All Languages Matter: Evaluating LMMs on Culturally Diverse 100 Languages (Supplementary Material)

A. ALM-Bench Categories

Our `ALM-bench` benchmark dataset consists of 19 categories. Among these categories, we include *Food, Lifestyle, Religion, Architecture, Sports, Customs* and *Heritage* from [**? ?**] with some minor modifications to it. We further include six additional categories including *Literature*, featuring prominent authors, poets, and their notable works; *Music*, showcasing traditional music and dance through visual samples; *Festivals*, covering major cultural celebrations; *Economy*, representing local industries and businesses; *Media*, highlighting cultural icons, entertainment figures, and popular TV shows; and *Notable Key Figures*, featuring influential historical leaders who serve as representatives for country-language association. Table A. 1 entails all our `ALM-bench` categories.

Following [**?**], we adopt an approach to group cultural attributes based on the country for each language. Additionally, we group culturally grounded elements representing shared knowledge, values, and objectives among the people in that culture that are collectively understood. We curated a culturally diverse dataset by collecting data from many cultural aspects (see Fig. A. 1) and manually curated the caption for each image from web-sources. We also show the most frequent words from these categories in Fig. A. 15.
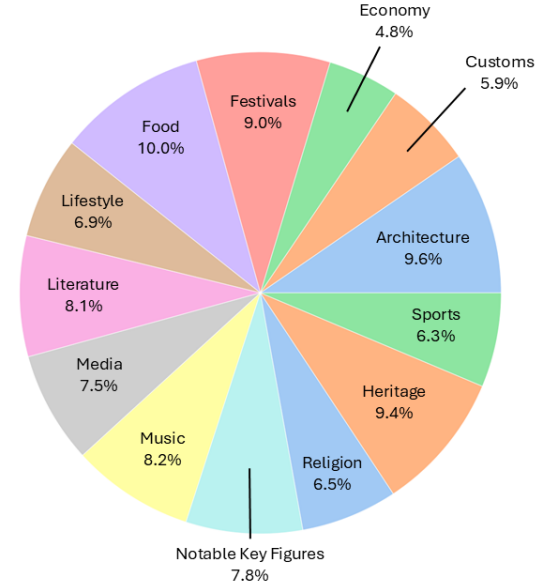
ALM-bench Categories

1. Indoor	11. Music
2. Outdoor	12. Customs
3. Food Items	13. Festivals
4. Memes	14. Heritage
5. Painting	15. Economy
6. Sketch	16. Media
7. Food	17. Architecture
8. Lifestyle	18. Sports
9. Religion	19. Notable Key Figures
10. Literature	

Table A. 1. The 19 categories present in our `ALM-bench` dataset. Note that *Food Items* and *Food* appear in both generic and cultural categories, but have a different context as described in Sec. E.

Fig. A. 1. A breakdown of the cultural categories from our `ALM-bench` is depicted. We ensure consistent samples across all subsets, except for the *Economy* category, where culturally unique images were challenging to find. These visual and question-answer samples are verified and filtered by the native speakers, removing any culturally irrelevant and redundant information.

B. Annotators Demographics

Our `ALM-bench` is the most comprehensive benchmark to date, capturing cultural nuances across 100 languages. Since building such large-scale datasets is challenging, especially due to the limited availability of high-quality native language samples, we enlisted over 60 volunteers to provide expert feedback on our curated examples. In addition to our in-house team, we collaborated with multilingual research communities to bring in external volunteers. These volunteers, representing 50 different countries, bring a deep understanding of their languages' cultural elements and insights unique to their country-language pairs.

Each annotator was given detailed instructions along with examples to guide them in curating high-quality datasets. We also hosted a UI interface as well as gave them

Hindi: Devanagari Script

Original Question:
इस चित्र में दिखाए गए मंदिर की विशेषता क्या है?

Original Answer:
कोणार्क सूर्य मंदिर पर पत्थर पर की गई जटिल नक्काशी और सूर्यदेव को समर्पित रथ का आकार प्रदर्शित है।

Corrected Question:
छवि में दिखाया गया मंदिर कौन सी विशिष्ट विशेषता प्रदर्शित करता है?

Corrected Answer:
कोणार्क सूर्य मंदिर जटिल पत्थर की नक्काशी और भगवान सूर्य को समर्पित एक रथ की आकृति को प्रदर्शित करता है।

Cultural Category: Architecture

Japanese: Japanese Script

Original Question:
画像に表示されている像に描かれた歴史的人物は誰ですか？

Original Answer:
徳川家康 (オプション：徳川家康、織田信長、豊臣秀吉、源頼朝)

Corrected Question:
画像に写っている像は、どの歴史上の人物を表していますか？

Corrected Answer:
徳川家康 (選択肢：徳川家康、織田信長、豊臣秀吉、源頼朝)

Cultural Category: Notable Key Figures

Spanish: Latin Script

Original Question:
¿Cuál es el monumento icónico que se muestra en la imagen conocido como?

Original Answer:
La Alcazar por Sevilla.

Corrected Question:
¿Cómo se conoce el Monumento icónico que se muestra en la imagen?

Corrected Answer:
El Alcázar de Sevilla.

Cultural Category: Heritage

Sanskrit: Devanagari Script

Original Question:
कस्मिन्हि दिने रामनवमी इति उत्सवः चित्रे आविष्क्रियते?

Original Answer:
चैत्रे उज्ज्वलपक्षस्य नवमः। (विकल्पाः : चैत्रे उज्ज्वलपक्षस्य नवमः।, दीपावली, मकरसंक्रान्तिः, चैत्रे कृष्णपक्षस्य पञ्चमी।)

Corrected Question:
चित्रे दर्शितं रामनवमी कस्मिन् दिने आचर्यते?

Corrected Answer:
चैत्रे उज्ज्वलपक्षस्य नवमी। (विकल्पाः : चैत्रे उज्ज्वलपक्षस्य नवमी।, दीपावली, मकरसंक्रान्तिः, चैत्रे कृष्णपक्षस्य पञ्चमी।)

Cultural Category: Food

Fig. A. 2. Qualitative examples of various mistakes in GPT-4o translations including *semantic, cultural, language and grammatical errors*. We employ expert human-feedback to rewrite the correct translations for all samples in our ALM−bench dataset.

the option to directly use the dataset. An example of our annotation platform is shown in Fig. A. 4. Annotators who made substantial contributions are recognized as co-authors of this paper to support high-quality language verification.

We summarize the annotator demographics in Fig. A. 3. To summarize, we show the Age and Gender distribution of our volunteers (*top-left*) indicating the presence of one-fourth of Female annotators and over 46% fall in the 18-25 age limit. We show their levels of language proficiency (*top-right*). Notably, over 80% of the annotators were either native speakers or bilingual, with an additional 14.8% at a proficient level in their chosen language. This high level of expertise ensures that our cultural samples have been meticulously reviewed, filtered, and refined by knowledgeable volunteers. Additionally, we pre-select annotators who are accustomed to the cultural values of their languages. Moreover, (*bottom-left*) depicts that over 88.5% are culturally familiar with their languages, and *bottom-right* highlights that 87.9% of the annotators have lived in the country where we have crafted the cultural samples, making them ideal participants for participating in this research effort.

C. Guidelines for Data Verification

For each country-language pair (e.g., Sri Lanka-Sinhala), QA pairs for general domains are created by translating the original English QA pairs from *LLaVA-Bench (In-the-Wild)* dataset [**?**]. For cultural domains, we curate the dataset by collecting images from online sources using targeted keywords (e.g., country name, language name, cultural cate-
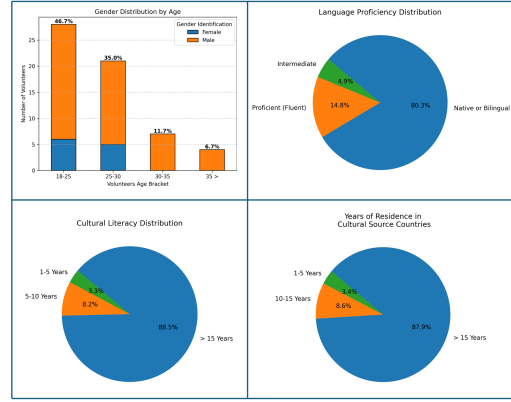

Fig. A. 3. **top-left.** depicts the *Age* and *Gender* distribution of our volunteers. We have one-fourth of the Female speakers and it is also shown that almost 48.3% of our volunteers fall in the age-bracket of 18-25. **top-right.** depicts the volunteer's native language proficiency in years. **bottom-left.** depicts the volunteer's Cultural Literacy proficiency in years. **bottom-right.** depicts the volunteer's native language proficiency in years.

gory) for each language. We also fetch meta-data along with the image, including image caption, image size, and image license. We use such information for post-processing to keep only high-quality images. Then, we perform manual filtering and blurring of personally identifiable data (PID). We discuss this step in detail in Sec. F. The cleaned images are used to generate and translate QA pairs using GPT-4o. To ensure quality, contributors with expertise in the respective native languages manually verify the cultural relevance of images, as well as the accuracy, relevance, and consistent translation of the generated QAs. This section outlines the instructions provided to contributors for verifying the relevance of questions generated for cultural domains and for reviewing the QA translations across both general and cultural domains. Additionally, we use GPT-4o as a judge to score the response and assess the accuracy of answers for cultural category images. This process is detailed further in Sec. J and K.

Each annotator is provided access to a dedicated verification platform that guides them through the list of QA pairs for the selected language (examples are presented in Sec D). Each QA pair includes the following components: Image, Category, English Question, English Answer, Translated Question, and Translated Answer. Contributors are required to follow the two steps outlined below.

Step 1 (Question Relevance): Verify that the question is phrased such that it cannot be answered without the use of a Vision-Language Model (VLM) analyzing the image. For example, a question like "Where is Aura Church located?" can be answered without any visual input. A more appropriate question would be, "Where is the church shown in

the image located?"

Step 2 (Translation Verification): If the translated question or answer is incorrect, provide the correct version in the designated space and classify the error into one of the following four categories.

> **Translation Error Types**
>
> - **Semantic error**: Translation hasn't captured the semantic meaning
> - **Cultural error**: Correct but unusual words uncommon in local language context are used
> - **Language error**: Translation is provided using characters from a different alphabet or most commonly not translated and still given in English
> - **Grammatical error**: Grammar-related errors present

D. Verification Platform and UI Examples

Fig. A. 4. User interface for translation verification hosted on Gradio, allowing contributors to classify incorrect entries and provide accurate translations.

E. Image Web Searching

We automated the process of web scraping for collecting cultural images. A custom URL, embedded with our encoded search query and licensing information, is passed to the Google Search Engine, and images are downloaded from the retrieved links. For each specific cultural domain within a given country-language pair, the search query is constructed as "`{language} {cultural domain description} in {country}`". For privacy and copyright issues, our benchmark only includes images that are either in the public domain or licensed under the Creative Commons license.

> **Cultural Domain Descriptions**
>
> - **Food**: Specific Dishes, Dining Dishes
> - **Lifestyle**: Daily Attire, Daily Life, Modern Lifestyle
> - **Literature**: Famous literature, authors and poets
> - **Music**: Traditional music and dance
> - **Religion**: Major religions and religious festivals
> - **Customs**: Social etiquette and traditional greetings
> - **Festivals**: Important cultural festivals and celebrations
> - **Heritage**: Popular historical heritage sites and iconic landmarks
> - **Economy**: Major economic industries
> - **Media**: Iconic Entertainment Figures and Popular TV Shows, Movies
> - **Architecture**: Traditional Art and Architecture
> - **Sports**: Famous traditional sports
> - **Notable key figures**: Famous historical leaders

F. Blurring the PIDs

Recent research highlights the importance of blurring personally identifiable details (PIDs) in image datasets to mitigate privacy risks and reduce bias [**?**]. Thus, we filter the collected cultural images for cultural relevance, and faces are blurred using a face detection model, except for the 'Media' and 'Notable Key Figures' categories where we have images of public figures or celebrities. The blurred images are then manually reviewed to identify any discrepancies, such as blurring personally identifiable data (PIDs) and removing watermark images, and any remaining issues are addressed when using the PicdeFacer tool.

G. Instruction Prompt for QA Generation

Fig. A. 5 illustrates the prompt messages used to guide GPT-4o in generating culturally relevant QA pairs. Each prompt includes an image, its manually added caption, and the search query used during image scraping. The model is instructed to generate two short QA pairs, two multiple-choice QA pairs, one true/false question, and one long QA pair. Detailed instructions ensure that the generated QAs are culturally grounded and specific to the image, avoiding answers based solely on common knowledge.

H. Instruction Prompt for QA Translations

After the curation of question-answer pairs using GPT-4o, we use the same open-source model to translate the English

```
prompt_messages = |

{ "role":"user", "content":[ { "type": "text", "text": search_query + "\n" + caption }, {"type": "image_url", "image_url": image_url" } ] },

{ "role":"system", "content": f """Here is an image and a caption I have on hand. I'd like you to generate two short questions and answers, two multiple choice questions
and answers, one true/false question, and one long question and answer. Refer to the caption for the context/hint. Take into account the cultural diversity of cultural category.

Follow the following rules while designing questions and answers:

    1.  The question must be answerable only by looking at the image.
    2.  Ensure that the questions are culturally relevant and specific to the image.
    3.  Provide answers that are concise, accurate, and directly related to the question.
    4.  You will also need to provide 1 correct option and 3 other incorrect options (distractors).
        For the distractors, choose options that are relevant, not obvious wrong answers.
    5.  The question must be answerable even without the multiple-choice.
        Example of the invalid question: ("What song is not performed by this musician" – not answerable if you don't know the choices).
    6.  Make sure the questions are written fluently in English.
    7.  Be mindful of cultural sensitivities and avoid stereotyping or misrepresenting cultural aspects.
    8.  Ensure there are variations in your questions. Identity questions are fine, eg "What is this", or "where is this". But additionally.
        For example, multi-hop reasoning, counting, referencing, or questions that require local commonsense knowledge to be answered.
    9.  Just generate these in English.
    10. For short questions and answers, don't keep it very short, include at least 2 sentences.
    11. Make the questions distinct and unique from each other.

Give the answers in the following JSON format and make sure to only output a valid JSON,

{ "short_questions": [ { "question": <question>, "answer": <answer> }],
  "multiple_choice_questions": [ { "question": <question>, "answer": <answer>, "options" <4 options> }],
  "true_false_question": { "question": <question>, "answer": <answer> },
  "long_question": { "question": <question>, "answer": <answer> }""" }

]
```

Fig. A. 5. Prompts used for generation of cultural QA pairs. *search_query* refers to the query used to search the image online. It includes language, country and cultural category details. *caption* is a manually added textual description specific to the image. *cultural category* indicates the domain to which the image belongs, selected from the 13 cultural domains in our ALM-bench dataset.

QA pairs into the native languages. We show the prompt that we used for this task in Fig. A. 6.

```
prompt_messages = |
{ "role":"system", "content": f "You're a helpful
assistant that translates English text to target_lang"},
{ "role":"user", "content": English_input}
]
```

Fig. A. 6. Prompts used for translating cultural QAs using GPT-4o. *English_input* refers to either the English question or answer to be translated into the *target_lang*, the desired target language.

I. Instruction Prompt for LMM Answer Generation

We conduct a comprehensive study by evaluating various state-of-the-art Large Multimodal Models (LMMs), including both open-weight and closed-weight models, on our ALM-bench benchmark. We used different prompts for each question type as shown in Fig. A. 7. We prompt all questions with English system instruction as highlighted by [**?**], suggesting that the use of prompts in English results in the best performance. Finally, we score each model's generated answer with the human-annotated ground truth an-

swers through a scoring system to assess each model's performance on the benchmark.

J. Prompts for GPT-Scoring

In addition to using GPT-4o as a judge in scoring the predicted answers, we also do scoring with multiple other LMMs to ensure a fair evaluation. Results are reported in Table A. 2.

Model Name	GPT-4o	Llama-3.1-8B-Instruct
GPT-4o	90.16%	90.34%
Gemini-1.5 Pro	80.21%	80.65%
GLM-4V-9B	71.35%	73.89%

Table A. 2. We evaluated decision-making questions (both True/-False and multiple-choice) across a sample of 20 randomly selected languages using the LLama-3.1-8B-Instruct model. This assessment aims to ensure consistency in performance on T/F and MCQs when scored using GPT-4o as a judge.

K. Guidelines for Error Analysis on GPT-4o Output

Each annotator is provided with an Excel file specific to a given country-language pair, containing the following columns: Image URL, Question, Ground Truth Answer,

```
messages =[{"role":"user", "content":[{"type": "image_url",
"image_url": image_url }, {"type": "text", "text": prompt}] }]
```

Multiple Choice Questions

```
prompt = (
  f "For the given the Multiple Choice Question Answer below, analyze the
question and answer strictly from one of the options below. Strictly answer the
choice only. No additional text." + question + choices
)
```

True/False Questions

```
prompt = (
question + f "The above question is a True/False question. Please provide the
answer as one word in target_language"
)
```

Long Questions

```
prompt = (
question + f "Answer the question in detail in target_language language"
)
```

Short Questions

```
prompt = (
question + f "Please provide brief, clear responses in target_language language"
)
```

Fig. A. 7. Prompts used to generate answers for cultural questions using multiple Large Multimodal Models evaluated in our study. Different prompts are designed for different question types. *question* refers to the cultural question associated with the given image, previously generated using GPT-4o. *choices* represents the four options provided in multiple choice question type, and *target_language* is the desired local language for the response.

```
messages =[{"role":"system", "content": " You are a helpful Assistant. Provide helpful
response to the user's question."},{"role": "user", "content": prompt_eval}]
```

True/False & Multiple Choice Questions

```
prompt_eval = (
    f "Evaluate the following answer based on Accuracy:"
    f "Question: {question}"
    f "Ground Truth: {ground_truth}"
    f "Model Prediction: {predicted_answer}"
    f "Match the meaning of ground truth with model prediction, if it matches give a 10. Otherwise 0."
    f "Strictly return only the numeric score, without any additional commentary." )
```

Long Questions

```
prompt_eval = (
    f "Evaluate the following answer based on Consistency, Fluency, and Relevance based on the
Ground Truth answer."
    f "A high score example when predicted response match closely with ground truth and a low
score example when predicted response lacks knowledge, or is not related to the ground truth."
    f "Question: {question}"
    f "Ground Truth: {ground_truth}"
    f "Model Prediction: {predicted_answer}"
    f "Provide a single overall score out of 10 based on these three criteria."
    f "Strictly return only the numeric score, without any additional commentary." )
```

Short Questions

```
prompt_eval = (
    f "Evaluate the following answer based on accuracy and correctness based on the Ground
Truth answer."
    f "A high score example when predicted response match closely with ground truth and a low
score example when predicted response lacks knowledge, or is not related to the ground truth."
    f "Question: {question}"
    f "Ground Truth: {ground_truth}"
    f "Model Prediction: {predicted_answer}"
    f "Provide a single overall score out of 10 based on these three criteria."
    f "Strictly return only the numeric score, without any additional commentary." )
```

Fig. A. 8. Prompts used to generate a score between 0 and 10, with GPT-4o acting as the evaluator to compare an LMM's predicted answers against ground truth answers. The terms *question*, *ground_truth*, and *predicted_answer* refer to the cultural question, the ground truth answer generated by GPT-4o and verified by experts, and the model's predicted answer, respectively.

Predicted Answer (all in the native language), and GPT-4o Score. The 'Predicted Answer' column records the response generated by GPT-4o when presented with the Question in the native language. The 'GPT-4o Score' column reflects the evaluation score assigned by GPT-4o, acting as a judge, based on the comparison between the Predicted Answer and the Ground Truth Answer.

Each annotator is required to complete two additional columns: **(1)** 'Is the GPT Score Justified?' with a binary response Yes/No, and **(2)** if the GPT-4o score is not justified, the 'Reason for GPT-4o Failure' column, where they select an appropriate reason from a predefined dropdown menu provided below.

Reason for GPT-4o Failure

- **Lack of Knowledge**: The model fails to find an answer to the question.
- **Lack of Cultural Understanding**: The model fails to understand the cultural aspect of the answer.
- **Language Error**: Some words in the GPT-4o's output are wrong for the language.
- **Reasoning Error**: The reason it gave does not match that language.
- **Translation Error**: A few native words were not properly translated.
- **Perceptual Error**: The model fails to understand where a certain entity is (for eg, the top of the image, the bottom of the image, etc).

L. Qualitative examples with different question types

Next, we present some qualitative examples showing various question types for each image sample from our dataset and in Fig. A. 9, A. 10, A. 11, A. 11, A. 13, and A. 14.

Multiple Choice Question Answers

Ques: تصویر میں کون سا جانور نمایاں طور پر دکھایا گیا ہے؟
(What animal is prominently featured in the image?)
Options:
A) گھوڑا **Horse**
B) اونٹ Camel
C) ہاتھی Elephant
D) Cattle مویشی

Ques: تصویر میں کون سے موسیقی کے آلات نظر آ رہے ہیں؟
(What kind of musical instruments are visible in the image?)
Options:
A) ڈھول **Drums**
B) Flute بانسری
C) Sitar ستار
D) Tabla طبلہ

True/False

Ques: لوگ تقریب کے دوران اپنے گھوڑوں کے ساتھ چل رہے ہیں جیسا کہ تصویر میں دکھایا گیا ہے؟
(People are walking with their horses during the event shown in the image?)
Ans: صحیح(TRUE)

Short Question Answer

Ques: تصویر میں کس قسم کا شو دکھایا گیا ہے؟
(What type of show is depicted in the image?)
Ans: گھوڑوں اور مویشیوں کا شو
(A horse and cattle show)

Long Question Answer

Ques: پاکستانی ثقافت میں گھوڑے اور مویشیوں کے شوز کی اہمیت پر تبادلہ خیال کریں، جن میں عموماً ہونے والے واقعات، ان کا مقامی برادریوں پر اثر اور وہ کس طرح ثقافتی ورثے کو فروغ دیتے ہیں۔
(Discuss the significance of horse and cattle shows in Pakistani culture, including the types of events that typically take place, their impact on local communities, and how they promote cultural heritage.)

Ans:
پاکستانی ثقافت میں گھوڑے اور مویشیوں کے شوز اہم تقریبات ہیں جو کمیونٹی کی زراعتی مہارت اور مویشی پالنے کی صلاحیتوں کو ظاہر کرتی ہیں۔ ان تقریبات میں عام طور پر مختلف مقابلے شامل ہوتے ہیں، جن میں بہترین نسل کے مقابلے، جانوروں کی پریڈ، اور پرفارمنس شوز جیسے گھوڑوں کا رقص شامل ہیں۔ یہ شوز کسانوں اور بریڈرز کو اپنے بہترین جانوروں کو پیش کرنے کا پلیٹ فارم فراہم کرتے ہیں، مقامی زرعی کاروبار کو فروغ دیتے ہیں اور مویشیوں کی نسل بڑھانے کے معیار کو بہتر بنانے کی حوصلہ افزائی کرتی ہیں۔ اقتصادی پہلو سے بڑھ کر، یہ تہوار ثقافتی مظہروں سے بھرپور ہوتے ہیں، جن میں روایتی موسیقی، رقص، اور شاندار سجاوٹیں شامل ہوتی ہیں۔ یہ کمیونٹی کو قریب لاتے ہیں اور ثقافتی ورثے کے تحفظ کی ترغیب دیتے ہیں، شرکاء اور ناظرین دونوں کے لئے ایک تفریحی اور تعلیمی تجربہ فراہم کرتے ہیں۔

Ques: یہ تقریب کس شہر میں ہو رہی ہے؟
(In which city is this event taking place?)
Ans: لاہور
(Lahore)

Language: Urdu
Category: Festivals

Multiple Choice Question Answers

Ques: 这个人头上戴着什么？
(What is this person wearing on her head?)
Options:
A) 凤冠 FengGuan
B) 贝雷帽 Beret
C) 头纱 Veil
D) 头盔 Helmet

Ques: 图片中描绘的是哪种文化服饰？
(Which cultural attire is depicted in the image?)
Options:
A) 现代汉服 Modern Chinese Hanfu
B) 和服 Kimono
C) 纱丽 Sari
D) 韩服 Hanbok

True/False

Ques: 图片中的衣服是传统汉服吗？
(Is the clothing in the image traditional Chinese Hanfu?)

Ans: 是的，是传统汉服。(True, it is traditional Chinese Hanfu.)

Short Question Answer

Ques: 图片中描绘的是什么风格的婚纱？
(What type of clothing is the person in this image wearing?)
Ans: 那个人穿着现代汉服。
(The person is wearing modern Chinese Hanfu clothing.)

Long Question Answer

Ques: 描述图片中的服装上的具体元素和图案，并解释现代汉服在当代中国社会中的文化意义
(Describe the specific elements and patterns seen on the outfit in the image and explain the cultural significance of modern Hanfu attire in contemporary Chinese society.)
Ans:
图片中的服装有一件白色外衣，宽袖口和一件红色内衣，内衣包括一条高腰裙子和一件有花卉图案的上衣。上衣上的花卉刺绣为整个服装增添了精致的细节。现代汉服虽然起源于传统中国文化，但在当代社会中得到了复兴。许多中国年轻人以穿戴汉服的方式来与自己的文化遗产建立联系，表达文化自豪感。这种流行的复苏在各种文化节日、社交媒体和汉服社区中也都有所体现，热爱汉服的人士在这些平台上分享对这种复杂而意义深远的服饰的热情。

(The outfit in the image features a white outer garment with wide sleeves and a red inner garment composed of a high-waisted skirt and a floral top. The floral embroidery on the top adds elaborate detail to the ensemble. Modern Hanfu attire, while rooted in traditional Chinese culture, has seen a revival in contemporary society. Many young people in China embrace Hanfu as a way to connect with their heritage and express cultural pride. This resurgence in popularity is also evident in various cultural festivals, social media, and Hanfu communities, where individuals share their passion for this intricate and significant form of attire.)

Ques: 这个人穿着的衣服主要是什么颜色？
(What colors are predominant in the person's outfit?)
Ans: 衣服主要是红色和白色。
(The outfit predominantly features red and white colors.)

Language: Chinese(Simplified)
Category: Lifestyle

Fig. A. 9. Qualitative examples of various question types in Urdu and Chinese Simplified Language.

Language: Hindi
Category: Heritage

Ques: चित्र में दिखाए गए किले का नाम क्या है?
(What is the name of the fort shown in the image?)
Options:
A) लाल किला **Red Fort**
B) अम्बर किला Amber Fort
C) आगरा किला Agra Fort
D) ग्वालियर किला Gwalior Fort

Ques: लाल किला किस शहर में स्थित है?
(In which city is the Red Fort located?)
Options:
A) नई दिल्ली **New Delhi**
B) जयपुर Jaipur
C) आगरा Agra
D) लखनऊ Lucknow

True/False

Ques: क्या चित्र में दिखाया गया किला लाल बलुआ पत्थर से बना है?
(Is the fort in the image constructed using red sandstone?)
Ans: सही (TRUE)

Short Question Answer

Ques: यह ऐतिहासिक स्थल कहाँ स्थित है?
(Where is this historic site located?)

Ans: नई दिल्ली, भारत।
(New Delhi, India.)

Long Question Answer

Ques: लाल किले के ऐतिहासिक महत्व पर चर्चा करें, जिसमें मुगल युग के दौरान इसकी भूमिका, इसकी वास्तुकला विशेषताएँ और आधुनिक भारत में इसका महत्व शामिल हैं।

(Discuss the historical significance of the Red Fort, including its role during the Mughal era, its architectural features, and its importance in modern India.)

Ans:

लाल किला, जो नई दिल्ली में स्थित है, का निर्माण मुगल सम्राट शाहजहाँ ने 17वीं सदी के मध्य में करवाया था। यह लगभग 200 वर्षों तक मुगल सम्राटों का मुख्य आवास रहा। किला मुगल वास्तुकला का प्रतीक है, जिसे इसकी विशाल लाल बलुआ पत्थर की दीवारों और जटिल डिजाइनों द्वारा पहचाना जाता है। मुगल युग के दौरान, यह एक राजनीतिक और सांस्कृतिक केंद्र के रूप में भारतीय इतिहास में महत्वपूर्ण भूमिका निभाता था। आधुनिक समय में, लाल किला उस स्थल के रूप में महत्वपूर्ण है जहाँ हर साल स्वतंत्रता दिवस पर भारत के प्रधानमंत्री राष्ट्रीय ध्वज फहराते हैं और भाषण देते हैं, जो भारत के इतिहास और संस्कृति में किले की स्थायी विरासत का प्रतीक है।

Ques: चित्र में दिखाए गए इस भवन का नाम क्या है?
(What is the name of this building depicted in the image?)

Ans: लाल किला। *(Red Fort.)*

Language: Sinhala
Category: Sports

Multiple Choice Question Answers

Ques: ඡායාරූපයේ දක්වා ඇති ශ්‍රී ලංකාවේ පුරාණ සටන් කුමය කුමක්ද?
(What ancient martial art from Sri Lanka is shown in the image?)
Options:
A) අංගම්පොර **Angampora**
B) කලාරිපයට්ටු Kalaripayattu
C) මුයි තායි Muay Thai
D) තයිකොන්ඩො Taekwondo

Ques: ඡායාරූපය අනුව මෙම සටන්කලාවේ භාවිතා වන අවි වර්ගය මින් කුමක්ද?
(Which of these weapons is used in angampora according to the image?)
Options:
A) ෂුරිකෙන් Shuriken
B) නන්චැකු Nunchaku
C) බෝ රීට Bo Staff
D) කඩු **Swords**

True/False

Ques: ඡායාරූපයේ පෙන්නුම් කර ඇති සම්මතය සම්ප්‍රදායික ශ්‍රී ලංකාවේ සටන් කලාවකි.
(The image shows a form of traditional Sri Lankan martial arts.)
Ans: ඔව්(TRUE)

Short Question Answer

Ques: ඡායාරූපයේ දක්වා ඇති සටන් කලාව කුමක්ද?
(What martial art is being demonstrated in the image?)
Ans: අංගම්පොර.
(Angampora.)

Long Question Answer

Ques: ඡායාරූපයේ දක්වා ඇති සටන්කලාවේ ප්‍රධාන සංරචක දෙක විස්තර කර සහ එම සංරචකවල වැදගත්කම පැහැදිලි කරන්න.
(Describe the two main components of angampora depicted in the image and explain their significance in the martial art.)

Ans:
අංගම්පොරයේ 'අංගම්' හා 'ඉල්ලාංගම්' යන ප්‍රධාන සංරචක දෙක අන්තර්ගත වේ. 'අංගම්' යනු අත්වලින් සටන් කිරීමයි. 'ඉල්ලංගම්' යනු කඩු සහ අනෙකුත් ආයුධ ප්‍රයෝජනයට ගැනීමයි. මෙම අංග ආත්මාරක්ෂාව සඳහා වැදගත් වනවා මෙන්ම, අභ්‍යාස කරන සටන් තමාගේ වේගවත්බව හා තම කුසලතාව වැඩිදියුණු කරනවා මෙන්ම, ආයුධ වලින් කරන සටන්, සටන් කලාවට නව මාන‍යක් එක් කරයි.

Ques: සටන් කටයුතුවලට අමතරව අංගම්පොර වල ප්‍රගුණ කරන ක්‍රියාව කුමක්ද?
(What distinct activity apart from combat is integrated into angampora?)
Ans: භාවනාව.
(Meditation.)

Fig. A. 10. Qualitative examples of various question types in Hindi and Sinhala Language.

Italian — Music

Language: Italian
Category: Music

Multiple Choice Question Answers

Ques: Quali abiti tradizionali indossano i ballerini nell'immagine?
(What traditional clothing are the dancers wearing in the image?)
Options:
A) Costumi tradizionali folcloristici italiani Traditional Italian folk costumes
B) Gonne scozzesi Scottish kilts
C) Kimoni giapponesi Japanese kimonos
D) Sari indiani ndian sarees)

Ques: Quale tipo di accompagnamento musicale è probabilmente presente in questo evento di danza tradizionale?
(What kind of musical accompaniment is likely present in this traditional dance event?)
Options:
A) Musica folcloristica dal vivo con strumenti come la fisarmonica o il violino. (Live folk music from instruments like the accordion or violin)
B) Rock band (Rock band)
C) Ensemble di jazz (Jazz ensemble)
D) Orchestra sinfonica classica (Classical symphony orchestra)

Short Question Answer

Ques: Che tipo di evento potrebbe rappresentare l'immagine?
(What kind of event does the image likely depict?)
Ans: L'immagine probabilmente rappresenta un evento di danza folcloristica italiana tradizionale.
(The image likely depicts a traditional Italian folk dance event.)

Long Question Answer

Ques: Descrivi gli elementi dell'abbigliamento e della danza tradizionali italiani raffigurati nell'immagine, compresi i dettagli sull'abbigliamento, i movimenti della danza e il significato culturale di tali performance.
(Describe the elements of traditional Italian dress and dance depicted in the image, including details about the attire, the dance moves, and the cultural significance of such performances.)

Ans:
L'immagine mostra ballerini in costumi tradizionali folcloristici italiani, che comprendono elementi come gilet ricamati, grembiuli colorati e gonne con bordi di pizzo per le donne, e cappelli a tesa larga e bandane per gli uomini. I movimenti della danza includono passi e gesti sincronizzati, spesso con l'uso di mani o fazzoletti. Queste performance sono culturalmente significative perché rappresentano tradizioni regionali, celebrano eventi storici e mantengono viva l'eredità culturale trasmettendola di generazione in generazione. Tali eventi sono spesso accompagnati da musica dal vivo, suonata con strumenti tradizionali come la fisarmonica o il violino, che arricchiscono l'atmosfera festosa.-

True/False

Ques: L'immagine mostra una performance di danza moderna?
Does the image show a modern dance performance?)
Ans: Falso (False)

Ques: Cosa tengono in mano i ballerini nell'immagine?
(What are the dancers in the image holding in their hands?)
Ans: I ballerini nell'immagine tengono fazzoletti in mano.
(The dancers in the image are holding handkerchiefs in their hands.)

German — Notable Key Figures

Language: German
Category: Notable Key Figures

Multiple Choice Question Answers

Ques: Welche politische Partei wurde von der Person auf dem Bild angeführt?
(Which political party was led by the person in the image?)
Options:
A) NSDAP (Nazi Party)
B) Kommunistische Partei (Communist Party)
C)Sozialdemokratische Partei (Social Democratic Party)
D)Christlich-Demokratische-Union(Christian Democratic Union)

Ques: Welcher Krieg wird mit der Führung der Person auf dem Bild in Verbindung gebracht?
(Which war is associated with the leadership of the person in the image?)
Options:
A) Zweiter Weltkrieg (World War II)
B) Erster Weltkrieg (World War I)
C) Kalter Krieg (Cold War)
D) Koreakrieg (Korean War)

Short Question Answer

Ques: Wer ist der Nazi-Führer auf diesem Bild?
(Who is the Nazi leader depicted in this image?)
Ans: Adolf Hitler (Adolf Hitler)

Ques: In welchem Zeitraum regierte der Führer auf dem Bild Deutschland?
(During what period did the leader in the image rule Germany?)
Ans: Adolf Hitler regierte Deutschland von 1933 bis 1945..
(Adolf Hitler ruled Germany from 1933 to 1945.)

Long Question Answer

Ques: Erörtern Sie die Auswirkungen und Konsequenzen der Herrschaft Adolf Hitlers auf Deutschland und die Welt, einschließlich der Ereignisse des Zweiten Weltkriegs und des Holocausts.
(Discuss the impact and consequences of Adolf Hitler's rule on Germany and the world, including the events of World War II and the Holocaust.)

Ans:

Die Herrschaft Adolf Hitlers hatte tiefgreifende und verheerende Auswirkungen auf sowohl Deutschland als auch die Welt. Als Führer des Nazi-Deutschlands führten Hitlers aggressive expansionistische Politiken direkt zum Ausbruch des Zweiten Weltkrieges im Jahr 1939. Sein Regime verübte den Holocaust, bei dem sechs Millionen Juden sowie Millionen anderer, die vom Nazi-Staat als unerwünscht angesehen wurden, systematisch in Konzentrationslagern ermordet wurden. Der Krieg selbst führte zum Tod von über 60 Millionen Menschen weltweit, zu massiver Zerstörung in ganz Europa und zu bedeutenden politischen, sozialen und wirtschaftlichen Folgen. Nach dem Krieg wurde Deutschland von den Alliierten besetzt und aufgeteilt, was schließlich zur Teilung in Ost- und Westdeutschland führte. Hitlers Handlungen veränderten die globale Geopolitik grundlegend, führten zur Gründung der Vereinten Nationen und bereiteten den Weg für den Kalten Krieg.

True/False

Ques: Hat die Person auf dem Bild eine bedeutende Rolle im Zweiten Weltkrieg gespielt?
Did the person in the image play a significant role in World War II?)
Ans: Wahr (TRUE)

Fig. A. 11. Qualitative examples of various question types in Italian and German Language.

Language: Spanish
Category: Customs

Multiple Choice Question Answers

Ques: ¿En qué ciudad se está llevando a cabo este evento?
(In which city is this event taking place?)
Options:
A) Barcelona, España (Barcelona, Spain)
B) Madrid, España (Madrid, Spain)
C) Sevilla, España (Seville, Spain)
D) Bilbao, España (Bilbao, Spain)

Ques: ¿Qué papel desempeña la persona con el traje tradicional en este evento?
(What role does the person in the traditional costume play in this event?)
Options:
A) Matador (Matador)
B) Picador (Picador)
C) Banderillero (Bandarillero)
D) Rejoneador (Rejoneador)

True/False

Ques: ¿El evento representado en la imagen es una corrida es una corrida de toros?
(Is the event depicted in the image a traditional Spanish bullfight?)
Ans: Verdadero (TRUE)

Long Question Answer

Ques: Explica la importancia de las corridas de toros en la cultura española, especialmente su historia, los roles involucrados en una corrida, y cómo es vista en la sociedad española contemporánea.
(Explain the significance of bullfighting in Spanish culture, particularly its history, the roles involved in a bullfight, and how it is viewed in contemporary Spanish society.)

Ans:
Las corridas de toros, conocidas como 'corrida de toros', han sido una parte significativa de la cultura española durante siglos, con orígenes que se remontan a la antigua Roma. Involucra una serie de etapas, cada una con roles específicos: el matador, que realiza la estocada final; los picadores, que debilitan al toro; y los banderilleros, que colocan palos coloridos (banderillas) en el toro. A pesar de sus profundas raíces históricas y su fuerte presencia en las tradiciones y festivales españoles, las corridas de toros se han convertido en un tema controvertido en la España moderna. Algunas regiones, como Cataluña, las han prohibido, reflejando un cambio en los valores sociales donde los derechos de los animales y las consideraciones éticas están tomando precedencia sobre las prácticas tradicionales.

Short Question Answer

Ques: ¿Qué tipo de evento se muestra en la imagen?
(What type of event is shown in the image?)
Ans: La imagen muestra una corrida de toros.
(The image shows a bullfight.)

Ques: ¿Qué objeto está sosteniendo el matador en la imagen?
(What item is the matador holding in the image?)
Ans: El matador está sosteniendo una capa.
(The matador is holding a cape.)

Language: Emirati Arabic
Category: Architecture

Multiple Choice Question Answers

Ques: من هو المعماري إلي صمم جناح الإمارات في الصورة؟
(Which architect designed the UAE Pavilion shown in the image?)
Options:
A) Santiago Calatrava سانتياغو كالاترافا
B) Frank Gehry فرانك جيري
C) Zaha Hadid زها حديد
D) Norman Foster نورمان فوستر

Ques: شو هو الإلهام المتوقع لتصميم جناح الإمارات؟
(What inspiration is most likely behind the design of the UAE Pavilion?)
Options:
A) Falcon wings أجنحة الصقر
B) Waves الأمواج
C) Mountains الجبال
D) Palm trees أشجار النخيل

True/False

Ques: هل يعكس تصميم جناح الإمارات عناصر إماراتية تقليدية؟
(Does the design of the UAE Pavilion reflect traditional Emirati elements?)
Ans: صح، يعكس (TRUE, it does)

Long Question Answer

Ques: تكلم عن كيف يعكس تصميم سانتياغو كالاترافا لجناح الإمارات في تطلعات وتراث دولة الإمارات الثقافي 2020إكسبو
(Discuss how Santiago Calatrava's architectural design of the UAE Pavilion at Expo 2020 reflects the aspirations and cultural heritage of the United Arab Emirates.)

Ans:
تصميم سانتياغو كالاترافا لجناح الإمارات في إكسبو 2020يُعتبر معجزة معمارية حديثة تعكس طموحات وتراث الإمارات الثقافي. الهياكل اللي تشبه الأجنحة مستوحاة من حركة الصقور في الطيران، اللي لها رمزية ثقافية ووطنية كبيرة في الإمارات، وتمثل القوة والتراث والروح الطموحة للدولة. التصميم المبتكر للجناح يعبر عن رؤية الإمارات المستقبلية، ويبين التزامها بالاستدامة والتكنولوجيا والتطور مع حفاظها على تقاليدها الثقافية العريقة. المزيج بين التصميم العصري والرمزية الثقافية يبرز الهوية الديناميكية للإمارات على مستوى العالم.

Short Question Answer

Ques: شو موجود في الصورة؟
(What is depicted in the image?)
Ans: الصورة تبين جناح الإمارات اللي صممه سانتياغو كالاترافا.
(The image shows Santiago Calatrava's UAE Pavilion.)

Ques: شو الميزة المعمارية البارزة في المبنى؟
(What unique architectural feature is prominent in the building?)
Ans: المبنى يتميز بهياكل تشبه الأجنحة.
(The building prominently features wing-like structures.)

Fig. A. 12. Qualitative examples of various question types in Spanish and Emirati Arabic Language.

<table>
<tr><td>

Language: **Saudi Arabic**
Category: **Media**

</td><td>

Multiple Choice Question Answers

Ques: من هو أحد الممثلين في المسلسل التلفزيوني سكّة سفر؟
(Who is one of the actors in the TV series Sekket Safar?)
Options:
A) **Mohammed Al-Shehri** محمد الشهري
B) Nasser Al Qasabi ناصر القصبي
C) Abdulmohsen Al-Nemr عبدالمحسن النمر
D) Rashed Essa راشد عيسى

Ques: ما هو نوع المسلسل التلفزيوني سكّة سفر؟
(What genre is the TV series Sekket Safar?)

Options:
A) **Comedy** كوميديا
B) Drama دراما
C) Horror رعب
D) Thriller إثارة

True/False

Ques: هل سكّة سفر هو مسلسل تلفزيوني غرض في عام 2022؟
(Is Sekket Safar a TV series that aired in 2022?)
Ans: صحيح (TRUE)

</td><td>

Long Question Answer

Ques: ناقش الأهمية الثقافية والترفيهية لمسلسل سكّة سفر في المملكة العربية السعودية، بالنظر إلى إنتاجه من قِبل MBC وتصويره لأنماط الحياة السعودية.
(Discuss the cultural and entertainment significance of the TV series Sekket Safar in Saudi Arabia, considering its production by MBC and the portrayal of Saudi lifestyles.)
Ans: سكّة سفر هو مسلسل كوميدي أنتجته MBC في عام 2022، يحمل أهمية ثقافية وترفيهية كبيرة في المملكة العربية السعودية. يصور المسلسل جوانب متنوعة من الحياة والثقافة السعودية، مقدماً للمشاهدين نظرة طريفة ومعبرة عن العادات المحلية والأعراف الاجتماعية والتجارب اليومية. يبرز المسلسل، الذي يشارك فيه ممثلون مثل محمد الشهري وسعد عزيز وصالح أبو عمرو، نمو صناعة الترفيه في المنطقة. إنتاج المسلسل من قبل شبكة رائدة مثل MBC يضمن جودة محتوى عالية تتفاعل مع الجمهور السعودي، مما يعكس قصصهم ويوفر لهم ترفيهًا مرتبطًا بهم. لذا، يسهم سكّة سفر في تنويع وإثراء المشهد الإعلامي السعودي.

</td></tr>
<tr><td colspan="3">

Short Question Answer

</td></tr>
<tr><td></td><td>

Ques: ما هو عنوان المسلسل التلفزيوني المعروض في الصورة؟
(What is the title of the TV series shown in the image?)
Ans: سكّة سفر
(Sekket Safar)

</td><td>

Ques: أي شبكة تلفزيونية أنتجت المسلسل التلفزيوني في الصورة؟
(Which TV network produced the TV series in the image?)
Ans: MBC
(MBC)

</td></tr>
<tr><td>

Language: **Afrikaans**
Category: **Economy**

</td><td>

Multiple Choice Question Answers

Ques: Watter stad in Suid-Afrika word op die beeld uitgebeeld?
(What South African city is depicted in the image?)
Options:
A) **Johannesburg (Johannesburg)**
B) Kaapstad (Cape Town)
C) Durban (Durban)
D) Pretoria (Pretoria)

Ques: In watter bedryf is Johannesburg veral bekend?

(Which industry is Johannesburg widely known for?)
Options:
A) **Mynbou (Mining)**
B) Landbou (Agriculture)
C) Toerisme (Tourism)
D) Mode (Fashion)

True/False

Ques: Die beeld toon die skyline van Johannesburg.
(The image represents the skyline of Johannesburg.)
Ans: Waar (TRUE)

</td><td>

Long Question Answer

Ques: Verduidelik die rol van Johannesburg in die Suid-Afrikaanse ekonomie, met spesifieke klem op sy historiese agtergrond in mynbou en ander belangrike ekonomiese sektore wat tot sy welvaart bydra.
(Explain the significance of Johannesburg in South Africa's economy, highlighting its historical background in mining and other major economic industries that contribute to its prosperity.)

Ans:

Johannesburg, dikwels beskryf as die ekonomiese spilpunt van Suid-Afrika, het vinnig gegroei na die ontdekking van goud in die laat 19de eeu en het steeds 'n noemenswaardige ekonomiese rol. Behalwe mynbou, het Johannesburg oor tyd gediversifiseer na ander sektore, insluitend finansies, vervaardiging, telekommunikasie en kleinhandel. Die stad huisves die Johannesburgse Effektebeurs (JSE), wat sy finansiële markposisie versterk. Met sy skyline as simbool van 'n moderne en dinamiese ekonomie, trek Johannesburg beleggers plaaslik en internasionaal, en speel dit 'n sleutelrol in Suid-Afrika se ekonomiese landskap.

</td></tr>
<tr><td colspan="3">

Short Question Answer

</td></tr>
<tr><td></td><td>

Ques: Watter stad word as hoofstad op die beeld vertoon?
(What is the main city shown in the image?)
Ans: Johannesburg
(Johannesburg)

</td><td>

Ques: Watter toringagtige struktuur is prominent op die beeld?
(Which structure is prominently featured with a tower-like design?)
Ans: Die Hillbrow Toring (The Hillbrow Tower)

</td></tr>
</table>

Fig. A. 13. Qualitative examples of various question types in Saudi Arabic and Afrikaans Language.

South Africa-Afrikaans
Category: Music

Short Question Answers

Waarop speel die Hausa-musikante in hierdie beeld?
(What are Hausa musicians in this image performing on?)

Hulle speel op verskeie tradisionele instrumente, insluitend dié wat van dierehorings gemaak is.

(They are performing on various traditional instruments, including those made from animal horns.)

Italy- Italian
Category: Food

True False

Descrivi gli ingredienti tipici e il metodo di preparazione della Zuppa di Minestrone raffigurata nell'immagine.

(The bread shown in the image is a type of Italian bread.)

Vero (True)

France-French
Category: Heritage

Multiple Choice Question Answers

Comment s'appelle le monument dans l'image?
(What is the monument in the image called?)

a) cathédrale Notre-Dame de Paris (Cathedral Notre-Dame of Paris)
b) Tour Eiffel (Eiffel Tower)
c) Arc de Triomphe), (Arc de Triomphe)
d) Pyramide du Louvre(Louvre Pyramid))

China-Chinese Simplified
Category: Media

Short Question Answers

(哪个电视频道播出 "中国好声音"？
(Which television network broadcasts 'The Voice of China'?)

浙江电视台
(Zhejiang Television)

India-Hindi
Category: Religion

Short Question Answers

इस चित्र में कौन सा त्यौहार मनाया जा रहा है?
(What festival is being celebrated in this image?)
मेवाड़ महोत्सव.
(The Mewar Festival.)

Germany-German
Category: Literature

True False

Befindet sich die Skulptur im Bild in Dresden?
(Is the sculpture in the image located in Dresden?.)

Falsch (False)

Bulgaria-Bulgarian
Category: Architecture

Multiple Choice Question Answers

Каква традиционна архитектурна особеност може да се види в сградата отляво?
(What traditional architectural feature can be observed in the building on the left?)

a) Сградата отляво се отличава с богато украсени дървени резби и цветни фрески. ✓ (The building on the left features ornate wooden carvings and colorful frescoes.
b) минималистични бетонни конструкции ✗
(Minimalist concrete structures)

Russia-Russian
Category: Festivals

Short Question Answers

Что изображено на картинке?
(What is depicted in the image?)

На изображении изображено празднование Масленицы с центральной фигурой, олицетворяющей Масленицу в традиционной одежде.

(The image images a celebration of Maslenitsa with a central figure representing Maslenitsa in traditional clothing.)

India-Marathi
Category: Notable key figures

Short Question Answers

शिवाजीचा पुतळा कोठे आहे?
(Where is this statue of Shivaji located?)

हा शिवाजीचा पुतळा भारतातील महाराष्ट्रातील रायगड किल्ल्यावर आहे
(This statue of Shivaji is located at Raigarh Fort in Maharashtra, India.)

Kazakhstan-Kazakh
Category: Economy

True False

Бәйтерек монументі Қазақстанда орналасқан ба?

(Is the Baiterek monument located in Kazakhstan?

ШЫН (True)

Sri Lanka-Sinhala
Category: Customs

Multiple Choice Question Answers

මෙම රූපයේ සිටින මධ්‍යම යුවළගේ වම් පසින් සාම්ප්‍රදායික ඇඳුමින් සැරසී සිටින පුද්ගලයන් බොහෝ විට නියෝජනය කරන්නේ කවුරුන්ද?

(Who are the people in traditional attire to the left of the central couple in this image most likely representing?)

a) සාම්ප්‍රදායික බෙර වාදකයන් සහ නර්තන ශිල්පීන් (Traditional drummers and dancers) ✓
b) විවාහ අමුත්තන් (Wedding guests) ✗
c) යුවළගේ පවුලේ සාමාජිකයන්(Family members of the couple) ✗

Bangladesh-Bengali
Sports

Short Question Answers

বাচ্চারা ছবিতে কি খেলা খেলছে?
(What game are the kids playing in the image?)

কাঁটার উপর ঝাঁপিয়ে পড়া।
(Jumping over thorns.)

Fig. A. 14. Some more qualitative examples of various question types from our benchmark.

Fig. A. 15. Word clouds depicting prominent concepts from 19 categories in our ALM-bench. For intriguing results, we plot and demonstrate the results on English samples of the plot representing both the cultural and generic elements for the entire 100 languages.

S#No	Language	Country	Script	Family	Subgrouping	Specification
1	Afrikaans	South Africa	Latin	Indo-European	Germanic	High
2	Albanian	Albania	Latin	Indo-European	Albanian	High
3	Amharic	Ethiopia	Ge'ez	Afro-Asiatic	Semitic	Low
4	Armenian	Armenia	Armenian	Indo-European	Armenic	Low
5	Assamese	India	Bengali	Indo-European	Indo-Aryan	Low
6	Azerbaijani	Azerbaijan	Latin	Turkic	Common Turkic	Low
7	Basque	Spain	Latin	Isolate	–	High
8	Belarusian	Belarus	Cyrillic	Indo-European	Balto-Slavic	Low
9	Bengali	Bangladesh	Bengali	Indo-European	Indo-Aryan	High
10	Bhojpuri	India	Devanagari	Indo-European	Indo-Aryan	Low
11	Bosnian	Bosnia	Latin	Indo-European	Balto-Slavic	High
12	Bulgarian	Bulgaria	Cyrillic	Indo-European	Balto-Slavic	High
13	Catalan	Spain	Latin	Indo-European	Italic	High
14	Cebuano	Philippines	Latin	Austronesian	Malayo-Polynesian	Low
15	Chinese Simplified	China	Chinese	Sino-Tibetan	Sinitic	High
16	Chinese Traditional	Hong Kong	Chinese	Sino-Tibetan	Sinitic	High
17	Croatian	Croatia	Latin	Indo-European	Balto-Slavic	High
18	Czech	Czech Republic	Latin	Indo-European	Balto-Slavic	High
19	Danish	Denmark	Latin	Indo-European	Germanic	High
20	Dutch	Netherlands	Latin	Indo-European	Germanic	High
21	Egyptian Arabic	Egypt	Arabic	Afro-Asiatic	Semitic	Low
22	Emirati Arabic	United Arab Emirates	Arabic	Afro-Asiatic	Semitic	High
23	English	United Kingdom	Latin	Indo-European	Germanic	High
24	Estonian	Estonia	Latin	Uralic	Finnic	High
25	Filipino	Philippines	Latin	Austronesian	Malayo-Polynesian	Low
26	Finish	Finland	Latin	Uralic	Finnic	High
27	French	France	Latin	Indo-European	Italic	High
28	Galician	Spain	Latin	Indo-European	Italic	Low
29	Georgian	Georgia	Georgian	Kartvelian	Kartvelian	Low
30	German	Germany	Latin	Indo-European	Germanic	High
31	Greek	Greece	Greek	Indo-European	Graeco-Phrygian	High
32	Gujarati	India	Gujarati	Indo-European	Indo-Aryan	Low
33	Hausa	Nigeria	Latin	Afro-Asiatic	Chadic	Low
34	Hawaiian	United States	Latin	Austronesian	Malayo-Polynesian	Low
35	Hebrew	Israel	Hebrew	Afro-Asiatic	Semitic	High
36	Hindi	India	Devanagari	Indo-European	Indo-Aryan	High
37	Hungarian	Hungary	Latin	Uralic	–	High
38	Icelandic	Iceland	Latin	Indo-European	Germanic	High
39	Igbo	Nigeria	Latin	Atlantic-Congo	Benue-Congo	Low
40	Indonesian	Indonesia	Latin	Austronesian	Malayo-Polynesian	High
41	Irish	Ireland	Latin	Indo-European	Celtic	Low
42	Italian	Italy	Latin	Indo-European	Italic	High
43	Japanese	Japan	Kanji/Kana	Japonic	Japanese Ryukyuan	High
44	Javanese	Indonesia	Latin	Austronesian	Malayo-Polynesian	Low
45	Kannada	India	Kannada	Dravidian	South Dravidian	Low
46	Kazakh	Kazakhstan	Cyrillic	Turkic	Common Turkic	High
47	Kinyarwanda	Rwanda	Latin	Atlantic-Congo	Benue-Congo	Low
48	Korean	South Korea	Hangul	Koreanic	Korean	High
49	Kurdish	Turkey	Arabic	Indo-European	Iranian	Low
50	Kyrgyz	Kyrgyzstan	Cyrillic	Turkic	Common Turkic	Low

S#No	Language	Country	Script	Family	Subgrouping	Specification
51	Lao	Thailand	Lao	Tai-Kadai	Kra-Dai	Low
52	Latin	Vatican City	Latin	Indo-European	Italic	Low
53	Latvian	Latvia	Latin	Indo-European	Balto-Slavic	High
54	Lithuanian	Lithuania	Latin	Indo-European	Balto-Slavic	High
55	Luxembourgish	Luxembourg	Latin	Indo-European	Germanic	Low
56	Macedonian	North Macedonia	Cyrillic	Indo-European	Balto-Slavic	High
57	Malagasy	Madagascar	Latin	Austronesian	Malayo-Polynesian	Low
58	Malay	Malaysia	Latin	Austronesian	Malayo-Polynesian	High
59	Malayalam	India	Malayalam	Dravidian	South Dravidian	Low
60	Maltese	Malta	Latin	Afro-Asiatic	Semitic	High
61	Marathi	India	Devanagari	Indo-European	Indo-Aryan	Low
62	Mongolian	Mongolia	Cyrillic	Mongolic-Khitan	Mongolic	Low
63	Myanmar (Burmese)	Myanmar	Myanmar	Sino-Tibetan	Burmo-Qiangic	Low
64	Nepali	Nepal	Devanagari	Indo-European	Indo-Aryan	Low
65	Norwegian	Norway	Latin	Indo-European	Germanic	Low
66	Odia (Oriya)	India	Oriya	Indo-European	Indo-Aryan	Low
67	Pashto	Pakistan	Arabic	Indo-European	Iranian	Low
68	Persian	Iran	Arabic	Indo-European	Iranian	High
69	Polish	Poland	Latin	Indo-European	Balto-Slavic	High
70	Portuguese	Portugal	Latin	Indo-European	Italic	High
71	Punjabi	Pakistan	Gurmukhi	Indo-European	Indo-Aryan	Low
72	Romanian	Romania	Latin	Indo-European	Italic	High
73	Russian	Russia	Cyrillic	Indo-European	Balto-Slavic	High
74	Sanskrit	India	Devanagari	Indo-European	Indo-Aryan	Low
75	Saudi Arabic	Saudi Arabia	Arabic	Afro-Asiatic	Semitic	High
76	Scots Gaelic	Scotland	Latin	Indo-European	Celtic	Low
77	Serbian	Serbia	Cyrillic	Indo-European	Balto-Slavic	Low
78	Shona	Zimbabwe	Latin	Atlantic-Congo	Benue-Congo	Low
79	Sindhi	Pakistan	Arabic	Indo-European	Indo-Aryan	Low
80	Sinhala	Sri Lanka	Sinhala	Indo-European	Indo-Aryan	Low
81	Slovak	Slovakia	Latin	Indo-European	Balto-Slavic	High
82	Slovenian	Slovenia	Latin	Indo-European	Balto-Slavic	High
83	Somali	Somalia	Latin	Afro-Asiatic	Cushitic	Low
84	Spanish	Spain	Latin	Indo-European	Italic	High
85	Sundanese	Indonesia	Latin	Austronesian	Malayo-Polynesian	Low
86	Swahili	Tanzania	Latin	Atlantic-Congo	Benue-Congo	High
87	Swedish	Sweden	Latin	Indo-European	Germanic	High
88	Tajik	Tajikistan	Cyrillic	Indo-European	Iranian	Low
89	Tamil	India	Tamil	Dravidian	South Dravidian	Low
90	Telugu	India	Telugu	Dravidian	South Dravidian	Low
91	Thai	Thailand	Thai	Tai-Kadai	Kam-Tai	High
92	Turkish	Turkey	Latin	Turkic	Common Turkic	High
93	Ukrainian	Ukraine	Cyrillic	Indo-European	Balto-Slavic	High
94	Urdu	Pakistan	Arabic	Indo-European	Indo-Aryan	Low
95	Uyghur	China	Arabic	Turkic	Common Turkic	Low
96	Uzbek	Uzbekistan	Latin	Turkic	Common Turkic	High
97	Vietnamese	Vietnam	Latin	Austroasiatic	Vietic	High
98	Welsh	United Kingdom	Latin	Indo-European	Celtic	Low
99	Yiddish	Israel	Hebrew	Indo-European	Germanic	Low
100	Yoruba	Nigeria	Latin	Atlantic-Congo	Benue-Congo	Low

Table A. 3. A comprehensive list of 100 languages, their associated country, language scripts, families, subgrouping, and the resource specification.

We thank reviewers for positive feedback: attempts to improve data quality incl. native speakers verification (**C9ZZ**, **ic3S**, **KCss**), being diverse, and incl. culturally relevant content (**C9ZZ**), comprehensive and valuable VQA benchmark (**ic3S**), significant efforts towards linguistic diversity, esp. for low-resource languages, extensive 13 distinct cultural categories (**KCss**). Our **code & dataset** will be made public. All suggested changes will be reflected in final draft.

C9ZZ: **On conflating "Language" and "Culture"**: We agree that the mapping is not one-to-one for each language and its respective culture. The suggestion to explore many-to-many mappings is indeed a valuable direction for future work. However, to ensure the curation of high-quality cultural samples on a larger scale for our multilingual multimodal benchmark, we adopted a simplified assumption by restricting mappings to a single <country-language> pair. The mappings are in the suppl. mat. (Tab. A.3), we will clarify this approach further in the final draft.

ic3S: **On using GPT-4o for automatic evaluation**: We have also evaluated a sample of 20 languages on True/False and MCQs using LLaMA-3.1-8B-Instruct to ensure reproducible and consistent results on *GPT-4o*, *Gemini-1.5-Pro*, and *GLM-4V-9B*. We show these results in the suppl. (Tab. A.2). As recommended, we conducted a **Human Evaluation** on a randomly selected subset of 100 samples across 8 languages. The results (Fig. 16) show the comparable scores of GPT-4o to human evaluation, verifying the effectiveness of LLM-as-a-judge in ALM-bench. **Higher performance of Long-VQAs compared to Short-VQAs**: Each question type was evaluated on different criteria as mentioned in L275-L286. While short questions are evaluated solely on *correctness*, long questions are assessed on *consistency*, *fluency*, and *relevancy*. The inclusion of *fluency* and *consistency* in evaluating long questions favor LLMs, as instruction tuning and RLHF alignment make the responses inherently more coherent and well-structured. Moreover, long responses allow the model to capture a broad range of information, increasing the likelihood of partial correctness (covered under relevancy) in a multilingual setting, even if some parts are incorrect. Existing work [6,25,43] also uses similar criteria to evaluate long questions. **Purpose of General VQA in ALM-Bench**: The inclusion of General VQA in ALM-bench establishes a baseline for evaluating multilingual visual understanding in 100 languages. It ensures models perform well on generic, universally relevant scenes (e.g., indoor and outdoor) w/o specific cultural nuances, allowing a comprehensive assessment of their visual and linguistic generalization capabilities. We also show **performance of general and cultural VQA separately** in Fig. 17. Since the general VQAs are only open-ended, we compare them with open-ended VQAs from the cultural part. Open-source
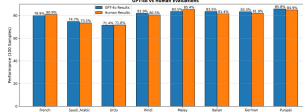

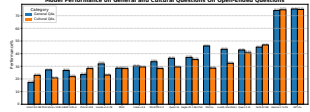Fig. A. 16. Human vs GPT-4o performance comparison.

Fig. A. 17. Results of general vs cultural open-ended VQAs.

Fig. A. 18. Examples of human verification step on a low-resource language, *Sinhala* to ensure cultural relevance in ALM-bench.

models, like Qwen-2-VL and LLaVA-OneVision, perform better on general questions, highlighting challenges with culturally specific ones. In contrast, closed-source models show near-identical performance across both, indicating better handling of cultural aspects. **To ensure the filtration of misclassified culture** category-language pair sample, we have two strategies embedded in our verification protocol: 1) Each QA pair is assigned a unique (country, language, cultural domain) triplet to prevent mismatches, ensuring correct associations with a single language. 2) Human annotators verify these mappings to ensure consistency. We will clarify this in the main text. **Minors**: 1) We will elaborate on country-specific prompts in the final version. Our experiments used structured triplets (language, country, cultural domain) in prompts. E.g., prompt = ques + "Provide brief, clear responses in *'lang'* language. The image represents *'cultural'* cat. in *'country'*." 2) The number of languages in each script in **Fig. 7** is similar. We will update it in the revised version.

KCss: **Performance variation in various cultural domains** stems from biases in LLM's training data (salewski2023context) and vision backbones (zhang2023vitaev2), which typically favor well-represented cultures. Language nuances and instruction tuning also impact performance, with closed-source models often outperforming open-source ones due to better alignment with diverse datasets. To **better utilize visual cues**, aligning vision and language components with fine-grained training samples improves LMM performance. In multilingual settings, linking visual representations of rich scenes with diverse, multilingual data can strengthen LMM's ability to interpret visual cues. **Leveraging error categories refines training and model arch.** Grammatical and semantic errors arise from limited multilingual data, highlighting the need for diverse datasets. Cultural errors reflect domain gaps, addressed by high-quality visual samples with detailed annotations. Incorporating these datasets during model training can significantly reduce errors. **Ensuring Cultural relevance in ALM-bench**: Our native experts verified and annotated misrepresented cultural concepts to ensure cultural accuracy across languages. We show an example in Fig. 18.

No.	Author Name	Affiliation	Email
1	Henok Biadglign Ademtew	Ethiopian Artificial Intelligence Institute	henokb2124@gmail.com
2	Yahya Hmaiti	University of Central Florida	yohan.hmaiti@ucf.edu
3	Amandeep Kumar	Johns Hopkins University	kumar.amandeep015@gmail.com
4	Kartik Kuckreja	MBZUAI	kartik.kuckreja@mbzuai.ac.ae
5	Mykola Maslych	University of Central Florida	mykola.maslych@ucf.edu
6	Wafa Al Ghallabi	MBZUAI	wafa.alghallabi@mbzuai.ac.ae
7	Mihail Mihaylov	MBZUAI	mihail.mihaylov@mbzuai.ac.ae
8	Chao Qin	MBZUAI	1746625542@qq.com
9	Abdelrahman M Shaker	MBZUAI	abdelrahman.youssief@mbzuai.ac.ae
10	Mike Zhang	Aalborg University	jjz@cs.aau.dk
11	Mahardika Krisna Ihsani	MBZUAI	mahardika.ihsani@mbzuai.ac.ae
12	Amiel Esplana	University of the West of England, Bristol	amiel2.esplana@live.uwe.ac.uk
13	Monil Gokani	IIT, Hyderabad	monilgokani08@gmail.com
14	Shachar Mirkin	Alpinference	shacharmirkin@gmail.com
15	Harsh Singh	MBZUAI	harsh.singh@mbzuai.ac.ae
16	Ashay Srivastava	University of Maryland, College Park	ashays06@umd.edu
17	Endre Hamerlik	HUN-REN Institute for CS and Control	hamerlik@sztaki.hu
18	Fathinah Asma Izzati	MBZUAI	fathinah.izzati@mbzuai.ac.ae
19	Fadillah Adamsyah Maani	MBZUAI	fadillah.maani@mbzuai.ac.ae
20	Sebastian Cavada	MBZUAI	sebastian.cavada@mbzuai.ac.ae
21	Jenny Chim	Queen Mary University of London, UK	jennychim@gmail.com
22	Rohit Gupta	University of Central Florida	rohitgupta.hpf@gmail.com
23	Sanjay Manjunath	MBZUAI	sanjay.manjunath@mbzuai.ac.ae
24	Kamila Zhumakhanova	MBZUAI	kamila.zhumakhanova@mbzuai.ac.ae
25	Feno Heriniaina Rabevohitra	Chongqing University	fenoheriniaina@gmail.com
26	Azril Amirudin	University of the People	azrilamirudin@my.uopeople.edu
27	Muhammad Ridzuan	MBZUAI	20020084@mbzuai.ac.ae
28	Daniya Kareem	MBZUAI	daniya.kareem@mbzuai.ac.ae
29	Ketan More	MBZUAI	ketan.more@mbzuai.ac.ae
30	Kunyang Li	University of Central Florida	kunyang.li@ucf.edu
31	Pramesh Shakya	University of Central Florida	pramesh.shakya@ucf.edu
32	Muhammad Saad	MBZUAI	muhammad.saad@mbzuai.ac.ae
33	Amirpouya Ghasemaghaei	University of Central Florida	aghaei.ap@ucf.edu
34	Amirbek Djanibekov	MBZUAI	amirbek.djanibekov@mbzuai.ac.ae
35	Dilshod Azizov	MBZUAI	dilshod.azizov@mbzuai.ac.ae
36	Branislava Jankovic	MBZUAI	branislava.jankovic@mbzuai.ac.ae
37	Naman Bhatia	Indian Institute of Technology, Hyderabad	naman.219311175@muj.manipal.edu
38	Alvaro Cabrera	MBZUAI	alvaro.berobide@mbzuai.ac.ae
39	Johan Obando-Ceron	University of Montreal, Mila	jobando0730@gmail.com
40	Olympiah Otieno	JKUAT	anikaolympiah@gmail.com
41	Fabian Farestam	ETHZ - ETH Zurich	ffarestam@student.ethz.ch
42	Muztoba Rabbani	University of California, Merced	mrabbani@ucmerced.edu
43	Sanoojan Baliah	MBZUAI	sanoojan.baliah@mbzuai.ac.ae
44	Santosh Sanjeev	MBZUAI	santosh.sanjeev@mbzuai.ac.ae
45	Abduragim Shtanchaev	MBZUAI	abduragim.shtanchaev@mbzuai.ac.ae
46	Maheen Fatima	Air University	231659@students.au.edu.pk
47	Thao Nguyen	MBZUAI	thao.nguyen@mbzuai.ac.ae
48	Amrin Kareem	MBZUAI	amrin.kareem@mbzuai.ac.ae
49	Toluwani Aremu	MBZUAI	toluwani.aremu@mbzuai.ac.ae
50	Nathan Xavier	Universidade Federal de Minas Gerais	nathanxavier@ufmg.br
51	Amit Bhatkal	MBZUAI	amitbhatkal12@gmail.com
52	Hawau Olamide Toyin	MBZUAI	hawau.toyin@mbzuai.ac.ae

Table A. 4. List of affiliations for all the volunteer co-authors who contributed to construct ALM-Bench.