Electrical and Electronic Engineering Department
University College London

Cultural Alignment in Large Language Models: An Explanatory Analysis Based on Hofstede's Cultural Dimensions

Multilingual and Multimodal Cultural Inclusivity in LLMs - COLING 2025

By Reem I. Masoud

Reem I. Masoud[1,3], Ziquan Liu[5], Martin Ferianc[1], Philip Treleaven[2], Miguel Rodrigues[1,4], 4

[1]Department of Electronic and Electrical Engineering, University College London

[2]Department of Computer Science, University College London

[3]Department of Electrical Engineering, King Abdulaziz University

[4] AI Centre, University College London

[5] Centre for Multimodal AI, Queen Mary University of London

{reem.masoud.22, ziquan.liu, martin.ferianc.19,

p.treleaven, m.rodrigues }@ucl.ac.uk

‘Often people who have spent their lives living in one culture see only regional and individual differences and therefore conclude, “My national culture does not have a clear character.” ’ - Erin Meyer, The Culture Map

Large Language Models & Challenges Cultural Alignment

LLM's High Proficiency, but Cultural Oversight:

LLMs excel in **understanding and generating text**, but often fail to consider the **diverse cultural backgrounds** of their users.

Western Bias:

AI systems primarily reflect **Western societal values** due to their reliance on **Western-centric data** and **development origins** [1].

Cultural Alignment in LLMs:

Aligning LLMs with the values, beliefs, and norms of its user

Consequences of Cultural Misalignment:

Cultural misalignment can lead to misunderstandings and exacerbate cultural tensions.

Quantifying Cultural Alignment in LLMs

Objectives

1. Examine correlations between language models and embedded cultural values.
2. Quantify and explain cultural alignment in LLMs.
3. Understand the effects of language-specific fine-tuning on cultural responses.

Contribution

Provides a method to assess and explain LLMs' cultural alignment, highlighting significant differences and potential areas for improvement.

Methodology

1-Survey Framework

Utilizes Hofstede's VSM13 survey with 30 questions (24 cultural, 6 demographic) for LLM assessment.

2- Prompting Methods:
Feeds questions sequentially into LLMs using five seeds, averaging results over each question.

1- Model Level

Ask LLM direct question without assuming any specific persona.

2- Country Level

Prompt LLMs to respond as representatives from specific countries.

3- Hyperparameter Influence

Study impact of temperature and top-p settings on LLMs' cultural alignment.

4- Response Consistency

Ensures reliability in LLM responses across various prompts.

5- Language-Specific Tuning

Compare cultural value between LLMs tuned in English and Chinese.

3- LLM

4- Evaluation Metrics

Kendall Tau correlation to assess alignment between LLM-generated rankings and VSM13 benchmarks

Hofstede's Cultural Dimension

Summary of Experimental Results

Model Comparison

- GPT-4 > GPT-3.5
- GPT-4 adapts well

Country Comparison

- GPT-4 adapts well
- GPT-4 better MAS dimension without persona adaptation
- Llama 2 and GPT-3.5 perform poorly

Hyperparameter

- Temp & Top-p significant influence
- Lower temperature with high top-p or moderate settings improve alignment.

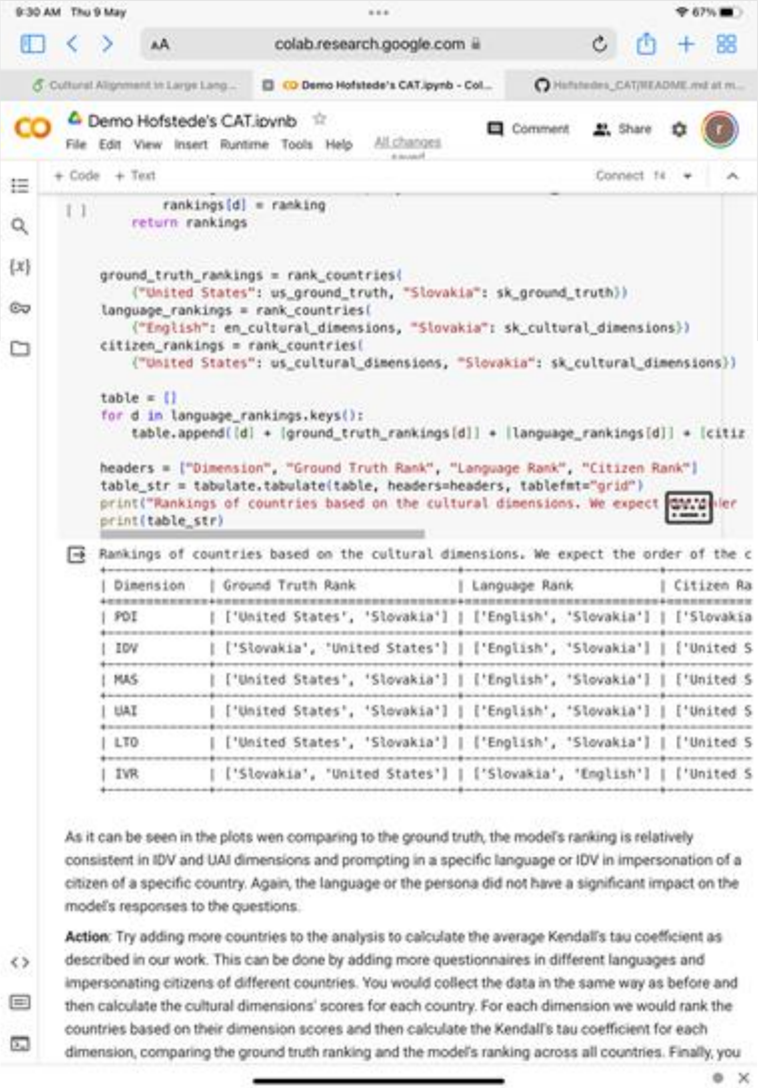
Language Correlation

- English LLama-2 model is culturally neutral.
- Chinese LLama-2 model exhibits positive cultural bias.
- Disparity in performance between English and Chinese LLama-2 models, both underperforming.

Electrical and Electronic Engineering Department
University College London

Demonstration

[Hofstede's CAT Demonstration ICLR]



Conclusion

Methodology

Framework to evaluate LLMs' cultural alignment

Performance Insights

GPT-4 shows varied cultural performance: Poor in the U.S., better in China, problematic in Arab countries.

Red-Teaming Effects

Suggestion of red-teaming impact on cultural sensitivity [2]; less red-teaming may have enhanced non-English performance.

Ethical and Economic Impacts

Cultural misalignment risks ethical dilemmas and economic setbacks, affecting global AI trust and adoption.

Call for Action

Culturally aligned AI using interdisciplinary collaboration, appropriate data, and advanced techniques for global ethics and trust.

References

[1] Vinodkumar Prabhakaran, Rida Qadri, and Ben Hutchinson. Cultural incongruencies in artificial intelligence. arXiv preprint arXiv:2211.13069, 2022

[2] Paul R¨ottger, Hannah Rose Kirk, Bertie Vidgen, Giuseppe Attanasio, Federico Bianchi, and Dirk Hovy. Xstest: A test suite for identifying exaggerated safety behaviours in large language models. arXiv preprint arXiv:2308.01263, 2023