

Title: “Which Humans? Replicating and Extending Cultural Psychology Analyses of LLMs to DeepSeek and Grok”

Introduction and Background

The research paper (*Which Humans?*) demonstrated that ChatGPT exhibits a distinct WEIRD (Western, Educated, Industrialized, Rich, Democratic) psychological bias, aligning most closely with U.S. cultural profiles and diverging from non-WEIRD populations. These findings suggest a “WEIRD in, WEIRD out” effect arising from the disproportionately WEIRD origin of internet-based training data.

This research proposal seeks to replicate and extend those findings by applying the same methodology to DeepSeek and Grok. The goal is to determine whether WEIRD bias is universal across LLM architectures or contingent upon training data composition.

Objective: Quantify and compare the WEIRD bias in DeepSeek and Grok relative to ChatGPT.

Research Questions

1. Does DeepSeek exhibit the same WEIRD psychological profile as ChatGPT?
2. Does Grok’s unique training source (Twitter/X) produce a distinct cultural or cognitive bias?
3. Is WEIRDness a systemic property of LLMs or a variable feature shaped by training data diversity?

Methodology: Each model (DeepSeek, Grok) will be administered the same instruments as in *Which Humans?*:

1. **World Values Survey (WVS):** 262 attitudinal variables, 1,000 samples per question, to map cultural alignment with 65 human nations.
2. **Triad Task:** 1,100 text-based analytic vs. holistic reasoning trials to assess cognitive style.
3. **Self-Concept Task (“I am...”):** 1,100 samples to evaluate self-construal (independent vs. interdependent).

Results will be analyzed via correlation and clustering to determine alignment with WEIRDness indices and to compare cross-model variance.

Expected Outcomes and Contribution

- **Replication Outcome:** If DeepSeek and Grok replicate ChatGPT's bias, it will confirm that WEIRD skew is systemic across LLM families.
- **Differentiation Outcome:** Divergence in Grok's profile would indicate that training data composition substantially shapes an LLM's emergent psychology.

Conclusion: This research offers the first cross-architectural, cross-data-source assessment of cultural bias in LLMs, advancing understanding of how training inputs shape models' cognitive and cultural behavior. It also underscores the ethical necessity of diversifying training corpora to represent global psychological diversity.

Appendix

The link of the paper "Which human?":

https://projects.iq.harvard.edu/sites/projects.iq.harvard.edu/files/culture_cognition_coevol_lab/files/which_humans_09222023.pdf

The link of the data set that is used in the paper "which human?":

<https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp>