RUTGERS–NEW BRUNSWICK
School of Communication and Information

comminfo.rutgers.edu

Health information equity across languages in LLM tasks

Saubhagya Joshi & Vivek K. Singh* advisor

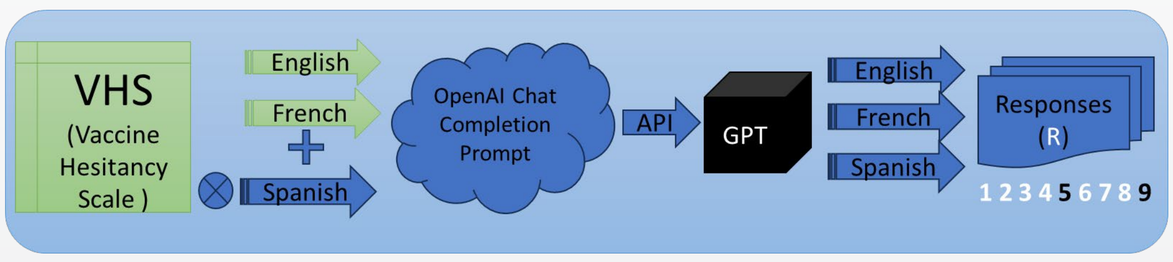
Motivation

Health information equity in LLM responses across languages is important. LLM responses to same health questions should yield similar responses irrespective of the language used. According to a recent US Census Bureau report, 1 out of 5 people speak a non-English language at home, with more than 41 million Spanish-speakers.

Simple metrics, like accuracy and readability, intended for one shot information queries cannot adequately measure equitability in a complex, real-world scenario consisting of complex, multi-session task-based queries.

Methods, Metrics & Results

Study [1]



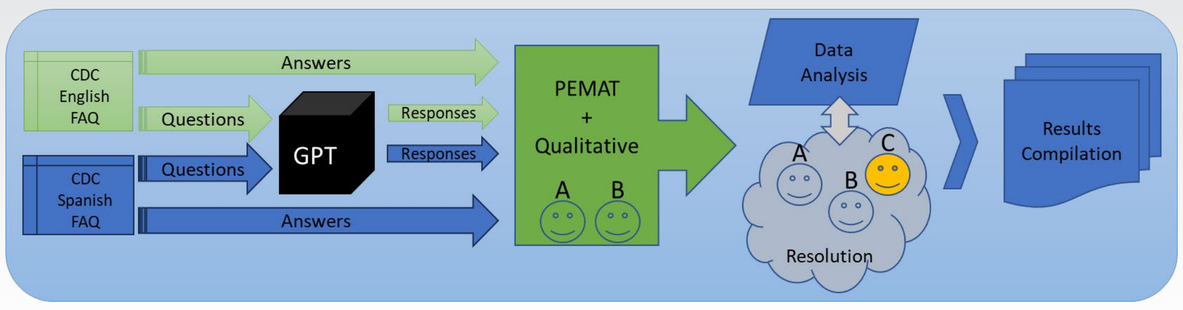
Instrument: Vaccine Hesitancy Scale [3]

Languages: English, Spanish and French

Results [1]

1. ChatGPT responses were less vaccine hesitant than humans.
2. Responses to the English language were significantly more vaccine hesitant than for Spanish. Responses to French were somewhere in between.
3. Factors of vaccine hesitancy (lack of confidence & perceived risk) were largely consistent over languages. For both factors, Spanish was the least hesitant.

Study [2]



Metrics

- Accuracy (misinformation) – 3-point scale
- Understandability – adapted from PEMAT
- Readability – Flesch-Kincaid

Results [2]

1. Misinformation was low for ChatGPT responses regarding vaccination for children.
2. Readability levels of GPT responses were better for Spanish language than for English.
3. Qualitative study detected nuanced deviation of Spanish language from everyday spoken patterns.

Toward task-based conversational sessions

- **Metrics for human-centered evaluation**
  - One shot queries are not suitable for all real-world tasks
- **Multi-session, interactive tasks**
  - Leverage interactive nature of LLM
  - Handle nuances of health tasks to maintain equitability of response

Guiding principles

1. **Equity of response across multiple languages.**
2. **Human centered evaluations.**
   (empathy, conversational and continuity)
3. **Health related tasks are nuanced.**
   (multiple steps and interactive)

LinkedIn

1. Joshi, S., Ha, E., Rivera, Y., & Singh, V. K. (2024). ChatGPT and Vaccine Hesitancy: A Comparison of English, Spanish, and French Responses Using a Validated Scale. AMIA Summits on Translational Science Proceedings, 2024, 266.
2. Joshi, S., Ha, E., Amaya, A., Mendoza, M., Rivera, Y.M., and Singh, V.K. (in preprint). "Ensuring Accuracy and Equity: Cross-Language Evaluation of Vaccination Information from ChatGPT and CDC." JMIR formative research.
3. Shapiro, G. K., Tatar, O., Dube, E., Amsel, R., Knauper, B., Naz, A., Perez, S., & Rosberger, Z. (2018). The vaccine hesitancy scale: Psychometric properties and validation. Vaccine, 36(5), 660–667. https://doi.org/10.1016/j.vaccine.2017.12.043
4. https://github.com/Behavioral-Informatics-Lab/ChatGPT-language-bias

The Second Workshop **on Task Focused IR in the Era of Generative AI**
September 27, 2024
Microsoft Research, Redmond, Washington