5/ 51.5/ 53.5	39.4/ 36.4/ 34.3	38.4/ 46.5/ 43.4	35.4/ 35.4/ 39.4	33.0/ 32.0/ 33.0	31.3/ 32.3/ 34.3	25.0/ 21.0/ 28.0	28.3/ 32.3/ 27.3	31.0/ 34.0/ 29.0	40.0/ 41.0/ 40.0
Religion and Spirituality	44.0/ 49.0/ 48.0	62.0/ 71.0/ 73.0	45.0/ 43.0/ 40.0	44.8/ 49.2/ 46.0	47.0/ 48.0/ 55.0	43.0/ 35.0/ 39.0	38.0/ 38.0/ 38.0	27.0/ 31.0/ 28.0	48.0/ 51.0/ 45.0	34.0/ 36.0/ 31.0	46.0/ 53.0/ 41.0
Social Welfare and Development	33.0/ 39.0/ 46.0	48.2/ 58.0/ 61.6	41.0/ 45.0/ 52.0	38.3/ 39.9/ 39.4	43.0/ 37.0/ 44.0	38.0/ 40.0/ 39.0	37.0/ 34.0/ 33.0	37.0/ 39.0/ 38.0	41.0/ 42.0/ 41.0	37.0/ 48.0/ 43.0	35.0/ 40.0/ 37.0
Sociology	22.6/ 25.2/ 32.2	38.0/ 45.6/ 47.2	35.0/ 28.0/ 33.0	31.2/ 30.0/ 29.4	29.0/ 23.0/ 27.0	22.0/ 19.0/ 24.0	28.3/ 30.2/ 30.8	21.0/ 22.0/ 30.0	27.0/ 34.0/ 34.0	35.1/ 34.2/ 26.1	28.8/ 32.4/ 35.3
Sports and Recreation	36.6/ 43.1/ 46.0	53.4/ 68.2/ 68.4	36.0/ 34.0/ 33.0	39.0/ 44.2/ 38.4	32.0/ 40.5/ 39.3	33.0/ 37.0/ 40.0	39.0/ 45.2/ 33.3	29.0/ 29.0/ 30.0	41.0/ 44.0/ 29.0	28.8/ 39.1/ 31.4	32.9/ 37.1/ 35.7
Technology and Innovation	37.0/ 40.0/ 51.0	54.4/ 64.0/ 67.8	37.0/ 42.0/ 36.0	43.4/ 49.2/ 49.2	40.0/ 43.0/ 45.0	46.0/ 46.0/ 47.0	40.0/ 46.0/ 48.0	34.0/ 38.0/ 37.0	50.0/ 45.0/ 43.0	37.0/ 38.0/ 40.0	39.0/ 36.0/ 45.0
Transportation and Logistics	42.0/ 38.0/ 36.0	43.1/ 52.3/ 53.8	31.0/ 25.0/ 34.0	31.6/ 35.6/ 37.2	29.3/ 30.3/ 31.3	41.0/ 38.0/ 36.0	30.6/ 34.7/ 33.7	25.2/ 30.3/ 26.3	27.0/ 27.0/ 27.0	34.0/ 39.0/ 38.0	30.0/ 27.0/ 27.0

Table 47: Detailed subject-wise evaluation for GOOGLE/GEMMA-2-9B-IT on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	47.0/ 55.0/ 53.0	66.4/ 68.8/ 69.4	46.5/ 58.6/ 54.5	58.1/ 61.6/ 63.6	51.0/ 55.0/ 55.0	50.5/ 59.6/ 60.6	48.0/ 50.0/ 48.0	37.0/ 39.0/ 35.0	51.0/ 55.0/ 59.0	44.0/ 49.0/ 48.0	51.0/ 49.0/ 55.0
Anthropology	54.0/ 62.0/ 56.0	67.0/ 72.0/ 73.0	52.0/ 61.0/ 54.0	55.6/ 55.6/ 57.6	48.0/ 53.0/ 48.0	52.0/ 58.0/ 53.0	54.0/ 60.0/ 60.0	45.0/ 41.0/ 43.0	49.0/ 54.0/ 56.0	49.0/ 57.0/ 56.0	46.0/ 48.0/ 50.0
Architecture and Design	45.0/ 57.0/ 55.0	67.0/ 70.0/ 68.0	49.0/ 55.0/ 56.0	53.0/ 61.0/ 60.0	45.0/ 42.0/ 47.0	52.0/ 54.0/ 54.0	53.0/ 58.0/ 60.0	28.0/ 29.0/ 28.0	48.0/ 45.0/ 49.0	48.0/ 54.0/ 56.0	39.0/ 43.0/ 48.0
Arts and Culture	50.3/ 53.9/ 59.4	70.2/ 78.2/ 77.0	52.0/ 62.0/ 55.0	54.9/ 60.5/ 61.3	52.8/ 59.5/ 63.8	54.0/ 48.0/ 49.0	57.9/ 59.6/ 64.5	30.0/ 25.0/ 24.0	40.0/ 53.0/ 51.0	45.5/ 50.3/ 49.1	45.7/ 47.2/ 51.3
Astronomy and Astrophysics	73.0/ 76.0/ 74.0	88.0/ 94.0/ 93.0	78.0/ 79.0/ 78.0	72.7/ 75.0/ 75.8	68.0/ 75.0/ 78.0	71.0/ 71.0/ 77.0	70.0/ 73.0/ 76.0	63.0/ 61.0/ 69.0	67.0/ 76.0/ 77.0	73.0/ 83.0/ 85.0	70.0/ 76.0/ 78.0
Biology	68.3/ 72.9/ 72.9	84.6/ 86.4/ 86.4	60.0/ 72.0/ 71.0	77.0/ 81.6/ 82.4	66.0/ 73.4/ 72.1	70.7/ 74.8/ 72.7	57.3/ 62.4/ 64.5	53.0/ 52.0/ 47.0	70.0/ 71.0/ 73.0	61.8/ 66.9/ 66.9	61.6/ 64.8/ 64.1
Business and Management	55.0/ 55.0/ 56.0	68.7/ 67.8/ 68.7	49.0/ 62.0/ 56.0	57.6/ 58.3/ 59.0	54.0/ 48.0/ 55.0	57.0/ 54.0/ 59.0	46.0/ 59.0/ 65.0	37.0/ 39.0/ 43.0	56.0/ 55.0/ 60.0	45.0/ 53.0/ 55.0	33.0/ 34.9/ 42.4
Chemistry	65.1/ 71.7/ 73.6	81.8/ 86.2/ 85.0	63.8/ 66.4/ 67.2	73.6/ 77.1/ 77.6	62.4/ 69.6/ 70.6	68.0/ 69.0/ 71.0	62.8/ 68.3/ 73.1	44.2/ 51.0/ 55.8	71.0/ 69.0/ 72.0	60.2/ 65.0/ 67.4	67.6/ 69.1/ 72.3
Computer Science	45.3/ 53.6/ 53.6	49.8/ 57.8/ 57.4	55.0/ 65.0/ 64.0	47.4/ 53.4/ 55.6	51.5/ 59.2/ 61.0	45.0/ 59.0/ 57.0	47.9/ 52.7/ 55.6	42.0/ 56.0/ 54.0	66.0/ 66.0/ 61.0	40.7/ 48.8/ 52.9	41.3/ 48.9/ 53.4
Defense and Security	60.0/ 61.0/ 69.0	81.0/ 86.0/ 86.0	50.0/ 61.0/ 60.0	55.5/ 58.3/ 63.5	51.0/ 60.0/ 66.0	47.0/ 54.0/ 61.0	52.0/ 65.0/ 65.0	37.0/ 38.0/ 43.0	49.0/ 67.0/ 66.0	58.0/ 62.0/ 60.0	53.0/ 66.0/ 69.0
Earth Sciences	58.0/ 67.0/ 66.0	77.0/ 83.0/ 83.4	55.0/ 54.0/ 55.0	60.7/ 63.7/ 67.7	69.0/ 71.0/ 76.0	59.0/ 64.0/ 64.0	63.0/ 65.0/ 63.0	40.4/ 37.4/ 41.4	54.0/ 62.0/ 63.0	64.0/ 64.0/ 67.0	62.0/ 66.0/ 64.0
Economics	63.3/ 69.1/ 74.8	74.5/ 79.3/ 79.1	55.0/ 56.0/ 57.0	64.7/ 68.2/ 69.8	64.0/ 72.0/ 74.0	53.0/ 54.0/ 54.0	52.8/ 56.3/ 54.2	41.0/ 38.0/ 42.0	59.0/ 60.0/ 57.0	57.1/ 59.5/ 63.5	55.7/ 55.7/ 62.3
Education	53.0/ 57.0/ 57.0	60.2/ 65.9/ 67.8	39.0/ 48.0/ 40.0	48.5/ 54.1/ 53.5	45.0/ 48.0/ 50.0	49.0/ 54.0/ 53.0	49.0/ 58.6/ 56.7	36.0/ 37.0/ 44.0	46.0/ 51.0/ 47.0	42.0/ 50.0/ 53.0	47.0/ 53.0/ 58.0
Energy and Power	59.0/ 63.0/ 71.0	72.3/ 80.4/ 83.8	65.0/ 67.0/ 70.0	61.0/ 69.2/ 72.5	62.0/ 73.0/ 70.0	63.0/ 75.0/ 76.0	64.0/ 63.0/ 65.0	42.0/ 53.0/ 56.0	58.0/ 58.0/ 61.0	54.0/ 59.0/ 60.0	63.0/ 67.0/ 66.0
Engineering	49.0/ 55.8/ 58.2	59.6/ 62.2/ 64.6	54.8/ 58.2/ 57.8	49.8/ 51.6/ 54.4	52.4/ 56.2/ 59.4	60.7/ 55.1/ 57.3	53.8/ 56.4/ 57.2	37.2/ 40.2/ 44.6	43.0/ 44.0/ 45.0	48.0/ 50.4/ 52.2	57.2/ 60.0/ 60.8
Environmental Science	52.0/ 67.0/ 64.0	75.0/ 78.0/ 79.4	56.0/ 62.0/ 57.0	65.7/ 68.3/ 69.3	58.6/ 59.4/ 66.4	58.0/ 60.0/ 57.0	54.2/ 65.3/ 66.3	41.4/ 42.4/ 48.5	57.0/ 64.0/ 64.0	51.0/ 55.0/ 58.0	53.1/ 53.6/ 58.7
Ethics and Human Rights	49.0/ 58.0/ 58.0	65.0/ 77.0/ 79.0	58.0/ 63.0/ 60.0	54.0/ 57.0/ 59.0	57.0/ 57.0/ 61.0	57.0/ 62.0/ 62.0	51.0/ 59.0/ 57.0	38.0/ 45.0/ 45.0	49.0/ 54.0/ 60.0	52.0/ 56.0/ 60.0	47.0/ 62.0/ 61.0
Finance and Investment	56.9/ 60.9/ 62.4	62.6/ 65.2/ 65.4	54.0/ 53.0/ 56.0	56.0/ 59.6/ 62.8	64.8/ 69.1/ 66.7	58.0/ 57.0/ 55.0	49.7/ 58.2/ 61.0	41.0/ 51.0/ 53.0	53.0/ 55.0/ 55.0	57.0/ 60.3/ 62.0	45.4/ 48.2/ 45.4
Food Science	56.0/ 66.0/ 70.0	82.0/ 82.0/ 81.0	59.0/ 66.0/ 68.0	69.7/ 74.8/ 75.8	67.0/ 64.0/ 64.0	64.0/ 67.0/ 71.0	65.0/ 72.0/ 72.0	47.0/ 52.0/ 54.0	64.0/ 67.0/ 72.0	63.0/ 69.0/ 66.0	63.0/ 67.0/ 76.0
Geography	59.2/ 60.9/ 67.6	68.6/ 74.4/ 75.6	60.0/ 63.0/ 57.0	58.6/ 64.9/ 64.5	58.5/ 61.2/ 66.2	53.5/ 56.6/ 54.5	54.4/ 56.8/ 58.0	45.0/ 40.0/ 49.0	47.0/ 56.0/ 58.0	54.3/ 57.1/ 59.4	51.6/ 56.8/ 59.9
Health and Medicine	71.0/ 72.0/ 73.0	83.0/ 87.8/ 87.6	65.0/ 70.0/ 70.0	75.1/ 76.5/ 80.6	71.0/ 75.0/ 76.0	71.0/ 71.0/ 75.0	58.0/ 70.5/ 69.6	53.0/ 58.0/ 58.0	70.0/ 73.0/ 70.0	62.1/ 64.7/ 62.9	65.4/ 68.2/ 66.4
History	57.9/ 61.6/ 68.8	66.5/ 70.9/ 70.3	53.0/ 62.0/ 61.0	61.9/ 64.9/ 62.1	63.0/ 65.2/ 70.0	51.0/ 58.0/ 57.0	49.4/ 52.8/ 53.5	31.0/ 41.0/ 46.0	52.0/ 60.0/ 61.0	55.6/ 62.2/ 61.8	48.9/ 53.7/ 54.0
Information Technology	79.0/ 75.0/ 80.0	83.0/ 84.0/ 86.0	67.0/ 79.0/ 77.0	75.4/ 75.4/ 82.1	77.0/ 79.0/ 83.0	77.0/ 81.0/ 76.0	69.0/ 69.0/ 68.0	64.0/ 64.0/ 62.0	73.0/ 72.0/ 74.0	67.0/ 69.0/ 73.0	67.0/ 80.0/ 78.0
International Relations	57.0/ 70.0/ 71.0	72.5/ 80.8/ 83.3	60.0/ 71.0/ 66.0	61.5/ 67.7/ 69.3	54.0/ 67.0/ 66.0	64.0/ 63.0/ 64.0	59.0/ 60.0/ 62.0	51.0/ 52.0/ 50.0	59.0/ 62.0/ 65.0	57.0/ 64.0/ 60.0	59.0/ 66.0/ 63.0
Language Studies	41.0/ 42.0/ 41.0	78.8/ 85.2/ 86.2	43.0/ 43.0/ 47.0	42.3/ 44.9/ 45.1	35.7/ 49.6/ 54.3	41.0/ 44.0/ 48.0	43.0/ 49.7/ 49.0	35.0/ 49.0/ 41.0	56.0/ 47.0/ 55.0	37.8/ 44.6/ 48.7	37.9/ 47.1/ 40.2
Law and Ethics	73.7/ 69.6/ 75.7	71.2/ 78.4/ 78.2	54.0/ 57.0/ 57.0	68.4/ 71.4/ 72.2	66.2/ 70.1/ 72.6	54.0/ 60.0/ 63.0	49.1/ 55.6/ 55.2	38.0/ 49.0/ 48.0	66.0/ 68.0/ 68.0	61.4/ 72.4/ 69.3	51.9/ 53.1/ 56.2
Literature and Linguistics	51.3/ 54.0/ 56.8	67.5/ 72.3/ 74.4	48.0/ 49.0/ 44.0	48.7/ 51.7/ 51.9	44.2/ 48.3/ 50.8	44.0/ 47.0/ 49.0	46.0/ 51.3/ 53.1	34.0/ 38.0/ 38.0	43.0/ 51.0/ 51.0	38.1/ 40.1/ 38.5	43.7/ 45.8/ 45.1
Logical Reasoning	44.5/ 48.4/ 49.6	44.0/ 49.8/ 49.8	41.6/ 46.8/ 46.1	43.7/ 45.9/ 47.7	43.1/ 44.4/ 45.6	35.0/ 39.0/ 36.0	37.6/ 38.2/ 40.1	34.0/ 33.0/ 29.0	35.0/ 42.0/ 38.0	37.9/ 42.2/ 42.6	44.8/ 46.0/ 44.3
Materials Science	62.0/ 61.0/ 68.0	71.5/ 76.8/ 76.8	44.0/ 53.0/ 49.0	56.1/ 58.3/ 61.8	55.0/ 54.0/ 58.0	55.0/ 60.0/ 56.0	58.0/ 51.0/ 62.0	37.0/ 30.0/ 40.0	56.0/ 59.0/ 58.0	52.0/ 58.0/ 59.0	53.0/ 59.0/ 59.0
Media and Communication	58.0/ 61.0/ 66.0	75.0/ 84.0/ 82.0	59.0/ 67.0/ 71.0	58.5/ 68.7/ 66.7	60.0/ 64.0/ 66.0	59.0/ 62.0/ 65.0	48.0/ 56.0/ 55.0	42.0/ 42.0/ 45.0	50.0/ 55.0/ 61.0	55.0/ 65.0/ 57.0	63.0/ 67.0/ 68.0
Music and Performing Arts	39.0/ 40.0/ 46.0	66.0/ 74.0/ 68.0	42.0/ 53.0/ 55.0	48.3/ 58.3/ 55.0	49.0/ 62.0/ 60.0	47.0/ 51.0/ 53.0	49.0/ 56.0/ 52.0	32.0/ 36.0/ 44.0	49.0/ 58.0/ 57.0	36.0/ 43.0/ 42.0	44.0/ 50.0/ 53.0
Physics	63.4/ 69.4/ 69.2	74.4/ 79.4/ 81.4	62.9/ 64.9/ 64.3	64.4/ 66.2/ 70.0	60.6/ 65.4/ 68.6	63.7/ 71.0/ 71.6	60.8/ 68.6/ 68.4	43.4/ 47.2/ 50.7	63.0/ 60.0/ 62.0	57.6/ 63.0/ 64.0	60.8/ 62.8/ 66.2
Politics and Governance	57.2/ 65.5/ 70.2	63.9/ 71.1/ 70.7	48.0/ 60.0/ 57.0	60.2/ 63.4/ 65.1	59.5/ 69.2/ 72.5	47.0/ 51.0/ 51.0	48.9/ 52.0/ 53.6	37.0/ 44.0/ 43.0	47.0/ 56.0/ 56.0	57.3/ 63.5/ 63.9	45.1/ 50.9/ 52.6
Psychology	58.0/ 61.0/ 59.0	81.0/ 92.0/ 89.0	55.0/ 54.0/ 58.0	57.4/ 65.1/ 65.1	60.0/ 65.0/ 64.0	60.0/ 61.0/ 60.0	54.0/ 56.0/ 60.0	53.0/ 48.0/ 43.0	62.0/ 66.0/ 65.0	53.0/ 59.0/ 60.0	70.0/ 76.0/ 71.0
Public Administration	47.0/ 41.0/ 46.0	64.6/ 70.7/ 66.7	50.5/ 58.6/ 60.6	55.6/ 54.5/ 53.5	41.4/ 42.4/ 43.4	47.0/ 49.0/ 49.0	40.4/ 46.5/ 46.5	39.0/ 43.0/ 42.0	49.5/ 63.6/ 62.6	47.0/ 44.0/ 45.0	47.0/ 53.0/ 51.0
Religion and Spirituality	58.0/ 64.0/ 64.0	73.0/ 82.0/ 88.0	60.0/ 65.0/ 66.0	63.7/ 70.6/ 68.5	66.0/ 70.0/ 69.0	61.0/ 67.0/ 68.0	62.0/ 64.0/ 67.0	36.0/ 39.0/ 37.0	68.0/ 71.0/ 69.0	64.0/ 63.0/ 62.0	64.0/ 61.0/ 69.0
Social Welfare and Development	57.0/ 56.0/ 58.0	61.6/ 73.2/ 67.9	53.0/ 61.0/ 64.0	56.5/ 62.2/ 64.8	45.0/ 49.0/ 54.0	47.0/ 51.0/ 51.0	43.0/ 56.0/ 55.0	40.0/ 35.0/ 40.0	57.0/ 58.0/ 60.0	47.0/ 56.0/ 53.0	56.0/ 52.0/ 49.0
Sociology	42.6/ 46.1/ 40.9	51.2/ 56.4/ 57.8	33.0/ 40.0/ 43.0	44.8/ 46.0/ 48.8	39.0/ 44.0/ 47.0	40.0/ 40.0/ 46.0	51.6/ 47.8/ 52.8	34.0/ 30.0/ 38.0	38.0/ 42.0/ 45.0	41.4/ 46.0/ 42.3	46.5/ 47.6/ 48.8
Sports and Recreation	56.4/ 68.3/ 69.8	74.2/ 82.0/ 83.0	47.0/ 64.0/ 60.0	57.6/ 63.8/ 69.4	54.5/ 62.9/ 68.0	54.0/ 50.0/ 57.0	65.0/ 69.5/ 71.2	38.0/ 36.0/ 35.0	46.0/ 65.0/ 65.0	55.1/ 59.6/ 59.6	52.4/ 59.1/ 61.4
Technology and Innovation	63.0/ 67.0/ 73.0	76.2/ 79.4/ 78.6	61.0/ 69.0/ 63.0	64.0/ 69.2/ 69.0	57.0/ 63.0/ 64.0	58.0/ 68.0/ 68.0	63.0/ 65.0/ 69.0	47.0/ 48.0/ 48.0	53.0/ 60.0/ 61.0	56.0/ 51.0/ 54.0	52.0/ 56.0/ 56.0
Transportation and Logistics	37.0/ 45.0/ 53.0	59.2/ 70.0/ 68.5	45.0/ 59.0/ 57.0	49.5/ 53.0/ 55.5	43.4/ 40.4/ 37.4	46.0/ 54.0/ 55.0	42.9/ 41.8/ 45.9	31.3/ 31.3/ 31.3	45.0/ 48.0/ 48.0	49.0/ 57.0/ 58.0	44.0/ 44.0/ 42.0

Table 48: Detailed subject-wise evaluation for GOOGLE/GEMMA-2-27B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	43.0/ 51.0/ 56.0	60.2/ 68.6/ 69.8	41.4/ 49.5/ 50.5	43.1/ 50.7/ 58.6	45.0/ 42.0/ 47.0	44.4/ 43.4/ 45.5	38.0/ 39.0/ 42.0	40.0/ 34.0/ 38.0	42.0/ 44.0/ 51.0	38.0/ 42.0/ 43.0	47.0/ 39.0/ 52.0
Anthropology	40.0/ 52.0/ 52.0	60.0/ 66.0/ 68.0	49.0/ 50.0/ 51.0	48.5/ 62.6/ 53.5	37.0/ 46.0/ 50.0	44.0/ 51.0/ 53.0	50.0/ 48.0/ 60.0	33.0/ 36.0/ 38.0	40.0/ 50.0/ 53.0	37.0/ 46.0/ 47.0	40.0/ 45.0/ 51.0
Architecture and Design	36.0/ 39.0/ 46.0	63.0/ 69.0/ 73.0	43.0/ 50.0/ 49.0	41.0/ 40.0/ 56.0	35.0/ 35.0/ 40.0	44.0/ 40.0/ 50.0	46.0/ 48.0/ 57.0	29.0/ 28.0/ 32.0	43.0/ 41.0/ 42.0	36.0/ 33.0/ 42.0	32.0/ 38.0/ 37.0
Arts and Culture	44.2/ 47.9/ 52.7	67.2/ 73.8/ 76.8	44.0/ 46.0/ 47.0	43.7/ 51.9/ 57.5	41.7/ 41.7/ 54.0	40.0/ 44.0/ 42.0	32.2/ 48.1/ 52.5	26.0/ 27.0/ 29.0	36.0/ 46.0/ 48.0	43.6/ 43.6/ 47.9	37.7/ 36.7/ 45.7
Astronomy and Astrophysics	58.0/ 64.0/ 73.0	75.0/ 93.0/ 91.0	55.0/ 71.0/ 79.0	52.3/ 66.4/ 72.7	60.0/ 66.0/ 73.0	59.0/ 53.0/ 61.0	53.0/ 60.0/ 77.0	57.0/ 55.0/ 68.0	54.0/ 65.0/ 75.0	54.0/ 61.0/ 74.0	64.0/ 62.0/ 74.0
Biology	53.9/ 62.7/ 68.0	72.8/ 86.6/ 86.4	48.0/ 49.0/ 53.0	54.1/ 66.3/ 81.4	41.4/ 48.8/ 62.3	44.4/ 55.6/ 60.6	43.0/ 49.5/ 58.2	34.0/ 38.0/ 52.0	54.0/ 62.0/ 74.0	41.4/ 55.1/ 61.5	46.2/ 52.8/ 64.8
Business and Management	47.0/ 52.0/ 56.0	57.0/ 62.8/ 69.0	40.0/ 46.0/ 47.0	44.0/ 49.9/ 57.9	34.0/ 44.0/ 54.0	40.0/ 45.0/ 52.0	46.0/ 45.0/ 52.0	38.0/ 39.0/ 45.0	40.0/ 48.0/ 55.0	46.0/ 44.0/ 56.0	34.9/ 34.9/ 47.2
Chemistry	51.9/ 62.1/ 70.9	69.9/ 82.6/ 85.4	49.1/ 58.6/ 69.0	57.9/ 66.5/ 77.1	48.7/ 49.4/ 65.0	49.0/ 57.0/ 70.0	45.1/ 54.4/ 67.5	46.2/ 45.2/ 58.7	50.0/ 59.0/ 66.0	46.0/ 58.2/ 65.0	47.9/ 53.6/ 70.1
Computer Science	39.1/ 47.4/ 57.3	46.8/ 55.2/ 58.4	44.0/ 55.0/ 65.0	42.0/ 50.0/ 51.0	42.0/ 51.5/ 61.0	31.0/ 48.0/ 55.0	41.4/ 47.3/ 53.8	44.0/ 49.0/ 55.0	39.0/ 43.0/ 53.0	40.1/ 50.0/ 61.1	36.3/ 44.8/ 49.8
Defense and Security	46.0/ 57.0/ 66.0	74.0/ 79.0/ 85.0	43.0/ 56.0/ 54.0	43.6/ 46.9/ 53.5	47.0/ 50.0/ 60.0	45.0/ 46.0/ 52.0	44.0/ 44.0/ 53.0	36.0/ 36.0/ 40.0	34.0/ 46.0/ 57.0	49.0/ 53.0/ 53.0	50.0/ 46.0/ 60.0
Earth Sciences	51.0/ 53.0/ 65.0	61.4/ 79.3/ 81.1	34.0/ 56.0/ 58.0	47.3/ 52.7/ 62.9	44.0/ 58.0/ 71.0	37.0/ 52.0/ 56.0	52.0/ 57.0/ 59.0	34.3/ 34.3/ 36.4	39.0/ 44.0/ 52.0	50.0/ 52.0/ 59.0	51.0/ 49.0/ 60.0
Economics	43.2/ 54.7/ 69.1	57.0/ 72.1/ 74.9	40.0/ 53.0/ 59.0	49.8/ 57.3/ 67.3	47.0/ 51.0/ 63.0	36.0/ 41.0/ 51.0	43.7/ 44.4/ 52.1	43.0/ 38.0/ 47.0	42.0/ 54.0/ 55.0	46.0/ 54.8/ 57.9	47.3/ 50.9/ 56.3
Education	43.0/ 43.0/ 54.0	50.3/ 60.2/ 63.1	34.0/ 46.0/ 51.0	36.4/ 42.0/ 49.4	40.0/ 36.0/ 44.0	45.0/ 44.0/ 51.0	42.0/ 42.0/ 44.6	30.0/ 33.0/ 40.0	38.0/ 48.0/ 47.0	39.0/ 41.0/ 47.0	51.0/ 40.0/ 50.0
Energy and Power	46.0/ 55.0/ 62.0	62.2/ 78.4/ 81.8	49.0/ 59.0/ 65.0	47.1/ 53.9/ 68.5	43.0/ 45.0/ 54.0	54.0/ 47.0/ 59.0	41.0/ 54.0/ 61.0	33.0/ 33.0/ 51.0	41.0/ 51.0/ 57.0	45.0/ 45.0/ 60.0	42.0/ 51.0/ 59.0
Engineering	39.0/ 46.0/ 53.2	50.6/ 57.8/ 62.8	41.0/ 46.0/ 52.0	41.4/ 41.0/ 52.4	42.0/ 43.8/ 52.4	41.4/ 45.7/ 51.7	40.4/ 43.0/ 55.8	31.6/ 32.7/ 45.8	38.0/ 42.0/ 52.0	40.2/ 42.0/ 45.6	39.4/ 45.8/ 59.6
Environmental Science	47.0/ 48.0/ 62.0	63.1/ 73.6/ 79.0	49.0/ 54.0/ 58.0	52.1/ 57.5/ 67.9	43.0/ 54.7/ 59.4	49.0/ 52.0/ 60.0	45.8/ 47.9/ 54.7	34.3/ 38.4/ 46.5	54.0/ 65.0/ 71.0	46.0/ 48.0/ 52.0	46.4/ 46.9/ 48.6
Ethics and Human Rights	45.0/ 45.0/ 48.0	54.0/ 70.0/ 76.0	43.0/ 48.0/ 60.0	49.0/ 53.0/ 57.0	45.0/ 42.0/ 50.0	47.0/ 49.0/ 52.0	43.0/ 43.0/ 50.0	37.0/ 33.0/ 35.0	48.0/ 41.0/ 49.0	41.0/ 45.0/ 49.0	51.0/ 47.0/ 51.0
Finance and Investment	42.1/ 52.0/ 59.4	56.6/ 65.2/ 66.8	51.0/ 54.0/ 61.0	44.6/ 48.0/ 58.8	54.3/ 56.2/ 67.9	46.0/ 50.0/ 57.0	45.8/ 48.0/ 56.5	39.0/ 42.0/ 44.0	40.0/ 41.0/ 50.0	45.2/ 48.6/ 57.5	47.5/ 43.9/ 48.9
Food Science	53.0/ 58.0/ 68.0	72.0/ 83.0/ 84.0	48.0/ 57.0/ 62.0	60.6/ 75.8/ 75.8	47.0/ 51.0/ 66.0	51.0/ 55.0/ 66.0	45.0/ 69.0/ 71.0	41.0/ 37.0/ 48.0	45.0/ 49.0/ 62.0	42.0/ 67.0/ 68.0	45.0/ 59.0/ 66.0
Geography	39.1/ 51.4/ 62.6	64.4/ 72.0/ 74.2	44.0/ 66.0/ 62.0	47.8/ 56.4/ 65.1	47.5/ 50.2/ 61.6	39.4/ 39.4/ 49.5	39.1/ 47.3/ 53.8	44.0/ 44.0/ 49.0	39.0/ 49.0/ 56.0	41.1/ 52.6/ 56.0	41.7/ 47.4/ 54.2
Health and Medicine	51.0/ 64.0/ 72.0	70.9/ 83.4/ 85.6	43.0/ 60.0/ 62.0	56.1/ 66.5/ 76.8	48.0/ 53.0/ 64.0	58.0/ 64.0/ 72.0	39.3/ 49.1/ 55.4	41.0/ 48.0/ 54.0	54.0/ 61.0/ 70.0	38.8/ 47.4/ 58.6	47.7/ 57.0/ 62.6
History	52.5/ 54.4/ 61.1	58.4/ 67.9/ 68.9	48.0/ 55.0/ 54.0	49.4/ 55.6/ 61.5	51.3/ 54.2/ 64.5	45.0/ 42.0/ 49.0	39.8/ 45.4/ 50.9	29.0/ 31.0/ 42.0	37.0/ 46.0/ 52.0	45.5/ 56.7/ 58.2	46.3/ 47.6/ 50.9
Information Technology	60.0/ 71.0/ 80.0	67.0/ 81.0/ 81.0	56.0/ 72.0/ 72.0	63.1/ 65.4/ 77.1	52.0/ 51.0/ 74.0	59.0/ 61.0/ 68.0	52.0/ 61.0/ 69.0	44.0/ 48.0/ 58.0	51.0/ 59.0/ 71.0	54.0/ 60.0/ 69.0	51.0/ 64.0/ 76.0
International Relations	51.0/ 57.0/ 67.0	66.7/ 78.3/ 80.8	50.0/ 63.0/ 60.0	48.6/ 56.4/ 63.8	42.0/ 52.0/ 60.0	44.0/ 46.0/ 52.0	46.0/ 57.0/ 60.0	47.0/ 45.0/ 56.0	49.0/ 49.0/ 54.0	54.0/ 54.0/ 64.0	43.0/ 45.0/ 49.0
Language Studies	24.0/ 39.0/ 51.0	59.0/ 77.8/ 83.6	32.0/ 41.0/ 49.0	39.7/ 40.5/ 48.1	27.9/ 26.4/ 43.4	29.0/ 26.0/ 42.0	30.2/ 28.9/ 41.6	41.0/ 40.0/ 47.0	50.0/ 53.0/ 54.0	34.5/ 40.5/ 50.9	32.8/ 37.9/ 39.7
Law and Ethics	58.1/ 63.5/ 73.0	61.2/ 73.6/ 74.8	50.0/ 56.0/ 54.0	52.4/ 57.2/ 63.0	51.0/ 51.0/ 65.6	43.0/ 51.0/ 52.0	41.8/ 43.5/ 46.6	29.0/ 36.0/ 36.0	63.0/ 63.0/ 65.0	55.9/ 54.3/ 59.8	38.8/ 39.4/ 48.8
Literature and Linguistics	40.5/ 46.0/ 50.4	58.5/ 69.3/ 70.3	37.0/ 48.0/ 48.0	39.5/ 44.5/ 51.3	36.7/ 39.2/ 50.0	33.0/ 28.0/ 39.0	38.9/ 44.2/ 45.1	38.0/ 42.0/ 39.0	39.0/ 44.0/ 42.0	38.9/ 37.2/ 38.5	30.3/ 33.8/ 40.8
Logical Reasoning	36.6/ 40.3/ 48.2	41.2/ 45.8/ 56.0	31.8/ 37.0/ 42.9	37.7/ 39.9/ 44.7	42.7/ 41.1/ 48.5	45.0/ 42.0/ 44.0	41.2/ 40.1/ 46.5	38.0/ 39.0/ 41.0	39.0/ 36.0/ 44.0	44.2/ 38.6/ 47.8	38.4/ 38.4/ 44.8
Materials Science	45.0/ 53.0/ 63.0	59.3/ 74.1/ 74.5	35.0/ 40.0/ 51.0	48.2/ 48.5/ 55.0	37.0/ 47.0/ 60.0	49.0/ 49.0/ 61.0	44.0/ 48.0/ 58.0	35.0/ 33.0/ 43.0	38.0/ 47.0/ 56.0	34.0/ 50.0/ 53.0	36.0/ 46.0/ 57.0
Media and Communication	50.0/ 54.0/ 59.0	68.0/ 81.0/ 79.0	46.0/ 55.0/ 59.0	47.6/ 59.2/ 68.0	32.0/ 45.0/ 64.0	47.0/ 45.0/ 54.0	34.0/ 44.0/ 52.0	30.0/ 34.0/ 49.0	50.0/ 47.0/ 55.0	41.0/ 44.0/ 52.0	52.0/ 55.0/ 56.0
Music and Performing Arts	40.0/ 42.0/ 48.0	55.0/ 60.0/ 62.0	44.0/ 41.0/ 44.0	37.8/ 39.1/ 47.7	37.0/ 33.0/ 44.0	32.0/ 29.0/ 38.0	40.0/ 41.0/ 49.0	28.0/ 30.0/ 29.0	50.0/ 49.0/ 47.0	30.0/ 40.0/ 38.0	36.0/ 36.0/ 52.0
Physics	46.0/ 55.8/ 65.6	62.0/ 76.0/ 79.6	44.0/ 48.8/ 61.0	51.0/ 58.0/ 68.4	44.0/ 49.0/ 66.4	46.3/ 48.9/ 63.2	43.6/ 49.0/ 67.6	37.1/ 37.9/ 49.6	44.0/ 46.0/ 57.0	44.0/ 49.8/ 59.8	45.2/ 47.8/ 63.0
Politics and Governance	49.4/ 60.4/ 69.0	56.7/ 66.7/ 67.9	38.0/ 42.0/ 45.0	43.8/ 53.4/ 61.2	43.7/ 47.8/ 61.5	42.0/ 46.0/ 43.0	39.3/ 42.4/ 44.7	38.0/ 35.0/ 42.0	43.0/ 41.0/ 53.0	43.1/ 50.2/ 60.6	43.9/ 44.6/ 47.4
Psychology	45.0/ 49.0/ 60.0	76.0/ 89.0/ 89.0	43.0/ 49.0/ 55.0	38.0/ 58.1/ 60.5	55.0/ 53.0/ 71.0	41.0/ 53.0/ 66.0	44.0/ 47.0/ 52.0	44.0/ 43.0/ 56.0	40.0/ 58.0/ 69.0	38.0/ 52.0/ 51.0	44.0/ 64.0/ 71.0
Public Administration	38.0/ 41.0/ 48.0	51.5/ 63.6/ 65.7	40.4/ 44.4/ 40.4	50.5/ 52.5/ 50.5	39.4/ 34.3/ 42.4	39.0/ 37.0/ 43.0	29.3/ 33.3/ 35.4	32.0/ 33.0/ 35.0	38.4/ 50.5/ 47.5	42.0/ 37.0/ 44.0	45.0/ 39.0/ 41.0
Religion and Spirituality	47.0/ 61.0/ 69.0	71.0/ 82.0/ 84.0	48.0/ 62.0/ 66.0	52.0/ 56.9/ 66.9	53.0/ 54.0/ 69.0	51.0/ 46.0/ 62.0	50.0/ 54.0/ 68.0	41.0/ 41.0/ 47.0	45.0/ 59.0/ 62.0	49.0/ 61.0/ 68.0	45.0/ 54.0/ 60.0
Social Welfare and Development	41.0/ 50.0/ 47.0	50.9/ 56.2/ 67.0	47.0/ 52.0/ 54.0	49.2/ 52.8/ 62.2	46.0/ 44.0/ 50.0	42.0/ 45.0/ 50.0	41.0/ 39.0/ 48.0	38.0/ 39.0/ 46.0	37.0/ 51.0/ 59.0	42.0/ 46.0/ 56.0	39.0/ 48.0/ 55.0
Sociology	34.8/ 32.2/ 40.0	44.4/ 56.8/ 56.8	27.0/ 27.0/ 32.0	37.8/ 38.0/ 46.0	24.0/ 35.0/ 41.0	29.0/ 39.0/ 38.0	37.7/ 39.6/ 43.4	34.0/ 35.0/ 39.0	28.0/ 38.0/ 42.0	32.4/ 36.9/ 48.6	31.2/ 41.2/ 41.2
Sports and Recreation	40.1/ 43.6/ 56.9	67.8/ 78.6/ 81.2	37.0/ 57.0/ 59.0	46.4/ 56.6/ 64.6	36.5/ 41.6/ 56.7	43.0/ 42.0/ 49.0	45.2/ 50.3/ 58.2	32.0/ 25.0/ 35.0	34.0/ 40.0/ 49.0	44.2/ 49.4/ 58.3	39.1/ 47.1/ 49.5
Technology and Innovation	55.0/ 52.0/ 65.0	63.6/ 74.0/ 79.4	49.0/ 50.0/ 54.0	47.0/ 53.8/ 66.6	47.0/ 47.0/ 56.0	45.0/ 52.0/ 60.0	55.0/ 54.0/ 63.0	46.0/ 45.0/ 54.0	43.0/ 57.0/ 55.0	34.0/ 36.0/ 55.0	44.0/ 42.0/ 58.0
Transportation and Logistics	45.0/ 35.0/ 38.0	56.1/ 62.3/ 68.5	40.0/ 36.0/ 49.0	43.5/ 45.7/ 54.9	32.3/ 29.3/ 41.4	45.0/ 42.0/ 42.0	37.8/ 44.9/ 45.9	31.3/ 33.3/ 37.4	46.0/ 48.0/ 49.0	40.0/ 50.0/ 55.0	41.0/ 33.0/ 39.0

Table 49: Detailed subject-wise evaluation for GOOGLE/GEMMA-2-27B-IT on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	29.0/ 29.0/ 31.0	49.1/ 48.1/ 48.9	32.3/ 26.3/ 30.3	39.4/ 41.4/ 42.2	37.0/ 29.0/ 30.0	29.3/ 38.4/ 34.3	30.0/ 33.0/ 34.0	30.0/ 29.0/ 33.0	31.0/ 25.0/ 27.0	26.0/ 26.0/ 34.0	24.0/ 27.0/ 25.0
Anthropology	35.0/ 37.0/ 36.0	50.0/ 54.0/ 49.0	35.0/ 37.0/ 34.0	44.4/ 40.4/ 44.4	36.0/ 36.0/ 42.0	32.0/ 36.0/ 36.0	36.0/ 43.0/ 41.0	38.0/ 37.0/ 39.0	35.0/ 32.0/ 32.0	43.0/ 41.0/ 43.0	37.0/ 36.0/ 35.0
Architecture and Design	26.0/ 31.0/ 31.0	48.0/ 51.0/ 50.0	24.0/ 22.0/ 24.0	41.0/ 40.0/ 40.0	35.0/ 37.0/ 32.0	31.0/ 34.0/ 31.0	35.0/ 32.0/ 39.0	24.0/ 26.0/ 28.0	22.0/ 23.0/ 22.0	29.0/ 36.0/ 36.0	34.0/ 24.0/ 29.0
Arts and Culture	23.6/ 24.9/ 23.6	53.6/ 55.6/ 54.0	35.0/ 35.0/ 38.0	43.9/ 39.5/ 40.5	35.6/ 31.3/ 33.1	34.0/ 29.0/ 29.0	37.2/ 35.5/ 38.2	26.0/ 27.0/ 22.0	28.0/ 28.0/ 30.0	33.3/ 29.1/ 27.9	27.1/ 21.1/ 21.6
Astronomy and Astrophysics	40.0/ 35.0/ 40.0	66.0/ 72.0/ 65.0	33.0/ 34.0/ 32.0	50.8/ 49.2/ 50.0	35.0/ 34.0/ 32.0	43.0/ 36.0/ 39.0	48.0/ 48.0/ 51.0	37.0/ 33.0/ 37.0	27.0/ 26.0/ 29.0	38.0/ 39.0/ 38.0	25.0/ 30.0/ 22.0
Biology	27.5/ 26.4/ 23.9	61.3/ 60.3/ 60.3	32.0/ 30.0/ 25.0	45.7/ 48.1/ 49.3	31.1/ 32.0/ 32.0	15.2/ 18.2/ 24.2	33.7/ 34.3/ 35.8	30.0/ 24.0/ 28.0	24.0/ 29.0/ 24.0	33.1/ 30.6/ 31.5	26.4/ 23.9/ 28.3
Business and Management	34.0/ 29.0/ 29.0	48.0/ 53.2/ 58.2	33.0/ 35.0/ 40.0	41.9/ 41.2/ 41.7	33.0/ 26.0/ 27.0	37.0/ 39.0/ 44.0	34.0/ 33.0/ 43.0	32.0/ 30.0/ 28.0	32.0/ 30.0/ 35.0	40.0/ 46.0/ 49.0	30.2/ 26.4/ 27.4
Chemistry	28.8/ 31.3/ 29.7	58.1/ 61.3/ 60.3	28.4/ 29.3/ 30.2	39.5/ 42.7/ 47.7	29.1/ 29.7/ 30.7	28.0/ 28.0/ 28.0	34.6/ 37.7/ 40.9	29.8/ 25.0/ 29.8	25.0/ 24.0/ 28.0	29.4/ 30.3/ 31.4	29.7/ 27.4/ 30.4
Computer Science	35.4/ 32.3/ 33.3	42.8/ 40.8/ 40.0	32.0/ 29.0/ 28.0	34.4/ 35.8/ 38.4	29.6/ 29.0/ 29.0	34.0/ 30.0/ 28.0	34.3/ 30.2/ 32.5	37.0/ 40.0/ 34.0	28.0/ 27.0/ 34.0	31.4/ 32.6/ 34.9	31.4/ 30.9/ 30.9
Defense and Security	32.0/ 40.0/ 36.0	58.0/ 64.0/ 63.0	30.0/ 32.0/ 25.0	39.3/ 40.8/ 42.6	28.0/ 31.0/ 35.0	28.0/ 33.0/ 28.0	31.0/ 38.0/ 36.0	30.0/ 30.0/ 32.0	33.0/ 37.0/ 30.0	26.0/ 28.0/ 25.0	27.0/ 30.0/ 27.0
Earth Sciences	43.0/ 43.0/ 43.0	56.1/ 58.0/ 57.3	22.0/ 28.0/ 30.0	38.9/ 40.3/ 42.3	32.0/ 31.0/ 32.0	36.0/ 42.0/ 35.0	33.0/ 30.0/ 31.0	30.3/ 27.3/ 25.2	26.0/ 21.0/ 18.0	32.0/ 39.0/ 41.0	22.0/ 29.0/ 28.0
Economics	36.0/ 35.2/ 35.2	50.4/ 52.2/ 52.4	39.0/ 35.0/ 29.0	43.8/ 42.5/ 46.6	37.0/ 35.0/ 37.0	31.0/ 28.0/ 32.0	33.1/ 33.8/ 38.0	30.0/ 28.0/ 36.0	27.0/ 19.0/ 24.0	32.5/ 39.7/ 30.9	30.5/ 27.0/ 32.9
Education	37.0/ 42.0/ 43.0	43.6/ 48.4/ 45.9	28.0/ 32.0/ 33.0	33.1/ 37.3/ 38.2	31.0/ 30.0/ 30.0	26.0/ 31.0/ 31.0	40.1/ 37.6/ 39.5	29.0/ 33.0/ 33.0	28.0/ 31.0/ 28.0	40.0/ 42.0/ 43.0	36.0/ 30.0/ 37.0
Energy and Power	27.0/ 33.0/ 31.0	56.1/ 58.1/ 57.4	34.0/ 33.0/ 30.0	39.0/ 42.0/ 43.7	25.0/ 20.0/ 24.0	37.0/ 38.0/ 37.0	38.0/ 41.0/ 46.0	29.0/ 34.0/ 32.0	24.0/ 27.0/ 27.0	24.0/ 27.0/ 25.0	28.0/ 32.0/ 35.0
Engineering	30.2/ 29.8/ 31.8	41.0/ 41.4/ 42.4	25.0/ 27.4/ 27.6	32.2/ 35.4/ 35.8	25.4/ 25.8/ 26.8	24.8/ 26.9/ 25.2	31.2/ 32.6/ 30.6	22.3/ 27.4/ 28.6	21.0/ 21.0/ 18.0	30.2/ 33.4/ 32.2	29.0/ 27.4/ 28.6
Environmental Science	31.0/ 31.0/ 30.0	54.7/ 59.5/ 59.3	25.0/ 34.0/ 33.0	44.3/ 43.3/ 46.5	33.6/ 28.1/ 28.1	39.0/ 33.0/ 35.0	32.1/ 35.8/ 34.7	30.3/ 26.3/ 27.3	34.0/ 31.0/ 35.0	42.0/ 37.0/ 44.0	31.8/ 34.1/ 34.1
Ethics and Human Rights	45.0/ 43.0/ 46.0	48.0/ 52.0/ 56.0	37.0/ 35.0/ 35.0	42.0/ 50.0/ 48.0	29.0/ 33.0/ 30.0	42.0/ 39.0/ 41.0	41.0/ 45.0/ 49.0	44.0/ 38.0/ 39.0	26.0/ 17.0/ 19.0	35.0/ 37.0/ 36.0	32.0/ 32.0/ 36.0
Finance and Investment	34.6/ 40.6/ 38.1	38.0/ 41.6/ 43.0	27.0/ 31.0/ 31.0	33.6/ 36.6/ 38.2	31.5/ 36.4/ 38.3	34.0/ 30.0/ 29.0	26.6/ 30.5/ 29.9	35.0/ 31.0/ 33.0	30.0/ 30.0/ 31.0	40.2/ 41.3/ 39.1	30.7/ 33.9/ 37.1
Food Science	30.0/ 31.0/ 36.0	56.0/ 63.0/ 60.0	27.0/ 27.0/ 31.0	60.6/ 52.5/ 60.6	30.0/ 30.0/ 29.0	39.0/ 37.0/ 33.0	37.0/ 47.0/ 43.0	23.0/ 32.0/ 34.0	18.0/ 29.0/ 25.0	33.0/ 39.0/ 36.0	24.0/ 33.0/ 25.0
Geography	31.8/ 34.6/ 33.0	53.8/ 55.4/ 56.2	27.0/ 29.0/ 30.0	38.6/ 42.2/ 38.4	31.5/ 29.7/ 32.4	25.2/ 27.3/ 25.2	34.9/ 38.5/ 37.3	35.0/ 37.0/ 32.0	25.0/ 30.0/ 29.0	33.1/ 33.7/ 32.6	26.6/ 24.0/ 27.1
Health and Medicine	36.0/ 31.0/ 35.0	62.1/ 63.7/ 63.1	33.0/ 31.0/ 32.0	49.5/ 51.3/ 54.1	29.0/ 33.0/ 29.0	32.0/ 36.0/ 32.0	30.4/ 37.5/ 38.4	33.0/ 25.0/ 34.0	23.0/ 27.0/ 22.0	46.6/ 40.5/ 45.7	32.7/ 26.2/ 33.6
History	29.1/ 28.0/ 31.7	45.6/ 49.4/ 50.0	31.0/ 28.0/ 30.0	44.0/ 45.4/ 44.8	28.6/ 30.4/ 34.1	31.0/ 30.0/ 40.0	35.7/ 34.9/ 36.4	33.0/ 27.0/ 29.0	26.0/ 17.0/ 22.0	36.0/ 38.6/ 36.4	28.1/ 28.4/ 29.2
Information Technology	39.0/ 40.0/ 44.0	57.0/ 56.0/ 62.0	32.0/ 36.0/ 40.0	52.5/ 55.3/ 58.7	33.0/ 31.0/ 28.0	37.0/ 41.0/ 43.0	54.0/ 51.0/ 58.0	40.0/ 41.0/ 39.0	34.0/ 36.0/ 36.0	42.0/ 40.0/ 43.0	39.0/ 40.0/ 39.0
International Relations	37.0/ 35.0/ 40.0	64.2/ 67.5/ 61.7	39.0/ 41.0/ 39.0	49.8/ 49.8/ 52.9	28.0/ 31.0/ 32.0	30.0/ 33.0/ 30.0	34.0/ 40.0/ 37.0	34.0/ 32.0/ 31.0	27.0/ 26.0/ 27.0	39.0/ 38.0/ 36.0	25.0/ 22.0/ 21.0
Language Studies	28.0/ 23.0/ 26.0	60.8/ 64.0/ 62.6	28.0/ 20.0/ 20.0	35.5/ 38.9/ 35.1	30.2/ 29.5/ 29.5	22.0/ 22.0/ 24.0	29.5/ 25.5/ 24.2	34.0/ 36.0/ 36.0	34.0/ 39.0/ 36.0	28.1/ 28.1/ 28.1	29.9/ 24.1/ 25.9
Law and Ethics	46.0/ 42.6/ 43.2	54.2/ 58.8/ 57.6	37.0/ 30.0/ 30.0	49.6/ 51.8/ 53.4	31.9/ 25.5/ 26.8	37.0/ 25.0/ 25.0	37.9/ 39.7/ 40.5	28.0/ 31.0/ 29.0	36.0/ 33.0/ 34.0	42.5/ 40.2/ 38.6	28.1/ 28.7/ 33.1
Literature and Linguistics	27.9/ 33.3/ 28.8	49.7/ 49.5/ 50.3	29.0/ 29.0/ 24.0	38.7/ 38.5/ 39.1	27.5/ 32.5/ 30.8	35.0/ 24.0/ 28.0	41.6/ 40.7/ 47.8	30.0/ 23.0/ 29.0	23.0/ 19.0/ 23.0	27.1/ 25.9/ 28.3	28.9/ 23.9/ 26.1
Logical Reasoning	32.2/ 32.4/ 32.2	33.4/ 32.6/ 35.4	22.7/ 26.0/ 22.7	31.5/ 29.9/ 30.7	31.1/ 29.0/ 28.2	34.0/ 28.0/ 31.0	26.5/ 27.0/ 28.1	25.0/ 24.0/ 25.0	28.0/ 27.0/ 26.0	35.1/ 33.5/ 33.1	26.2/ 27.5/ 27.5
Materials Science	35.0/ 28.0/ 24.0	54.4/ 52.5/ 54.0	21.0/ 22.0/ 27.0	41.9/ 40.4/ 46.7	24.0/ 30.0/ 29.0	31.0/ 29.0/ 34.0	35.0/ 41.0/ 42.0	24.0/ 30.0/ 31.0	34.0/ 26.0/ 24.0	29.0/ 32.0/ 36.0	22.0/ 23.0/ 23.0
Media and Communication	36.0/ 33.0/ 39.0	58.0/ 68.0/ 63.0	34.0/ 40.0/ 40.0	41.5/ 38.1/ 47.6	37.0/ 40.0/ 43.0	51.0/ 46.0/ 48.0	36.0/ 41.0/ 45.0	31.0/ 28.0/ 33.0	22.0/ 35.0/ 33.0	38.0/ 38.0/ 38.0	37.0/ 36.0/ 34.0
Music and Performing Arts	29.0/ 27.0/ 27.0	46.0/ 52.0/ 51.0	39.0/ 30.0/ 33.0	38.4/ 34.4/ 32.5	28.0/ 30.0/ 32.0	35.0/ 36.0/ 37.0	41.0/ 40.0/ 41.0	35.0/ 22.0/ 32.0	25.0/ 31.0/ 28.0	27.0/ 27.0/ 27.0	29.0/ 25.0/ 28.0
Physics	32.8/ 31.8/ 33.2	50.4/ 53.6/ 53.0	29.5/ 31.5/ 32.0	38.4/ 40.8/ 42.6	29.4/ 28.2/ 29.6	26.3/ 28.4/ 30.0	32.0/ 35.0/ 35.2	28.8/ 28.5/ 28.1	30.0/ 30.0/ 31.0	31.0/ 32.0/ 32.8	27.4/ 27.2/ 28.2
Politics and Governance	35.7/ 36.1/ 34.9	46.5/ 50.3/ 48.9	26.0/ 25.0/ 27.0	43.4/ 42.0/ 44.6	32.0/ 33.6/ 35.2	40.0/ 40.0/ 34.0	39.0/ 38.8/ 38.8	39.0/ 38.0/ 40.0	24.0/ 27.0/ 29.0	36.5/ 38.6/ 39.8	27.8/ 29.9/ 29.9
Psychology	39.0/ 37.0/ 39.0	63.0/ 65.0/ 67.0	35.0/ 36.0/ 34.0	49.6/ 45.0/ 48.8	37.0/ 35.0/ 41.0	41.0/ 36.0/ 37.0	27.0/ 39.0/ 41.0	35.0/ 34.0/ 34.0	22.0/ 25.0/ 23.0	33.0/ 33.0/ 38.0	32.0/ 29.0/ 23.0
Public Administration	29.0/ 33.0/ 35.0	42.4/ 49.5/ 46.5	34.3/ 38.4/ 37.4	41.4/ 46.5/ 47.5	33.3/ 36.4/ 30.3	26.0/ 35.0/ 32.0	30.3/ 24.2/ 30.3	37.0/ 29.0/ 30.0	29.3/ 29.3/ 27.3	33.0/ 39.0/ 37.0	33.0/ 29.0/ 33.0
Religion and Spirituality	21.0/ 24.0/ 29.0	61.0/ 68.0/ 63.0	26.0/ 27.0/ 25.0	41.9/ 45.2/ 46.8	27.0/ 28.0/ 33.0	23.0/ 28.0/ 28.0	49.0/ 41.0/ 46.0	26.0/ 23.0/ 24.0	26.0/ 22.0/ 23.0	32.0/ 27.0/ 32.0	16.0/ 15.0/ 17.0
Social Welfare and Development	27.0/ 23.0/ 29.0	50.9/ 51.8/ 55.4	34.0/ 38.0/ 30.0	40.9/ 40.9/ 44.0	25.0/ 33.0/ 37.0	28.0/ 28.0/ 34.0	33.0/ 32.0/ 34.0	32.0/ 29.0/ 32.0	17.0/ 15.0/ 21.0	34.0/ 43.0/ 45.0	34.0/ 34.0/ 33.0
Sociology	26.1/ 28.7/ 24.3	40.2/ 42.0/ 41.6	25.0/ 26.0/ 26.0	34.8/ 38.4/ 39.4	29.0/ 25.0/ 25.0	33.0/ 25.0/ 27.0	28.9/ 34.0/ 30.8	23.0/ 29.0/ 31.0	29.0/ 25.0/ 26.0	27.0/ 35.1/ 35.1	28.2/ 27.7/ 27.1
Sports and Recreation	31.2/ 32.7/ 34.6	58.2/ 61.0/ 61.4	25.0/ 33.0/ 31.0	43.8/ 45.8/ 48.0	35.4/ 37.6/ 37.1	35.0/ 35.0/ 36.0	36.2/ 40.1/ 39.0	32.0/ 29.0/ 38.0	28.0/ 28.0/ 30.0	35.3/ 39.7/ 37.8	29.0/ 28.6/ 30.5
Technology and Innovation	34.0/ 41.0/ 36.0	54.2/ 57.2/ 57.6	33.0/ 37.0/ 34.0	41.4/ 43.2/ 46.2	24.0/ 34.0/ 31.0	36.0/ 42.0/ 40.0	45.0/ 47.0/ 48.0	35.0/ 36.0/ 33.0	25.0/ 25.0/ 26.0	37.0/ 42.0/ 40.0	28.0/ 31.0/ 35.0
Transportation and Logistics	30.0/ 36.0/ 27.0	40.8/ 48.5/ 45.4	31.0/ 27.0/ 25.0	30.0/ 34.1/ 36.9	25.2/ 25.2/ 30.3	25.0/ 28.0/ 26.0	26.5/ 30.6/ 33.7	33.3/ 25.2/ 25.2	24.0/ 22.0/ 26.0	28.0/ 30.0/ 32.0	30.0/ 26.0/ 30.0

Table 50: Detailed subject-wise evaluation for COHEREFORAI/AYA-23-35B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	bn	en	gu	hi	kn	ml	mr	or	pa	ta	te
Agriculture	51.0/ 60.0/ 59.0	50.5/ 71.8/ 72.2	43.4/ 52.5/ 54.5	39.6/ 63.8/ 64.2	40.0/ 61.0/ 64.0	54.5/ 61.6/ 62.6	44.0/ 54.0/ 54.0	51.0/ 60.0/ 58.0	46.0/ 59.0/ 58.0	48.0/ 54.0/ 59.0	42.0/ 58.0/ 60.0
Anthropology	53.0/ 67.0/ 69.0	55.0/ 72.0/ 71.0	47.0/ 63.0/ 62.0	49.5/ 62.6/ 67.7	57.0/ 70.0/ 75.0	47.0/ 65.0/ 62.0	58.0/ 70.0/ 71.0	54.0/ 59.0/ 61.0	45.0/ 65.0/ 60.0	53.0/ 61.0/ 60.0	49.0/ 56.0/ 55.0
Architecture and Design	51.0/ 63.0/ 71.0	52.0/ 72.0/ 71.0	47.0/ 64.0/ 58.0	41.0/ 59.0/ 59.0	52.0/ 63.0/ 63.0	49.0/ 65.0/ 62.0	54.0/ 68.0/ 69.0	44.0/ 55.0/ 54.0	45.0/ 59.0/ 65.0	52.0/ 63.0/ 57.0	45.0/ 57.0/ 61.0
Arts and Culture	48.5/ 69.1/ 67.3	59.4/ 81.2/ 79.6	48.0/ 73.0/ 67.0	45.5/ 67.3/ 68.7	54.6/ 73.6/ 74.2	50.0/ 67.0/ 66.0	44.3/ 61.2/ 61.2	50.0/ 54.0/ 56.0	48.0/ 68.0/ 69.0	47.3/ 63.0/ 63.6	46.2/ 64.3/ 66.8
Astronomy and Astrophysics	60.0/ 75.0/ 78.0	60.0/ 96.0/ 88.0	67.0/ 87.0/ 82.0	57.0/ 77.3/ 75.8	63.0/ 80.0/ 82.0	66.0/ 83.0/ 81.0	60.0/ 73.0/ 74.0	67.0/ 77.0/ 76.0	58.0/ 77.0/ 71.0	68.0/ 80.0/ 82.0	57.0/ 74.0/ 73.0
Biology	48.9/ 68.7/ 66.9	61.5/ 88.6/ 86.2	38.0/ 72.0/ 69.0	50.7/ 76.5/ 78.0	51.6/ 75.8/ 78.3	50.5/ 64.6/ 65.7	43.0/ 61.2/ 61.8	47.0/ 51.0/ 52.0	43.0/ 73.0/ 76.0	45.9/ 62.1/ 63.4	45.0/ 60.4/ 63.5
Business and Management	45.0/ 56.0/ 57.0	51.4/ 66.6/ 66.6	36.0/ 65.0/ 64.0	46.1/ 61.6/ 61.8	43.0/ 62.0/ 67.0	49.0/ 57.0/ 59.0	39.0/ 53.0/ 57.0	41.0/ 49.0/ 50.0	48.0/ 68.0/ 63.0	44.0/ 57.0/ 59.0	46.2/ 51.9/ 50.0
Chemistry	58.8/ 72.0/ 71.7	62.1/ 85.4/ 84.8	59.5/ 72.4/ 69.0	52.5/ 75.1/ 76.5	53.3/ 69.6/ 71.2	60.0/ 71.0/ 71.0	53.3/ 66.2/ 68.3	44.2/ 60.6/ 55.8	49.0/ 69.0/ 73.0	51.9/ 61.1/ 63.8	53.9/ 68.3/ 69.8
Computer Science	39.1/ 44.8/ 49.5	45.4/ 52.8/ 53.4	51.0/ 46.0/ 52.0	46.6/ 48.8/ 52.4	40.2/ 51.5/ 54.4	39.0/ 45.0/ 50.0	44.4/ 43.2/ 50.9	38.0/ 38.0/ 47.0	53.0/ 53.0/ 47.0	41.3/ 44.2/ 50.0	36.3/ 40.8/ 47.5
Defense and Security	60.0/ 76.0/ 84.0	70.0/ 87.0/ 82.0	53.0/ 77.0/ 73.0	51.7/ 72.0/ 73.9	54.0/ 65.0/ 75.0	53.0/ 76.0/ 76.0	51.0/ 69.0/ 75.0	51.0/ 61.0/ 67.0	52.0/ 71.0/ 72.0	53.0/ 66.0/ 68.0	60.0/ 76.0/ 76.0
Earth Sciences	46.0/ 61.0/ 68.0	60.9/ 81.8/ 80.0	42.0/ 58.0/ 55.0	42.3/ 67.3/ 68.9	52.0/ 75.0/ 74.0	51.0/ 64.0/ 63.0	47.0/ 63.0/ 70.0	50.5/ 46.5/ 57.6	48.0/ 70.0/ 69.0	55.0/ 71.0/ 69.0	47.0/ 70.0/ 71.0
Economics	61.2/ 72.7/ 75.5	55.6/ 76.7/ 75.5	47.0/ 70.0/ 67.0	49.0/ 69.0/ 69.3	58.0/ 75.0/ 77.0	57.0/ 66.0/ 67.0	50.7/ 59.2/ 58.5	58.0/ 58.0/ 57.0	51.0/ 66.0/ 69.0	54.0/ 69.0/ 69.0	47.3/ 59.9/ 61.7
Education	48.0/ 60.0/ 57.0	49.7/ 67.2/ 65.0	38.0/ 55.0/ 57.0	47.6/ 57.7/ 59.8	54.0/ 62.0/ 61.0	37.0/ 51.0/ 52.0	46.5/ 60.5/ 61.8	43.0/ 48.0/ 53.0	32.0/ 59.0/ 50.0	41.0/ 51.0/ 55.0	43.0/ 61.0/ 57.0
Energy and Power	58.0/ 67.0/ 72.0	56.8/ 83.8/ 79.7	62.0/ 73.0/ 75.0	48.5/ 67.1/ 70.2	51.0/ 78.0/ 77.0	47.0/ 76.0/ 72.0	57.0/ 66.0/ 69.0	46.0/ 62.0/ 65.0	50.0/ 71.0/ 69.0	47.0/ 66.0/ 67.0	50.0/ 69.0/ 67.0
Engineering	39.6/ 54.4/ 58.4	47.6/ 66.0/ 66.4	40.4/ 51.4/ 52.6	38.4/ 52.0/ 53.4	37.8/ 52.8/ 57.4	37.2/ 47.9/ 45.3	36.8/ 52.8/ 57.4	35.7/ 42.9/ 44.4	36.0/ 41.0/ 46.0	39.4/ 44.6/ 45.6	39.6/ 50.4/ 52.6
Environmental Science	47.0/ 65.0/ 74.0	57.1/ 79.4/ 80.0	50.0/ 65.0/ 63.0	49.3/ 69.7/ 70.1	56.2/ 68.8/ 70.3	46.0/ 66.0/ 63.0	45.8/ 54.2/ 62.1	47.5/ 53.5/ 54.5	55.0/ 67.0/ 68.0	54.0/ 62.0/ 58.0	46.4/ 55.9/ 59.2
Ethics and Human Rights	52.0/ 60.0/ 63.0	57.0/ 75.0/ 75.0	49.0/ 61.0/ 61.0	46.0/ 60.0/ 61.0	51.0/ 62.0/ 65.0	51.0/ 69.0/ 71.0	52.0/ 62.0/ 64.0	47.0/ 63.0/ 62.0	44.0/ 58.0/ 60.0	45.0/ 59.0/ 66.0	50.0/ 63.0/ 65.0
Finance and Investment	56.4/ 63.9/ 65.8	56.8/ 69.4/ 68.0	55.0/ 63.0/ 67.0	51.0/ 63.6/ 65.8	54.3/ 66.7/ 66.0	44.0/ 52.0/ 55.0	50.8/ 63.3/ 65.0	42.0/ 54.0/ 52.0	47.0/ 57.0/ 53.0	52.0/ 61.5/ 65.9	45.0/ 50.0/ 53.9
Food Science	41.0/ 70.0/ 77.0	56.0/ 88.0/ 87.0	51.0/ 73.0/ 74.0	50.5/ 74.8/ 70.7	48.0/ 65.0/ 68.0	55.0/ 77.0/ 75.0	50.0/ 68.0/ 72.0	49.0/ 60.0/ 61.0	46.0/ 65.0/ 70.0	47.0/ 69.0/ 68.0	54.0/ 71.0/ 65.0
Geography	53.1/ 69.3/ 71.5	56.0/ 78.2/ 75.8	48.0/ 64.0/ 65.0	45.8/ 67.9/ 68.3	55.2/ 77.6/ 79.0	52.5/ 70.7/ 67.7	53.8/ 61.5/ 65.1	52.0/ 63.0/ 63.0	51.0/ 67.0/ 71.0	52.6/ 59.4/ 67.4	51.6/ 66.1/ 65.1
Health and Medicine	54.0/ 68.0/ 69.0	61.1/ 83.4/ 82.6	52.0/ 71.0/ 66.0	50.9/ 79.0/ 79.8	56.0/ 79.0/ 77.0	59.0/ 74.0/ 76.0	45.5/ 63.4/ 62.5	56.0/ 67.0/ 69.0	68.0/ 84.0/ 83.0	59.5/ 67.2/ 68.1	52.3/ 64.5/ 66.4
History	57.1/ 73.1/ 78.7	55.0/ 75.5/ 70.3	63.0/ 76.0/ 78.0	54.2/ 72.4/ 74.8	59.3/ 77.3/ 80.2	51.0/ 68.0/ 65.0	50.2/ 61.0/ 65.8	55.0/ 58.0/ 59.0	51.0/ 77.0/ 78.0	54.5/ 65.8/ 68.7	51.7/ 61.4/ 61.6
Information Technology	60.0/ 81.0/ 83.0	69.0/ 82.0/ 78.0	60.0/ 75.0/ 73.0	59.8/ 78.8/ 81.0	56.0/ 77.0/ 77.0	57.0/ 83.0/ 83.0	55.0/ 70.0/ 78.0	51.0/ 68.0/ 71.0	64.0/ 75.0/ 78.0	60.0/ 73.0/ 69.0	63.0/ 76.0/ 74.0
International Relations	59.0/ 76.0/ 78.0	65.0/ 82.5/ 82.5	54.0/ 70.0/ 69.0	62.6/ 72.4/ 72.0	56.0/ 63.0/ 66.0	62.0/ 75.0/ 74.0	66.0/ 74.0/ 74.0	55.0/ 64.0/ 64.0	57.0/ 69.0/ 70.0	57.0/ 75.0/ 73.0	51.0/ 68.0/ 71.0
Language Studies	40.0/ 38.0/ 40.0	59.4/ 82.4/ 82.0	33.0/ 41.0/ 47.0	38.5/ 52.5/ 48.9	39.5/ 51.9/ 51.9	30.0/ 41.0/ 41.0	34.9/ 51.0/ 50.3	42.0/ 47.0/ 46.0	44.0/ 50.0/ 53.0	35.2/ 39.3/ 44.2	31.6/ 48.9/ 48.9
Law and Ethics	74.3/ 84.5/ 85.8	65.8/ 80.2/ 80.4	62.0/ 70.0/ 63.0	61.2/ 75.6/ 76.8	58.6/ 78.3/ 74.5	53.0/ 63.0/ 64.0	45.3/ 62.5/ 60.3	56.0/ 66.0/ 68.0	66.0/ 82.0/ 78.0	63.0/ 68.5/ 68.5	60.0/ 63.7/ 68.1
Literature and Linguistics	51.3/ 68.5/ 76.6	55.7/ 74.6/ 75.9	45.0/ 67.0/ 62.0	42.7/ 60.3/ 63.9	48.3/ 66.7/ 65.8	45.0/ 57.0/ 58.0	45.1/ 58.4/ 62.8	45.0/ 53.0/ 56.0	47.0/ 68.0/ 69.0	41.7/ 44.1/ 43.7	38.7/ 52.1/ 58.5
Logical Reasoning	37.4/ 45.2/ 48.9	38.2/ 49.2/ 50.2	33.8/ 50.0/ 48.7	36.1/ 40.1/ 43.7	41.1/ 47.7/ 49.0	38.0/ 44.0/ 46.0	36.2/ 37.6/ 42.3	29.0/ 41.0/ 48.0	33.0/ 41.0/ 36.0	39.8/ 40.2/ 43.0	35.6/ 40.8/ 41.6
Materials Science	54.0/ 67.0/ 67.0	50.6/ 75.3/ 77.2	38.0/ 51.0/ 51.0	46.5/ 61.1/ 61.8	41.0/ 64.0/ 65.0	45.0/ 60.0/ 63.0	44.0/ 66.0/ 67.0	40.0/ 53.0/ 52.0	35.0/ 61.0/ 63.0	45.0/ 51.0/ 56.0	40.0/ 70.0/ 70.0
Media and Communication	52.0/ 65.0/ 69.0	63.0/ 77.0/ 80.0	50.0/ 64.0/ 68.0	49.7/ 65.3/ 63.9	47.0/ 70.0/ 71.0	50.0/ 63.0/ 67.0	50.0/ 66.0/ 65.0	55.0/ 61.0/ 63.0	49.0/ 62.0/ 61.0	53.0/ 62.0/ 64.0	45.0/ 66.0/ 65.0
Music and Performing Arts	49.0/ 60.0/ 67.0	58.0/ 82.0/ 78.0	42.0/ 62.0/ 58.0	46.4/ 64.2/ 63.6	48.0/ 59.0/ 68.0	41.0/ 56.0/ 56.0	48.0/ 63.0/ 66.0	49.0/ 58.0/ 62.0	52.0/ 67.0/ 64.0	37.0/ 45.0/ 45.0	46.0/ 63.0/ 66.0
Physics	59.8/ 68.6/ 69.8	60.2/ 81.8/ 78.2	54.0/ 62.7/ 67.1	45.6/ 68.0/ 67.0	56.2/ 70.8/ 71.6	57.9/ 60.5/ 62.6	53.6/ 66.2/ 69.6	53.5/ 57.6/ 63.2	58.0/ 61.0/ 61.0	52.8/ 62.6/ 63.8	55.0/ 64.0/ 67.6
Politics and Governance	54.5/ 77.2/ 79.6	54.1/ 76.1/ 71.5	52.0/ 71.0/ 69.0	52.6/ 71.1/ 74.5	45.8/ 70.9/ 76.1	46.0/ 71.0/ 72.0	43.5/ 55.6/ 57.3	64.0/ 56.0/ 57.0	51.0/ 64.0/ 68.0	45.2/ 63.1/ 61.8	44.6/ 52.8/ 57.1
Psychology	50.0/ 66.0/ 64.0	59.0/ 85.0/ 83.0	40.0/ 65.0/ 66.0	37.2/ 57.4/ 60.5	43.0/ 64.0/ 59.0	46.0/ 62.0/ 60.0	45.0/ 55.0/ 58.0	52.0/ 59.0/ 62.0	51.0/ 68.0/ 69.0	43.0/ 58.0/ 61.0	49.0/ 72.0/ 65.0
Public Administration	41.0/ 52.0/ 62.0	44.4/ 67.7/ 64.6	33.3/ 58.6/ 52.5	44.4/ 68.7/ 68.7	38.4/ 51.5/ 57.6	45.0/ 59.0/ 60.0	43.4/ 47.5/ 50.5	51.0/ 52.0/ 59.0	34.3/ 51.5/ 55.6	42.0/ 53.0/ 56.0	41.0/ 48.0/ 50.0
Religion and Spirituality	59.0/ 74.0/ 77.0	53.0/ 88.0/ 84.0	63.0/ 73.0/ 72.0	53.6/ 74.6/ 74.6	58.0/ 76.0/ 75.0	59.0/ 74.0/ 76.0	57.0/ 73.0/ 75.0	61.0/ 66.0/ 65.0	58.0/ 73.0/ 75.0	54.0/ 71.0/ 73.0	60.0/ 71.0/ 71.0
Social Welfare and Development	44.0/ 61.0/ 60.0	47.3/ 75.0/ 77.7	47.0/ 58.0/ 57.0	48.7/ 62.2/ 63.7	42.0/ 55.0/ 57.0	34.0/ 47.0/ 48.0	47.0/ 61.0/ 60.0	50.0/ 46.0/ 51.0	51.0/ 63.0/ 65.0	49.0/ 59.0/ 59.0	55.0/ 67.0/ 64.0
Sociology	33.0/ 52.2/ 49.6	46.0/ 60.8/ 60.2	37.0/ 47.0/ 45.0	37.0/ 49.2/ 51.0	41.0/ 47.0/ 46.0	26.0/ 47.0/ 46.0	35.2/ 48.4/ 50.3	36.0/ 37.0/ 38.0	36.0/ 46.0/ 50.0	39.6/ 46.0/ 46.0	36.5/ 47.1/ 48.8
Sports and Recreation	49.0/ 69.8/ 77.7	65.0/ 87.2/ 87.0	47.0/ 78.0/ 74.0	54.4/ 73.4/ 77.0	43.8/ 74.2/ 80.3	53.0/ 78.0/ 75.0	52.0/ 74.6/ 74.6	52.0/ 64.0/ 68.0	36.0/ 72.0/ 73.0	53.8/ 70.5/ 73.1	46.2/ 68.1/ 70.0
Technology and Innovation	55.0/ 64.0/ 62.0	58.6/ 81.6/ 80.0	50.0/ 64.0/ 66.0	53.8/ 68.8/ 71.6	48.0/ 71.0/ 69.0	50.0/ 65.0/ 61.0	56.0/ 69.0/ 65.0	48.0/ 56.0/ 60.0	44.0/ 62.0/ 62.0	47.0/ 60.0/ 65.0	52.0/ 57.0/ 58.0
Transportation and Logistics	41.0/ 54.0/ 58.0	56.9/ 65.4/ 66.9	38.0/ 55.0/ 53.0	53.6/ 65.6/ 67.2	47.5/ 53.5/ 59.6	41.0/ 56.0/ 54.0	45.9/ 45.9/ 52.0	45.5/ 40.4/ 49.5	40.0/ 57.0/ 58.0	50.0/ 55.0/ 57.0	43.0/ 47.0/ 54.0

Table 51: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.1-70B on MILU across different languages. The results reported are X / Y / Z where X denote the 0-shot, Y denotes 1-shot and Z denotes 5-shot performance respectively.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	57.0	52.1	52.5	55.9	58.0	62.6	47.0	54.0	58.0	53.0	59.0
Anthropology	68.0	51.0	57.0	53.5	65.0	60.0	64.0	58.0	61.0	63.0	64.0
Architecture and Design	65.0	48.0	63.0	58.0	59.0	63.0	59.0	55.0	53.0	59.0	63.0
Arts and Culture	61.2	52.2	68.0	61.9	71.2	69.0	53.5	54.0	74.0	58.2	62.3
Astronomy and Astrophysics	80.0	70.0	84.0	71.9	79.0	76.0	67.0	76.0	74.0	74.0	73.0
Biology	64.4	55.7	68.0	70.7	69.3	64.6	57.3	56.0	67.0	65.3	61.3
Business and Management	58.0	53.6	60.0	55.5	52.0	57.0	53.0	43.0	60.0	58.0	56.6
Chemistry	64.6	61.1	69.0	64.3	64.0	56.0	57.3	46.2	64.0	59.4	64.1
Computer Science	41.1	38.0	53.0	41.0	47.3	41.0	42.6	43.0	49.0	45.9	39.5
Defense and Security	70.0	69.0	72.0	70.6	65.0	66.0	70.0	62.0	68.0	65.0	72.0
Earth Sciences	70.0	59.3	57.0	61.7	75.0	67.0	61.0	49.5	69.0	73.0	72.0
Economics	69.1	53.2	65.0	59.3	70.0	58.0	51.4	53.0	61.0	64.3	56.9
Education	57.0	55.4	45.0	56.2	60.0	49.0	50.3	55.0	44.0	51.0	48.0
Energy and Power	74.0	60.1	75.0	62.0	70.0	70.0	59.0	60.0	70.0	62.0	67.0
Engineering	52.0	47.6	52.4	45.6	51.8	45.3	47.0	42.9	43.0	44.2	50.0
Environmental Science	67.0	57.3	64.0	63.9	71.9	67.0	54.2	58.6	63.0	63.0	63.7
Ethics and Human Rights	65.0	55.0	58.0	52.0	56.0	62.0	53.0	59.0	59.0	57.0	60.0
Finance and Investment	57.9	51.6	57.0	55.0	68.5	60.0	55.4	53.0	48.0	55.9	52.1
Food Science	66.0	48.0	57.0	61.6	61.0	61.0	55.0	59.0	68.0	63.0	65.0
Geography	71.5	56.0	65.0	61.7	73.5	70.7	53.8	48.0	68.0	65.1	66.1
Health and Medicine	74.0	63.9	72.0	68.7	78.0	68.0	52.7	69.0	75.0	70.7	62.6
History	74.9	53.0	71.0	68.3	76.6	72.0	59.5	55.0	68.0	64.7	60.4
Information Technology	85.0	52.0	65.0	72.1	76.0	68.0	68.0	66.0	61.0	70.0	69.0
International Relations	69.0	66.7	69.0	65.4	63.0	65.0	66.0	61.0	66.0	68.0	63.0
Language Studies	38.0	57.2	37.0	44.3	40.3	39.0	37.6	43.0	50.0	36.7	39.1
Law and Ethics	81.1	63.4	70.0	72.0	74.5	60.0	57.8	61.0	78.0	76.4	66.2
Literature and Linguistics	64.0	50.1	61.0	52.9	65.8	51.0	62.0	54.0	64.0	47.0	58.5
Logical Reasoning	47.7	31.8	47.4	40.1	46.5	43.0	36.8	30.0	46.0	44.2	41.3
Materials Science	64.0	56.3	51.0	57.0	53.0	55.0	61.0	53.0	59.0	56.0	56.0
Media and Communication	68.0	57.0	67.0	59.2	65.0	64.0	53.0	63.0	61.0	68.0	69.0
Music and Performing Arts	62.0	54.0	53.0	57.0	63.0	59.0	58.0	52.0	61.0	53.0	63.0
Physics	60.4	55.8	57.7	57.6	64.4	59.0	56.6	49.6	54.0	58.2	59.0
Politics and Governance	75.7	54.3	66.0	63.0	68.8	62.0	46.4	60.0	63.0	64.7	55.9
Psychology	66.0	59.0	66.0	60.5	66.0	60.0	54.0	51.0	56.0	61.0	60.0
Public Administration	55.0	44.4	54.5	56.6	43.4	53.0	41.4	53.0	51.5	53.0	51.0
Religion and Spirituality	79.0	53.0	76.0	64.9	71.0	73.0	66.0	54.0	68.0	67.0	69.0
Social Welfare and Development	55.0	58.9	60.0	60.6	50.0	48.0	50.0	49.0	64.0	57.0	54.0
Sociology	47.0	42.2	46.0	49.0	50.0	36.0	48.4	39.0	44.0	46.0	46.5
Sports and Recreation	63.4	64.2	68.0	63.8	68.0	61.0	63.8	45.0	66.0	68.6	64.8
Technology and Innovation	65.0	59.2	67.0	59.6	64.0	63.0	63.0	56.0	61.0	58.0	57.0
Transportation and Logistics	48.0	50.8	48.0	53.0	51.5	51.0	44.9	46.5	50.0	43.0	42.0

Table 52: 0-shot subject-wise evaluation for META-LLAMA/LLAMA-3.1-70B-INSTRUCT on MILU across different languages.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	58.0	49.9	61.6	56.7	60.0	55.6	46.0	56.0	62.0	52.0	55.0
Anthropology	68.0	52.0	65.0	55.6	61.0	60.0	69.0	58.0	64.0	62.0	63.0
Architecture and Design	69.0	51.0	61.0	60.0	57.0	62.0	64.0	58.0	60.0	62.0	60.0
Arts and Culture	66.7	50.2	68.0	60.7	73.0	64.0	57.9	53.0	68.0	58.2	63.3
Astronomy and Astrophysics	81.0	73.0	86.0	70.3	76.0	75.0	70.0	74.0	73.0	77.0	71.0
Biology	65.1	54.9	63.0	67.1	68.0	62.6	57.6	56.0	65.0	61.8	62.6
Business and Management	58.0	56.4	57.0	56.9	58.0	55.0	55.0	44.0	61.0	49.0	57.6
Chemistry	63.5	60.9	67.2	61.3	55.9	63.0	64.6	50.0	61.0	60.2	61.9
Computer Science	39.6	41.2	47.0	43.8	47.9	37.0	42.6	42.0	44.0	41.9	31.8
Defense and Security	70.0	70.0	76.0	68.7	64.0	66.0	73.0	61.0	71.0	68.0	73.0
Earth Sciences	64.0	58.2	54.0	59.9	70.0	68.0	59.0	48.5	66.0	72.0	68.0
Economics	66.2	51.0	65.0	61.7	72.0	54.0	54.9	56.0	60.0	64.3	54.5
Education	53.0	51.9	53.0	57.1	53.0	47.0	49.0	54.0	42.0	47.0	53.0
Energy and Power	67.0	62.2	71.0	62.7	66.0	66.0	68.0	60.0	66.0	60.0	66.0
Engineering	51.4	46.2	52.0	46.8	49.0	44.4	51.2	42.9	42.0	44.0	49.6
Environmental Science	63.0	56.3	65.0	62.9	71.1	62.0	55.3	54.5	66.0	69.0	60.3
Ethics and Human Rights	61.0	61.0	59.0	58.0	55.0	56.0	54.0	59.0	58.0	58.0	65.0
Finance and Investment	60.4	56.2	63.0	58.4	63.0	53.0	59.9	51.0	48.0	59.2	51.1
Food Science	66.0	55.0	61.0	60.6	61.0	59.0	56.0	57.0	61.0	62.0	59.0
Geography	72.6	57.0	64.0	61.7	73.1	71.7	58.0	49.0	71.0	64.0	65.6
Health and Medicine	64.0	63.5	74.0	69.3	75.0	71.0	54.5	70.0	76.0	68.1	65.4
History	72.8	51.8	74.0	71.0	75.1	65.0	59.9	60.0	70.0	66.5	59.9
Information Technology	77.0	55.0	66.0	67.0	74.0	68.0	70.0	63.0	66.0	59.0	70.0
International Relations	67.0	65.8	69.0	63.4	66.0	69.0	63.0	61.0	63.0	66.0	61.0
Language Studies	45.0	58.0	38.0	44.7	34.1	40.0	34.9	46.0	53.0	35.6	36.2
Law and Ethics	81.1	62.2	69.0	71.4	66.9	65.0	59.1	65.0	75.0	75.6	64.4
Literature and Linguistics	64.0	47.9	61.0	54.5	62.5	52.0	60.2	53.0	61.0	48.2	60.6
Logical Reasoning	42.5	31.8	47.4	37.5	40.2	42.0	36.8	32.0	36.0	43.4	39.6
Materials Science	65.0	58.2	52.0	57.2	59.0	53.0	62.0	44.0	54.0	52.0	56.0
Media and Communication	61.0	59.0	71.0	61.9	60.0	64.0	54.0	60.0	57.0	58.0	70.0
Music and Performing Arts	61.0	55.0	51.0	59.6	65.0	56.0	62.0	50.0	62.0	47.0	65.0
Physics	61.8	60.2	57.1	57.0	58.2	53.2	58.4	55.2	52.0	56.2	58.6
Politics and Governance	74.1	54.9	70.0	66.7	68.4	61.0	48.3	61.0	65.0	63.9	54.5
Psychology	64.0	56.0	65.0	56.6	58.0	61.0	49.0	48.0	57.0	66.0	57.0
Public Administration	53.0	44.4	57.6	54.5	38.4	56.0	42.4	56.0	49.5	50.0	53.0
Religion and Spirituality	77.0	57.0	76.0	67.7	69.0	73.0	67.0	56.0	67.0	70.0	71.0
Social Welfare and Development	54.0	58.0	59.0	60.6	55.0	50.0	52.0	47.0	64.0	55.0	50.0
Sociology	45.2	41.4	48.0	50.0	43.0	37.0	47.2	41.0	47.0	50.4	44.1
Sports and Recreation	68.8	62.0	72.0	65.6	73.6	68.0	67.2	53.0	66.0	66.0	67.6
Technology and Innovation	67.0	60.6	64.0	59.2	58.0	61.0	63.0	55.0	56.0	57.0	53.0
Transportation and Logistics	50.0	50.8	52.0	53.6	52.5	54.0	51.0	50.5	50.0	46.0	41.0

Table 53: 1-shot subject-wise evaluation for META-LLAMA/LLAMA-3.1-70B-INSTRUCT on MILU across different languages.

Topic	*bn*	*en*	*gu*	*hi*	*kn*	*ml*	*mr*	*or*	*pa*	*ta*	*te*
Agriculture	60	57.1	54.5	47.1	60	58.6	46	52	56	53	55
Anthropology	69	63	59	53.5	65	62	57	59	61	56	63
Architecture and Design	68	63	64	41	59	65	48	47	62	58	60
Arts and Culture	68.5	66.6	66	46.7	73	77	50.3	62	65	62.4	73.4
Astronomy and Astrophysics	74	70	77	59.4	75	74	60	70	67	75	65
Biology	69	69.9	58	51.5	65.6	70.7	45.1	51	71	60.8	59.8
Business and Management	65	59.4	61	51	63	56	47	53	62	62	48.1
Chemistry	66.2	65.1	57.8	56.1	60.5	69	51.7	39.4	63	60.2	61.9
Computer Science	44.8	44.8	48	46	49.1	49	44.4	38	52	40.1	43.5
Defense and Security	77	68	67	60.7	73	76	52	71	69	68	71
Earth Sciences	65	60	55	46.1	72	69	58	52.5	59	64	67
Economics	69.1	62.6	63	57.3	71	65	57.8	57	69	65.1	61.1
Education	59	54.5	48	48.5	52	49	54.1	47	52	47	53
Energy and Power	73	58.8	71	50.5	72	74	52	58	64	61	67
Engineering	57	59.2	51.2	43	52.2	51.7	42.4	41.4	45	48.4	54.6
Environmental Science	66	64.1	67	49.9	64.8	67	44.2	59.6	71	67	65.9
Ethics and Human Rights	64	62	73	58	66	68	53	61	67	67	68
Finance and Investment	68.8	61.8	75	53.6	67.9	57	62.7	59	51	60.3	55
Food Science	61	66	56	56.6	69	70	49	67	60	58	67
Geography	65.9	60.6	75	46.8	74.4	67.7	53.2	73	69	65.1	63
Health and Medicine	76	68.3	67	56.7	73	78	50	74	71	66.4	68.2
History	74.7	60.2	80	53.6	77.3	74	52.4	66	68	72	63.2
Information Technology	80	70	71	61.5	77	75	71	69	67	64	70
International Relations	75	75	66	64.2	70	72	61	64	69	70	65
Language Studies	43	68	41	38.5	42.6	41	39.6	49	56	41.9	36.2
Law and Ethics	79	66.6	61	61	72.6	67	50.9	64	77	71.7	70.6
Literature and Linguistics	72.1	63.1	67	50.5	65	64	59.3	54	71	46.6	60.6
Logical Reasoning	52.1	45.2	52.6	42.9	53.5	46	37.9	37	42	47.8	46
Materials Science	62	59.3	46	48.5	55	65	45	47	47	46	49
Media and Communication	57	66	71	54.4	72	66	52	56	59	60	68
Music and Performing Arts	56	65	61	51.7	64	63	57	60	66	55	69
Physics	68.2	63.8	60.2	48.2	60.8	55.3	54.8	52.8	56	58.4	59
Politics and Governance	74.9	59.7	67	54.8	69.6	68	48.9	51	56	65.1	61.6
Psychology	67	61	60	41.9	56	63	41	58	68	53	67
Public Administration	61	54.5	51.5	50.5	50.5	55	47.5	48	55.6	56	51
Religion and Spirituality	77	58	76	49.2	81	81	53	67	74	72	75
Social Welfare and Development	62	59.8	68	54.9	55	57	56	51	65	53	64
Sociology	50.4	48	45	39.4	52	45	42.8	45	55	50.4	50
Sports and Recreation	72.8	73.6	70	58.2	71.4	75	61.6	65	65	67.3	67.6
Technology and Innovation	62	63.4	65	54.2	65	69	54	60	55	53	61
Transportation and Logistics	46	53.1	60	51.1	50.5	58	38.8	52.5	58	50	51

Table 54: Detailed subject-wise evaluation for META-LLAMA/LLAMA-3.1-405B on MILU across different languages. The results reported are for 0-shot experiments.