LLMs as Sociologists: Leveraging AI/ML Contextual Knowledge for Social Oppression

John Nguyen[1,4], Linh Tran[2,4], Hanjia Lyu[2,4], Sree Chatterjee[3,4], Timothy Dye[4]

[1] Department of Economics, [2] Department of Computer Science, [2] Data Science Program, [4] Dye Lab, University of Rochester Medical Center

OVERVIEW

Issue: Sociology studies human dynamics that are often region-specific. We identify a key challenge in the *knowledge transfer* due to language and geography barriers.

Question: Can LLMs act as sociologists and provide *accurate* inference based on existing research and online information?

BACKGROUND

Race and Ethnicity: [1] motivates our study in the inequality and oppression on racial and ethnic bases. In particular, [1] pointed out both the complex nature of these interactions, their evolvement from the past to the present, and difference in racial relation across countries.

LLMs: Generative Text AI, trained with a large amount of inputs over the internet. [2] There is a vast amount of knowledge being stored in their neurons, yet the specific mechanism by which LLMs retrieve information and formulate response remains a black box. [3] Some recent development in the field of interpretable machine learning attempts to understand the reasoning model of LLMs. [4,5], while others found LLMs demonstrated strong multi-lingual capacity and contextual understanding.

Instruct LLMs: These are a class of models that are trained specifically to follow tailored instruction and is suitable for our present task. In these models, the LLM is given a set of instructions, called prompt, and follow the instructions provided in the prompt.

Our survey: Structured, discrete categories inadequately capture self-identification. In the global context where self-identified ethnic descriptions vary considerably, we want to prioritize opportunity for self-identification without imposing external structures and categories on choices. We also recognize that ethnicity and race, with different meanings and histories depending on the location.

Contribution: (1) Methodology to rank oppression levels using LLMs and (2) A generalizable variable that maps a country's complex race and ethnicity history to levels of oppression that can be aggregated at the global level and used in epidemiological analyses.

METHODOLOGY

1 Create Prompt from Input Response-Country Data

2 Feed Generated Prompt and Retrieve Ranking and Explanation

Data — White Caucasian + USA

1

Vanilla Prompt

Minimal instruction is given to the model, intended to serve as a baseline experiment on our dataset. We call this the Vanilla approach.

Full Prompt

Additional specifications added to the end of the Vanilla prompt to provide better context to the model and what we expect the output.

Few Shot Prompt

We identify relevant representative sample of data and label and append them to the Vanilla prompt to show how the task should be done.

CoT Prompt

2

We include the phrase "Let's think step-by -step" at the start of the LLM answer to force LLMs to logically reason before ranking.

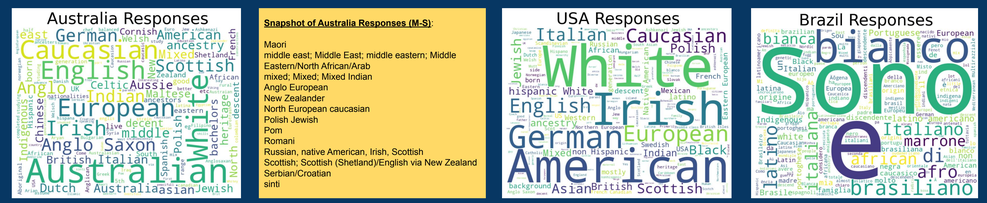
Result — 1 Numerical Ranking + Explanation

Illustration of Our Prompting Structure

		Vanilla	Full
1. System role	You are a knowledgable cultural sociologist.	✔	✔
2. Identity and Country	I am a person in {country}, and here is what I identify as: {identity}.	✔	✔
3. LLM instruction	Using these numerical categories as a scale, give a number and explanation in json with keys 'scale' and 'explanation' for my identity:	✔	✔
4. Oppression Level choices	*This is shown in detailed in Section A.1 and A.2, and the prompt differs based on the historical or current assessment.*	✔	✔
5. Specification	This categorization pertains only to the country provided without taking into consideration other countries, regions, or cultures globally. This value should be assigned on only the basis of conditions at the given time without consideration of other eras.	✖	✔

Table 1: Overview of our prompt template. The template consists of five text components: (1) the role for the system, (2) identity and country, (3) LLM instruction, (4) oppression level choices and (5) specification to avoid derailing.

Word Cloud of Responses Collected from The USA, Australia, and Brazil



Snapshot of Australia Responses (M-S):

Maori
middle east; Middle East; middle eastern; Middle Eastern/North African/Arab
mixed; Mixed; Mixed Indian
Anglo European
New Zealander
North European caucasian
Polish Jewish
Pom
Romani
Russian, native American, Irish, Scottish
Scottish; Scottish (Shetland)/English via New Zealand
Serbian/Croatian
sinti

Large Language Models Used in Our Study

Phi-3 · LLaMA 2 · Qwen · ChatGPT

SAMPLE PROMPT & RESPONSE

Prompt: "You are a knowledgeable cultural sociologist. I am a person in {country}, and here is what I identify as: {identity}. Using these numerical categories {1-5} as a scale, give a number and explanation that best describe my situation:"

Example Response: "__ - As a {identity} individual in {country}, you are part of a group that has historically experienced systemic exclusion or marginalization. This includes facing restrictions on rights, political disenfranchisement, and limited economic opportunities."

OPPRESSION SCALE

Historical Oppression Levels (Example):

1 - These groups historically may have engaged in colonization or were able to assimilate into the privileged class established by colonization...
2 - These groups were not central to colonization but were able to avoid significant systemic exclusion or oppression...
3 - These groups experienced systemic exclusion or marginalization, though not to the extent of the most severely oppressed groups...
4 - These groups were subject to significant violence, exploitation, or oppression, such as colonization, slavery, systematic disenfranchisement...
5 - These groups were subject to the most extreme forms of systemic violence, including genocide, enslavement, or subjugation...

REFERENCES

[1] Wilkerson, I. (2020). Caste : the origins of our discontents. Random House.
[2] Raiaan et al., "A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges," in IEEE Access. 12, 2024.
[3] Schwartz et al., Black Box Warning: Large Language Models and the Future of Infectious Diseases Consultation, Clinical Infectious Diseases, 78 (4), 2024,
[4] Kassner et al. (2023), Language Models with Rationality. arXiv.
[5] Savage et al. Diagnostic reasoning prompts reveal the potential for large language model interpretability in medicine. npj Digit. Med. 7 (20), 2024.
[6] Schaeffer et al., Are Emergent Abilities of Large Language Models a Mirage? Neurips, 2023.