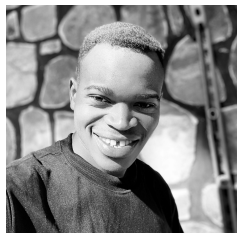
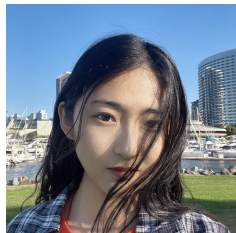


Grounding Multilingual Multimodal LLMs With Cultural Knowledge



Jean Nyandwi



Yueqi Song



Simran Khanuja



Graham Neubig



AI that serves the world should know more than Hollywood and Wikipedia – yet most open-source multimodal LLMs don't.



Qwen3-VL



Molmo



Aya Vision



Issues in Multimodal LLMs

- Multimodal LLMs have deficits in understanding world cultures and long-tail entities.
- Even frontier models lags behind when it comes to cultures and open-source models are way behind.
- The cultural gap in MLLM:
 - **Cultural nuances scarcity** in existing large multimodal datasets(LAION, LLaVA-Ins, etc...)
 - Existing cultural datasets are **small, limited cultural domains, lack of long-tail entities coverage.**



Qwen3-VL



Molmo



Aya Vision



CulturalGround

CulturalGround addresses major issues in existing cultural-centric datasets:

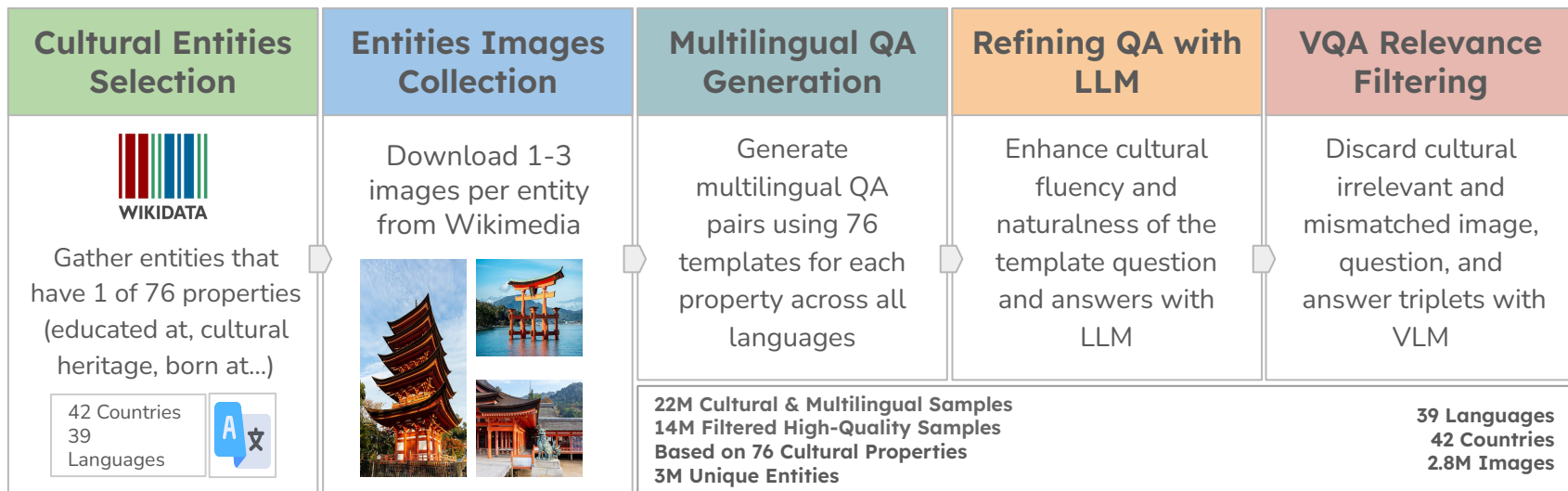
- **Multilingual coverage:** 39 languages
- **Long-tail entities** coverage
- **Visual grounded:** 1-3 images per entity
- A wide array of **culture domains**(people, places, schools, museums, heritages, etc..)
- **Size:** 30M, the largest open-source cultural VQA dataset.

Grounding Multilingual Multimodal LLMs With Cultural Knowledge

Jean de Dieu Nyandwi **Yueqi Song** **Simran Khanuja** **Graham Neubig**
{jeandedi, yueqi, skhanuja, gneubig}@andrew.cmu.edu
Carnegie Mellon University

<https://neulab.github.io/CulturalGround/>

CulturalGround Curation Pipeline



CulturalGround curation pipeline automatically create questions and answers about diverse set of entities from select regions.

CulturalGround Curation MCQs



Using selected cultural entities and multilingual QAs as **seed**, we also generate multiple choice questions(MCQs) subset grounded in cultures using LLM.

CulturalGround MCQs
8M Cultural & Multilingual Samples
6M Filtered High-Quality Samples
Based on 76 Cultural Properties

A Walkthrough: Wikidata Homepage

The screenshot displays the Wikidata homepage with a network diagram in the background. The diagram consists of nodes connected by lines, with labels like 'open', 'multilingual', 'collaborative', 'linked', and 'structured' placed near specific nodes. The main content area features the title 'Welcome to Wikidata' and a description: 'the free knowledge base with 119,105,424 data items that anyone can edit.' Below this is a list of links: 'Introduction • Project Chat • Community Portal • Help' and a prompt 'Want to help translate? Translate the missing messages.' The top navigation bar includes the Wikidata logo, a search bar with 'Akagera National Park' entered, and links for 'English', 'Donate', 'Create account', and 'Log in'. The bottom of the page has two buttons: 'Welcome!' and 'Learn about data'.

WIKIDATA

Items Search

English Donate Create account Log in

Main Page Discussion Read View source View history Tools

Welcome to Wikidata

the free knowledge base with 119,105,424 data items that anyone can edit.

[Introduction](#) • [Project Chat](#) • [Community Portal](#) • [Help](#)

Want to help translate? [Translate the missing messages.](#)

Appearance hide

Text

☐ Small

☒ Standard

☐ Large

Width

☒ Standard

☐ Wide

Welcome!

Learn about data

Entity in Wikidata: Akagera National Park

Akagera National Park (Q116715)

[Item](#) [Discussion](#)

[Read](#) [View history](#) [Tools](#) ▼

national park in Rwanda [edit](#)

[In more languages](#)
[Configure](#)

Language	Label	Description	Also known as
default for all languages	Akagera National Park	—	
English	Akagera National Park	national park in Rwanda	
Kinyarwanda	Pariki y'Akagera	No description defined	
French	parc national de l'Akagera	parc national au Rwanda	"Parc national de l'Akagera"

[All entered languages](#)

Statements

instance of

national park

[edit](#)

~ 0 references

[+ add reference](#)


park

[edit](#)


> 1 reference

**Property
P31**

image

[edit](#)

[Parc National d'Akagera.jpg](#)
1,600 × 1,063; 271 KB
[P](#)
[0 references](#)

[edit](#)

[KAGERA National Park Entry Point.jpg](#)
4,032 × 2,268; 4.13 MB
[P](#)
[0 references](#)

Entity in Wikidata: Akagera National Park

country	<div><div><div></div><div></div></div><div>Rwanda</div></div> <div>► 1 reference</div>
located in the administrative territorial entity	<div><div><div></div><div></div></div><div>Eastern Province</div></div> <div>▼ 0 references</div>
shares border with	<div><div><div></div><div></div></div><div>Mutara Game Reserve</div></div> <div>► 2 references</div>


visitors per year	<div><div><div></div><div></div></div><div>34,000</div></div> <div>point in time2022</div> <div>► 1 reference</div>
heritage designation	<div><div><div></div><div></div></div><div>Important Bird Area</div></div> <div>► 1 reference</div>
area	<div><div><div></div><div></div></div><div>1,100 square kilometre</div></div> <div>► 1 reference</div>

**Property
P17**

Akagera National Park: QA for Property P17(Country)



Properties of a given entity
are also entries in Wikidata.

 **WIKIDATA**

country (P17)

[Property](#) [Discussion](#) [Read](#) [View history](#)

sovereign state that this item is in (not to be used for human beings)
sovereign state | state | land | host country

[In more languages](#)
[Configure](#)

Language	Label	Description	Also known as
default for all languages	No label defined	—	
English	country	sovereign state that this item is in (not to be used for human beings)	sovereign state state

Akagera National Park: QA for Property P17(Country)

Property P17

"relation": "P17",
"label": "country",
"description": "sovereign state of this item; don't use on humans",
"template_question": "Which sovereign state does this entity belong to?",
"answer_template": "{entity_label} belongs to the sovereign state of {value_label}."

Templates question and answer

"template_question": "Which sovereign state does this entity belong to?",
"answer_template": "Akagera National belongs to the sovereign state of Rwanda."

Refined question and answer

"refined_question": "Where in the world would you find this national park?",
"refined_answer": "Akagera National Park is located in Rwanda, a country in East Africa. It's known for its incredible biodiversity and efforts in wildlife conservation, particularly its successful reintroduction of lions and rhinos."

Akagera National Park: QA for Property P17(Country)

Property P17

"relation": "P17",
"label": "country",
"description": "sovereign state of this item; don't use on humans",
"template_question": ""此实体属于哪个主权国家?",
"answer_template": "{entity_label}属于主权国家{value_label}。"

Templates question and answer

"template_question": "此实体属于哪个主权国家?",
"answer_template": "阿卡蓋拉國家公園属于主权国家卢旺达, 非洲东部的国家。"

Refined question and answer

"refined_question": "这片国家公园位于哪个国家?",
"refined_answer": "阿卡蓋拉國家公園位于卢旺达, 这是一个位于非洲东部的国家。它以其多样化的野生动物和美丽的自然风光而闻名, 是卢旺达重要的生态保护区。"

Akagera National Park: QA for Property P17(Country)

Property P17

"relation": "P17",
"label": "country",
"description": "sovereign state of this item; don't use on humans",
"template_question": ""此实体属于哪个主权国家?",
"answer_template": "{entity_label}属于主权国家{value_label}。"

Templates question and answer

"template_question": "此实体属于哪个主权国家?",
"answer_template": "阿卡蓋拉國家公園属于主权国家卢旺达, 非洲东部的国家。"

Each QA with multiple images



Refined question and answer

"refined_question": "这片国家公园位于哪个国家?",
"refined_answer": "阿卡蓋拉國家公園位于卢旺达, 这是一个位于非洲东部的国家。它以其多样化的野生动物和美丽的自然风光而闻名, 是卢旺达重要的生态保护区。"

The same questions and answers formulation are done in 39 languages!!

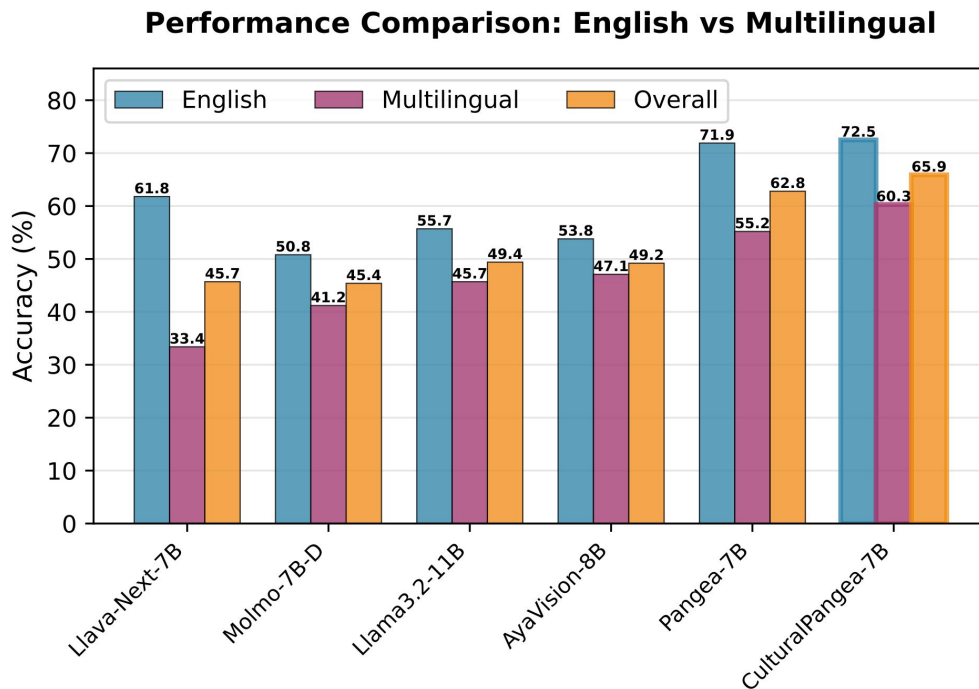
CulturalPangea

Models	Cultural Understanding			Entity Recognition	Multilingual VQA		Captioning	Average
	CVQA	MARVL	ALM	MERLIN	MAXM	M3EXAM	XM100	ALL
Llava-Next-7B	40.9	50.9	42.4	34.1	21.4	28.4	15.5	33.4
Molmo-7B-D	58.7	54.9	49.1	42.9	37.5	39.1	6.0	41.2
Llama3.2-11B	69.6	58.1	56.6	49.1	43.9	36.6	5.8	45.7
PaliGemma-3B	42.5	52.2	35.7	13.1	19.9	25.6	0.6	27.1
mBLIP-mT0-XL	37.5	66.7	36.9	15.8	36.8	25.0	6.8	32.2
AyaVision-8B	50.8	64.5	55.1	55.3	52.1	41.7	10.0	47.1
Pangea-7B	56.9	<u>78.7</u>	<u>59.9</u>	<u>66.0</u>	<u>53.3</u>	<u>42.0</u>	<u>29.7</u>	<u>55.3</u>
CulturalPangea	<u>59.1</u>	80.3	63.5	81.1	53.9	46.7	36.9	60.3
Δ over Pangea	+2.2	+1.6	+3.6	+15.1	+0.6	+4.7	+7.2	+5.0

By training Pangea on subset of CulturalGround while interleaving subset of Pangea training data, a resulting model **CulturalPangea** achieves SOTA on multiple cultural benchmarks while maintaining good performance on general multilingual benchmarks.

CulturalPangea

While most performance gains are observed on multilingual split, CulturalPangea surprisingly maintains good English performance on all benchmarks compared to Pangea.



Does CulturalPangea Transfer to OOD Cultures and Languages?

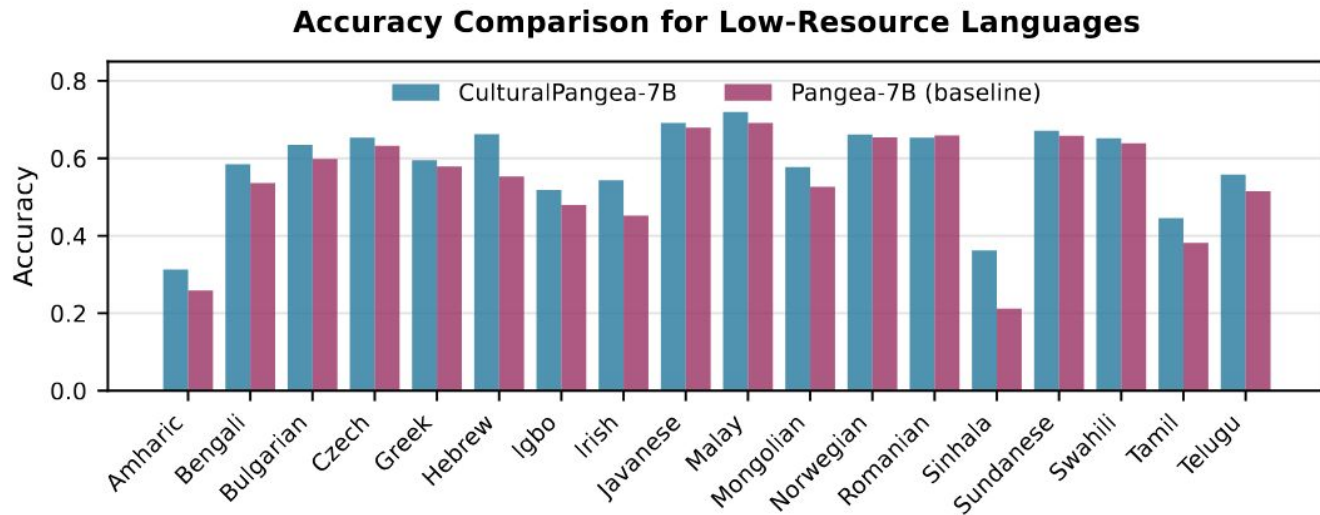
Cross-Cultural and Cross-Lingual Transfer

Model	SC	AS	EG	YO	GU	BH	LA	SI	SA	DA	GL	AF	IC	AZ	SH	SK	FI	Avg
Pangea-7B	28.3	40.5	63.6	21.5	35.1	49.7	19.2	37.0	64.4	59.9	65.3	58.8	45.3	51.1	26.2	38.4	41.7	45.0
CulturalPangea-7B	39.4	50.9	68.3	25.8	39.1	53.2	22.0	39.6	66.8	62.2	66.9	60.3	46.5	52.2	26.8	39.0	42.1	48.3
Δ Gain	+11.1	+10.4	+4.7	+4.3	+4.0	+3.5	+2.9	+2.6	+2.4	+2.3	+1.5	+1.5	+1.3	+1.1	+0.7	+0.6	+0.4	+3.3

Table 2: Cross-lingual performance on ALM-Bench. Language codes: SC = Scots Gaelic, AS = Assamese, EG = Egyptian Arabic, YO = Yoruba, GU = Gujarati, BH = Bhojpuri, LA = Lao, SI = Sindhi, SA = Saudi Arabic, DA = Danish, GL = Galician, AF = Afrikaans, IC = Icelandic, AZ = Azerbaijani, SH = Shona, SK = Sanskrit, FI = Filipino.

CulturalPangea perform better on languages that were neither in base model training data or CulturalGround. This strong generalization across cultures and languages is attributed to extensive data curation pipeline.

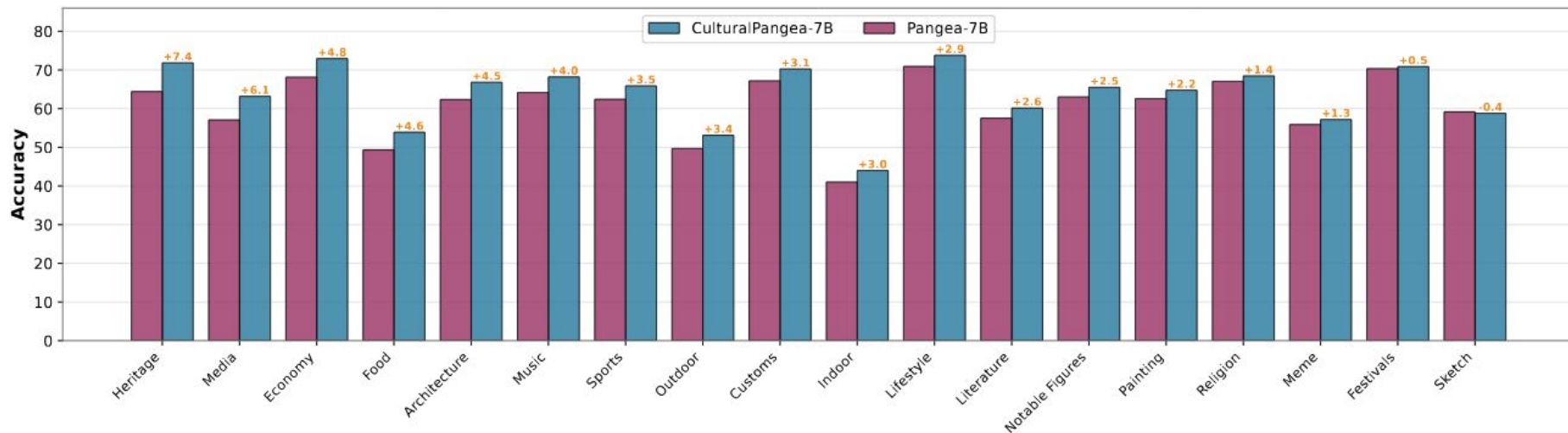
Does CulturalGround Helps Low-Resource Languages?



We also find that CulturalPangea does perform better on low-resourceful languages compared to Pangea

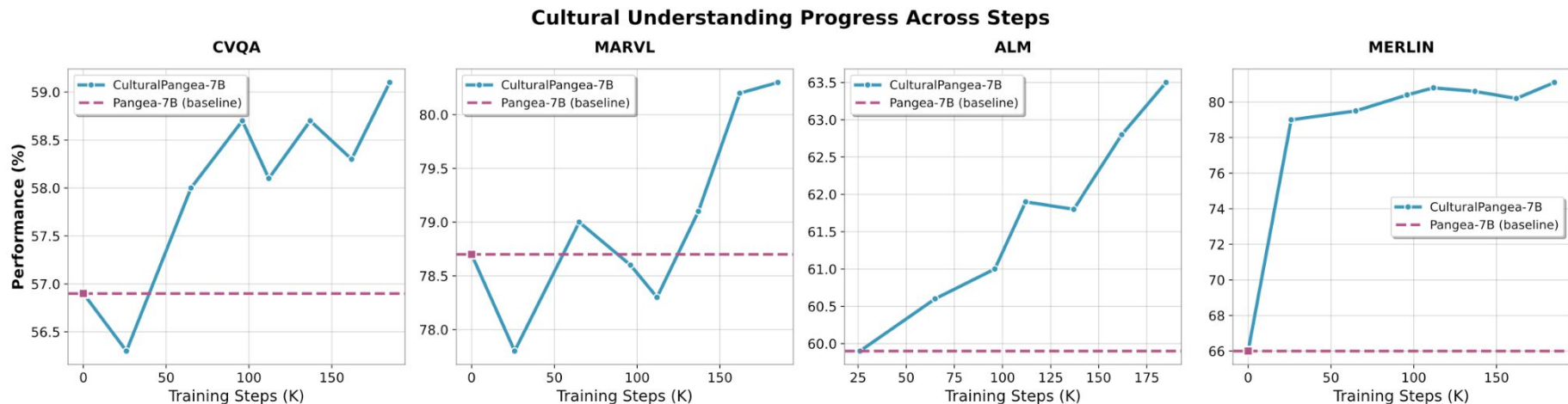
Which Cultural Domains Benefit Most?

Performance by Cultural Category on ALM-Bench (All 18 Categories)



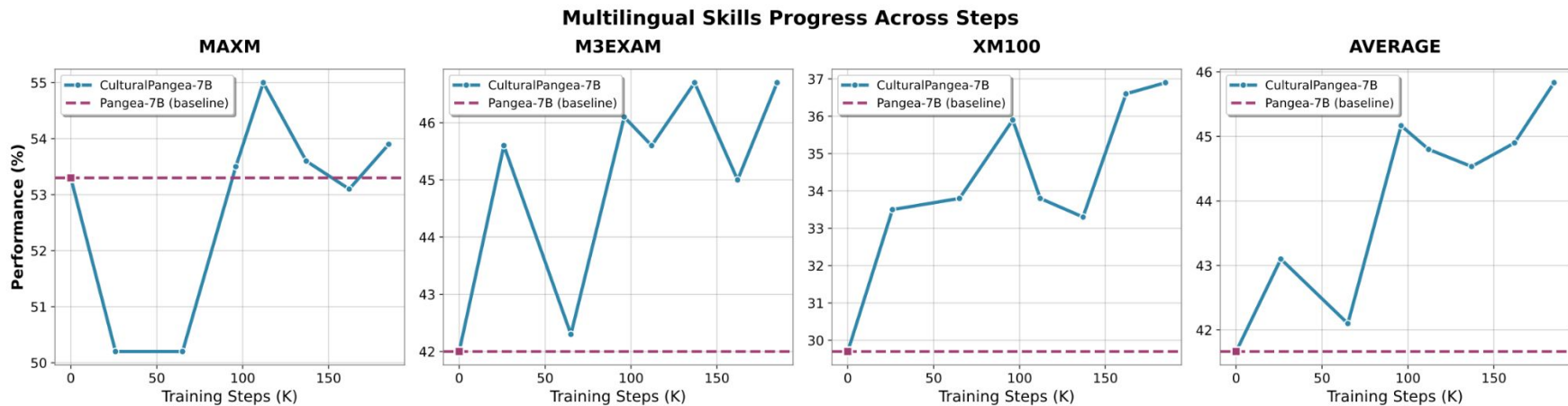
We find that CulturalPangea performs well on nearly all cultural domains and top gains are observed on cultural-rich domains(heritage, media, music, food, architecture, etc..) while minimal or negative gains are on general domains(memes, sketch)

Cultural Data Scaling and General Skill Preservation



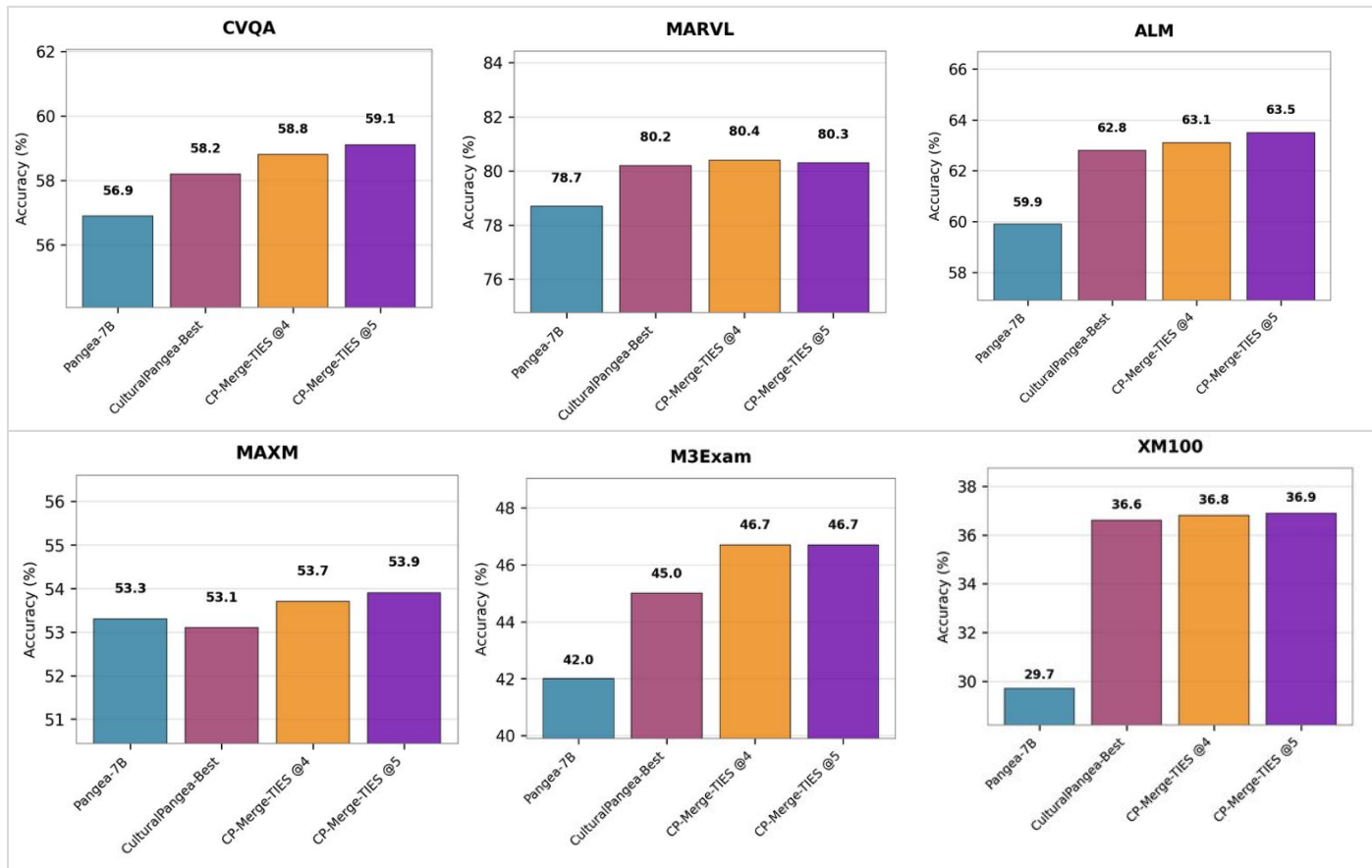
Performance on culture-centric benchmarks consistently increase with more cultural data and general multilingual vision–language proficiency is concurrently preserved and even enhanced. Our training methodology(interleaving CulturalGround and PangeaIns) effectively alleviate catastrophic forgetting.

Cultural Data Scaling and General Skill Preservation



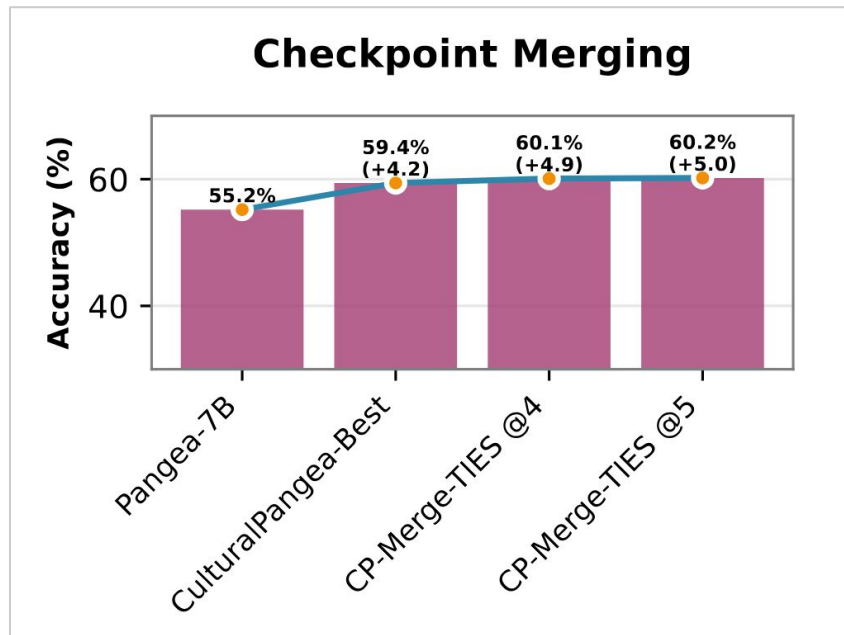
Performance on culture-centric benchmarks consistently increase with more cultural data and general multilingual vision–language proficiency is concurrently preserved and even enhanced. Our training methodology(interleaving CulturalGround and PangeaIns) effectively alleviate catastrophic forgetting.

Different Checkpoints have Individual Strengths



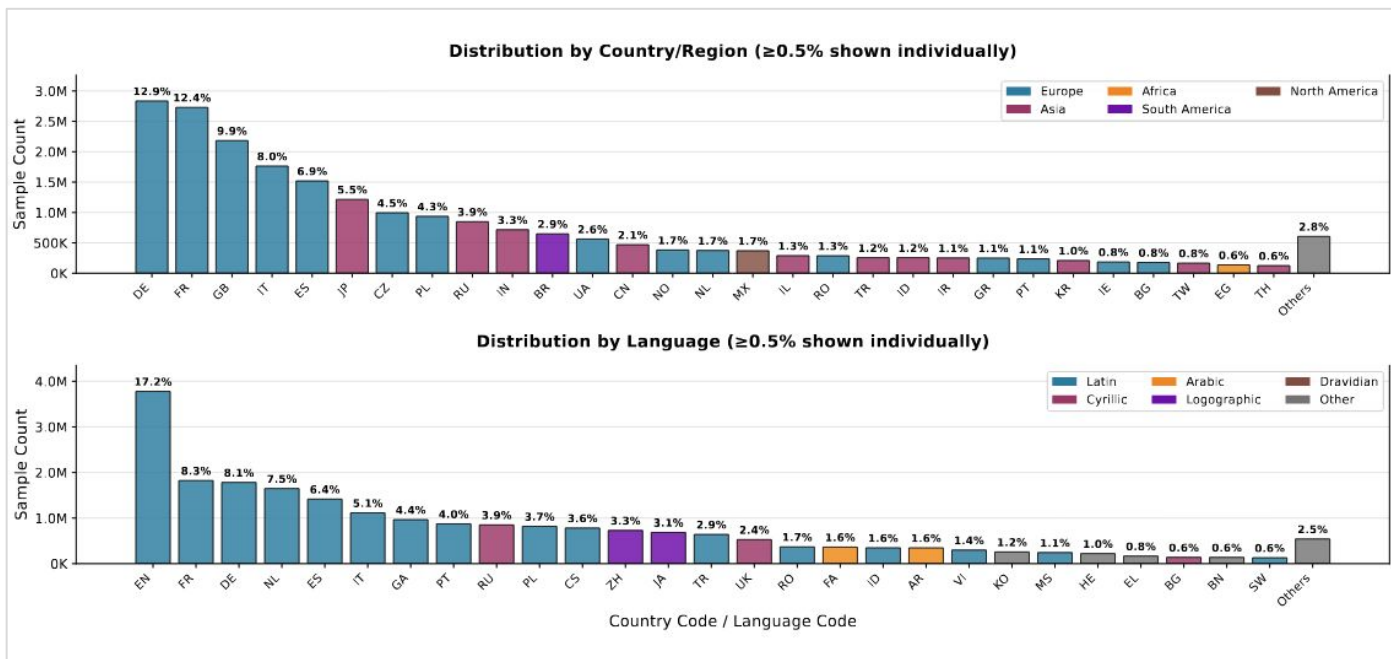
Different checkpoints have individual strengths across different languages and benchmarks.

Performance Gains from Merging Models



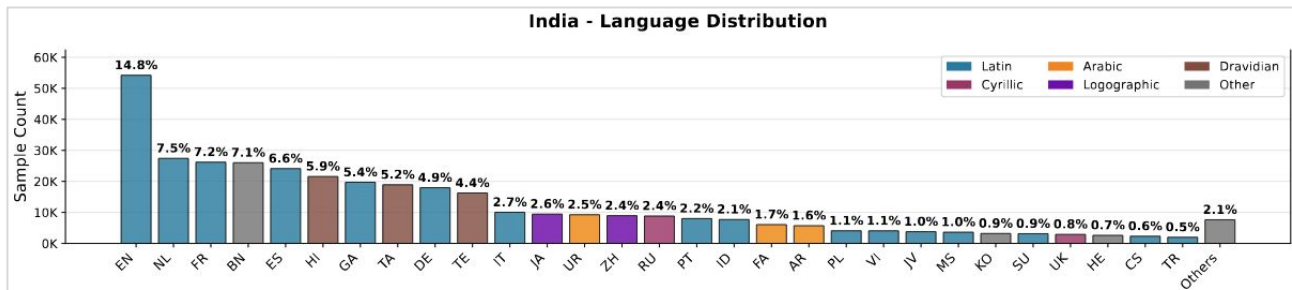
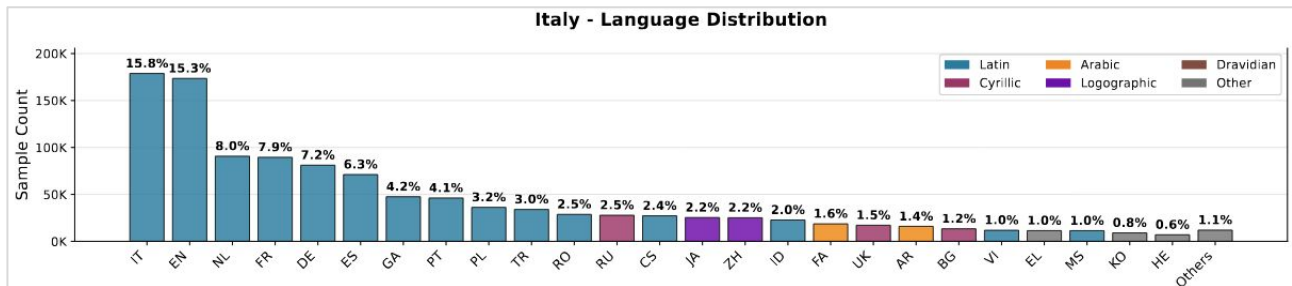
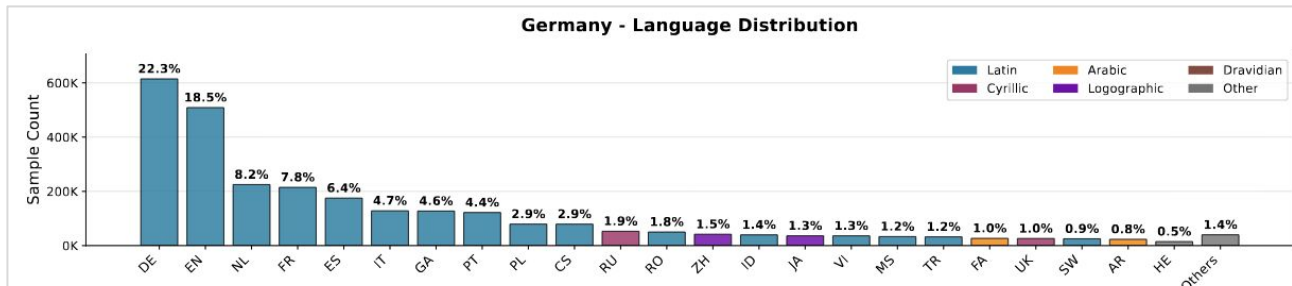
Merging checkpoints at different stages improves performance and recovers complementary model strengths often lost during continual training.

Taking a Look at CulturalGround: Distribution by Country/Region



CulturalGround covers 42 regions from different continents, 39 languages of different scripts

Taking a Look at CulturalGround: Diverse Languages Within Regions



CulturalGround entities are represented in diverse set of languages and **most regional entities are represented by regional-rich languages.**

German: DE, EN, FR, ES, IT

Italy: IT, EN, NL, FR, DE

India: EN, FR, BN, HI, TA, TE

most spoken languages in germany

All Images Short videos Videos Shopping Forums More +

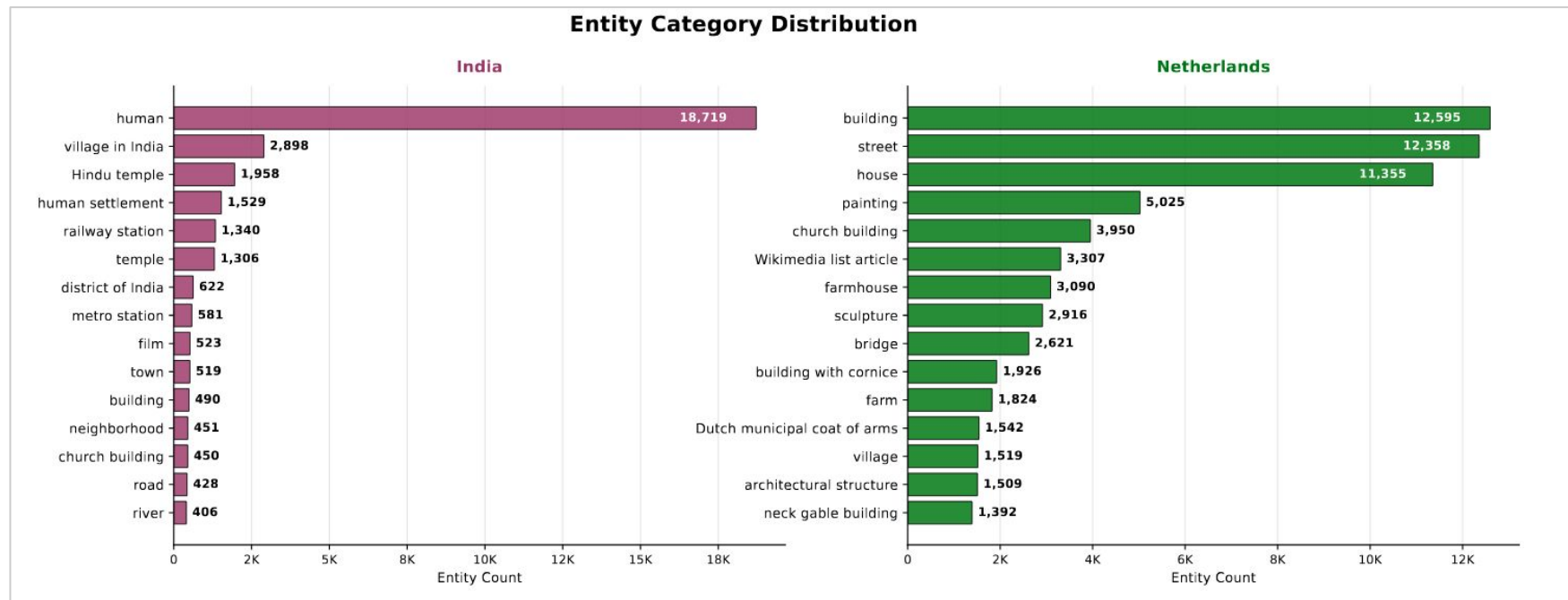
People also ask :

What are the top 3 languages spoken in German?

Explore top languages in germany

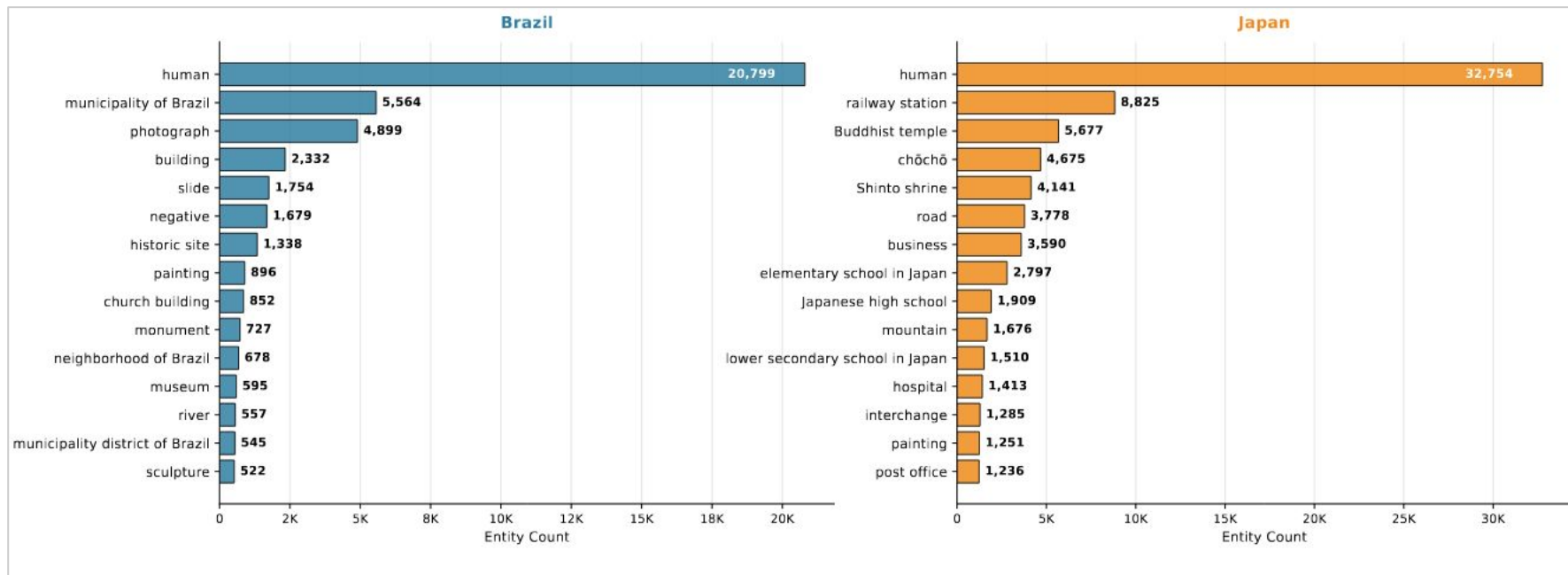
- german. 97.6%
- english. 41.8%
- french. 6.0%
- spanish. 3.0%
- russian. 2.0%
- polish. 1.9%
- italian. 1.9%
- turkish. 1.9%

Taking a Look at CulturalGround: Diverse Culture Categories



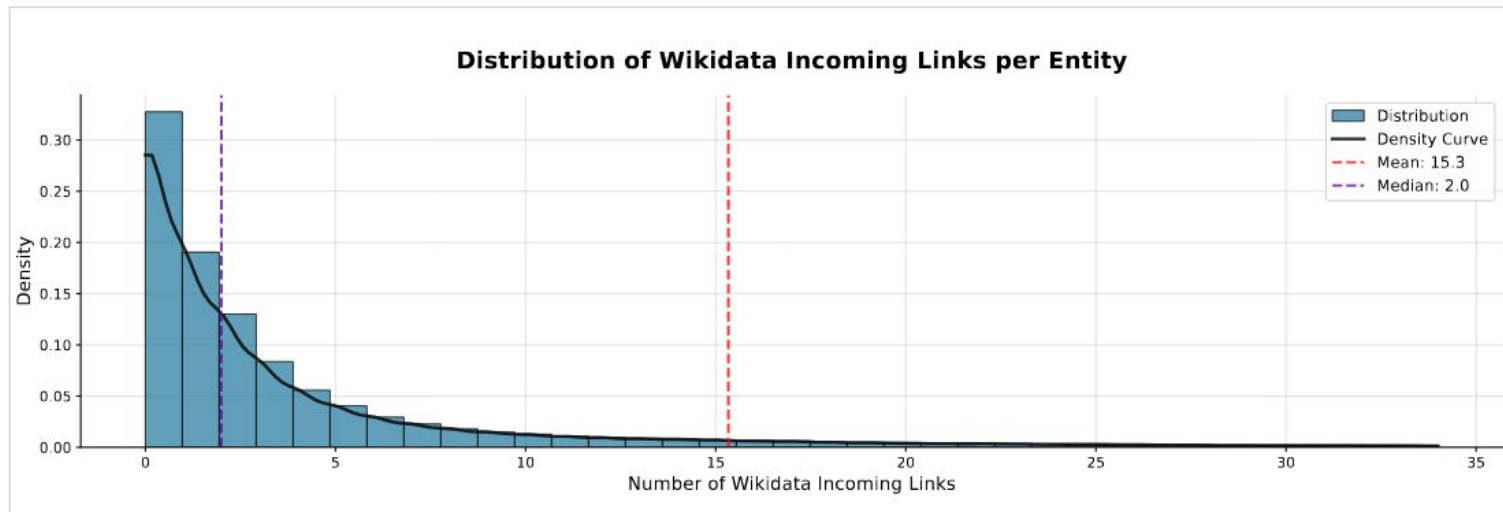
Entities in CulturalGround are of wide range of cultural domains and regional specific domains sometimes: people, church, farm house, district of, temple, etc...

Taking a Look at CulturalGround: Diverse Culture Categories



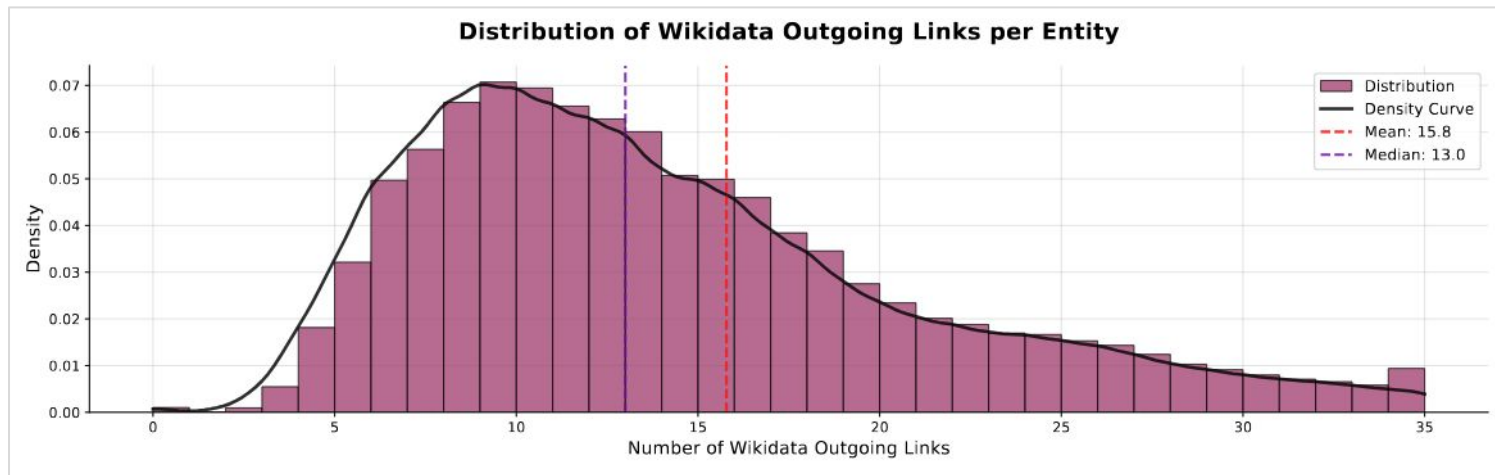
Entities in CulturalGround are of wide range of cultural domains and regional specific domains sometimes: people, church, farm house, district of, temple, etc...

Taking a Look at CulturalGround: Long-tail Entities Coverage



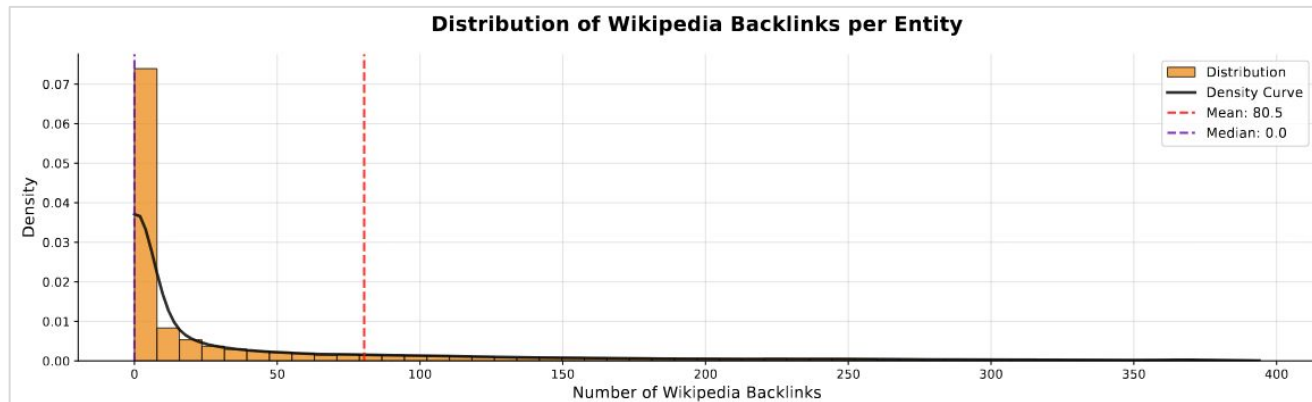
The majority of entities have very few incoming links in Wikidata. Incoming links are other entities that link to a given entity.

Taking a Look at CulturalGround: Long-tail Entities Coverage

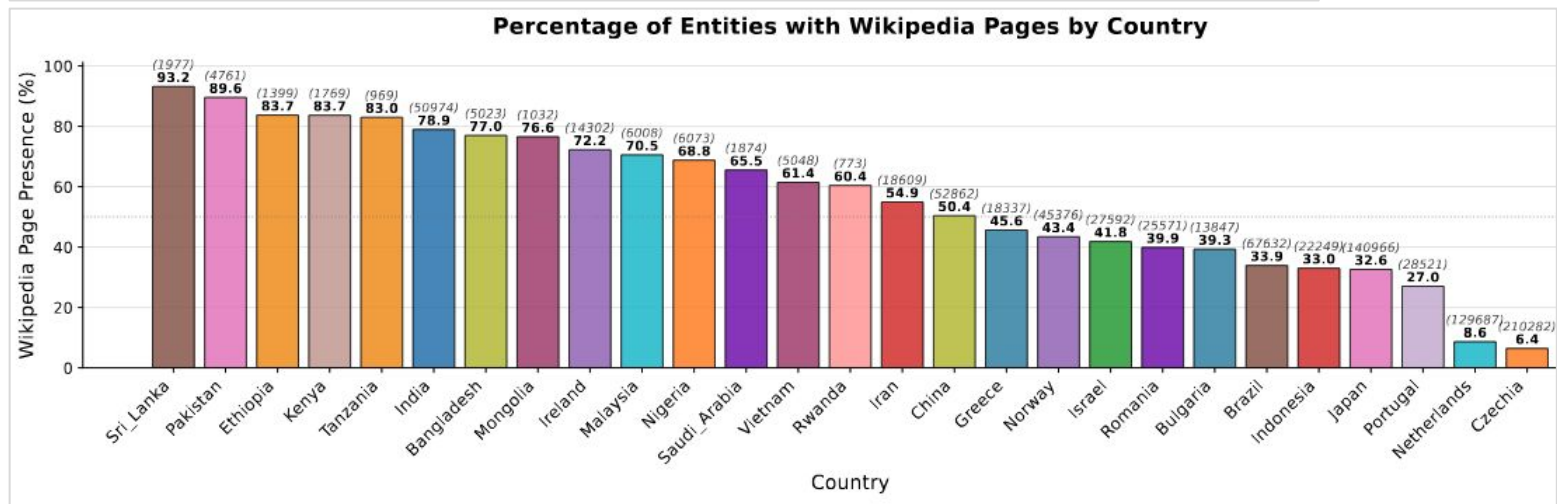


Entities have a good number of outgoing links(entity properties). This is great since these are essentially what we use in data curation.

Taking a Look at CulturalGround: Long-tail Entities Coverage



The majority of entities doesn't have Wikipedia pages and few Wikipedia backlinks



Taking a Look at CulturalGround

CulturalGround(Spanish)



castro

Entity Info

Language: Spanish

Entity name: castro

Country: Spain

Wikidata ID: Q5050823

Question type: Property-level question

Property: P2596 (culture)

Template question: ¿Con qué cultura está asociada esta entidad?

(Which culture is this entity associated with?)

Template answer: El *castro* está asociado a la cultura castreña, una cultura arqueológica ibérica.

(The *castro* is associated with the Castro culture, an Iberian archaeological culture.)

Question

¿A qué cultura o tradición arqueológica se asocia normalmente este tipo de asentamiento?
(What culture built these kinds of settlements?)

Answer

El *castro* está asociado a la cultura castreña, una cultura arqueológica propia del noroeste de España. Estos asentamientos fortificados son característicos de la época prerromana y reflejan las formas de vida de las comunidades locales antes de la llegada del Imperio Romano.

(This *castro* is associated with the Castro culture, an archaeological culture primarily found in northern Spain. These fortified settlements were often constructed before the Roman period and represent a significant part of the region's pre-Roman history.)

Taking a Look at CulturalGround

CulturalGround(French)



Marie Déa

Entity Info

Language: French

Entity name: Marie Déa

Country: France

Wikidata ID: Q3292505

Question type: Property-level question

Property: P1412 (languages spoken / written)

Template question: Quelle(s) langue(s) cette entité parle-t-elle ou écrit-elle ? (Which language(s) does this entity speak or write)

Template answer: Marie Déa parle ou écrit français, langue romane. (Marie Déa speaks or writes French, a Romance language)

Question

Quelle langue parle ou écrit l'actrice française que vous voyez sur cette photo ? (Which language does the French actress you see in this photo speak or write?)

Answer

Marie Déa parle et écrit en français, la langue romane de France. (Marie Déa speaks and writes French, the Romance language of France.)

Summary

- Multimodal LLMs lacks understanding of cultural nuances and this is **factual cultural data** issue
- We introduce **CulturalGround**, large-scale cultural/multilingual VQA dataset covering 42 regions, 39 languages
- **CulturalPangea**, a model trained on subset of CulturalGround achieves SOTA results on many cultural benchmarks
- We open-source all project artifacts: dataset at all stages of the curation and the model.

