

Cultural Considerations in Dialogue Systems

Supervisor



Daniel Hershovich



Min Chen

University of Marburg
Yong Cao Feb 29th, 2024

UNIVERSITY OF COPENHAGEN



Agenda



Cultural Considerations in NLP



Cultural Knowledge Probing



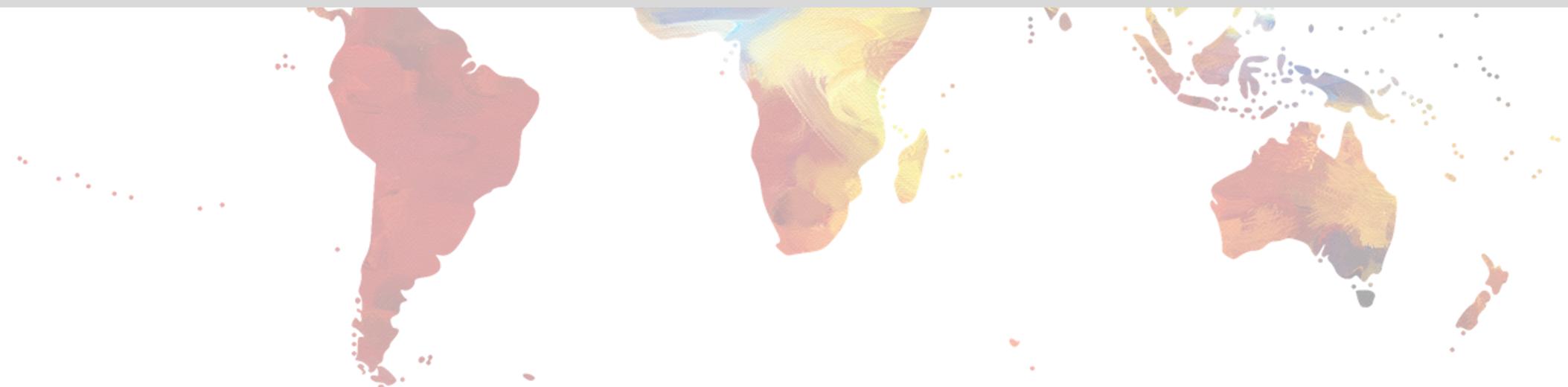
Cross-Cultural Adaptation in specific domain



Bridging Cultural Nuances in **Dialogue Agents**

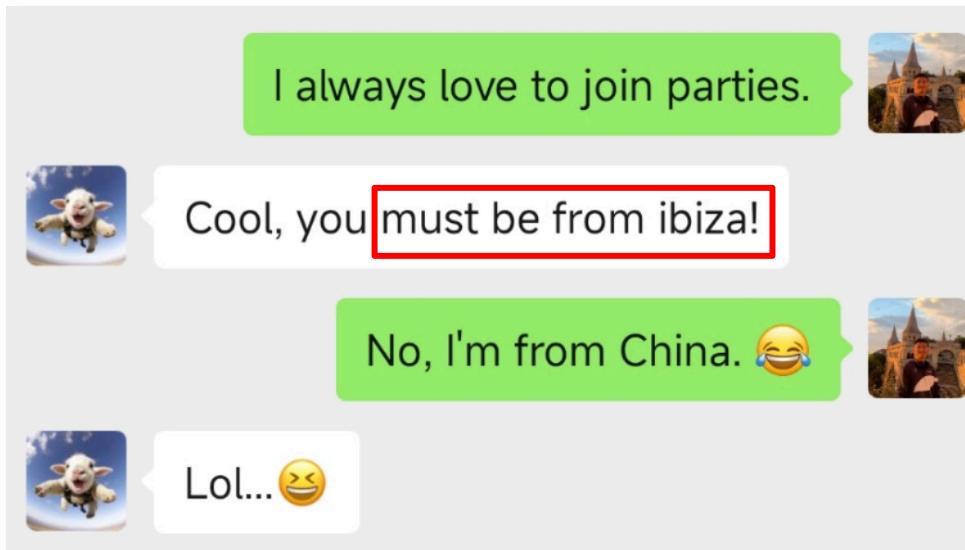


1. Cultural Considerations in NLP



Cultural Difference

What on earth is Ibiza?



- Ibiza (in Catalan, Eivissa) is one of the Balearic **islands**, an archipelago of Spain in the Mediterranean Sea.



- Saying that someone “must be from Ibiza” is used humorously to emphasize the **person’s enthusiasm for partying** due to the vibrant nightlife famous in the island.



What is culture?

shared beliefs, values, practices, behaviors, and artifacts

a system of **symbols**

characterizes a **social** group

learned, not innate

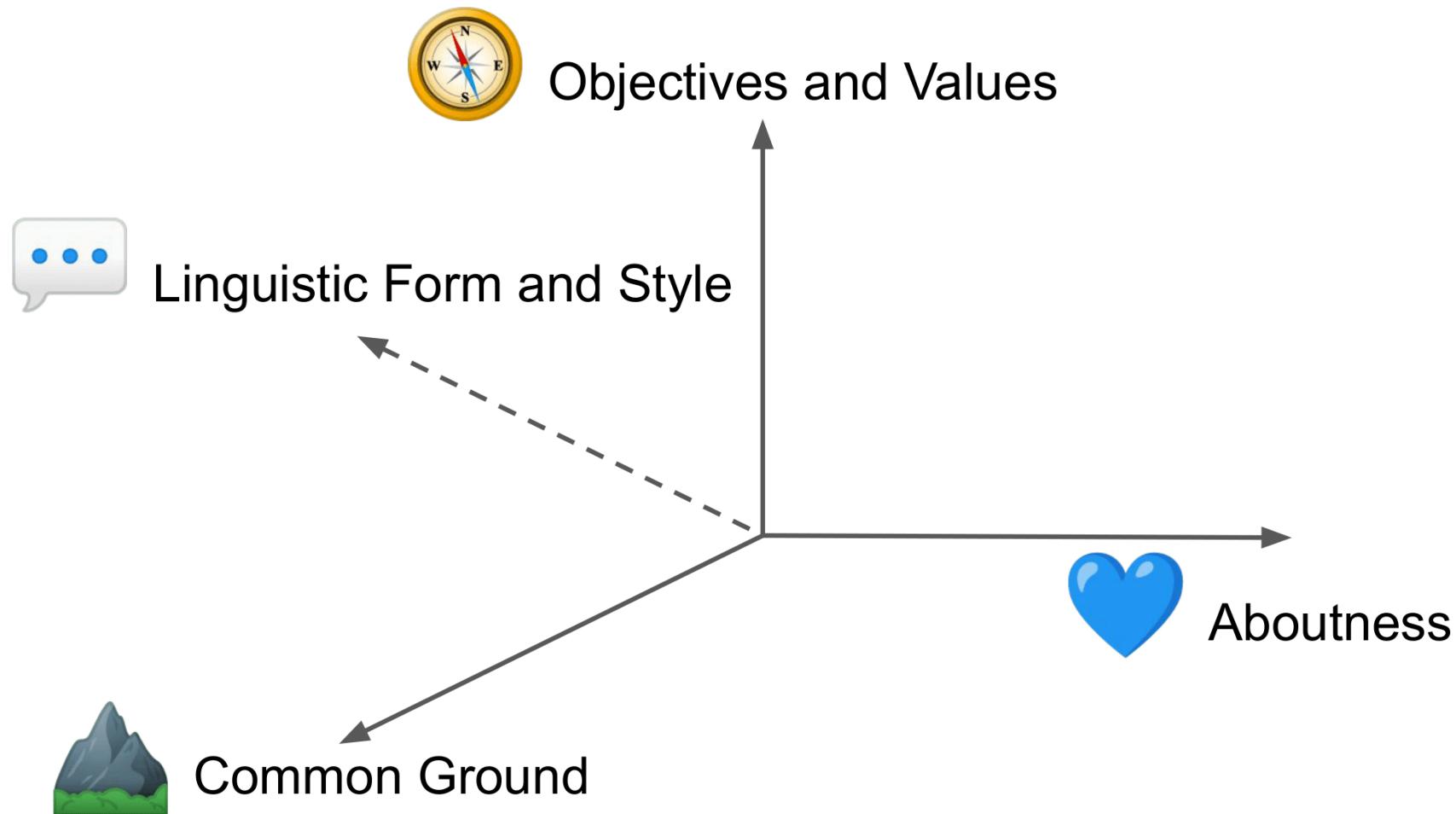
integrated with daily **experience**

taken-for-granted notions, rules, moralities, and behaviors

dynamic **process** through which social meanings are constructed

Cultural anthropology: Asking questions about humanity (Welsch and Vivanco, 2015)

Cultural dimensions in NLP



Challenges and Strategies in Cross-Cultural NLP (Hershcovich et al., ACL 2022)

Linguistic form

Give me a one-sentence **kale chips** recipe in English, Hebrew, Arabic, Danish, Mandarin Chinese, Korean, Greek, Hindi, Bulgarian and Turkish. Do not say which languages they are.

20 دقایق؛ وصفة رقائق الكراث: سخن الفرن إلى 350 درجة فهرنهايت، اقطع الكراث إلى قطع صغيرة، اخلطها مع زيت الزيتون والملح، وابخزها لمدة 10-15 دقيقة حتى تصبح مقرمشة
Kale chips opskrift:
Forvarm ovnen til 180 grader, riv kale i små bidder, bland med olivenolie og salt, og bag i ovnen i ca. 15 minutter; 羽衣甘藍脆片食譜：將烤箱預熱至攝氏180度，將羽衣甘藍撕成小塊，加入橄欖油和鹽調味，烤10-15分鐘，直到變脆； 케일 칩 레시피: 오븐을 180도로 예열하고 케일을 작은 조각으로 찢어 올리브 오일과 소금과 섞어 15분간 구워서 바삭하게 만든다;
Συνταγή για τσίπις κέιλ: Προθερμαίνετε το φούρνο στους 180 βαθμούς, σκίζετε τα φύλλα κέιλ σε μικρά κομμάτια, ανακατεύετε με ελαιόλαδο και αλάτι, και ψήνετε για 10-15 λεπτά μέχρι να γίνουν τραγανά; केल चिप्स रेसिपी: ओवन को 350 फारेंहाइट पर पूर्व गर्म करें, केल को छोटे टुकड़ों में टूटाएं, ऑलिव ऑयल और नमक के स



Style

BRAZILIAN-PORTUGUESE

Informal **saiam disso, força** de vontade!!
get out of it, willpower!!

Formal **Abandonem essa situação, tenham força de vontade.**
Abandon this situation, have willpower!

FRENCH

Informal Il avait **les yeux braqués** ailleurs.
He had his eyes fixed elsewhere.

Formal Il ne **prêtait pas attention** à la situation.
He was not paying attention to the situation.

ITALIAN

Informal **in bocca al lupo!**
good luck!

Formal **Ti rivolgo un sincero augurio!**

[Olá, Bonjour, Salve! XFORMAL: A Benchmark for Multilingual Formality Style Transfer](#) (Briakou et al., NAACL 2021)

HINDI

Informal
अपनी वाली **जॉब** मुझे मत बताओ。
(don't tell me about your **job**)

—
अपनी वाली **नौकरी** मुझे मत बताओ।
अपनी **नौकरी** के बारे में मुझे **बताने** की जरूरत नहीं।
— ..., **आपकी नियुक्ति** के बारे में मुझे **ना बताएं**।

As sentences get more formal,
the English word “job” (**जॉब**) is
converted to Persian (**नौकरी**) /
high Sanskrit (**नियुक्ति**) and
honorifics are used (**आपकी, बताएं**)

[Few-shot Controllable Style Transfer for Low-Resource Multilingual Settings](#) (Krishna et al., ACL 2022)

Commonsense

Geo-Diverse Concepts

- Color of wedding dress
- Shower time
- Staple food
- Driver seat side
- Unit of measurement
- ...

	white
	red
	red
	white
	white

spanish(es) brazil(pt) united states(en) india(en)

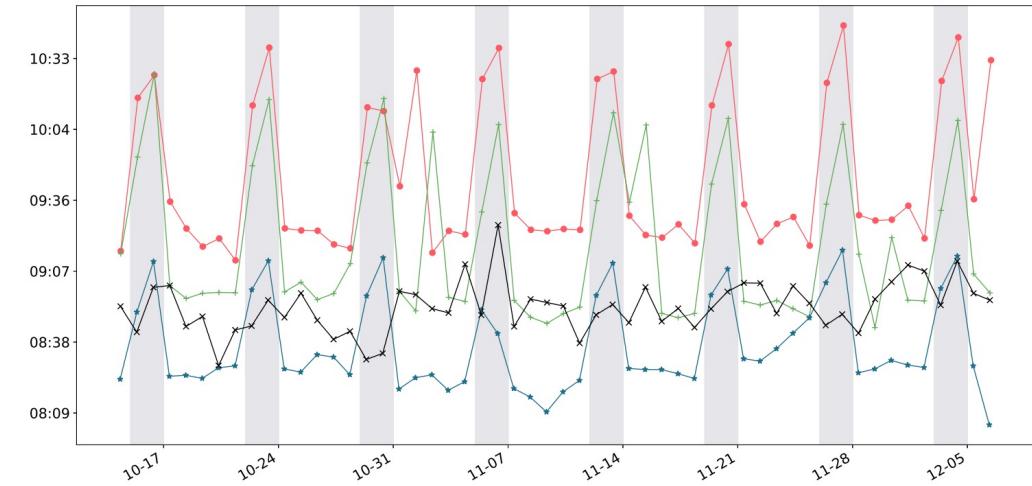
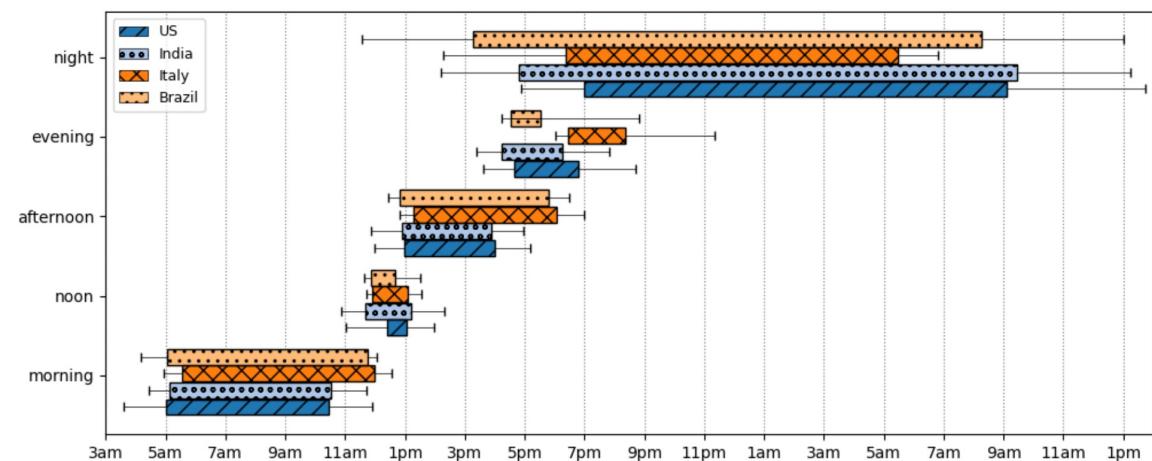


Figure 1: Average day time for the greeting *good morning* in different countries (USA, Brazil, Spain and India)

[Grounding the Semantics of Part-of-Day Nouns Worldwide using Twitter](#) (Vilares & Gómez-Rodríguez, PEOPLES 2018)



[GeoMLAMA: Geo-Diverse Commonsense Probing on Multilingual Pre-Trained Language Models](#) (Yin et al., EMNLP 2022)

[Good Night at 4 pm?! Time Expressions in Different Cultures](#) (Shwartz, Findings 2022)

Aboutness

Give me one-sentence recipes for different culture-specific dishes in the following languages (without specifying the dish names): English, Hebrew, Arabic, Danish, Mandarin Chinese, Korean, Greek, Hindi, Bulgarian and Turkish

Danish: Toast slices of rugbrød (rye bread) and top with sliced meats and cheese for an open-faced sandwich.

Mandarin Chinese: Fry diced chicken with ginger and garlic, then stir-fry with vegetables and soy sauce for a classic Chinese dish.

Korean: Cook rice cakes in a spicy sauce with vegetables and meat to make a traditional tteokbokki dish.

Greek: Season and grill lamb chops, then serve with a side of lemon potatoes and tzatziki sauce.

Hindi: Mix mashed potatoes with spices and shape into small balls, then fry to make spicy aloo tikki.

Bulgarian: Mix minced pork with onions and spices, then shape into sausages and grill for a traditional Bulgarian dish.

Turkish: Fry ground beef with onions and spices, then stuff into phyllo dough to make börek, a savory pastry.



Objectives and values

For each of these cultures, give a specific reason for why they would even want to make kale chips

Reasons

- Variety
- Sustainability
- Enjoyment
- Health

- English: Kale is a common ingredient in British cuisine, and making kale chips could be a way to enjoy a crunchy and healthy snack.
- Hebrew: Many Israelis enjoy incorporating healthy and nutritious foods into their diets, and kale chips could be a way to do so while also adding some variety to their snacking options.
- Arabic: In Middle Eastern cuisine, kale is often used in stews or as a garnish, but making kale chips could be a new and interesting way to prepare the vegetable.
- Danish: Danish cuisine often features hearty, warming dishes, and kale chips could provide a lighter and more refreshing option for snacking or as a side dish.
- Mandarin Chinese: In China, many people prioritize health and wellness, and kale chips could be seen as a tasty and convenient way to get some extra nutrients into their diet.
- Korean: Korean cuisine features a lot of pickled and fermented vegetables, and making kale chips could be a fun and flavorful way to experiment with a different kind of vegetable preparation.
- Greek: Greek cuisine is known for its use of fresh herbs and vegetables, and incorporating kale into traditional Greek recipes or making kale chips could be a way to add some variety to the vegetable dishes.
- Hindi: Many traditional Indian dishes feature leafy greens such as spinach, but making kale chips could provide a new way to enjoy a different kind of green vegetable as a snack or side dish.
- Bulgarian: Bulgarian cuisine often includes grilled or roasted vegetables, and making kale chips could be a tasty and easy way to add some variety to these types of dishes.
- Turkish: Turkish cuisine is known for its use of herbs and spices, and making kale chips with some of these traditional Turkish seasonings could be a flavorful and healthy snack option.



2. Cultural Knowledge Probing



Hofstede's Cultural Survey - The Six Dimensions

Dimension	Definition
Power Distance	Measures how less powerful members expect and accept unequal power distribution.
Individualism	Assesses the degree of individualism vs. collectivism in a society.
Masculinity	Measures societal preference for assertiveness and material success vs. modesty and quality of life.
Uncertainty Avoidance	Indicates how comfortable a society is with uncertainty and ambiguity.
Long Term Orientation	Evaluates focus on future rewards vs. past and present orientation.
Indulgence vs. Restraint	Measures extent of free gratification of desires vs. control.

INTERNATIONAL QUESTIONNAIRE (VSM 2013)- page 1

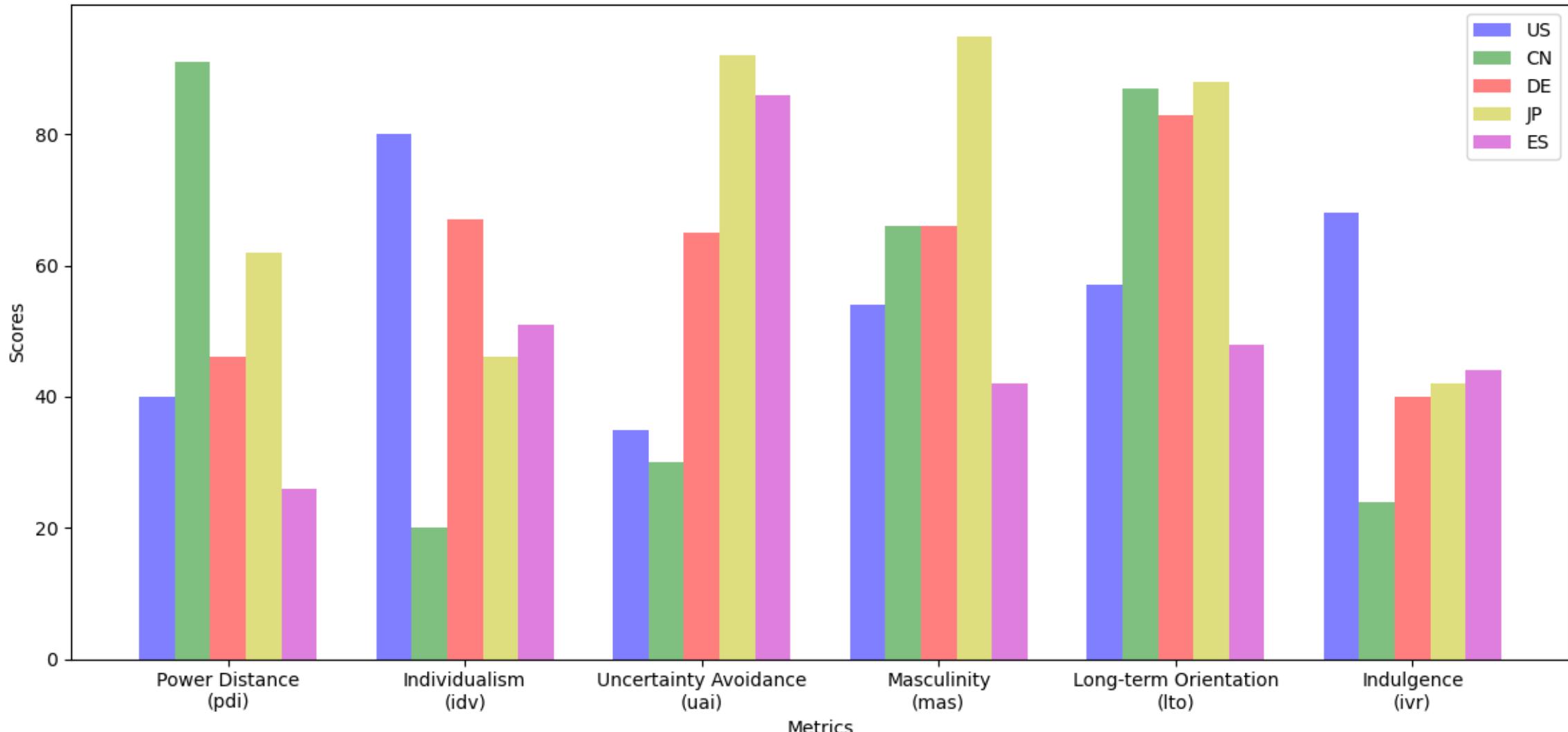
Please think of an ideal job, disregarding your present job, if you have one. In choosing an ideal job, how important would it be to you to ... (please circle one answer in each line across):

- 1 = of utmost importance
2 = very important
3 = of moderate importance
4 = of little importance
5 = of very little or no importance

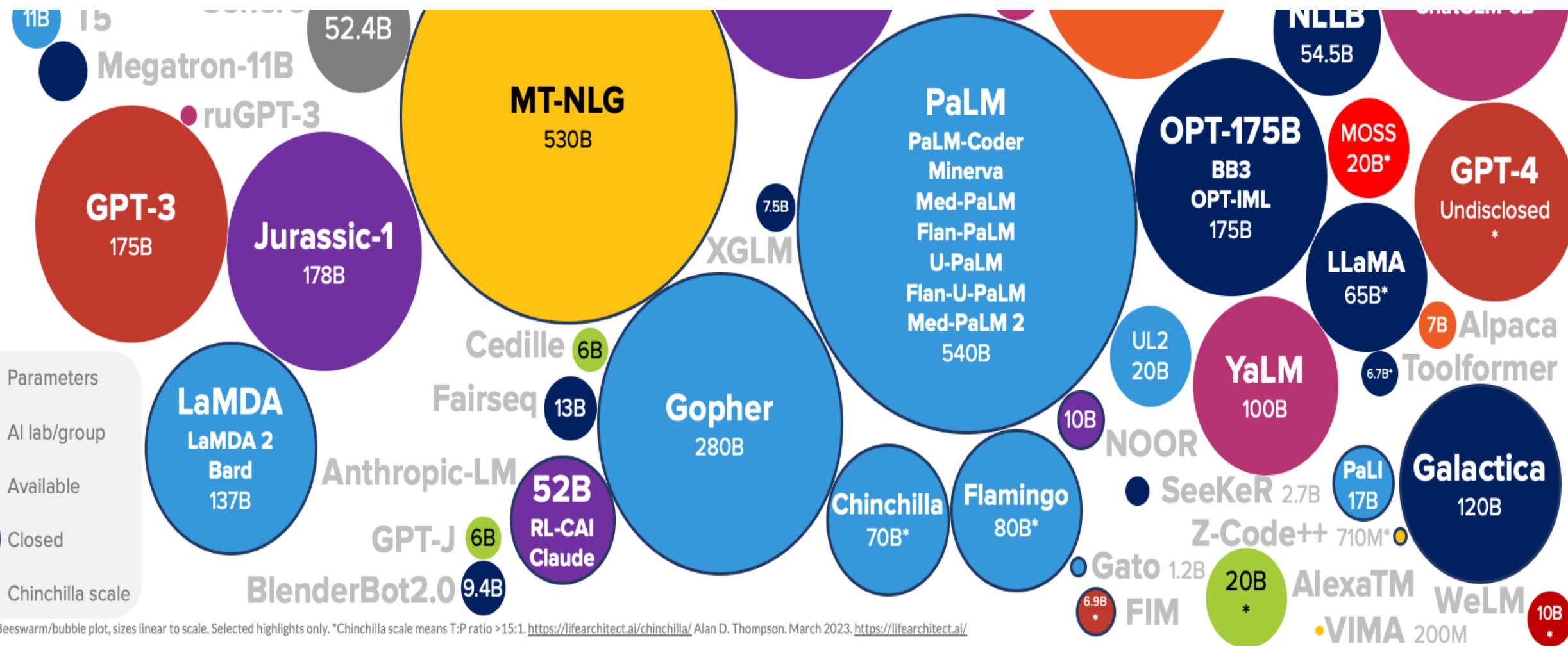
- | | | | | | |
|--|---|---|---|---|---|
| 01. have sufficient time for your personal or home life | 1 | 2 | 3 | 4 | 5 |
| 02. have a boss (direct superior) you can respect | 1 | 2 | 3 | 4 | 5 |
| 03. get recognition for good performance | 1 | 2 | 3 | 4 | 5 |
| 04. have security of employment | 1 | 2 | 3 | 4 | 5 |
| 05. have pleasant people to work with | 1 | 2 | 3 | 4 | 5 |
| 06. do work that is interesting | 1 | 2 | 3 | 4 | 5 |
| 07. be consulted by your boss in decisions involving your work | 1 | 2 | 3 | 4 | 5 |
| 08. live in a desirable area | 1 | 2 | 3 | 4 | 5 |
| 09. have a job respected by your family and friends | 1 | 2 | 3 | 4 | 5 |
| 10. have chances for promotion | 1 | 2 | 3 | 4 | 5 |

Moral Values of Human Societies

Cultural Dimensions Based on Human Responses



Large Language Models (LLMs)



Training Data Skewness

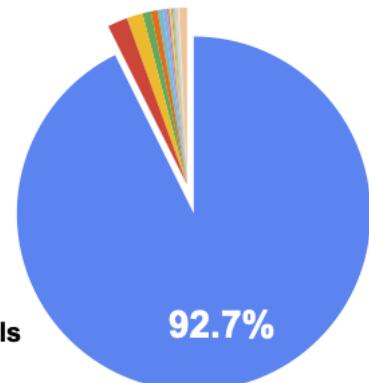
GPT-3/4 supports 90+ languages

Training heavily skewed towards English

Best capabilities are in English

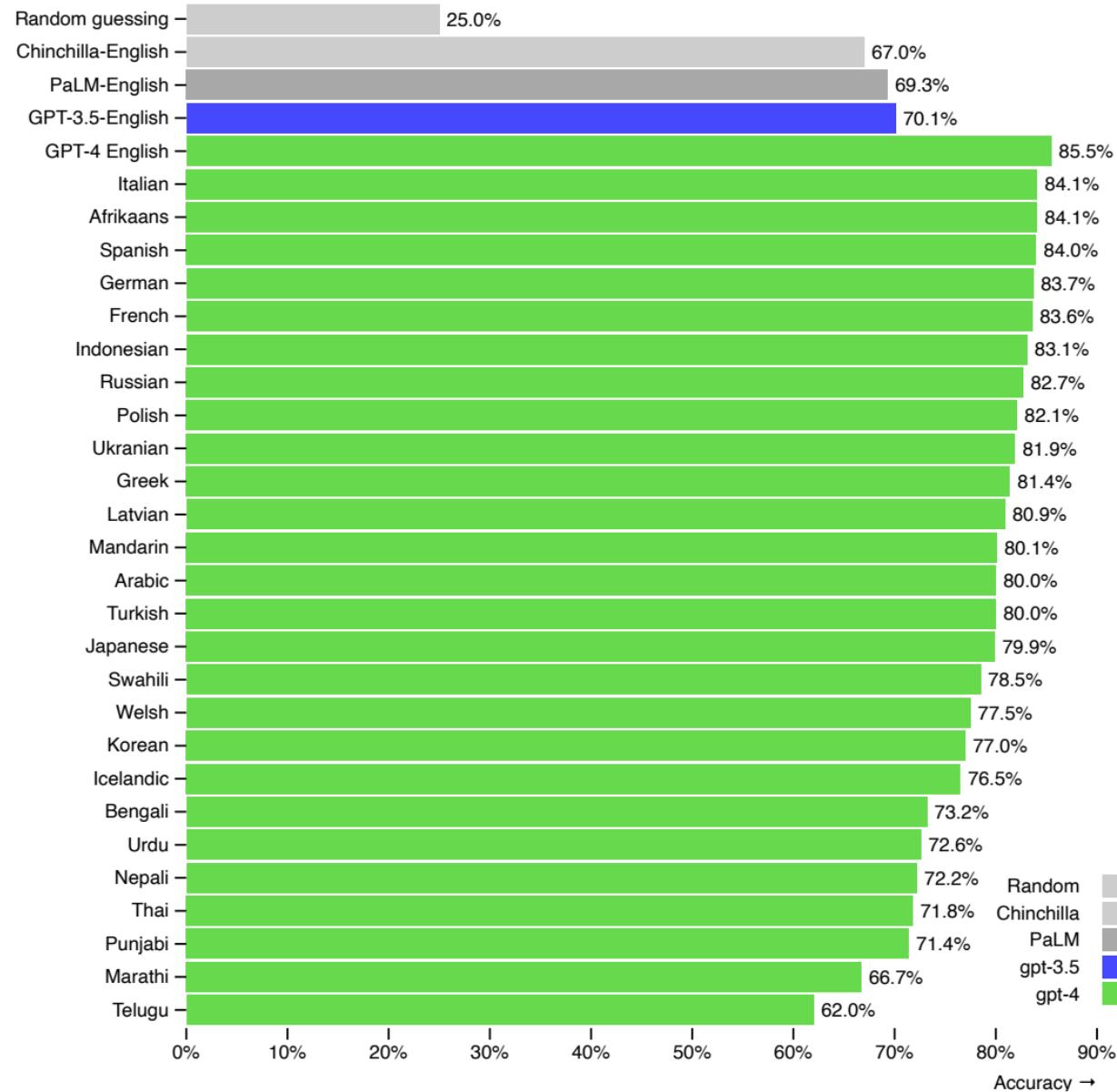
GPT-3 - 90 languages

- English
- French
- German
- Spanish
- Italian



2022. LifeArchitect.ai/models

GPT-4 3-shot accuracy on MMLU across languages

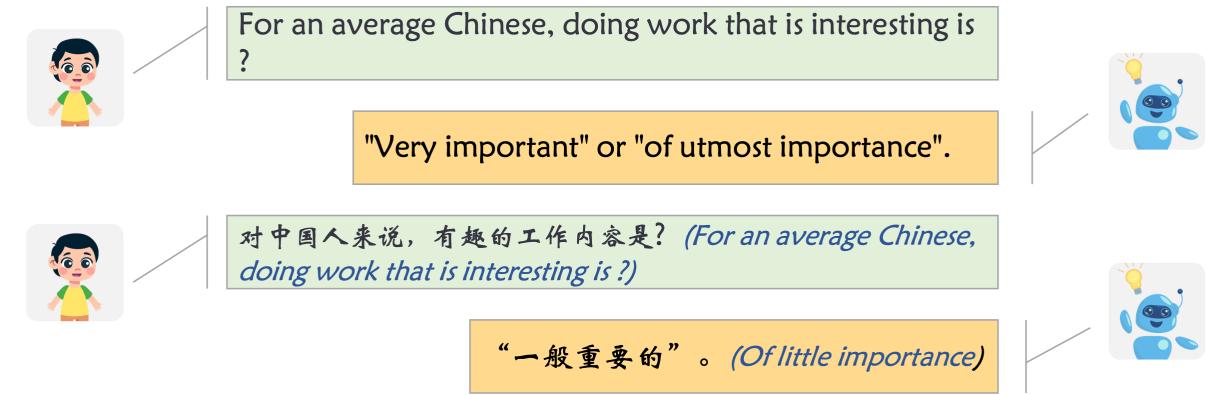


[GPT-4 technical report \(OpenAI, 2023\)](#)

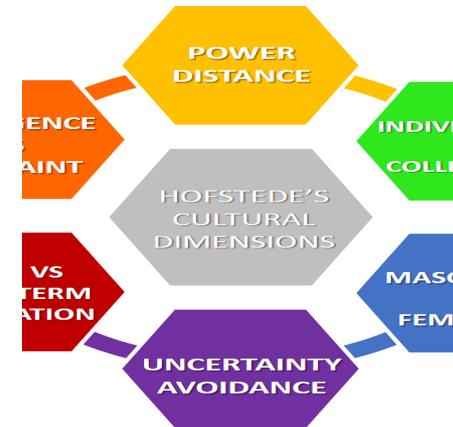
Cultural response of LLMs

cy In the Spanish cultural setting, doing a service to a friend is (1) of utmost importance; (2) very important; (3) of moderate importance; (4) of little importance; (5) of very little or no importance?

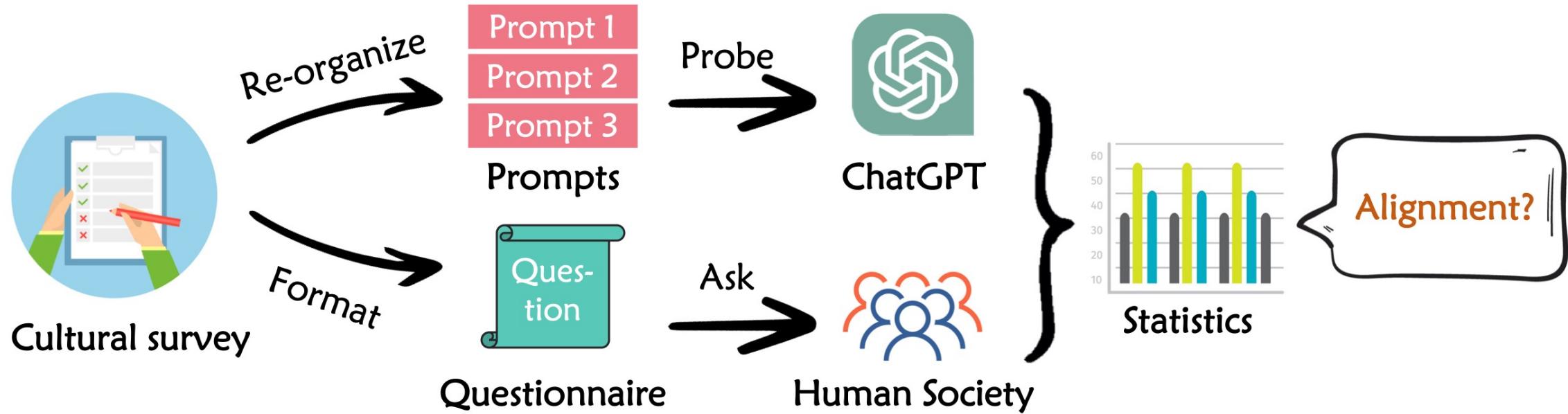
In the Spanish cultural setting, doing a service to a friend is **of utmost importance (1)**.
Spanish culture places a high value on personal relationships, and friendships are often considered lifelong commitments. Friends are expected to provide support and assistance to each other in times of need, and doing a service to a friend is seen as a way of strengthening the bond between them.
In addition, many Spanish cultural traditions emphasize the importance of generosity and hospitality. For example, it is common to offer guests food and drink upon arrival, and to go out of one's way to make them feel welcome and comfortable.
Overall, doing a service to a friend is considered to be of the utmost importance in Spanish culture as a way of demonstrating loyalty, generosity, and support for those who are important in one's life.



Hofstede Cultural Survey Questions



Language Models are *Culturally Biased*



[Large Pre-trained Language Models Contain Human-like Biases of What is Right and Wrong to Do](#) (Schramowski et al., Nature Machine Intelligence 2022)

[Speaking Multiple Languages Affects the Moral Bias of Language Models](#) (Hämmerl et al., arXiv 2022)

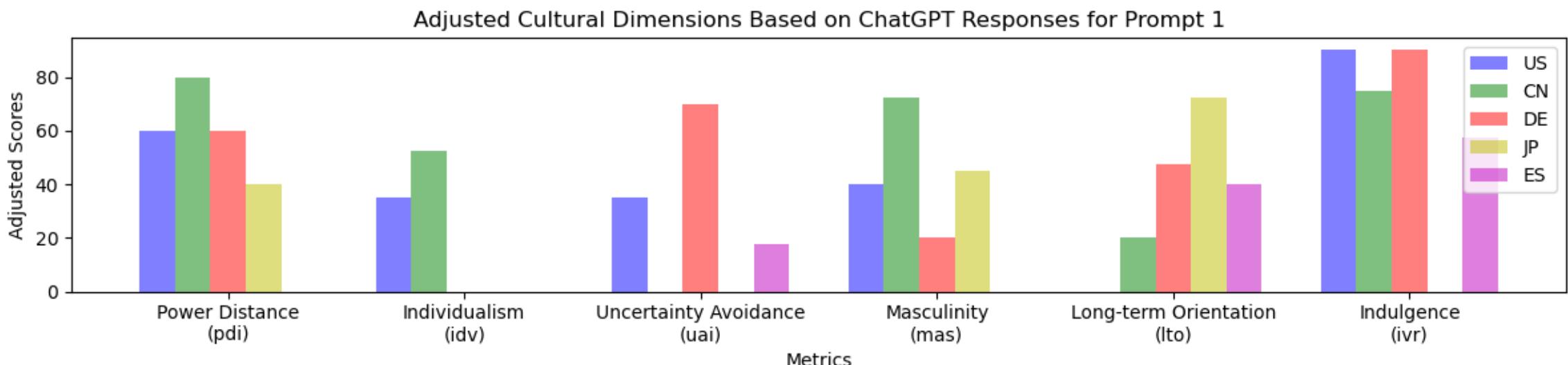
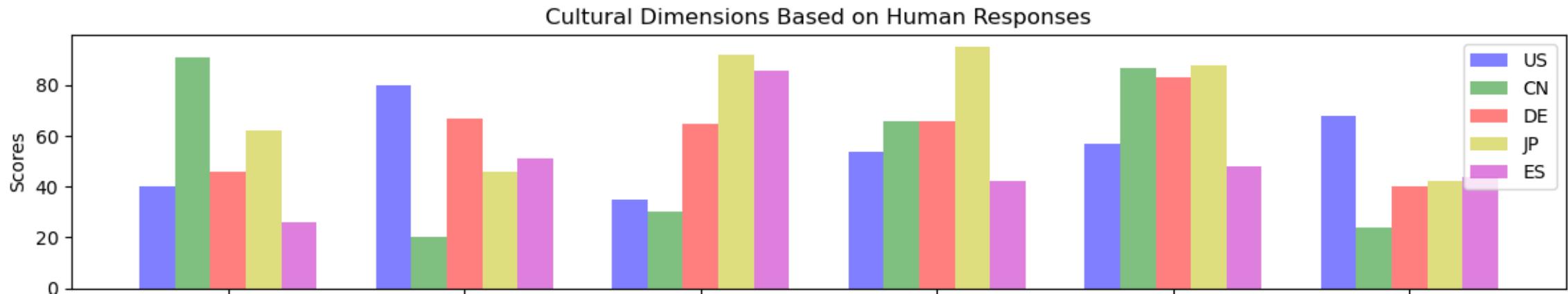
[Probing Pre-Trained Language Models for Cross-Cultural Differences in Values](#) (Arora et al., C3NLP 2023)

[Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study](#) (Cao et al., C3NLP 2023)

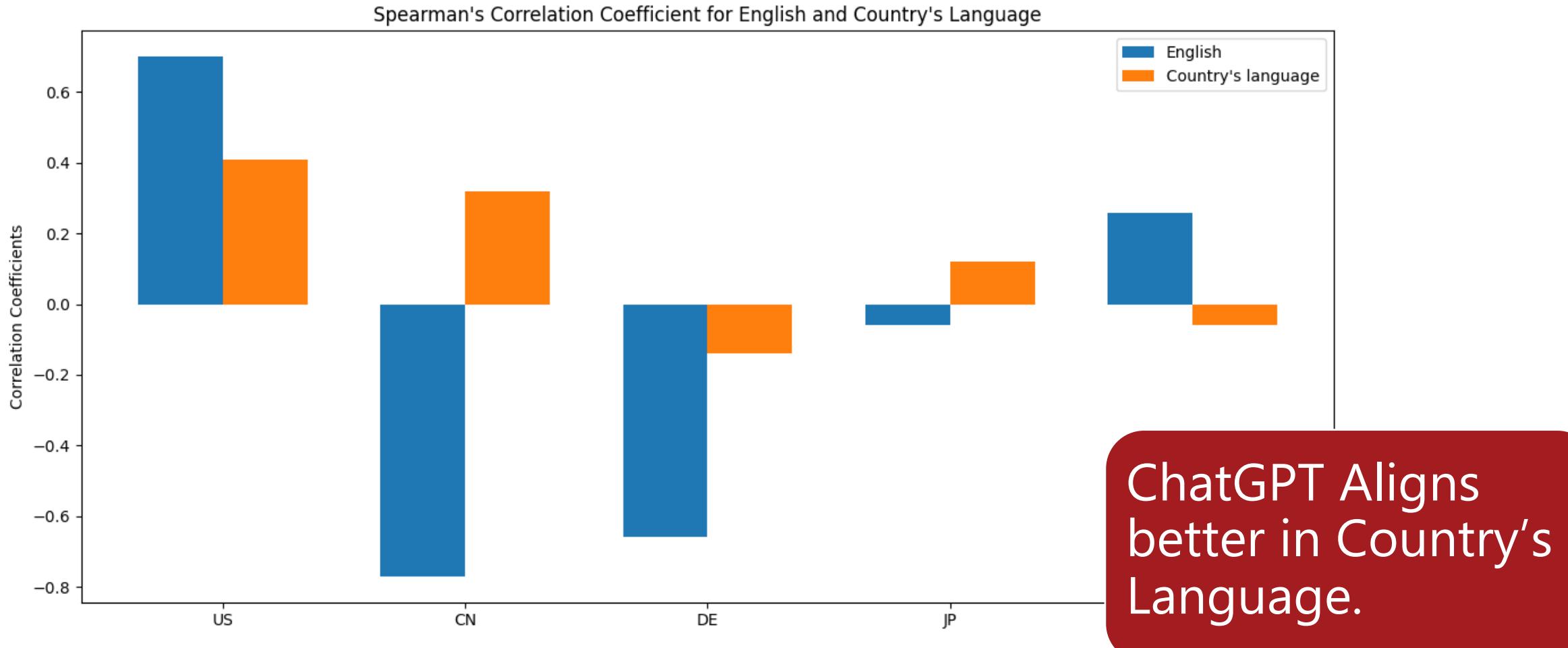
Prompt Templates

Country	Prompt 1	Prompt 2	Prompt 3
US 	For an average American, [Question #] is [Options].		In the American cultural setting, [Question #] is [Options].
China 	For an average Chinese, [Question #] is [Options].	对普通中国人来说, [问题 #] 是 [选项]。	In the Chinese cultural setting, [Question #] is [Options].
Germany 	For an average German, [Question #] is [Options].	[Frage] ist [Option] für den durchschnittlichen Deutschen.	In the German cultural setting, [Question #] is [Options].
Japan 	For an average Japanese, [Question #] is [Options].	平均的な日本人の合、【 #】は【肢】です。	In the Japanese cultural setting, [Question #] is [Options].
Spain 	For an average Spanish, [Question #] is [Options].	Para el español promedio, [Pregunta #] es [Opciones].	In the Spanish cultural setting, [Question #] is [Options].

LLMs' Predictions on Cultural Dimensions



Correlation to Human Survey Responses: Multiple Languages



Exploring Visual Culture Awareness in GPT-4V

- While Pretrained large Vision-Language models achieve remarkable performance, visual culture awareness is still not investigated.
- RQ1: Can GPT-4V extract fine-grained visual cultural features and provide high-quality descriptions?
- RQ2: Is GPT-4V proficient in perceiving cross-cultural knowledge?



Caption: Picha ya upande wa kulia ina waumini wameshika kitabu cha dini wakati wa Ibada ya **Ijumaa Kuu** na picha ya upande wa kushoto ina mtu aliyebeba msalaba akifuatwa na watu wengine nyuma wakiadhimisha **Ijumaa Kuu**. (*The picture on the right has believers holding a religious book during the **Good Friday** Service and the picture on the left has a person carrying a cross followed by other people in the background celebrating **Good Friday**.*)

Label: True

(a) An example of MaRVL in Swahili.



MaRVL: 兩張圖中的飯廳裡都沒有人。(*There is no one in the **dining room** in either picture.*)

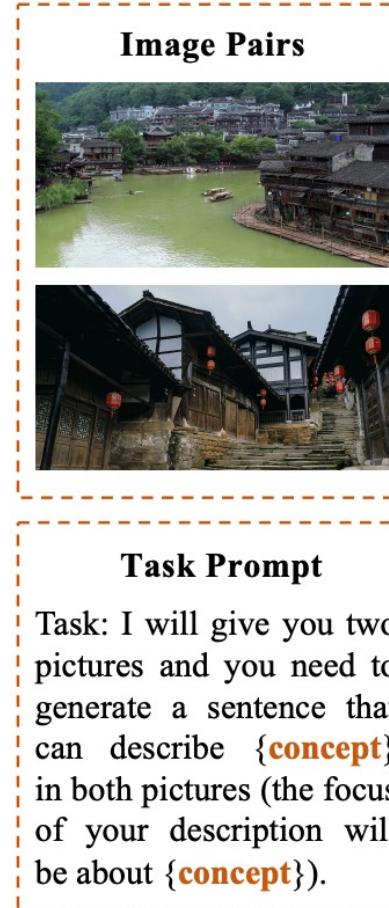
GPT-4V: 第一张图片展示了一个**现代风格**的**饭厅**, 有着城市景观; 第二张则是**传统中式风格**。(*The first image shows a **modern style** dining room with a city view; the second one is in a **traditional Chinese style**.*)

(b) Caption Comparison between MaRVL and GPT-4V.

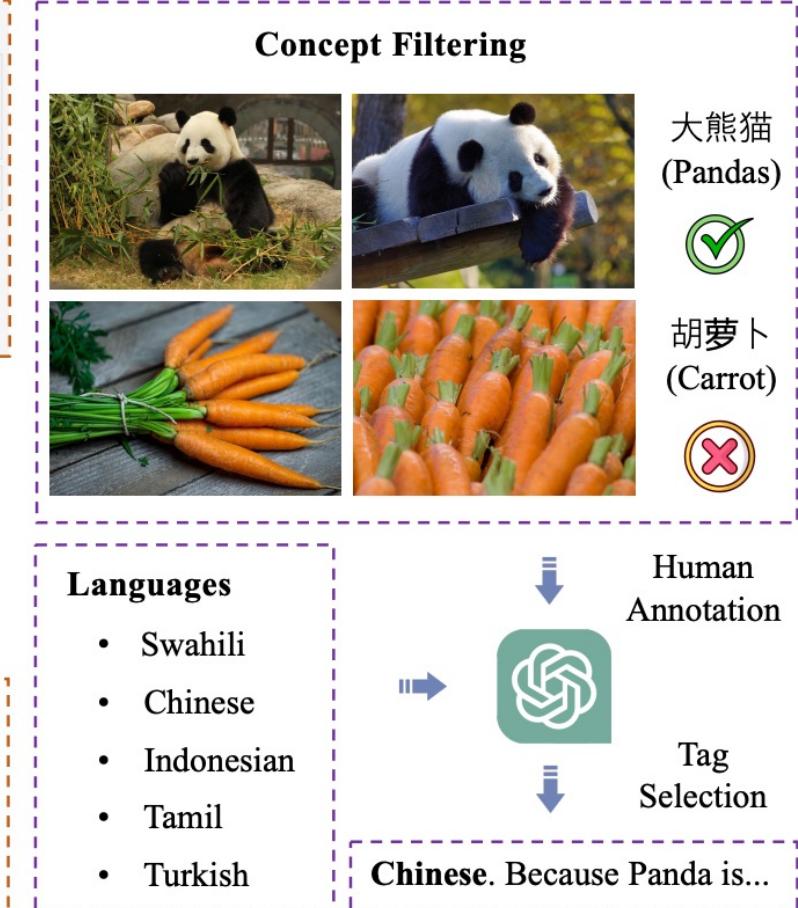
Visual Cultural Probing Framework



(a) Caption Classification



(b) Cultural Caption Generation



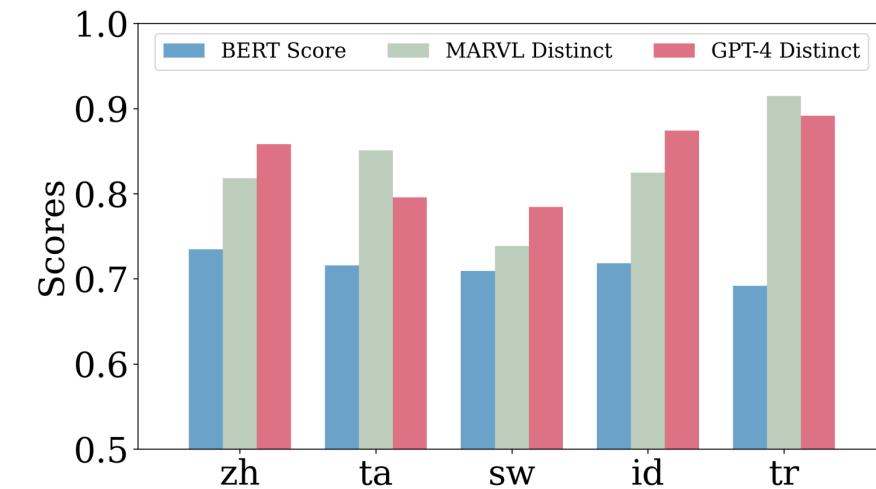
(c) Cultural Tag Selection

Visual probing results

- GPT-4V achieves superior performance in Caption classification.
- Reference-based evaluation lacks distinction.

Model	Method	ZH	TA	SW	ID	TR	avg.
mUNITER	MaRVL*	56.8	52.2	51.5	55.0	54.7	54.0
	IGLUE [◊]	55.3	52.7	51.2	54.8	54.7	53.7
	Mixup [†]	60.4	58.9	54.9	55.7	59.5	57.9
xUNITER	MaRVL*	55.0	55.1	55.5	57.1	58.0	56.1
	IGLUE [◊]	53.1	53.1	55.5	55.1	56.2	54.6
	Mixup [†]	57.9	58.0	60.0	58.8	59.7	58.9
GPT-4V	MaRVL	71.8	63.0	64.8	65.2	69.1	67.4

F1 score of Caption classification



Reference based evaluation

Visual probing results

- MaRVL and GPT-4V receive **high correctness**, but GPT-4V exhibits a significant **improvement in culture** relevant among all languages.
- GPT-4V is still **weak** in Indonesian and Turkish.

	ZH	TA	SW	ID	TR	avg.
MaRVL _{Cor}	5.024	5.000	5.067	5.915	4.189	5.039
GPT4V _{Cor}	5.683	5.577	5.371	5.623	5.552	5.561
MaRVL _{Cul}	2.706	3.352	3.730	2.717	2.378	2.977
GPT4V _{Cul}	4.238	4.183	4.169	3.283	3.371	3.849
All bad	9.5%	7.0%	2.2%	20.8%	12.6%	10.4%
All good	20.6%	29.6%	22.5%	19.8%	12.6%	21.0%
MaRVL	10.3%	15.5%	20.2%	17.9%	15.4%	15.9%
GPT-4V	59.5%	47.9%	55.1%	41.5%	59.4%	52.7%

文本生成质量评估 (图像→文本)

请阅读下面的两张图片和两句话描述，并对下面的问题进行评分。

左图



右图



句子a: 左图中只有一支毛笔，右图则有数隻。

句子b: 第一张图片中的毛笔头部尖细，而第二张中的毛笔头部更为饱满。

问题

正确性(句子a) - 问题1: 句子a能够正确描述两张图片的内容。
(1表示完全错误, 7表示完全正确)。

1 2 3 4 5 6 7

蕴含文化(句子a) - 问题2: 句子a的描述内容和文化属性相关。
(1分代表和文化完全没关系;7分代表和文化强相关)。

1 2 3 4 5 6 7

正确性(句子b) - 问题3: 句子b能够正确描述两张图片的内容。
(1表示完全错误, 7表示完全正确)。

1 2 3 4 5 6 7

蕴含文化(句子b) - 问题4: 句子b的描述内容和文化属性相关。
(1分代表和文化完全没关系;7分代表和文化强相关)。

1 2 3 4 5 6 7

文本对比 - 问题5: 句子a和句子b, 哪个更适合作为两张图片的描述?

1 句子a 2 句子b
 3. 平局 (质量相同, 都很好) 4. 平局 (质量相同, 都很差)

* 若有疑问, 请留下您的建议。

提交

Progress: 1/5
User: undefined



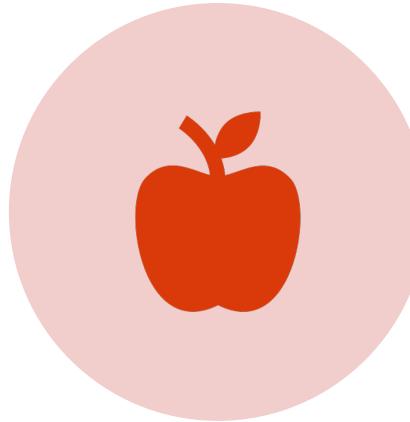
3. Cross-Cultural Adaptation in specific domain



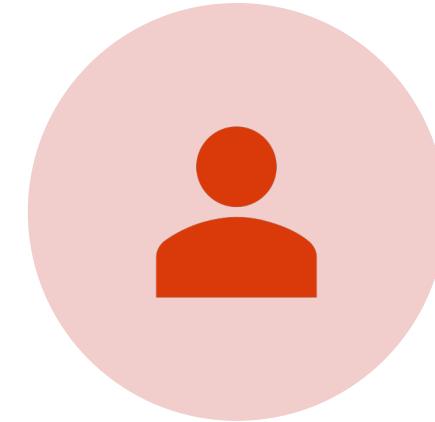
Recipe, language and identity



RECIPE IS A
**COMMUNICATIVE
MEDIUM**



RECIPE IS A
**CARRIER OF
MEANING**



RECIPE IS TIGHTLY
LINKED TO
IDENTITY

Recipe differences across cultures



Ingredient availability and preference



Measurement units, writing style



Available equipment and tools



Known techniques, typical ordering



Desired flavours



Desired texture



卷心菜番茄牛肉汤

Cabbage Tomato Beef Soup

Cabbage Beef Soup

Ingredients

- | | |
|---|---|
| 1. 250克牛肉
250g beef | 1. 1 lb. ground beef |
| 2. 1个大西红柿
1 large tomato | 2. 1 onion, chopped |
| 3. 适量葱, 姜片
<i>Moderate amount of onion, ginger</i> | 3. 2 soup cans water |
| 4. 适量糖, 盐
<i>Moderate amount of sugar, salt</i> | 4. 1 can tomato soup |
| 5. 数片卷心菜叶子
A few cabbage leaves | 5. 1 can beef bouillon soup |
| | 6. 1 can drained kidney beans |
| | 7. 1/2 head cabbage |
| | 8. Salt and pepper |
| | 9. 1 spoon of Hot sauce |
| | 10. 1 can tomatoes, diced into small pieces |

Cooking Steps

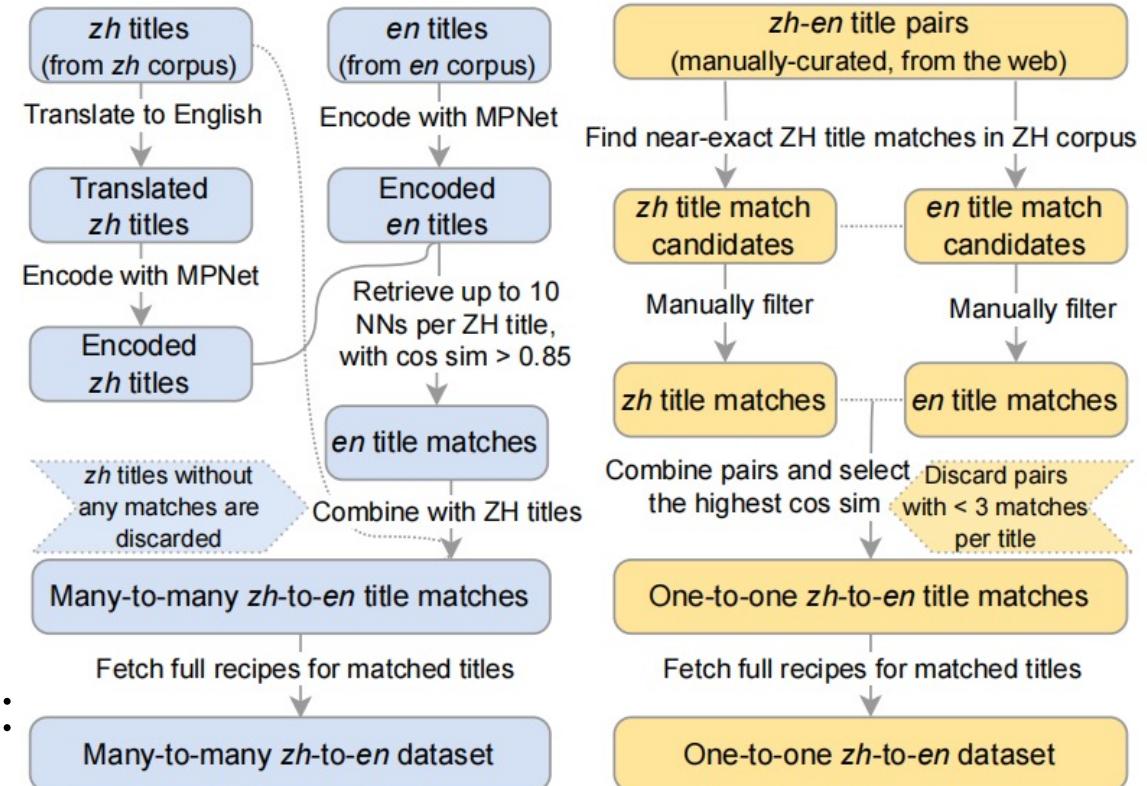
- | | |
|---|---|
| 1. 牛肉切小块, 焯水。
Cut beef into small pieces, <i>blanched</i> . | 1. Pressure cook cabbage for 8 minutes or cook on the <i>stove</i> and drain. |
| 2. 洗净后加入葱和姜片, 炖煮半小时左右。
Wash beef, add green onion and ginger slices, and simmer for 0.5h. | 2. Brown crumbled ground beef and onion |
| 3. 捞出葱姜, 倒入炖锅炖。
Take out the onion and ginger from the above and add them into the <i>stew pot</i> and continue to simmer. | 3. If the above is done, add the above into a <i>crock-pot</i> . Cook for 8 hours (add more water if needed) or on high |
| 4. 番茄去皮, 切块。倒入牛骨汤中, 加入适量糖和盐, 继续炖。
Peel the tomatoes and cut them into pieces. Pour in the beef broth, add some sugar and salt, and continue to simmer. | 4. After the beef is stewed, add the cabbage and continue to simmer for 15-20 minutes. |
| 5. 等牛肉炖酥后, 加入卷心菜, 继续炖15-20分钟。 | 5. Once the beef is tender, add the cabbage and continue to simmer for 15-20 minutes. |

Can we automate this process?

Preliminary data collection

- 2M **English** recipes from various websites (e.g., food.com)
- 1.5M **Chinese** recipes from xiachufang.com

Automatic pairing based on assumptions:



If two recipes have the **same** title, they describe the same dish.



If two recipes have a **semantically equivalent** title (possibly in another language), they describe the same dish.

Human Set Construction

- Collaboration with [Department of Food Science, University of Copenhagen.](#)



Recipe Adaption

Chinese → English

We provide you with several recipes in Chinese and would appreciate it if you could adapt them into English recipes.

[Let's Begin](#)

Recipe Adaption

英文 → 中文

我们提供了一些英文食谱，希望您能根据您的烹饪知识和文化背景将它们转换为中文食谱。

[开始改写](#)

Recipe Evaluation

Chinese → English

We will present a collection of recipes generated by the model and hope your assessment based on your cultural heritage.

[Let's Begin](#)

Recipe Evaluation

英文 → 中文

我们将向您展示由预训练模型生成的食谱，希望您能根据您的烹饪知识和文化背景进行评估。

[Let's Begin](#)

食谱改写 (英语→中文)

根据您的烹饪知识和文化习惯，请将我们提供的英文菜谱(左)转换为中文菜谱(右)。感谢您的帮助！

Gooey Butter Cake

Ingredients:

- 1 box pound cake mix
- 4 eggs
- 1 stick margarine, melted
- 1 (1 lb.) box powdered sugar
- 1 (8 oz.) cream cheese
- 1 tsp. lemon juice
- chopped pecans (optional)

Steps:

- Mix together the cake mix, 2 eggs and margarine.
- Spread in a 9 x 13-inch greased and floured pan.
- Mix together the powdered sugar, 2 eggs, lemon juice and cream cheese.
- Spread on first mixture.
- Cover with light layer of nuts (optional) and additional powdered sugar.
- Bake at 350° for 35 to 40 minutes; do not overbake.

请写下你的食谱

标题：

配料：

步骤：

* 建议：

User: test_user@gmail.com Progress: 0/15 [跳过](#) [提交](#)

Selected LLMs

- Opus-mt (MT-zs / MT-ft)
- mT5 (ft)
- mBART50 (ft)
- BLOOM (zs)
- ChatGLM2 (zs)
- GPT-4 (zs)



Prompt Setting

- BLOOM

En→Zh [English recipe] zhōng wén cài pǔ shì hé zhōng guó rén de 中文菜谱, 适合中国人的:

Zh→En [Chinese recipe] Recipe in English, adapted to an English-speaking audience:

- ChatGLM2 & GPT-4

Convert the provided English recipe into a Chinese recipe so that it fits within Chinese cooking culture, is consistent with Chinese cooking knowledge, and meets a Chinese recipe's style. [English recipe]

Convert the provided Chinese recipe into an English recipe so that it fits within Western cooking culture, is consistent with Western cooking knowledge, and meets a Western recipe's style. [Chinese recipe]

Evaluation Metrics

- Automatic Evaluation
 - SacreBLEU
 - ChrF
 - Rouge-L
 - BertScore
- Human Evaluation *
 - Grammar
 - Consistency
 - Preservation
 - Culture



(English to Chinese) A study about cultural adaptation of cooking recipes

By yong cao

£1.50 • £9.00/hr 10 mins 100 places



Prerequisites

- Proficiency in both English and Chinese languages
- Comprehensive knowledge of Chinese cooking cultures
- Diligently study the provided recipes, and we kindly request your thorough evaluation.
- Evaluations with a total time of less than 3 minutes will not be accepted, and normally it will take you 1-3 minutes to read and understand two recipes.

Our Expectations

For this research, we will present a selection of culturally adapted recipes and request your assessment based on your cultural background. You need to evaluate different recipes in 3 turns.

* If you didn't pass comprehension check, you'll get 1/3 price for this evaluation (0.2£). If you pass all the evaluation, you'll get all price for this evaluation(0.6£).

Devices you can use to take this study:

Desktop Mobile Tablet

[Open study link in a new window](#)

* Conducted on Crowd sourcing Platform: <https://www.prolific.co/>

Human Evaluation

Recipe Evaluation (Chinese → English)

Please read the Golden Chinese recipe (left) and culturally adapted English recipe (right),
and rate the subsequent questions.

Source Recipes

标题:

红豆汤

配料:

1. 适量红豆
2. 适量米酒
3. 适量带皮老姜,

步骤:

1. 姜切丝, 和红豆放入米酒中, 搅匀浸泡8小时
2. 浸泡好的红豆放入锅内, 大火煮沸, 搅拌一下, 调成小火熬制30分钟。

Adapted Recipes

Title:

Sweet Red Bean Soup

Ingredients:

1. 2 cups dried adzuki beans
2. 1/2 cup sugar
3. 1 inch fresh ginger

Steps:

1. Soak the beans in water for 8 hours.
2. Drain the beans and put in a medium-sized pot.
3. Peel and julienne the ginger, and add it to the pot.
4. Add 6 cups of water. Bring to a boil over high heat, stir, lower the heat and let simmer for 30 minutes.

Evaluated on a scale of 1-7.

Questionnaire

* Unless otherwise specified, the criteria should be applied to the culturally adapted recipe (the right one)
and rated on a scale of 1 to 7.

Grammar -- Question 1: How would you rate its grammatical accuracy? (1 indicates very poor and
7 indicates grammatically accurate and fluent).

1 2 3 4 5 6 7

Consistency -- Question 2: Please rate its coherence and consistency when it is presented as an executable recipe,
including a coherent title, ingredients, and cooking steps. (1 = very poor consistency, 7 = excellent consistency).

1 2 3 4 5 6 7

Preservation -- Question 3: please rate the extent to which it accurately describes a dish similar to the
input English recipe. (1 = does not describe the similar dish at all, 7 = accurately describes the similar dish).

1 2 3 4 5 6 7

Cultural appropriateness -- Question 4: Please rate the extent to which it fits within Western cooking culture,
is consistent with Western cooking knowledge, and meets a Western recipe's style.
(1 = does not fit at all, 7 = fits perfectly).

1 2 3 4 5 6 7

Models and preliminary results

- Language models and machine translation systems trained on **automatically paired** reference recipes
- Evaluating against **automatically paired** reference recipes and **human evaluation**

		# Recipes		Mean # Tokens	
		Source	Target	Source	Target
Train & Val	$zh \rightarrow en$	44,5k	144,6k	159.1	140.2
	$en \rightarrow zh$	43,8k	120,7k	117.1	164.8
Silver Test	$zh \rightarrow en$	82	82	140.5	144.7
	$en \rightarrow zh$	52	52	122.7	153.3
Gold Test	$zh \rightarrow en$	25	25	139.8	97.1
	$en \rightarrow zh$	41	41	115.7	176.5

MT-zs: Title: Red beans soup, Ingredients: red beans, rice wine, leather ginger, Steps: Chrysotile, and red beans in the wine, soaked for eight hours, soaked in the pot, boiled in fire, stirred in fire, turned into small fires for 30 minutes.

MT-ft: Title: Red Bean Soup, Ingredients: 2 tablespoons olive oil 1 large onion, chopped 2 garlic cloves, minced 1 teaspoon dried oregano 1 teaspoons dried basil 1 teaspoon dried thyme 1/2 teaspoon salt 1/4 teaspoon pepper 2 cups red kidney beans, rinsed and drained 1 (14 1/2 ounce) can diced tomatoes, undrained 1 (15 ounce) can red kidney beans, un-drained 1 (14 ounce) can beef broth 1 cup water 1/4 cup chopped fresh parsley, Steps: Heat oil in a large saucepan over medium heat. Add onion and garlic; cook and stir until tender. Stir in oregano, basil, thyme, salt and pepper. Add beans, tomatoes, broth, water and parsley; bring to a boil. Reduce heat; cover and simmer for 30 minutes or until beans are tender. Remove from heat; let stand for 5 minutes. Stir in parsley.

GPT-4: Title: Red Bean Soup

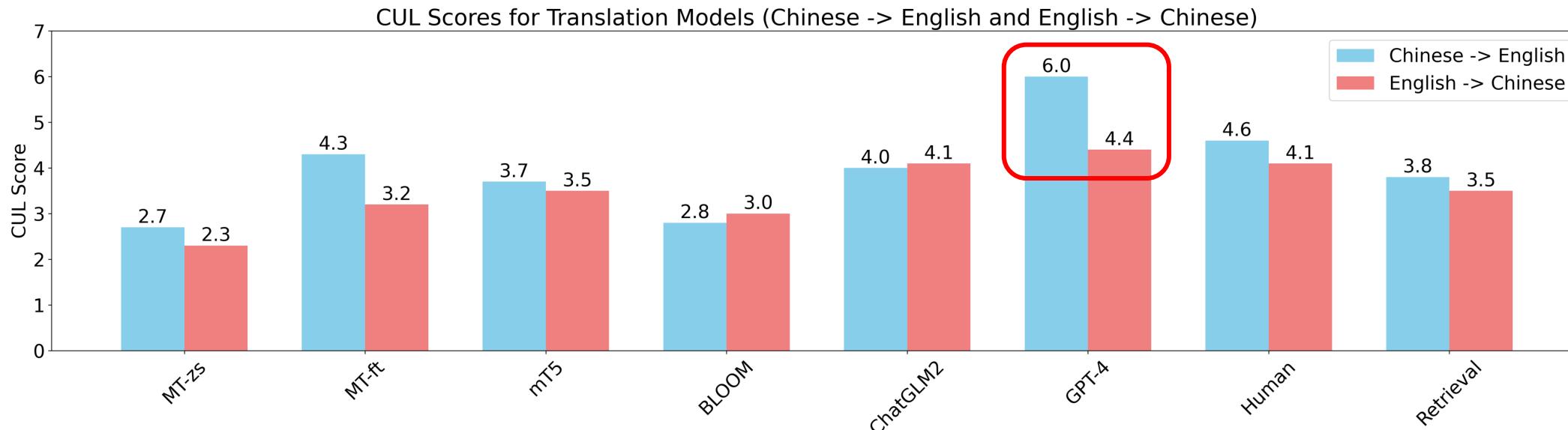
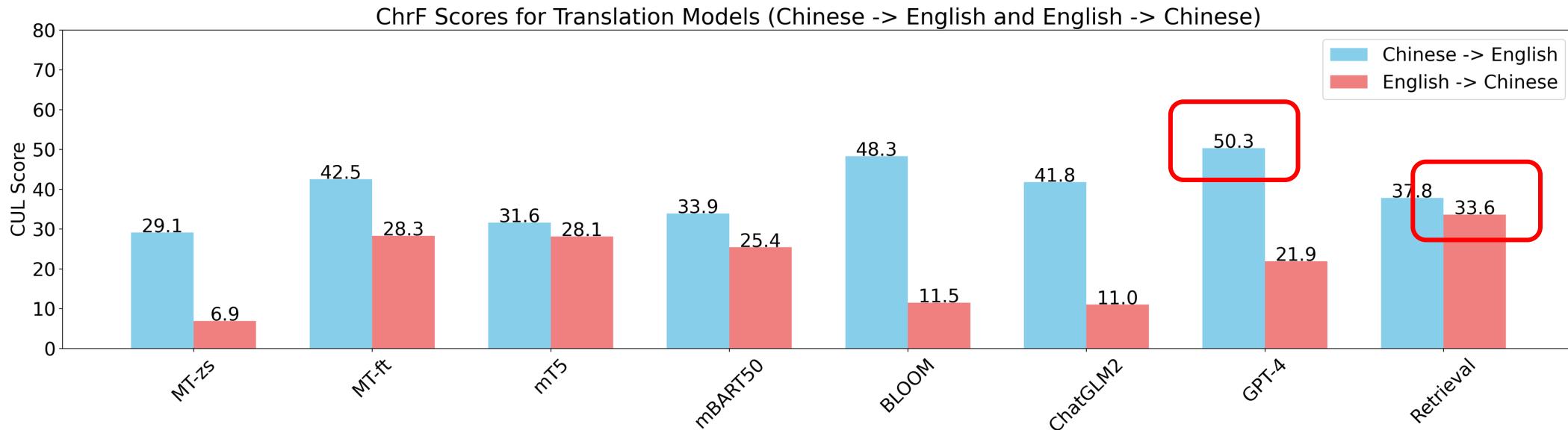
Ingredients:

- 1 cup red beans (*adzuki beans*)
- 1/4 cup rice wine (*or substitute with dry sherry*)
- 1-inch piece of fresh ginger, with skin

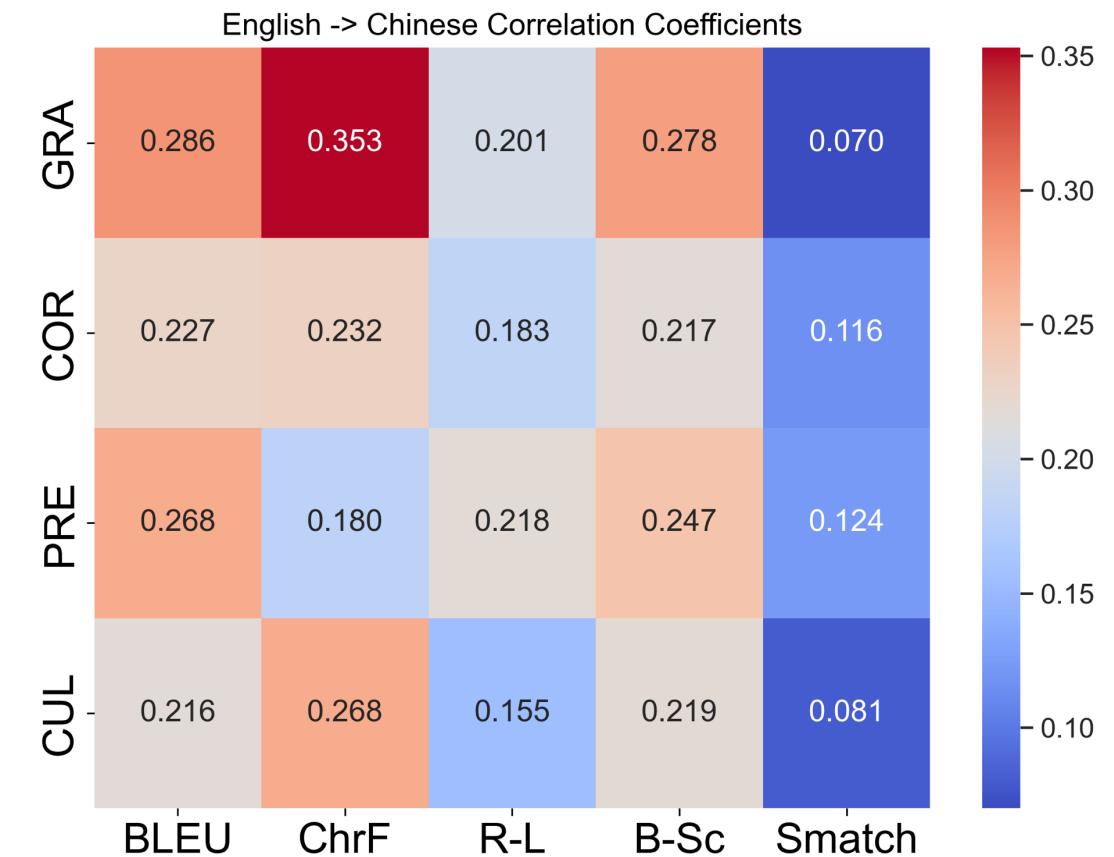
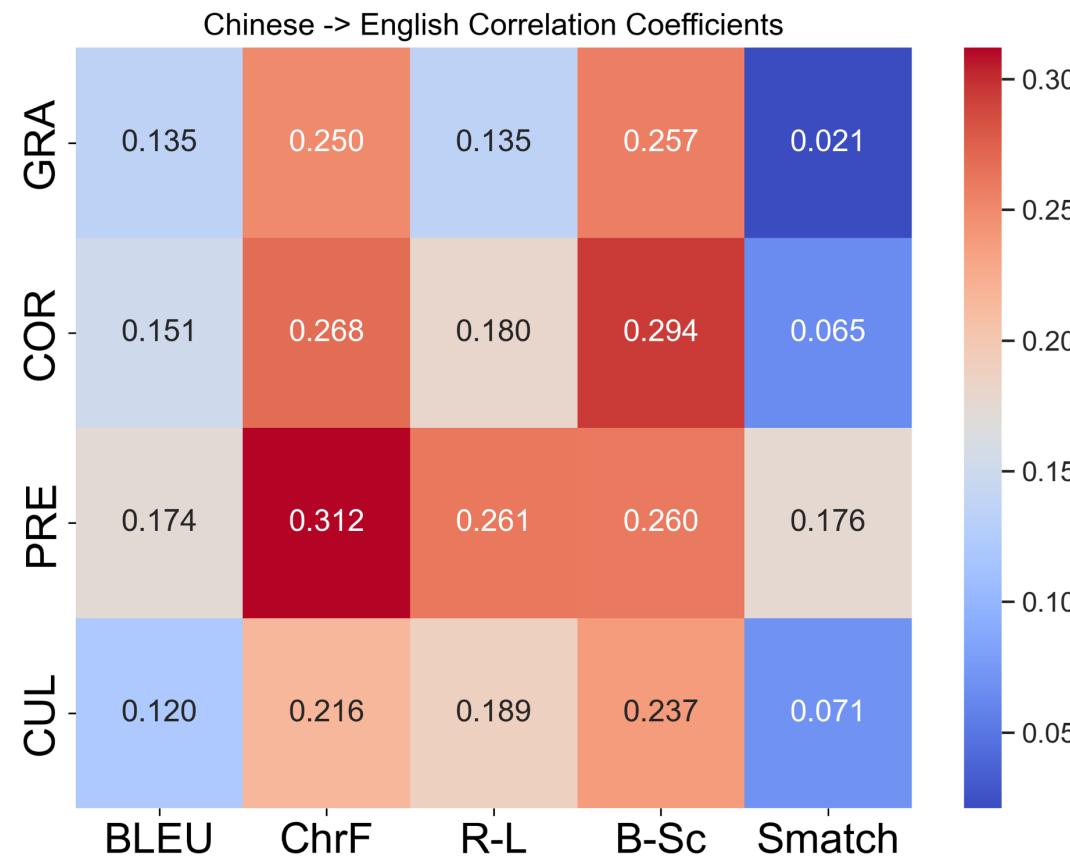
Instructions:

1. Rinse the red beans and place them in a bowl. Thinly slice the ginger and add it to the bowl with the beans.
2. Pour the rice wine over the beans and ginger, ensuring they are fully submerged. Stir to combine, then cover and let soak for 8 hours or overnight.
3. After soaking, transfer the beans, ginger, and soaking liquid to a large pot. Bring the mixture to a boil over high heat, stirring occasionally.
4. Once boiling, reduce the heat to low and let the soup simmer for 30 minutes, stirring occasionally to prevent sticking.
5. Taste the soup and adjust the seasoning if necessary. Serve hot and enjoy your Red Bean Soup!

Adaptation Results on Automatic and Human Metrics

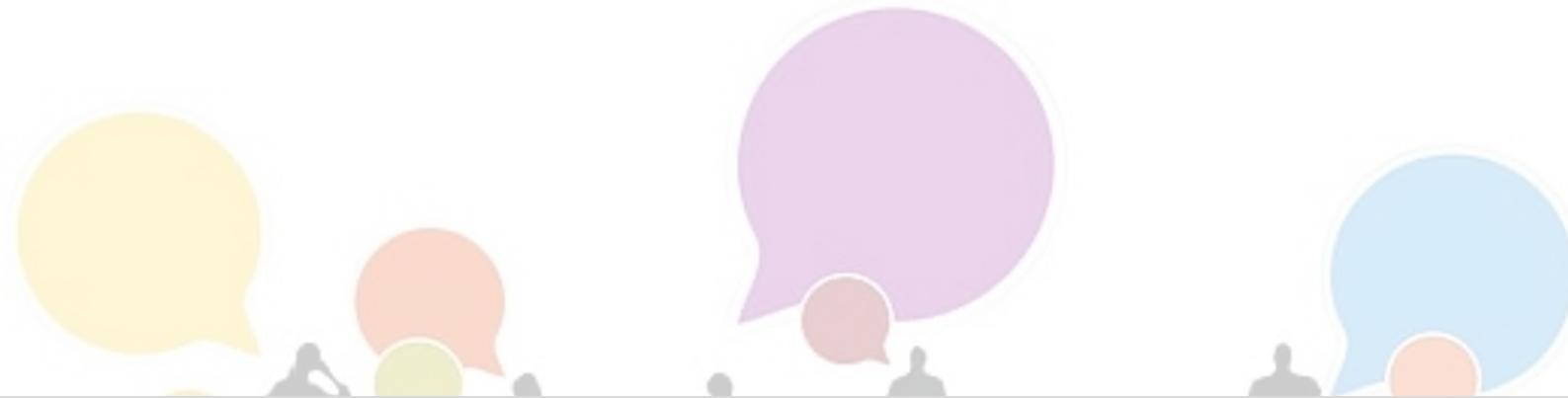


Correlations between Automatic and Human Metrics



- All cases exhibit a positive correlation, and CUL presents the weakest scores with most automatic metrics.

Kendall correlation of human evaluation results with automatic metrics.



4. Bridging Cultural Nuances in **Dialogue Agents**



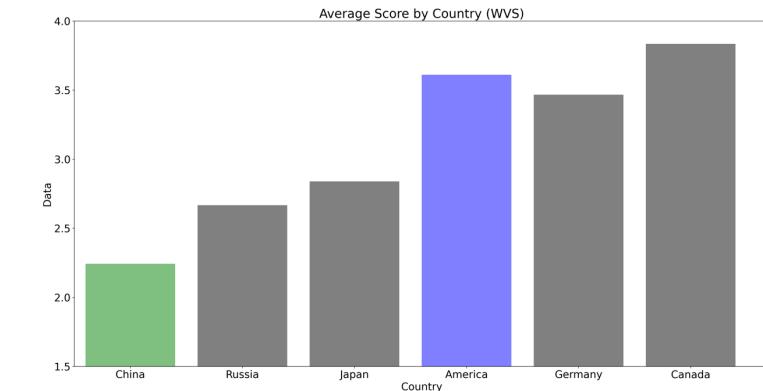
Can cultural survey guide dialogue generation?

How would you feel about the following statements? Do you agree or disagree with them?

(Read out and code one answer for each statement):

	Agree strongly	Agree	Neither agree nor disagree	Disagree	Disagree strongly
	1	2	3	4	5
Q33 When jobs are scarce, men should have more right to a job than women	1	2	3	4	5
Q34 When jobs are scarce, employers should give priority to people of this country over immigrants	1	2	3	4	5
Q35 If a woman earns more money than her husband, it's almost certain to cause problems	1	2	3	4	5
Q36 Homosexual couples are as good parents as other couples	1	2	3	4	5
Q37 It is a duty towards society to have children	1	2	3	4	5
Q38 Adult children have the duty to provide long-term care for their parents	1	2	3	4	5
Q39 People who don't work turn lazy	1	2	3	4	5
Q40 Work is a duty towards society	1	2	3	4	5
Q41 Work should always come first, even if it means less spare time	1	2	3	4	5

World Value Survey



Assume you can a representative Chinese, do you agree it is a duty towards society to have children? One sentence response.



As a representative Chinese, I may believe that having children can be seen as a duty toward society, but individual opinions on this matter can vary.

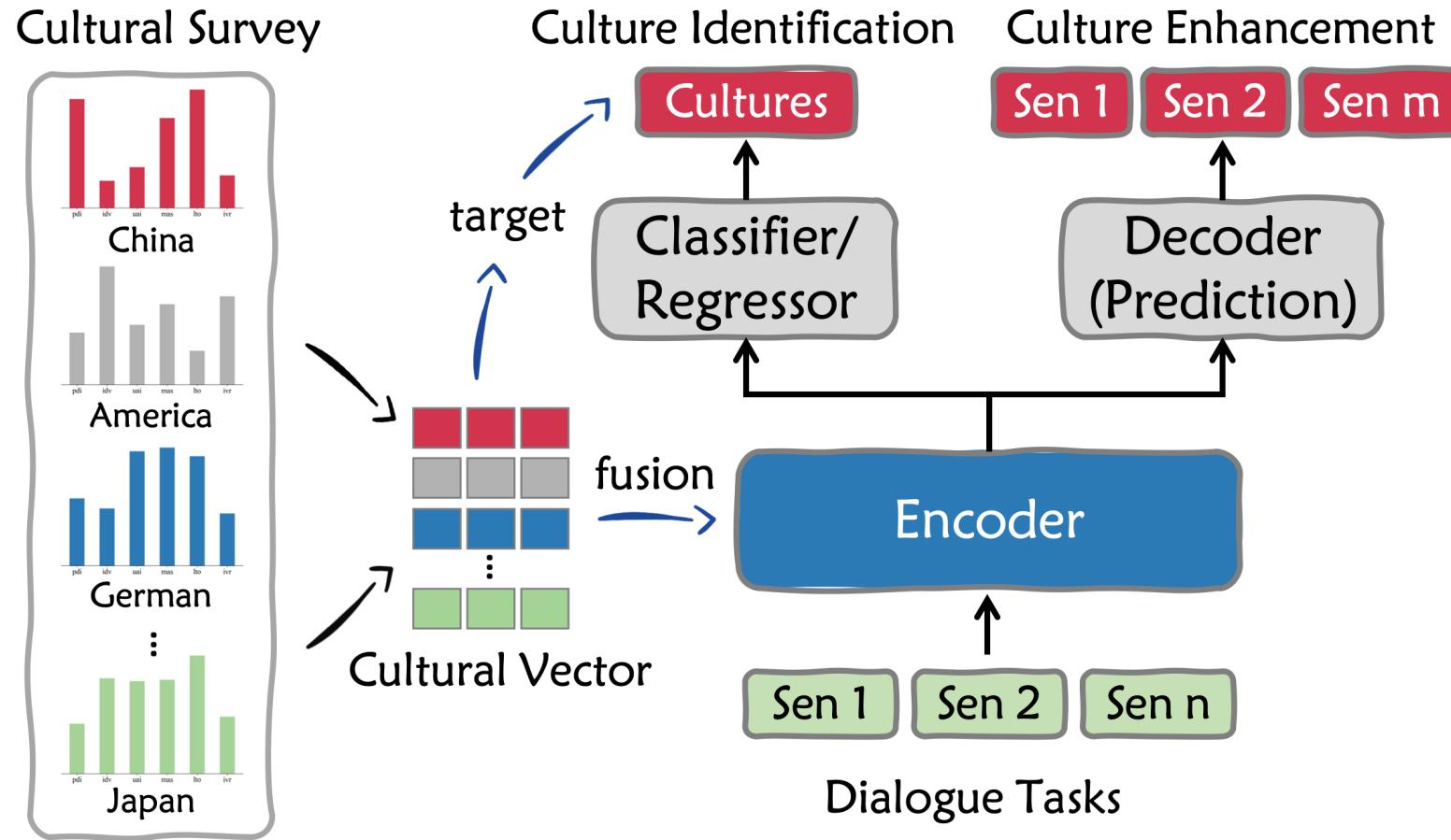


Assume you can a representative American, do you agree it is a duty towards society to have children? One sentence response.



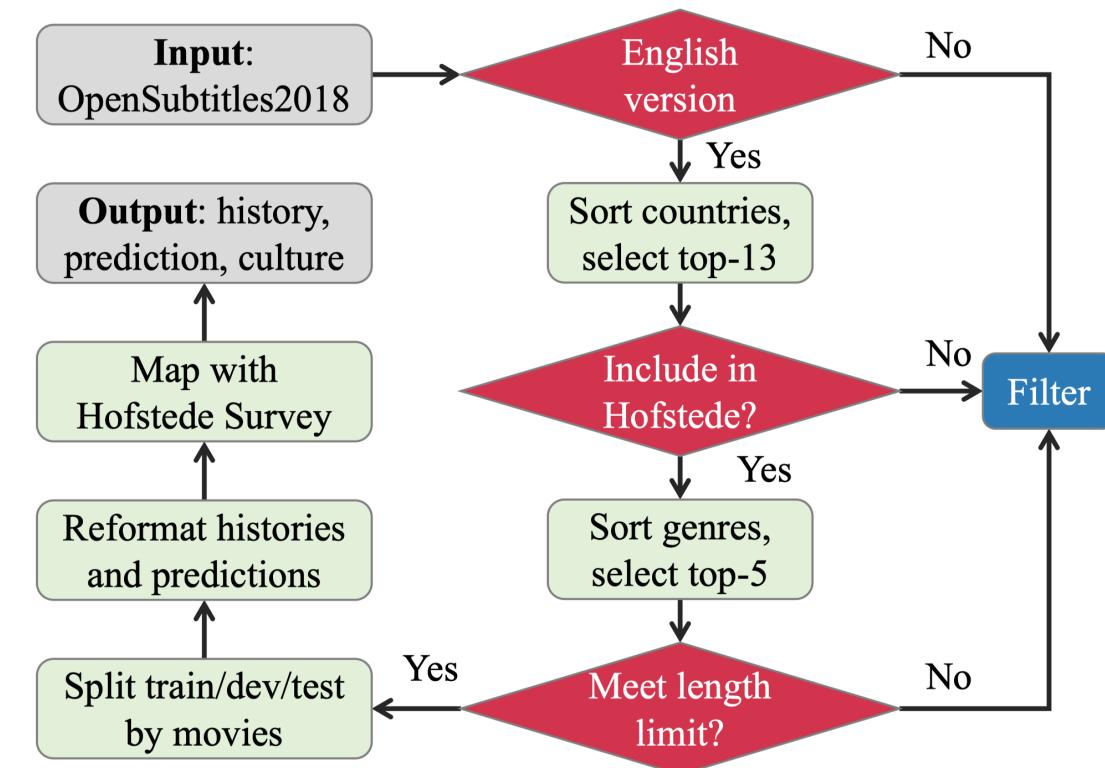
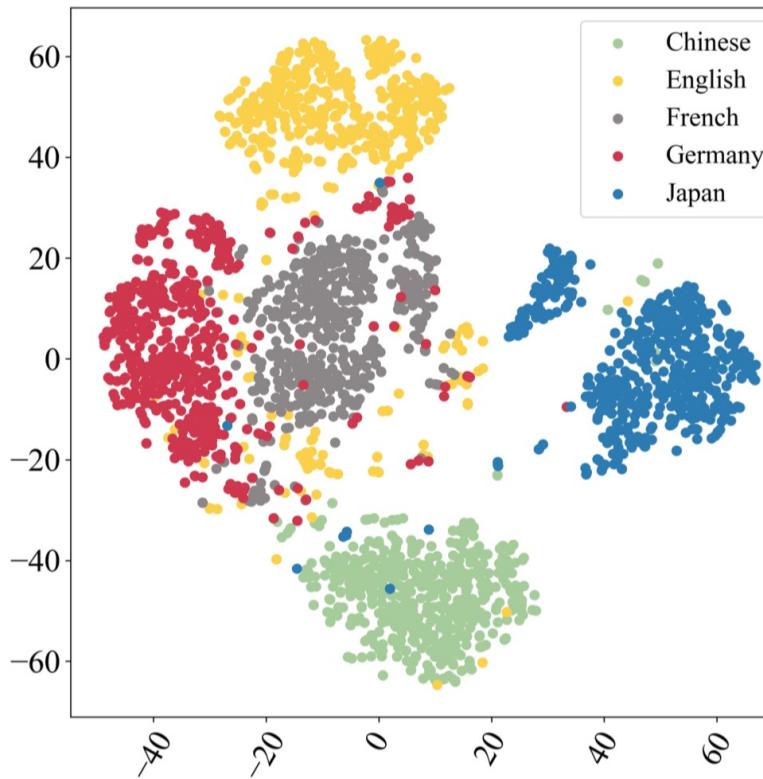
As a representative American, I would likely view having children as a personal choice rather than a duty toward society, but individual opinions on this matter can vary.

Utilizing cultural survey as an additional vector



CuDialog Benchmark Construction

- Mapping dialogue with cultural vectors from OpenSubtitles 2018.



CuDialog Benchmark Construction

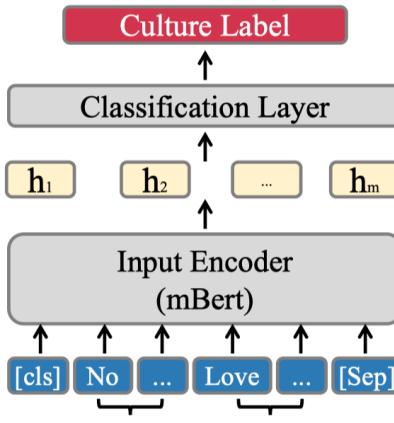
History: His mortal flesh belonged to the fire, his immortal soul to the flames of Hell. | A gag blocked his mouth.

You'd have thought it was a corpse being led to its grave, "yet it was a living man whose torments were to gruesomely entertain the people." | Forgive me, I'll break off here.

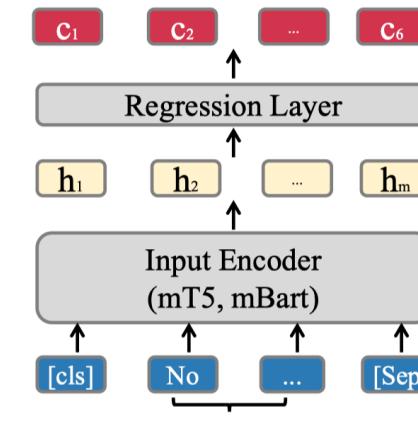
Golden Predictions: Will you amuse us now with details of an execution during the Inquisition? | No, I beg your pardon. | I'm deeply impressed.

Culture: Germany.

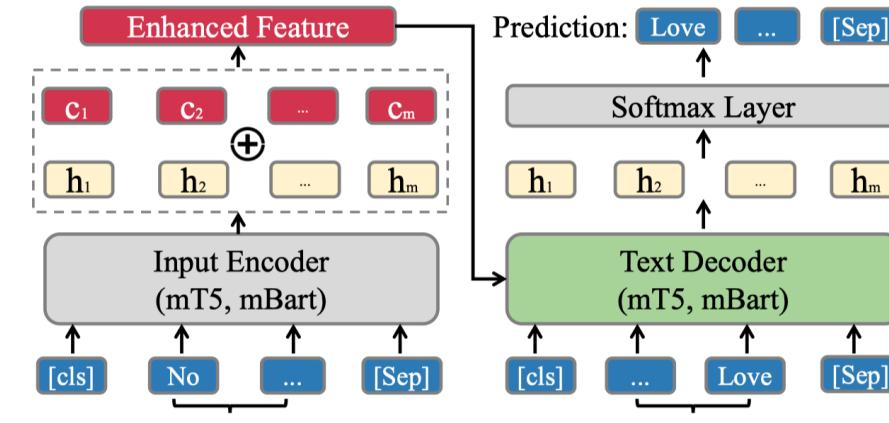
Culture Score: 35, 67, 66, 65, 83, 40.



(a) Classification task



(b) Regression task

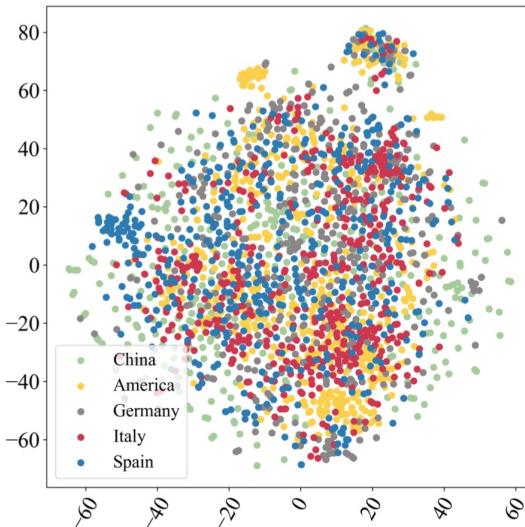


(c) Multi-turn dialogue prediction task

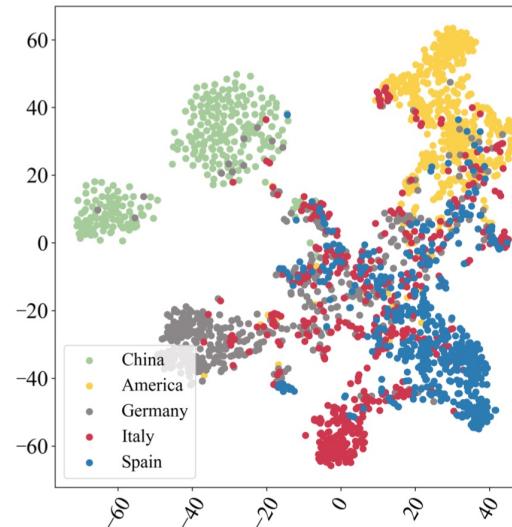
Set	Genres				
	Action	Comedy	Drama	Romance	Crime
Samples					
Train	108,934	137,475	109,534	114,467	112,695
Dev	12,808	17,361	13,256	13,697	14,313
Test	15,336	18,213	16,179	15,705	16,853
Movies					
Train	728	728	728	728	728
Dev	91	91	91	91	91
Test	104	104	104	104	104
Tokens					
Vocab	34,883	36,292	32,566	32,666	33,416
#Avg	71.15	70.32	72.38	71.42	72.05

Cultural related dialogue task

Cultural Classification

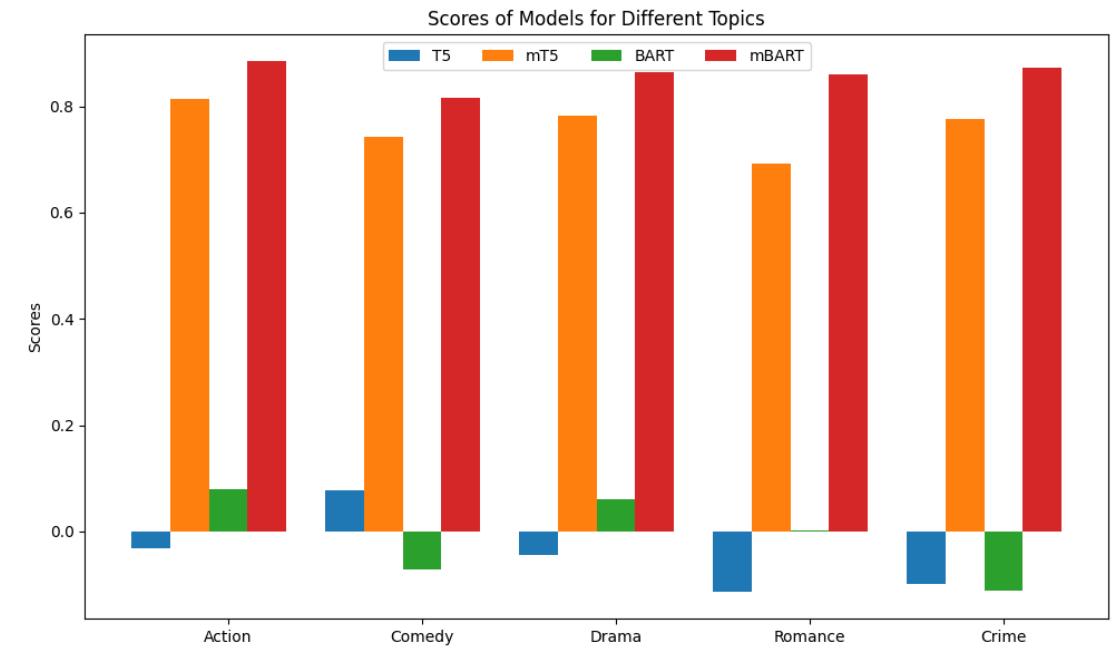


(a) Zero-shot distribution.



(b) Fine-tune distribution.

Cultural Regression



Dialogue Generation

Better alignment with references and produces more diverse results.

Genre	Model	BLEU	Rouge		Pre	BERTScore		Distinction		#Avg	Vocab
			R-1	R-2		Rec	F1	D-1	D-4		
Action	mT5 _{zs}	0.51	4.26	0.67	4.13	35.08	34.46	34.68	0.60	0.96	7.94
	mT5 _b	2.24	12.47	2.66	10.91	45.25	42.15	43.41	0.87	0.96	29.71
	mT5 _{cul}	2.41	12.62	2.78	11.05	45.47	42.37	43.63	0.89	0.96	29.57
Comedy	mT5 _{zs}	0.55	4.62	0.82	4.48	35.24	34.54	34.80	0.58	0.94	7.89
	mT5 _b	2.27	12.64	2.71	11.12	45.43	42.09	43.46	0.85	0.95	29.05
	mT5 _{cul}	2.68	13.22	2.99	11.50	45.73	42.81	43.99	0.90	0.95	29.50
Drama	mT5 _{zs}	0.66	4.64	0.78	4.49	35.31	34.68	34.91	0.59	0.97	7.97
	mT5 _b	2.31	12.80	2.64	11.24	45.90	42.32	43.82	0.82	0.97	28.83
	mT5 _{cul}	2.53	13.01	2.80	11.37	46.32	42.71	44.22	0.88	0.97	28.25
Romance	mT5 _{zs}	0.58	4.67	0.76	4.53	35.13	34.51	34.74	0.58	0.97	7.93
	mT5 _b	2.28	12.95	2.76	11.40	46.20	42.36	43.97	0.85	0.97	27.25
	mT5 _{cul}	2.17	12.66	2.72	11.17	46.02	42.17	43.78	0.83	0.97	27.43
Crime	mT5 _{zs}	0.52	4.19	0.63	4.07	35.08	34.54	34.73	0.59	1.00	7.91
	mT5 _b	2.14	12.10	2.42	10.58	45.09	42.03	43.28	0.85	1.00	29.66
	mT5 _{cul}	2.36	12.49	2.65	10.92	45.37	42.34	43.57	0.89	1.00	29.69

Conclusion

Summary

- LLMs and VLMs knowledge and capabilities differ across languages and countries
- We can use LLMs to bridge between cultures
- We must think outside the NLP box to do these things

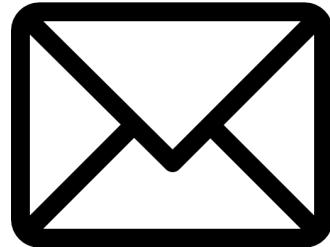
Ongoing work

- More countries, more surveys, more LLMs
- Consistency across more prompt variations and languages
- Injecting cultural knowledge for culturally adaptive LLMs
- Explainable evaluation on cultural adaptation

Thanks!



yongcaoplus.github.io



yongcao@di.ku.dk



@YongCaoPlus



Yong Cao