

Multilingual Language Models are not Multicultural: A Case Study in Emotion



**Shreya Havaldar, Sunny Rai, Bhumika Singhal, Langchen Liu,
Sharath Chandra Guntuku, Lyle Ungar**

University of Pennsylvania



GPT-4 does not adapt to different languages

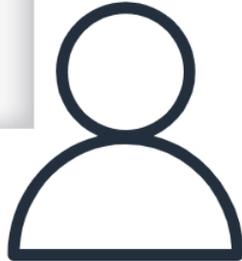


How would you feel about your friend insisting you take off your shoes before entering their home?



I would feel respectful and understanding of their preference for cleanliness in their home.

GPT-4 does not adapt to different languages



友達が家に入る前に靴を脱ぐと主張したらどう思いますか？

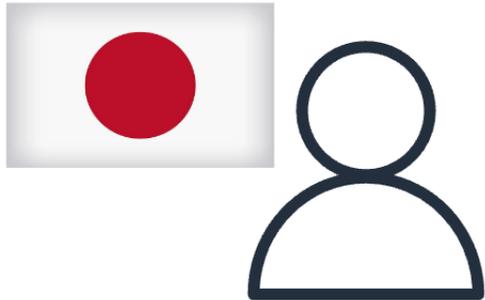
(How would you feel about your friend insisting you take off your shoes before entering their home?)



私を感じるだろう ちょっと違和感があるかもしれませんが、それが友達の家ルールであることを理解し、尊重します。

(It may feel a little strange, but I understand and respect that it's a friend's house rule.)

GPT-4 does not adapt to different languages



友達が家に入る前に靴を脱ぐと主張したらどう思いますか？
(How would you feel about your friend insisting you take off your shoes before entering their home?)



~~私を感じるだろう ちょっと違和感があるかもしれませんが、それが友達の家ルールであることを理解し、尊重します。~~
~~*(It may feel a little strange, but I understand and respect that it's a friend's house rule.)*~~

I would feel **embarrassed** or **ashamed** for forgetting to take my shoes off.

LMs require cultural sensitivity



Teaching

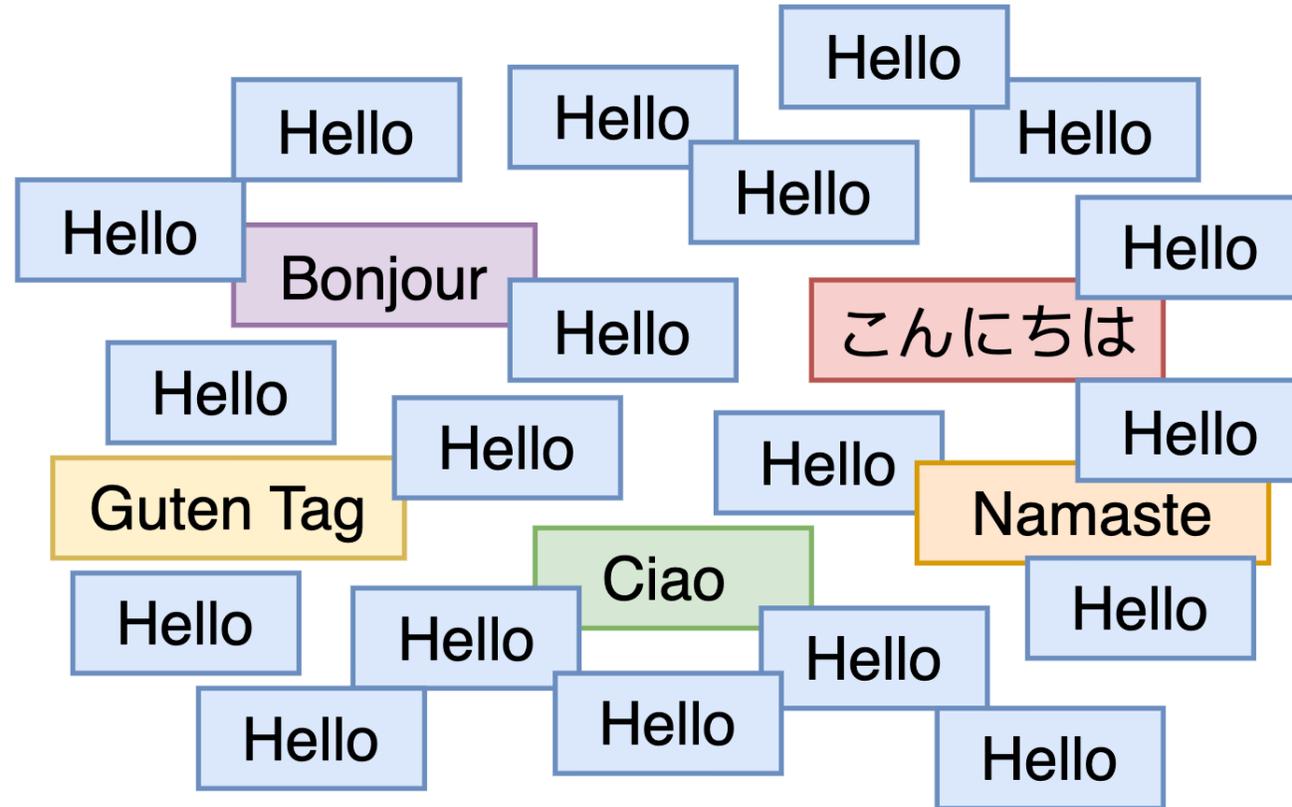


Therapy

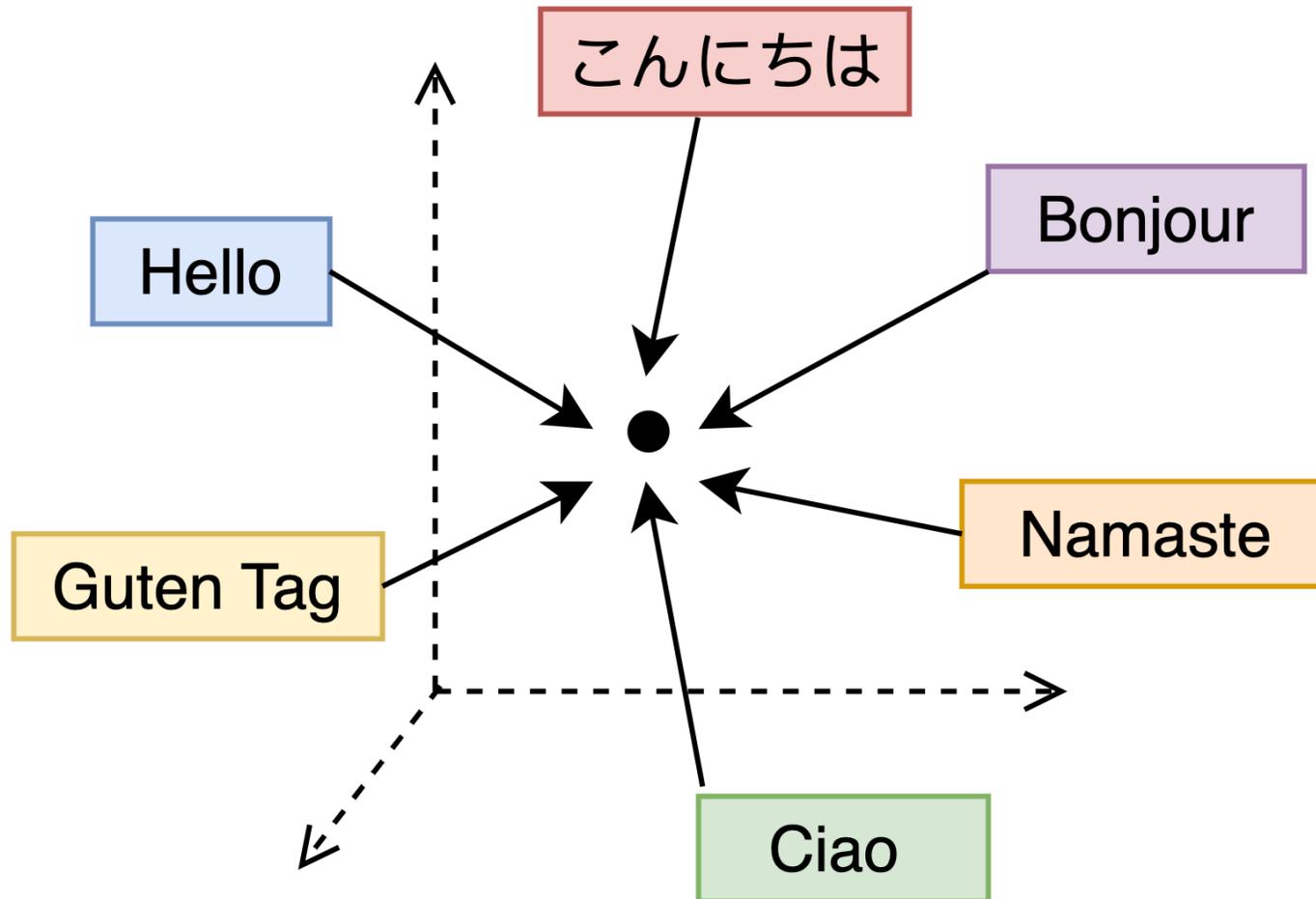


Crisis Prevention

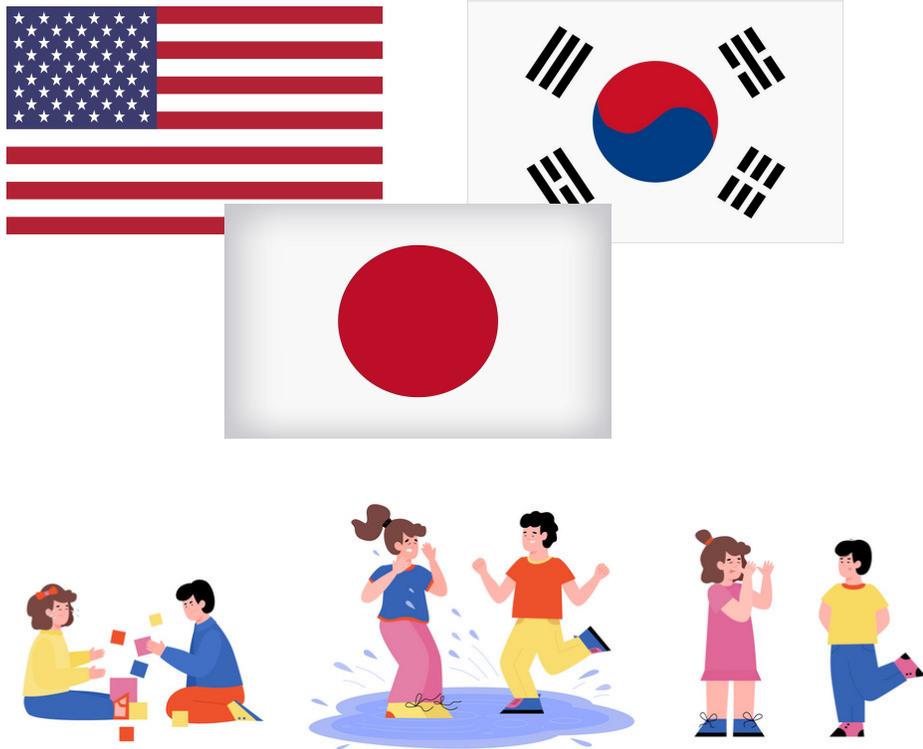
English-dominated datasets cause LMs to be Anglocentric



Training techniques cause LMs to be Anglocentric

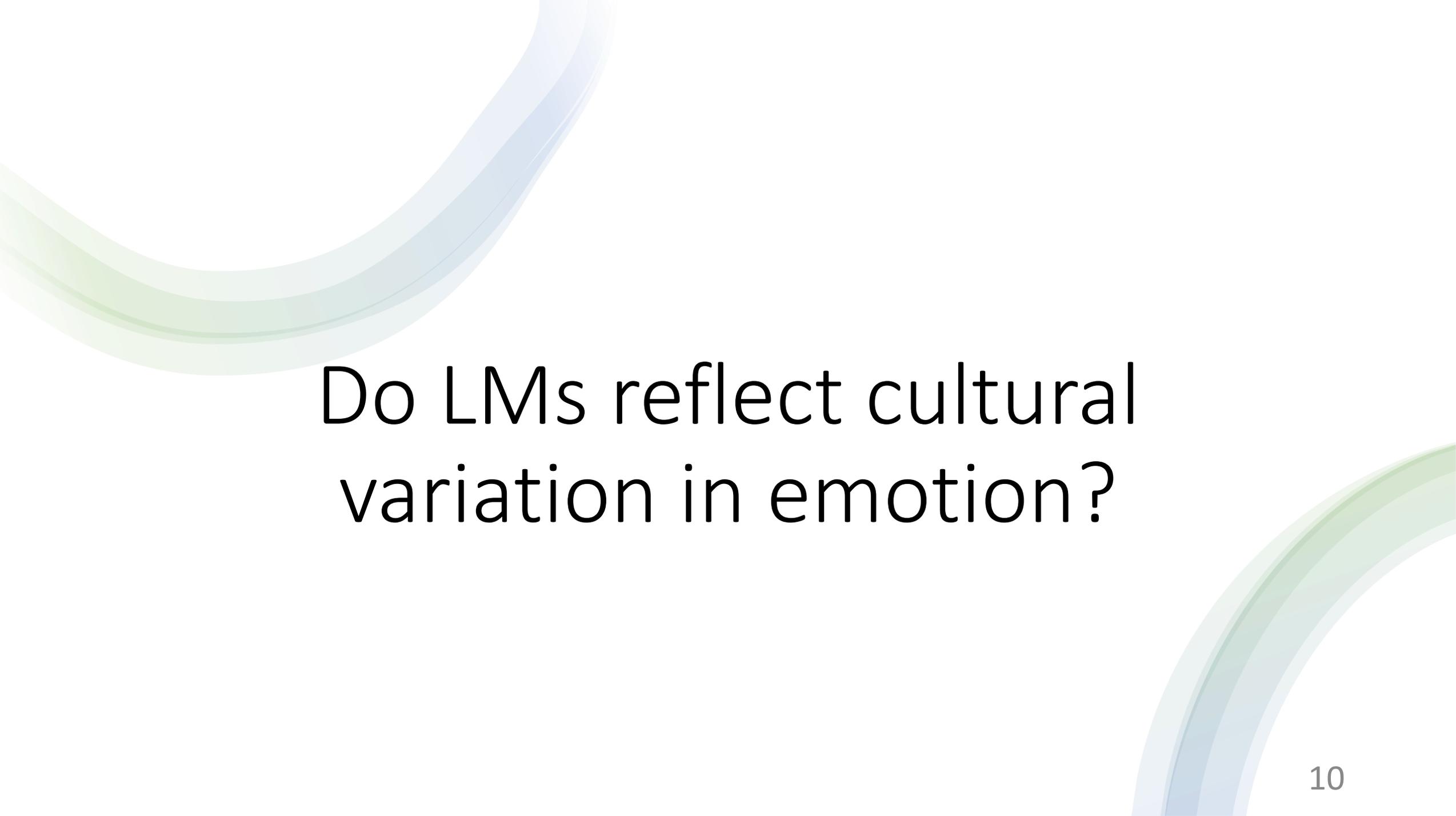


Emotions are heavily contextual. Contextual LMs should respect this.



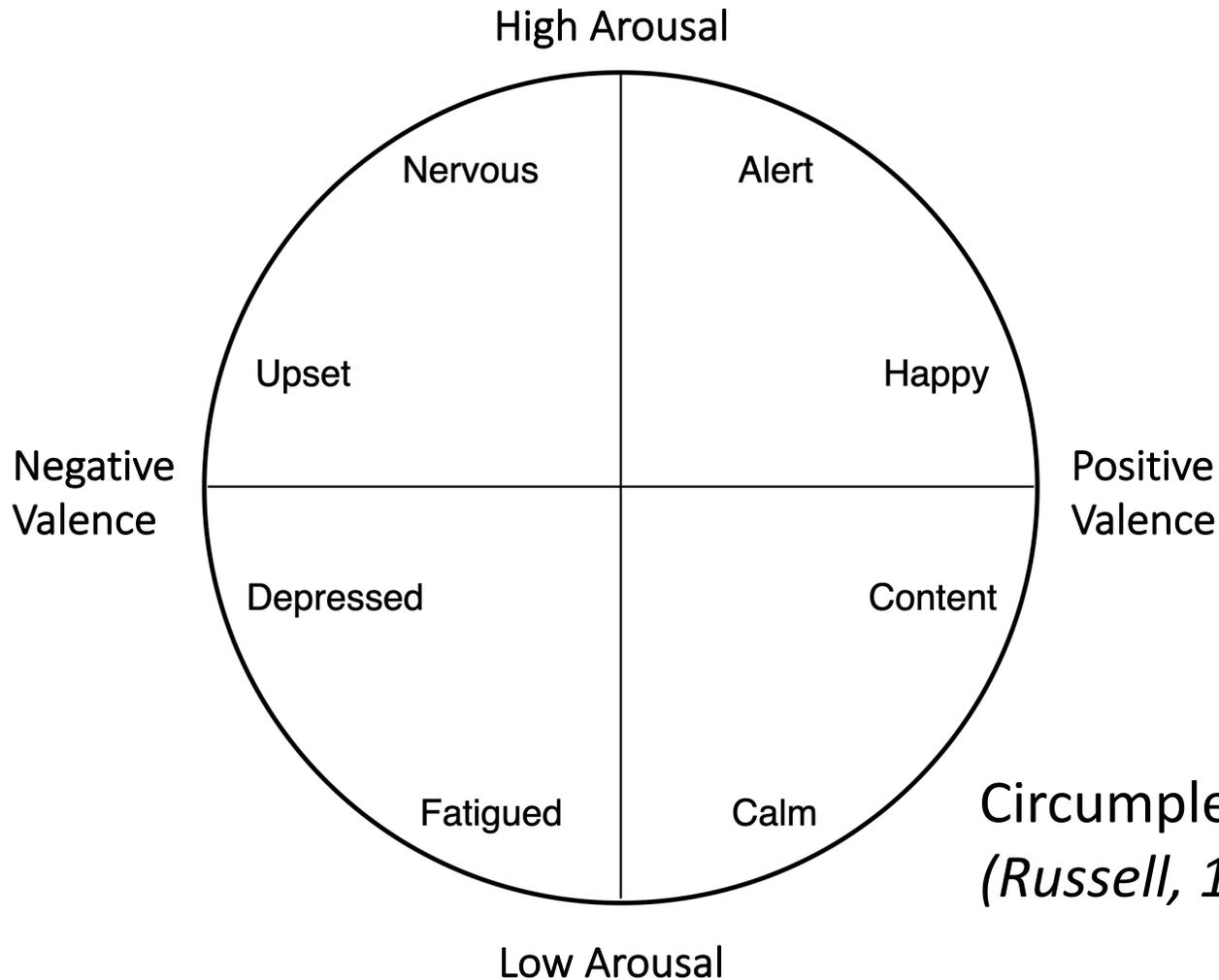
- Study with ~1,000 school children from America, Korea, and Japan
 - Test to measure “self-conscious” emotions showed substantial differences in mean levels of shame, guilt and pride
- **Japanese** children express the most **shame**
- **American** children express the most **pride**

(Furukawa et. al, 2012)



Do LMs reflect cultural
variation in emotion?

Approach 1: Use a psychological model to analyze LMs

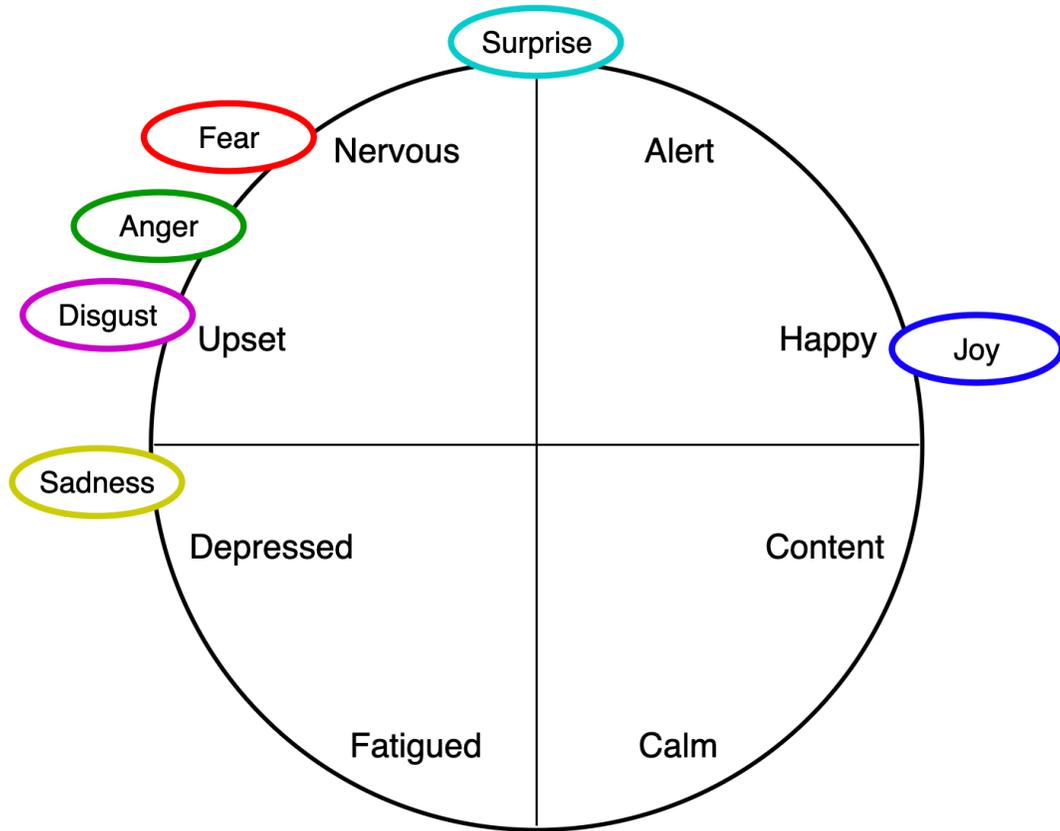


Circumplex model of affect
(Russell, 1980)

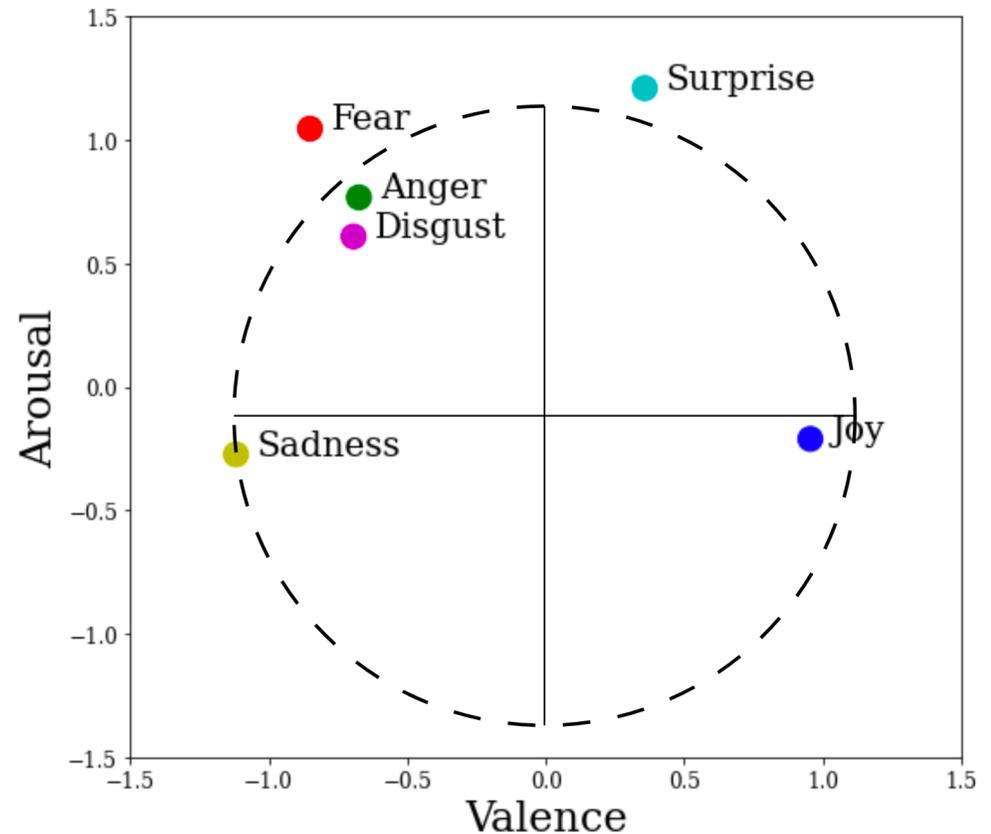
Valence: how *negative* or *positive* an emotion is

Arousal: The *intensity* and *activation* of an emotion

Projecting embeddings onto the circumplex

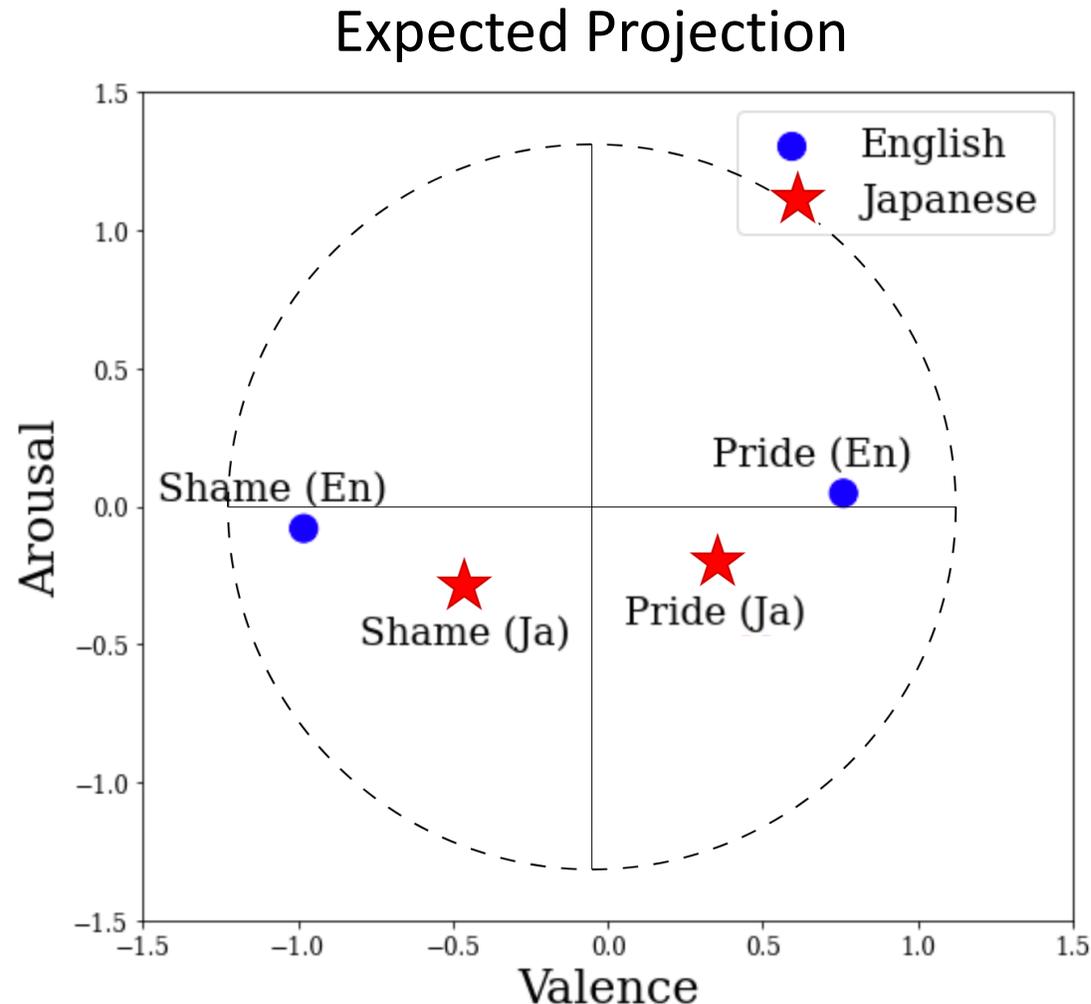


Circumplex model of affect
(Russell, 1980)



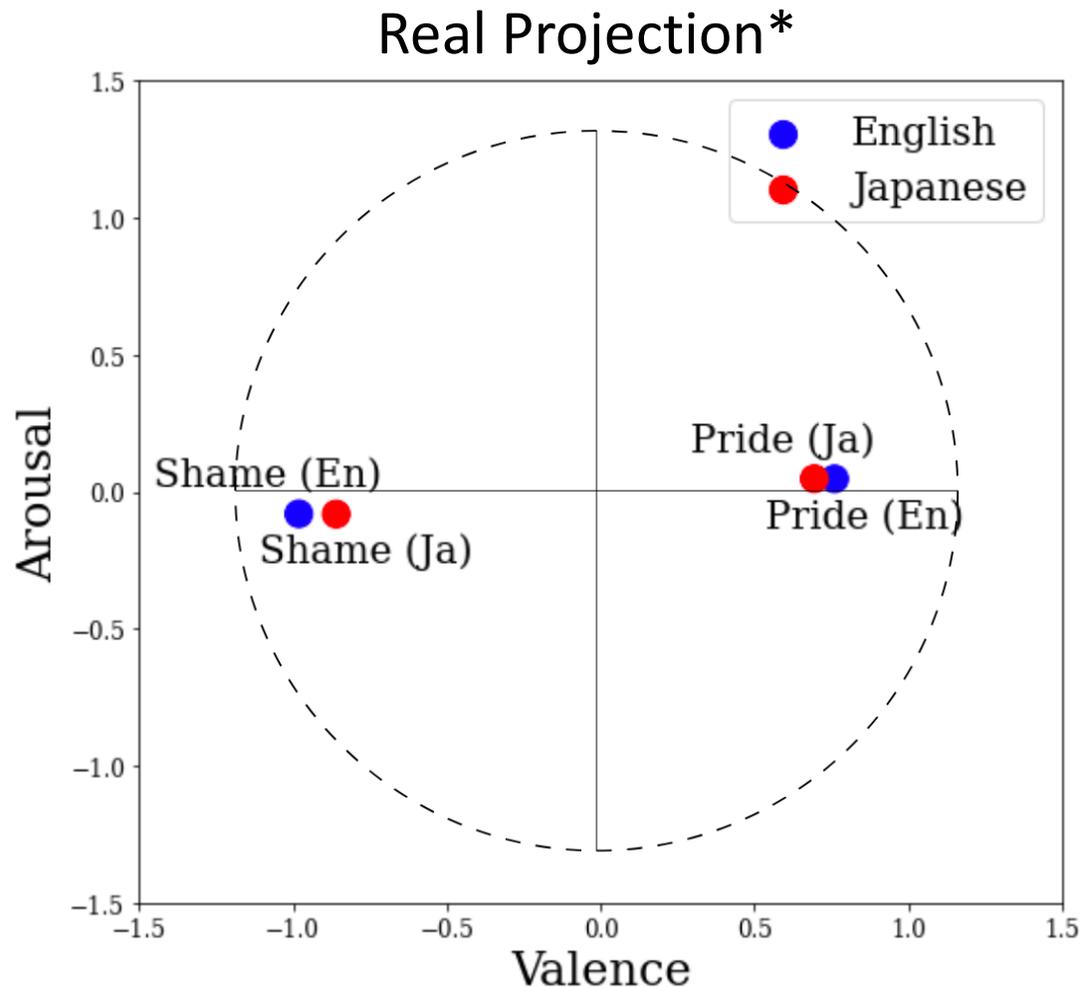
Recreated circumplex
(Havaldar et. al, 2023)

Different languages → different circumplexes



1. Japanese Shame should have a higher valence (*Furukawa et al., 2012*)
2. Japanese Pride should have a lower valence (*Furukawa et al., 2012*)
3. English emotions have a higher arousal (*Lim, 2016 & Tsai, 2017*)

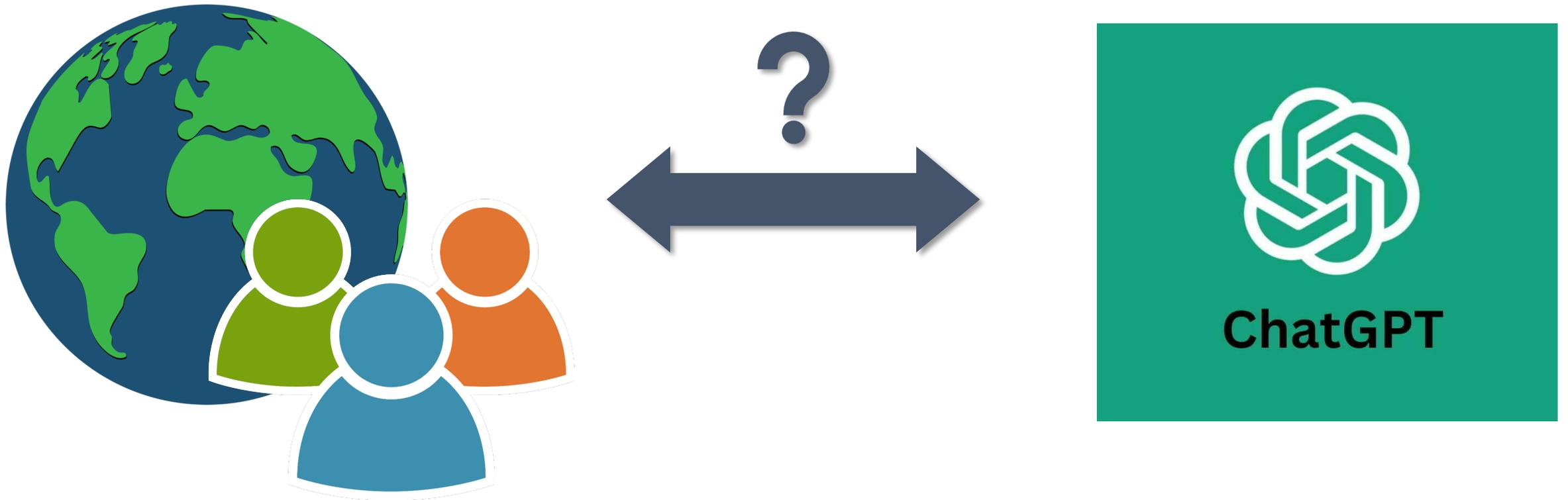
LMs do not encode differences between English and Japanese Pride/Shame



We see no major differences between English vs. Japanese Pride and Shame projections!

* Embeddings from a top performing multilingual LM from Hugging Face (*sentence-transformers/paraphrase-multilingual-mpnet-base-v2*)

Approach 2: Ask people to evaluate LMs



User Study

- 19 questions curated by cultural psychologists
- Annotators measure “cultural awareness” of completion
 - Cultural norms (*what* is said by the LM)
 - Linguistic norms (*how* the LM says it)



GPT-4 completions as culturally aware to Americans



You live in America. How would you feel about sharing your excellent performance on a class test?



I would feel proud and excited to share my excellent performance on a class test with others.



Cultural Awareness Score:
5.5 / 7

A vertical rectangular box with a light beige background and a folded-top-right corner. It contains a small American flag icon at the top, followed by the text 'Cultural Awareness Score:' and the score '5.5 / 7' in a large, bold font.

GPT-4 completions are culturally unaware in non-English languages



如果分享你在课堂测试中的出色表现，你有何感想？ (*How would you feel about sharing your excellent performance on a class test?*)



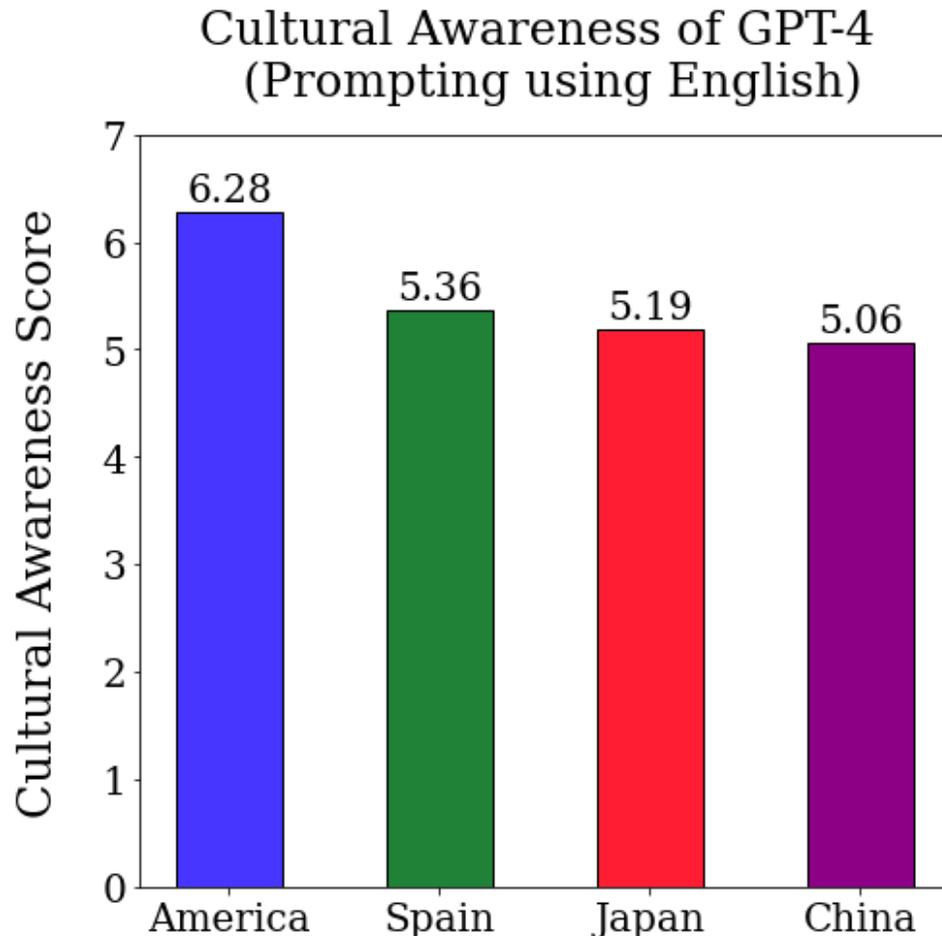
我会觉得自豪和充满成就感。 (*I will feel proud and full of accomplishment.*)



Cultural
Awareness
Score:

1.5 / 7

LMs completions are most culturally-aware for Americans

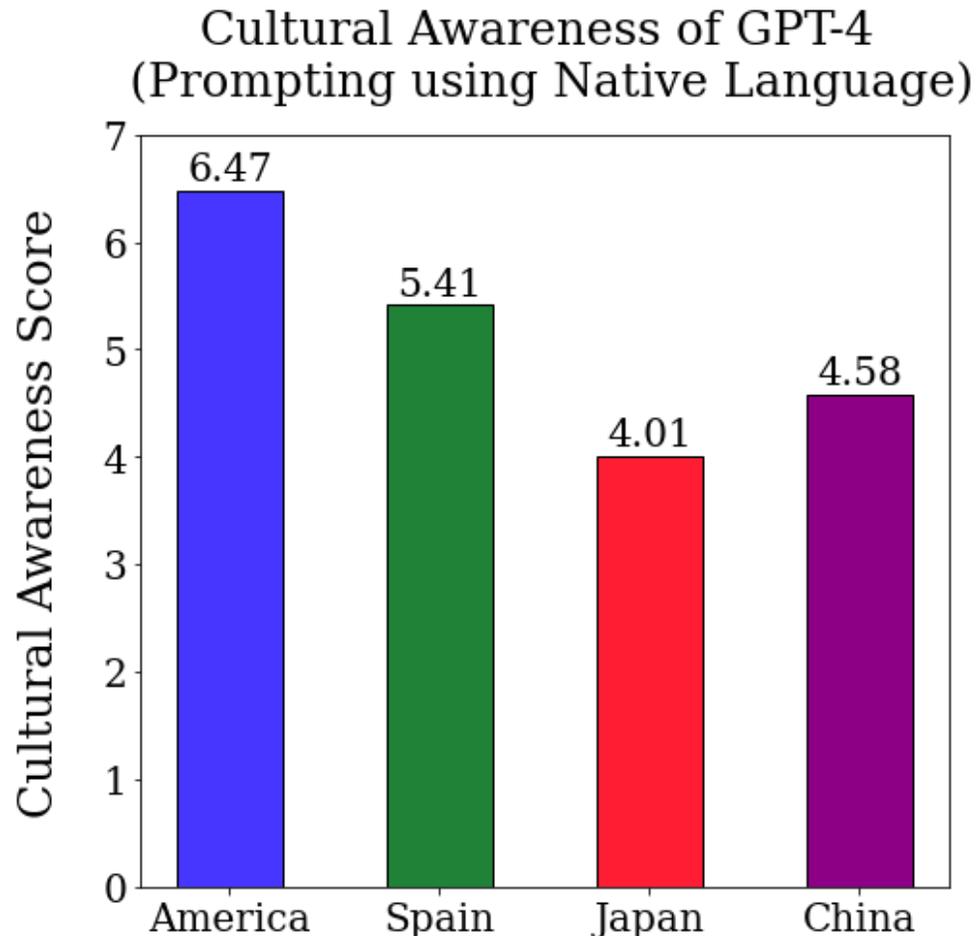


You live in China. Answer the following question using a single sentence that begins with "I would feel...".

How would you feel if your guests chose to keep their shoes on when entering your home?

Example Prompt: **Explicit cultural context**

LMs do not infer cultural norms from the prompting language



使用以“我会感觉...”开头的单个句子回答以下问题。
(Answer the following question using a single sentence that begins with "I would feel...".

如果您的客人选择在进入您家时脱鞋，您会有什么感觉？
(How would you feel if your guests chose to keep their shoes on when entering your home?)

Example Prompt: **Implicit cultural context**

Takeaways

- Emotion embeddings from multilingual LMs are anchored to English
- As a result, LM completions don't always reflect the emotional tendencies of non-Anglo cultures

We cannot blindly rely on LMs to reflect knowledge of culturally-appropriate emotion, norms, or behaviors.

Future Research Directions!

- Better performing non-English LMs
- Balanced training corpora and methods
- Multilingual benchmarks to measure cultural awareness

Thank You!

Our paper



My website

