

# Using AI for Cognitive Health Measures

**in a Multilingual  
Survey Context**

Midwest Association for Public Opinion Research Annual Conference, 2024

**Sunghee Lee, Stephanie Morales, Joanne Tian**

University of Michigan

# Acknowledgements

- This study is supported by the National Institute on Aging [grant number: R01 AG082080, PI: Lee]
- Our multilingual team
  - Co-Investigators: Mengyao Hu, Thuy Nguyen, Elizabeth Vazquez
  - Research Assistants: Felix Baez-Santiago, Valeria Castañeda, Leng Seong Che, Jay Kim, Phuong Nghi Nguyen, Yao Sun, Gloria Zhou

# Alzheimer's Disease and Related Dementias (ADRD)

- Population aging around the globe
  - Increase in ADRD
- ADRD imposes a significant burden for individuals, families and societies
  - Projected to cost the US >\$1.1 trillion by 2050
- Large window of time between preclinical stage and the onset of ADRD provides a chance to address modifiable risks
- Early ADRD detection through cognition measures

# Cognition and Cognitive Measures

- Cognition is a complex latent construct with various domains and subdomains within
  - Memory: Long-term memory; Short-term memory; Working memory; Sensory memory; Episodic memory; Semantic memory
  - Attention: Divided attention; Selective attention; Sustained attention
  - Orientation
  - Executive function
  - Processing speed
  - ⋮

# Cognitive Measures in Multi-lingual Web Survey Contexts

- Increasing racial, ethnic, and linguistic diversity among the aging population
- Translation of cognitive measures becomes necessary for inclusive data collection; but is resource intensive and difficult
- E.g., A measure for processing speed:

Next, we would like you to type the sentence below as quickly as you can. When you are ready, please begin typing in the box, then select “Next” when you are finished.

“The quick brown fox jumps over the lazy dog.”

→ Can AI help?

# This study: Background

- A qualitative preparation stage for a population-based national Web survey of  $\geq 40$  y.o. adults with an oversample of Latino, Chinese, Korean and Vietnamese Americans
- Instruments in English, Spanish, Chinese, Korean, and Vietnamese
- Focus measures: Cognitive measures in the Health and Retirement Study applicable to the Web survey setting

# This study: Use of AI

- RAs who do not speak the target language collected GPT translation data
- ChatGPT 3.5 and UMGPT 4 with 4 different prompts
  - Prompt 1: Hello, [ChatGPT/UMGPT] can you please translate the following into [LANGUAGE]?
  - Prompt 2: Prompt 1 +  
We will ask this in a Web survey.
  - Prompt 3: Prompt 2 +  
to measure cognitive health.
  - Prompt 4: Prompt 3 +  
in the domain of [COGNITION DOMAIN].

# This study: Evaluation

- RAs native or fluent in the target language provided critique
- Investigators whose mother tongue is the target language provided critique
- Consolidated the critiques



# Results: Overall Impression

- Question length: GPT struggle with translating long questions
- **Language:** Different issues emerge across languages
- **ChatGPT vs. UMGPT:** Performance differs by language
- **Prompts:** Performance differs by language
- **Adaptation:** Not able to make adaptations to the pangram

# Results: By Language

- **Spanish**
  - Mostly on target
  - Verb translation issues
  - Inconsistent conversational connotation across items
- **Chinese**
  - Overall, reasonable translation quality
  - Poor quality of Likert response scale translation
- **Korean**
  - Noticeable struggle with long questions (better to break down into separate sentences in Korean)
  - Some prompts altered the source Likert response scale in the translation
- **Vietnamese**
  - More struggles than other languages

# Results: ChatGPT vs. UMGPT

- UMGPT better in Spanish, Chinese, and Korean
- ChatGPT better in Vietnamese
  
- The core technology is the same between the two
- Training data differs as intended audience differs
  - ChatGPT: General use; Trained on a mixture of licensed data, data created by human trainers, and publicly available data; Free version updated to September 2021 data
  - UMGPT: For University of Michigan students, faculty, and staff; Training data up to October 2023

# Results: By Prompts

- Spanish
  - No difference
- Chinese
  - Prompt 2 better
- Korean
  - Prompt 1 resulted in conversational translations, not suitable for interview settings
  - Prompt 2-4 perform similarly
- Vietnamese
  - Prompt 3 better

# Implications

- Despite imperfectness, investigators unanimously agreed on the benefit of having AI-generated translation version as a reference material in their own question translation
- Adaptation is another matter...

# Where it is headed to

- More AI experiments
  - Different platforms
  - Different prompts
  - Different topics
- Test through cognitive interviews in all languages
- Implement in a web survey

**Thank you!**      ¡Gracias!  
谢谢!      감사합니다!      Cảm ơn!

Questions?

[sungheel@umich.edu](mailto:sungheel@umich.edu)