JamC-QA: 日本固有の知識を問う多肢選択式質問応答ベンチマークの構築

岡 照晃 [1] 柴田 知秀 [1] 吉田 奈央 [1]
[1]SB Intuitions 株式会社
{teruaki.oka, tomohide.shibata, nao.yoshida}@sbintuitions.co.jp

概要

日本語大規模言語モデルの開発競争が活発化する中、日本の文化や風習に特化した難度の高い評価用ベンチマークが必要になっている。本稿では、現在構築している日本語の多肢選択式質問応答ベンチマーク JamC-QA について述べる。JamC-QA は日本の文化や風習といった国内独自の知識を問う問題を既存のベンチマークの翻訳でなく、1 から作成しており、問題数は 2024 年 12 月現在、1,045 問である。評価実験では、JamC-QA を用いることで、日本固有の知識問題に関するモデル性能の差を確認できた。これは既存の日本語ベンチマークでは見えなかったものである。またスコア向上の余地もあり、解くべき難しさもまだ十分に含むことがわかった。

1 はじめに

ChatGPT [1] や Llama シリーズ [2, 3] など、英語を中心に学習した大規模言語モデル（以下、**多言語モデル**）は学習用データに日本語テキストをほとんど含まないにもかかわらず、日本語ベンチマークで高いスコアを示している。例えば、JCommonsenseQA（以下、**JComQA**） [4] には言語横断的な一般常識を問う問題が多く含まれることから図 1 の比較でも多言語モデル [3, 5, 6] は日本語継続事前学習モデル [7, 8] や日本語でスクラッチ学習したモデル [9] [1)2)] と同等のスコアを達成している。一方で日本国内の多くの組織、アカデミアや企業は日本語に強い言語モデルの開発を続けている。その意義として、日本固有の知識の獲得がある [10]。獲得された知識を評価するには、日本の文化や風習に合わせた難度の高いベンチマークが必要だが、高性能なモデルに合わせた難度の高いベンチマーク開発も英語が主で行われている。英語ベンチマークを機械・人手で翻訳して用いることもあるが、翻訳の不自然さや英語圏との文化差から、日本固有の知識の評価に適さない。

同じ問題意識から、非ヨーロッパ圏を中心に翻訳ではなく、ベンチマークを独自に 1 から構築する流れがある。例えば英語の標準ベンチマーク MMLU [11] に対し、中国語版の CMMLU [12] や韓国語版の KMMLU [13] が翻訳を使わずスクラッチから構築された。日本語でも JMMLU[3)] が構築されたが、ほとんどの問題が MMLU の翻訳（**翻訳問題**）である。翻訳問題は数学や天文学、国際法や世界宗教といったどの国にも関わる言語横断的な知識問題であり、日本特有の文化や風習を問うものではない。日本でスクラッチから作られた問題（**日本問題**）も一部あるが、JMMLU の日本問題は JComQA と同様に難度が低く、図 1 の比較でも日本問題のみに限定した場合、スコアは頭を打っている。

そこで本稿では、現在構築している日本語の多肢選択式質問応答ベンチマーク JamC-QA について述べる。JamC-QA は MMLU や JMMLU と同じく多肢選択式質問応答だが、既存ベンチマークの翻訳でなく 1 から問題を作成している。難度は JMMLU の日本語問題よりも高いものを目指し、日本の文化、風習を問う問題に重点を置いた。その結果、図 1 の一番左の JComQA ではほとんど差の見えない Llama-3.1-70B と Llama-3.1-405B も一番右に示す JamC-QA を使った評価では 10pt 以上の差が確認できる。また Llama の継続事前学習モデル Swallow が Llama シリーズのスコアから伸び悩んでいる一方、より多くの日本語のテキストでスクラッチから事前学習した Sarashina シリーズが日本固有の知識を問う問題で高いスコアを示すこともわかった。

1） https://huggingface.co/sbintuitions/sarashina2-70b
2） https://huggingface.co/sbintuitions/sarashina2-8x70b
3） https://github.com/nlp-waseda/JMMLU

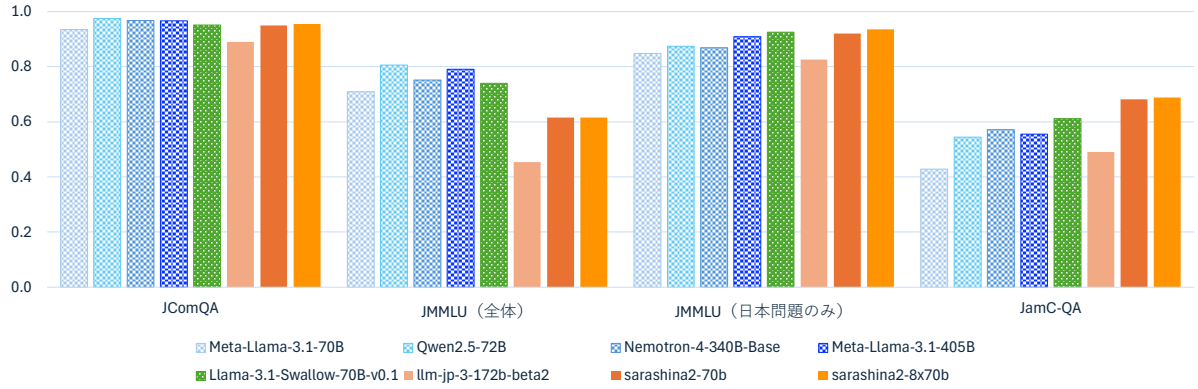

図 1　ベンチマーク別の Exact match スコアの比較: 日本語のベンチマークでスコア上位のモデルを以下の 4 つのベンチマークで評価した。JComQA、JMMLU 全体（翻訳問題＆日本問題）、JMMLU 日本問題のみ、JamC-QA。比較したモデルは左から Meta-Llama-3.1-70B [3]、Qwen2.5-72B [5]、Nemotron-4-340B-Base [6]、Meta-Llama-3.1-405B [3]、Llama-3.1-Swallow-70B-v0.1 [7, 8]、llm-jp-3-172b-beta2 [9]、sarashina2-70b、sarashina2-8x70b である。棒グラフの色は青系統が多言語モデル、緑色が継続事前学習モデル、燈系統が日本語でスクラッチ学習したモデルを意味している。

2　関連研究

MMLU [11] は 57 科目の幅広い知識を問う多肢選択式質問応答であり、英語の標準的ベンチマークである。STEM（科学・技術・工学・数学）、人文学、社会科学、その他の 4 カテゴリの下に初等数学から大学数学、機械学習、法律や経済学などの科目が設けられ、全 14,270 問からなる。しかし大規模言語モデルの性能向上に伴い、MMLU もスコアが既に頭を打っており、より難度の高いベンチマーク構築を目的とした研究が行われている [14, 15]。MMLU は選択肢の数が少なく（4 択）、知識があれば推論能力を必要としない問題が多い。そこで構築されたのが MMLU-Pro である [14]。MMLU-Pro では選択肢数を 4 から最大 10 に増やし、知識だけでなく推論を必要とする問題を作成している。MMLU-Pro+ [15] では、MMLU-Pro 中の選択肢 X と GPT-4o で自動生成した選択肢 Y を組み合わせた *Both X and Y are correct* という選択肢を導入し、さらなる複雑化を試みている。また MMLU の正解には誤りや曖昧性が多く含まれる。MMLU-Redux は MMLU 内の 3,000 問を人手で確認・修正することで構築された [16]。

英語で作成された MMLU を機械翻訳し、複数言語の評価に用いる研究もある [17]。しかし翻訳した MMLU を評価に使うことの問題は以前から指摘されており、翻訳の質と、原言語（英語圏）の文化的背景が翻訳先で通用しないことから、中国語 [12]、韓国語 [13]、トルコ語 [18]、インド語 [19]、ロシア語 [20] といった非ヨーロッパ圏の言語でそれぞれ独自の MMLU 構築が試みられている。日本語でも JMMLU が構築されたが、56 科目中 52 科目が MMLU を日本語に機械翻訳した後、人手で修正した問題で、日本独自に作られたのは、日本史、日本地理、公民、熟語の 4 科目のみである。

3　日本語ベンチマークの構築

日本固有の知識を問う新しいベンチマーク Japanese Multiple Choice QA (**JamC-QA**) の構築について述べる。JamC-QA の問題形式は JComQA、MMLU と同じく多肢選択式質問応答とし、選択肢の数は MMLU と同じ 4 択とした。言語横断的な知識を問う問題は JMMLU に含まれていることから、日本の文化や風習に関する質問を中心に問題を作成している。問題には、『文化』、『風習』、『風土』、『行政』、『日本史』、『法律』、『医療』の 7 つのカテゴリを設定した。各カテゴリに含まれる質問と選択肢の具体例を表 1 に示す。

『文化』は助数詞、元号や四季の言葉といった日本の文化的背景知識を要するカテゴリである。伝統文化に限らず漫画・アニメといったサブカルチャーも『文化』に含めた。『風習』は催事やマナーに関する知識を中心とした。『文化』が実際の行動を伴わない知識（e.g., 和暦と西暦の対応）に対し、『風習』は箸使いのように実際の行動を伴う知識と定義した。『風土』は日本全体でなく、地方ごとの知識を問うカテゴリである。都道府県内、市町村内に閉じた知見や方言といった日本全体では共通しない知識を問う問題が含まれている。『日本史』は既に JMMLU に学校教育レベルの問題が用意されていることから、学校教育に含まれない裾野知識を問うカ

表 1　JamC-QA の質問と選択肢の具体例

カテゴリ	質問と選択肢の例（**太字**は正解選択肢）			
文化	Q. 西暦 1989 年 1 月 7 日を和暦で書くと何年何月何日？			
	A. 昭和 62 年 1 月 7 日	B. 昭和 63 年 1 月 7 日	**C. 昭和 64 年 1 月 7 日**	D. 平成元年 1 月 7 日
風習	Q. 正月飾りの門松をしまうことを一般的に何と呼ぶか？			
	A. 松納め	B. 門松終い	C. 松終い	D. 門納め
風土	Q.「神戸」を「かんべ」と読む土地は次のうちどこ？			
	A. 愛知県田原市神戸町	B. 兵庫県神戸市	C. 群馬県高崎市神戸町	D. 岡山県津山市神戸
日本史	Q. 肉食禁止令を出した人は誰？			
	A. 天武天皇	B. 弘文天皇	C. 昭和天皇	D. 後光明天皇
行政	Q. コンビニエンスストアで発行できない証明書は次のうちどれ？			
	A. 住民票	B. 戸籍謄本	C. 住民票記載事項証明書	**D. 戸籍記載事項証明書**
法律	Q.「占有離脱物横領罪」が成立した場合の罰金は何万円以下か。			
	A. 3 万円	B. 5 万円	**C. 10 万円**	D. 20 万円
医療	Q. 日本の市町村が予防接種を行う定期接種の対象疾患である A 類疾病でないものはどれか			
	A. 破傷風	B. ポリオ	C. 結核	**D. インフルエンザ**

表 2　JamC-QA の問題数の内訳

カテゴリ	フィルタリング前	フィルタリング後
文化	372	246
風習	224	145
風土	270	181
行政	208	136
日本史	185	130
法律	208	132
医療	129	75
計	1,596	1,045

テゴリとした。また『行政』では身近な行政サービスに関する知識を扱う一方、日本の法律に関わる専門知識は『法律』カテゴリに区別した。『医療』も国内外で共通の知識問題は避け、日本国内のみで通用する知識とした。

問題は 11 名のアノテータが翻訳でなく独自に 1 から作成した。この際の注意事項として、Wikipedia のようなウェブ上のまとまった知識源からの問題作成を禁止した。これには、モデルサイズや学習トークン数、モデルアーキテクチャに関係なく、特定時期の特定の知識源（e.g., Wikipedia）を事前学習に使っているか否かがスコアに影響することを避ける意図がある。一方で、作成した問題の真偽確認のためであればこれらの閲覧は可とした。アノテータは正解選択肢 1 つと誤答選択肢 3 つを作成するが、この際、正解選択肢の番号に偏りが出ないよう（A,B,C,D のうち、A ばかりが正解にならないよう）、評価時には選択肢をシャッフルして用いた。MMLU の 10,000 問規模を目指し、アノテータ 1 人当たり 7〜10 問/日の速度で問題を作成している。問題数は 2024 年 12 月 執筆時点で、1,596 問である。

MMLU-Pro [14] では難度を高めるため、過度に簡単な質問を除外するフィルタリングを行なっている。具体的には 8 つの弱いモデルを用意し、4 つ以上が正解した問題を除外している。JamC-QA の構築でも問題の難度を高くするため、同様のフィルタリングを実施する。フィルタリングには次の 8 つのモデルを用いた: Llama-2-7b [2]、Meta-Llama-3.1-8B [3]、Minitron-8B-Base[21]、Qwen2.5-1.5B [5]、Yi-1.5-9B [22]、gemma-2-baku-2b [23]、llm-jp-3-3.7b [9]、sarashina2.1-1b [4)]。8 つのモデルのうち、4 つ以上のモデルが正解した問題を除外した。その結果、1,596 → 1,045 問になった。カテゴリ別の問題数の内訳を表 2 に示す。

除外された問題の中でも、すべての弱いモデルが正解したものを確認したところ、

Q.「東北四大祭り」のひとつに数えられる「仙台七夕まつり」は、どこの県で行われるか（『風土』）
A. 山形県, B. 青森県, C. 秋田県, D. **宮城県**

Q. 母の日に送る赤いカーネーションの花言葉は次のうちどれ?（『文化』）
A. **母への愛**, B. 尊敬, C. 結束, D. 偉大

のように質問文中の「仙台」や「母の日」から専門的な知識がなくとも容易に正解選択肢を選ぶことができる問題であった。

4　結果と分析

JamC-QA を使い、パラメータサイズ 70B 規模以上で日本語ベンチマークスコア上位のモデルを評価した。比較に用いたモデルは 図 1 でも使用している 8 つである。またアノテータが作成した全問題でなく、フィルタリング後の 1,045 問を評価に使用した。推論は greedy decoding で行い、fewshot は固定、

4)　https://huggingface.co/sbintuitions/sarashina2.1-1b

表 3　JamC-QA のカテゴリ別 Exact match スコアの比較

モデル	ALL	文化	風習	風土	行政	日本史	法律	医療
Meta-Llama-3.1-70B [3]	0.428	0.362	0.345	0.365	0.537	0.454	0.530	0.533
Qwen2.5-72B [5]	0.544	0.541	0.510	0.459	0.610	0.477	0.644	0.653
Nemotron-4-340B-Base [6]	0.571	0.602	0.517	0.508	0.581	0.538	0.629	0.667
Meta-Llama-3.1-405B [3]	0.555	0.508	0.469	0.503	0.559	0.592	0.644	0.773
Llama-3.1-Swallow-70B-v0.1 [7, 8]	0.611	0.565	0.579	0.525	0.640	0.569	**0.735**	**0.840**
llm-jp-3-172b-beta2 [9]	0.491	0.480	0.517	0.453	0.471	0.485	0.500	0.600
sarashina2-70b	0.681	0.687	**0.738**	0.641	**0.669**	**0.677**	0.659	0.720
sarashina2-8x70b	**0.688**	**0.724**	0.731	**0.663**	0.654	0.646	0.659	0.733

使用したテンプレートは Appendix A に示す。カテゴリごとの Exact match スコアを表 3 に示す。

表 3 を見ると、全カテゴリをまとめて評価した ALL のスコア（マイクロ平均）が最も高いのは sarashina2-8x70b である。しかしカテゴリ別では法律と医療のカテゴリで Llama-3.1-Swallow-70B-v0.1 が最も高いスコアである。これは法律や医療カテゴリにはフィルタリングを行なった後も「国際郵便で送れるものはどれ（『法律』）」「第三大臼歯を含めた永久歯の本数は次のうちどれ?（『医療』）」といった日本国内に限定しない、言語横断的な知識を問う問題が含まれたことから、Llama からの継続事前学習が有効に働いたと考えられる。しかし JamC-QA ではそうした言語横断的な問題よりも日本固有の知識の評価を目指している。そのためフィルタリングに使用するモデルの数やサイズの検討を行い、より精度の高いフィルタリングを実現した上で再度評価を行う必要がある。

sarashina2-8x70b の ALL のスコアは 0.688 とまだ約 3 割の問題が解けていない。sarashina2-8x70b が解けなかった問題の例を示す。

Q. 雛人形の仕丁の表情で当てはまらないものは?（『文化』）
A. **喜び**, B. 泣き, C. 怒り, D. 笑い

Q. 回転灯は法律で基準を定められていますが、停止表示灯の色は次のうちどれ?（『法律』）
A. **紫色**, B. 青色, C. 赤色, D. 黄色

前者は正解「喜び」を「笑い」と誤ったニアミスであるが、後者では正解の「紫色」に対し、パトカー、救急車や消防車などの緊急自動車で見る機会の多い「赤色」を選んでいた。

3 節の後半で説明した弱いモデルを使ったフィルタリングの効果を確認するため、フィルタリング前後の Exact match スコアを比較した。図 2 を見ると、フィルタリング前後でモデル間のスコアの順位に影響を与えることなくスコアの絶対値が下がっており、過度に簡単な問題が除外できていることがわかる。

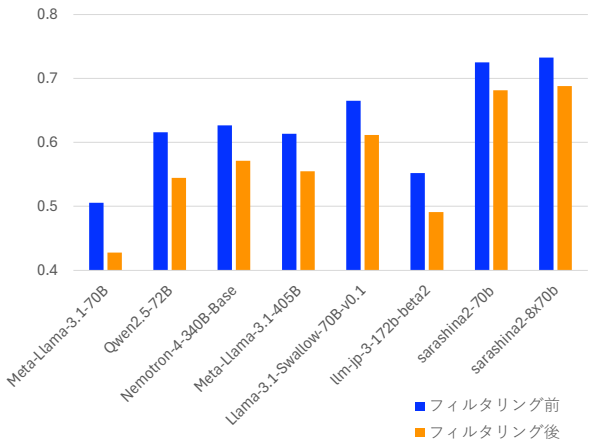

図 2　フィルタリング前後の Exact match スコアの比較

5　おわりに

本稿では日本固有の知識を問う新しいベンチマーク JamC-QA 構築について述べた。JamC-QA は質問に対し、4 つの選択肢の中から正解を選ぶ多肢選択式質問応答であり、日本の文化や風習に特化した問題を作成をしている。難度も既存のベンチマークより高いものを目指した。その結果、既存のベンチマークでは見られなかったモデル間のスコアの差を確認できた。問題作成は継続して行なっており、10,000 問の規模での公開を目指している。単に問題数を増やすだけでなく、フィルタリング時に除外された過度に簡単な問題をアノテータにフィードバックすることで、質問や選択肢の複雑化を行い、ベンチマーク全体の難度を高めていくことも行なっていく。また現状の問題カテゴリは粗く、例えば『文化』には伝統文化、サブカルチャー、日本語の知識を問う問題が混在している状態である。サブカテゴリを設け、問題管理を細かく行なっていくことも今後の課題である。

参考文献

[1] OpenAI. GPT-4 Technical Report. **arXiv:2303.08774**, 2024. https://arxiv.org/abs/2303.08774.

[2] Kevin Stone Peter Albert Amjad Almahairi Yasmine Babaei Nikolay Bashlykov Soumya Batra Prajjwal Bhargava Shruti Bhosale et al. Hugo Touvron, Louis Martin. Llama 2: Open Foundation and Fine-Tuned Chat Models. **arXiv:2307.09288**, 2023. https://arxiv.org/abs/2307.09288.

[3] Abhinav Jauhri Abhinav Pandey Abhishek Kadian Ahmad Al-Dahle Aiesha Letman Akhil Mathur Alan Schelten Alex Vaughan et al. Aaron Grattafiori, Abhimanyu Dubey. The Llama 3 Herd of Models. **arXiv:2407.21783**, 2024. https://arxiv.org/abs/2407.21783.

[4] 栗原健太郎, 河原大輔, 柴田知秀. JGLUE: 日本語言語理解ベンチマーク. 自然言語処理, Vol. 30, No. 1, pp. 63–87, 2023.

[5] Qwen. Qwen2.5 Technical Report. **arXiv:2412.15115**, 2024. https://arxiv.org/abs/2412.15115.

[6] Nvidia. Nemotron-4 340B Technical Report. **arXiv:2406.11704**, 2024. https://arxiv.org/abs/2406.11704.

[7] 藤井一喜, 中村泰士, Mengsay Loem, 飯田大貴, 大井聖也, 服部翔, 平井翔太, 水木栄, 横田理央, 岡崎直観. 継続事前学習による日本語に強い大規模言語モデルの構築. 言語処理学会第 30 回年次大会 (NLP2024), pp. 2102–2107, 2024.

[8] 水木栄, 飯田大貴, 藤井一喜, 中村泰士, Mengsay Loem, 大井聖也, 服部翔, 平井翔太, 横田理央, 岡崎直観. 大規模言語モデルの日本語能力の効率的な強化: 継続事前学習における語彙拡張と対訳コーパス活用. 言語処理学会第 30 回年次大会 (NLP2024), pp. 1514–1519, 2024.

[9] LLM-jp. LLM-jp: A Cross-organizational Project for the Research and Development of Fully Open Japanese LLMs. **arXiv:2407.03963**, 2024. https://arxiv.org/abs/2407.03963.

[10] 齋藤幸史郎, 水木栄, 大井聖也, 中村泰士, 塩谷泰平, 前田航希, Ma Youmi, 服部翔, 藤井一喜, 岡本拓己, 石田茂樹, 高村大也, 横田理央, 岡崎直観. Llm に日本語テキストを学習させる意義. 研究報告自然言語処理（NL）, No. 2024-NL-261(12), pp. 1–15, 2024.

[11] Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. Measuring Massive Multitask Language Understanding. **arXiv:2009.03300**, 2021. https://arxiv.org/abs/2009.03300.

[12] Haonan Li, Yixuan Zhang, Fajri Koto, Yifei Yang, Hai Zhao, Yeyun Gong, Nan Duan, and Timothy Baldwin. CMMLU: Measuring Massive Multitask Language Understanding in Chinese. **arXiv:2306.09212**, 2024. https://arxiv.org/abs/2306.09212.

[13] Guijin Son, Hanwool Lee, Sungdong Kim, Seungone Kim, Niklas Muennighoff, Taekyoon Choi, Cheonbok Park, Kang Min Yoo, and Stella Biderman. KMMLU: Measuring Massive Multitask Language Understanding in Korean. **arXiv:2402.11548**, 2024. https://arxiv.org/abs/2402.11548.

[14] 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. MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark. **arXiv:2406.01574**, 2024. https://arxiv.org/abs/2406.01574.

[15] Saeid Asgari Taghanaki, Aliasgahr Khani, and Amir Khasahmadi. MMLU-Pro+: Evaluating Higher-Order Reasoning and Shortcut Learning in LLMs. **arXiv:2409.02257**, 2024. https://arxiv.org/abs/2409.02257.

[16] Aryo Pradipta Gema, Joshua Ong Jun Leang, Giwon Hong, Alessio Devoto, Alberto Carlo Maria Mancino, Rohit Saxena, Xuanli He, Yu Zhao, Xiaotang Du, Mohammad Reza Ghasemi Madani, and Claire Barale et al. Are We Done with MMLU? **arXiv:2406.04127**, 2024. https://arxiv.org/abs/2406.04127.

[17] Viet Dac Lai, Chien Van Nguyen, Nghia Trung Ngo, Thuat Nguyen, Franck Dernoncourt, Ryan A. Rossi, and Thien Huu Nguyen. Okapi: Instruction-tuned Large Language Models in Multiple Languages with Reinforcement Learning from Human Feedback. **arXiv:2307.16039**, 2023. https://arxiv.org/abs/2307.16039.

[18] Arda Yüksel, Abdullatif Köksal, Lütfi Kerem Şenel, Anna Korhonen, and Hinrich Schütze. TurkishMMLU: Measuring Massive Multitask Language Understanding in Turkish. **arXiv:2407.12402**, 2024. https://arxiv.org/abs/2407.12402.

[19] Sshubam Verma, Mohammed Safi Ur Rahman Khan, Vishwajeet Kumar, Rudra Murthy, and Jaydeep Sen. MILU: A Multi-task Indic Language Understanding Benchmark. **arXiv:2411.02538**, 2024. https://arxiv.org/abs/2411.02538.

[20] Alena Fenogenova, Artem Chervyakov, Nikita Martynov, Anastasia Kozlova, Maria Tikhonova, Albina Akhmetgareeva, Anton Emelyanov, Denis Shevelev, Pavel Lebedev, and Leonid Sinev et al. MERA: A Comprehensive LLM Evaluation in Russian. **arXiv:2401.04531**, 2024. https://arxiv.org/abs/2401.04531.

[21] Saurav Muralidharan, Sharath Turuvekere Sreenivas, Raviraj Joshi, Marcin Chochowski, Mostofa Patwary, Mohammad Shoeybi, Bryan Catanzaro, Jan Kautz, and Pavlo Molchanov. Compact Language Models via Pruning and Knowledge Distillation. **arXiv:2407.14679**, 2024. https://arxiv.org/abs/2407.14679.

[22] 01. AI. Yi: Open Foundation Models by 01.AI. **arXiv:2403.04652**, 2024. https://arxiv.org/abs/2403.04652.

[23] Kei Sawada, Tianyu Zhao, Makoto Shing, Kentaro Mitsui, Akio Kaga, Yukiya Hono, Toshiaki Wakatsuki, and Koh Mitsuda. Release of Pre-Trained Models for the Japanese Language. In **Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)**, pp. 13898–13905, 5 2024.

A JamC-QA 評価用テンプレート

以下はタスクを説明する指示と、追加の背景情報を提供する入力の組み合わせです。要求を適切に満たす回答を書いてください。
指示:: 質問と回答の選択肢を入力として受け取り、選択肢から回答を選択してください。回答の他には何も含めないことを厳守してください。

質問:: 西暦 1989 年 1 月 7 日を和暦で書くと何年何月何日?, 選択肢::
昭和 62 年 1 月 7 日
昭和 63 年 1 月 7 日
昭和 64 年 1 月 7 日
平成元年 1 月 7 日, 回答:: 昭和 64 年 1 月 7 日
質問:: 正月飾りの門松をしまうことを一般的に何と呼ぶか?, 選択肢::
松納め
門松終い
松終い
門納め, 回答:: 松納め
質問:: 「神戸」を「かんべ」と読む土地は次のうちどれか?, 選択肢::
愛知県田原市神戸町
兵庫県神戸市
群馬県高崎市神戸町
岡山県津山市神戸, 回答:: 愛知県田原市神戸町
質問:: 肉食禁止令を出した人は誰?, 選択肢::
弘文天皇
天武天皇
昭和天皇
後光明天皇, 回答:: 天武天皇
質問:: コンビニエンスストアで発行できない証明書は次のうちどれ, 選択肢::
住民票
戸籍謄本
住民票記載事項証明書
戸籍記載事項証明書, 回答:: 戸籍記載事項証明書
質問:: 「占有離脱物横領罪」が成立した場合の罰金は何万円以下か。, 選択肢::
3 万円
5 万円
10 万円
20 万円, 回答:: 10 万円
質問:: 日本で定期接種が実施される A 類疾病ではないものはどれか?, 選択肢::
破傷風
ポリオ
結核
インフルエンザ, 回答:: インフルエンザ
質問:: {{ QUESTION }}, 選択肢::
{{ CHOICE_A }}
{{ CHOICE_B }}
{{ CHOICE_C }}
{{ CHOICE_D }}, 回答::