Examining Patient-Large Language Model Interactions Using the PromptWise Paradigm for Medical Education

Satvik Tripathi, University of Pennsylvania; Rithvik Sukumaran; Suhani Dheer; Tessa S. Cook, MD, PhD, CIIP, FSIIM

Introduction/Background

With the rise of large language models (LLMs) for general-purpose use, researchers have begun studying how they might improve patient care. In earlier work, we proposed the PromptWISE (Prompt engineering for Well-structured, Interactive, and Supportive Education) paradigm to educate patients on using LLMs to understand their medical issues. PromptWISE helps patients engineer higher-quality prompts that can enhance their medical experience and reduce the burden on medical professionals.

Methods/Intervention

We applied the six-point PromptWISE guidelines to answer a set of 25 questions patients might ask LLMs about their health or medical care. Using Amazon Mechanical Turk, we conducted an IRB-approved survey (n=1074) to compare a pair of LLM-generated responses, one from a simple prompt and the other from a PromptWISE-designed prompt. GPT-4 provided all text generations. Volunteers picked the better response based on three criteria: clarity, information, and relevance. We also collected demographic information, including gender, race, age bracket, income bracket, education level, and healthcare employment status. Statistical analyses were performed to determine the generalizability and reproducibility of our results.

Results/Outcome

The demographic reporting indicated a diverse cohort of volunteers, reducing any reporting biases. In our analysis, volunteers overwhelmingly (n=837) chose responses generated from PromptWISE prompts over those generated from non-PromptWISE prompts ($p < 0.0001$). We also found that non-PromptWISE prompts lacked essential details, leading to inaccurate or irrelevant responses.

Conclusion

The results demonstrate the tangible impact of prompt engineering on patient-LLM interactions. Querying LLMs with prompts crafted following our guidelines yielded more comprehensive and precise responses while also refraining from giving any medical advice. Finally, volunteers demonstrated a strong preference for responses to PromptWISE prompts, further indicating the impact and need for prompt engineering when interacting with LLMs.

Statement of Impact

PromptWISE significantly improves patient-LLM interactions by generating clearer, more informative responses. This study underscores the importance of prompt optimization in enhancing patient engagement and accuracy when utilizing LLMs for medical information retrieval.
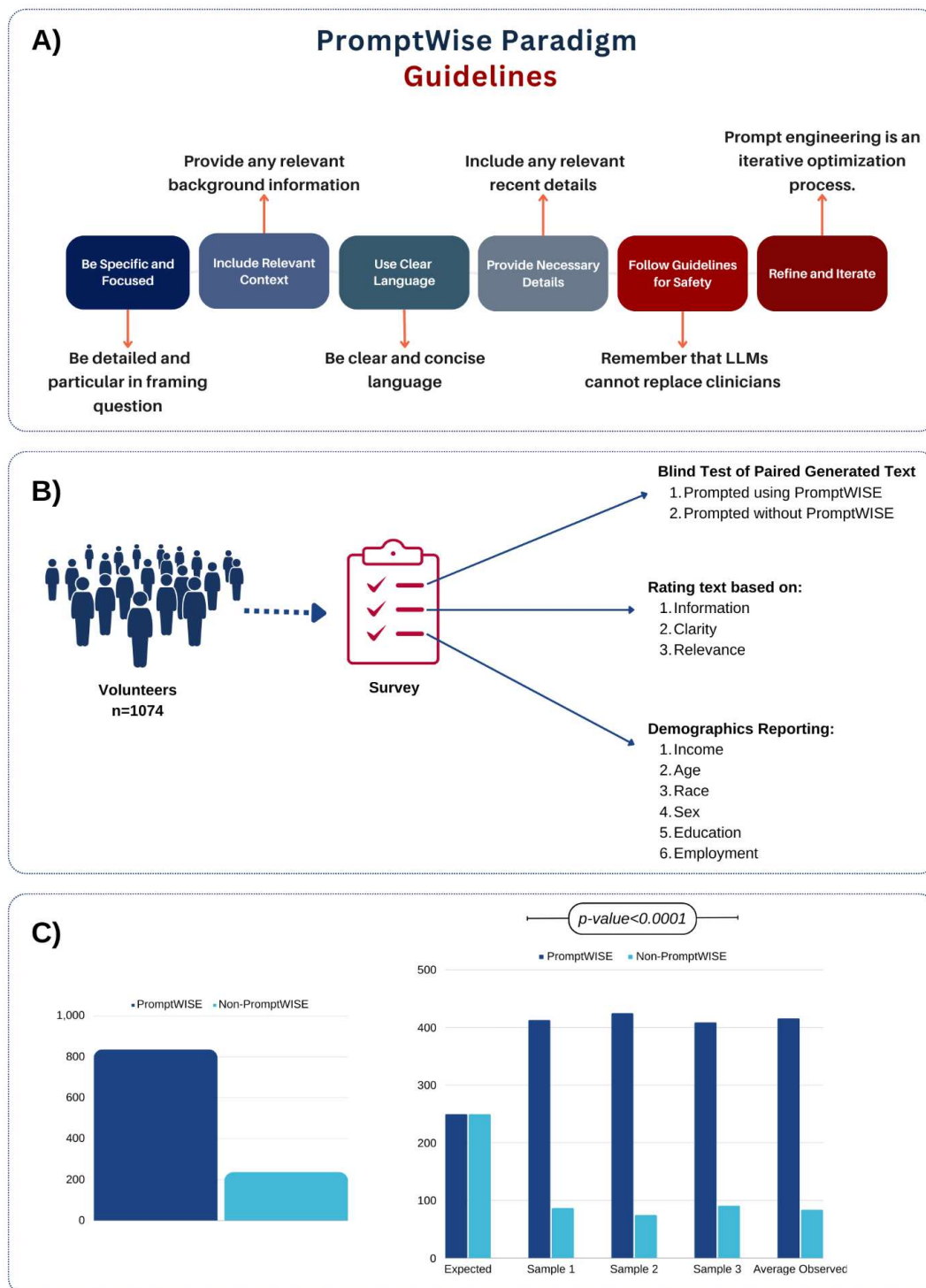
Figure 1. A) Graphical representation of PromptWise paradigm for patients to write better and more explainable prompts. B) Overflow of the blind survey study conducted to assess the effectiveness of PromptWISE prompts. C) The left bar graph shows the overall distribution of patient responses in the survey, and the right bar graph represents the results of the Chi-Square test.
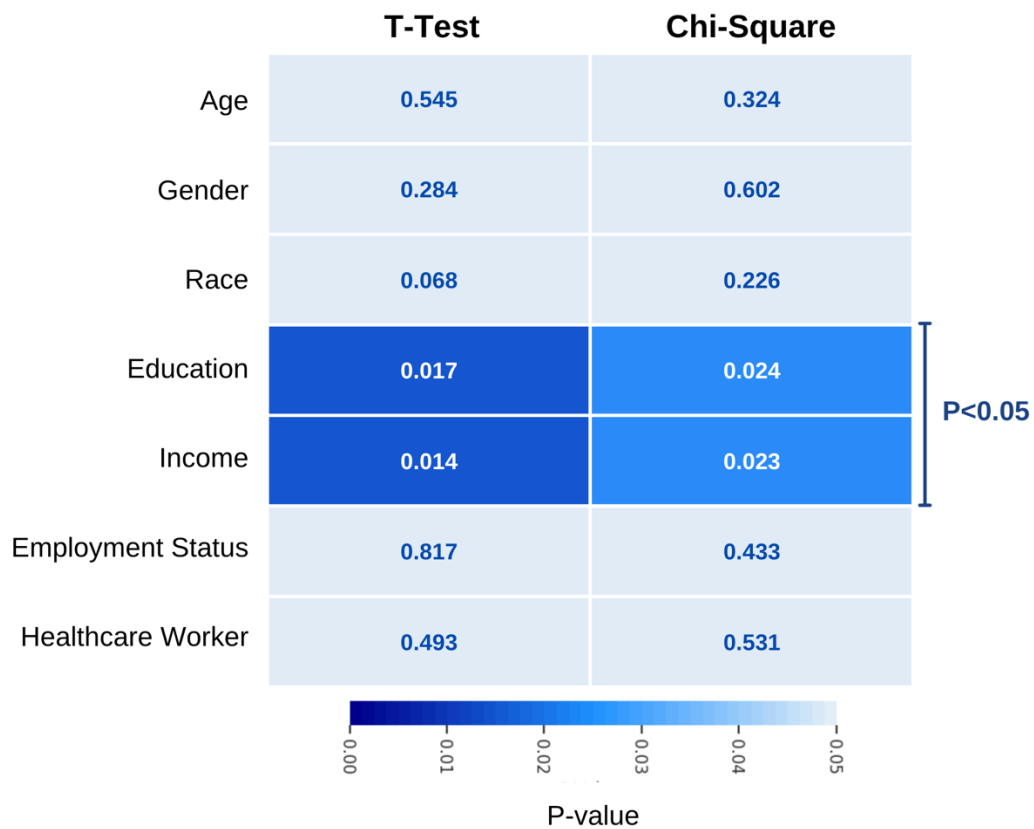
Fig.2. Statistical analysis for the significance of demographic variables