AI Models for Dysphonia Patient Education

RUTGERS HEALTH | New Jersey Medical School

Rebecca A. Ho, BA[1], Esha Shah, BA[1], John Sebastian De Armas, MD[1], Kenneth Yan, MD, PhD[1], Rachel Kaye, MD FACS[1]

[1]Rutgers New Jersey Medical School (Newark, NJ)

INTRODUCTION

- Chat Generative Pre-Trained Transformer (ChatGPT) and Google AI Overview (GAO) have been increasingly utilized in the medical field.
- AI model responses have not been compared with gold standard laryngology guidelines.
- This study aims to evaluate and compare ChatGPT, GAO, and Dysphonia International (DI) responses to patient queries related to dysphonia.

METHODS

- Two board-certified otolaryngologists (KY, RK) and an otolaryngology resident (SD) each blindly rated ChatGPT, GAO, and DI's responses to questions regarding spasmodic dysphonia, muscle tension dysphonia, and vocal tremor.
- Questions were derived from patient information webpages and subsections for each pathology on Dysphonia International.
- Responses were graded on a 5-point Likert scale for comprehensiveness, accuracy, appropriateness to patients, and patient safety.
- Readability was assessed with Flesch-Kincaid Grade Level (FKGL), Flesch Reading Ease Score (FRES), Gunning Fog Index (GFI), and average word count.
- One-way ANOVA and Tukey HSD were used for statistical analysis; significance was set at $p<0.05$.
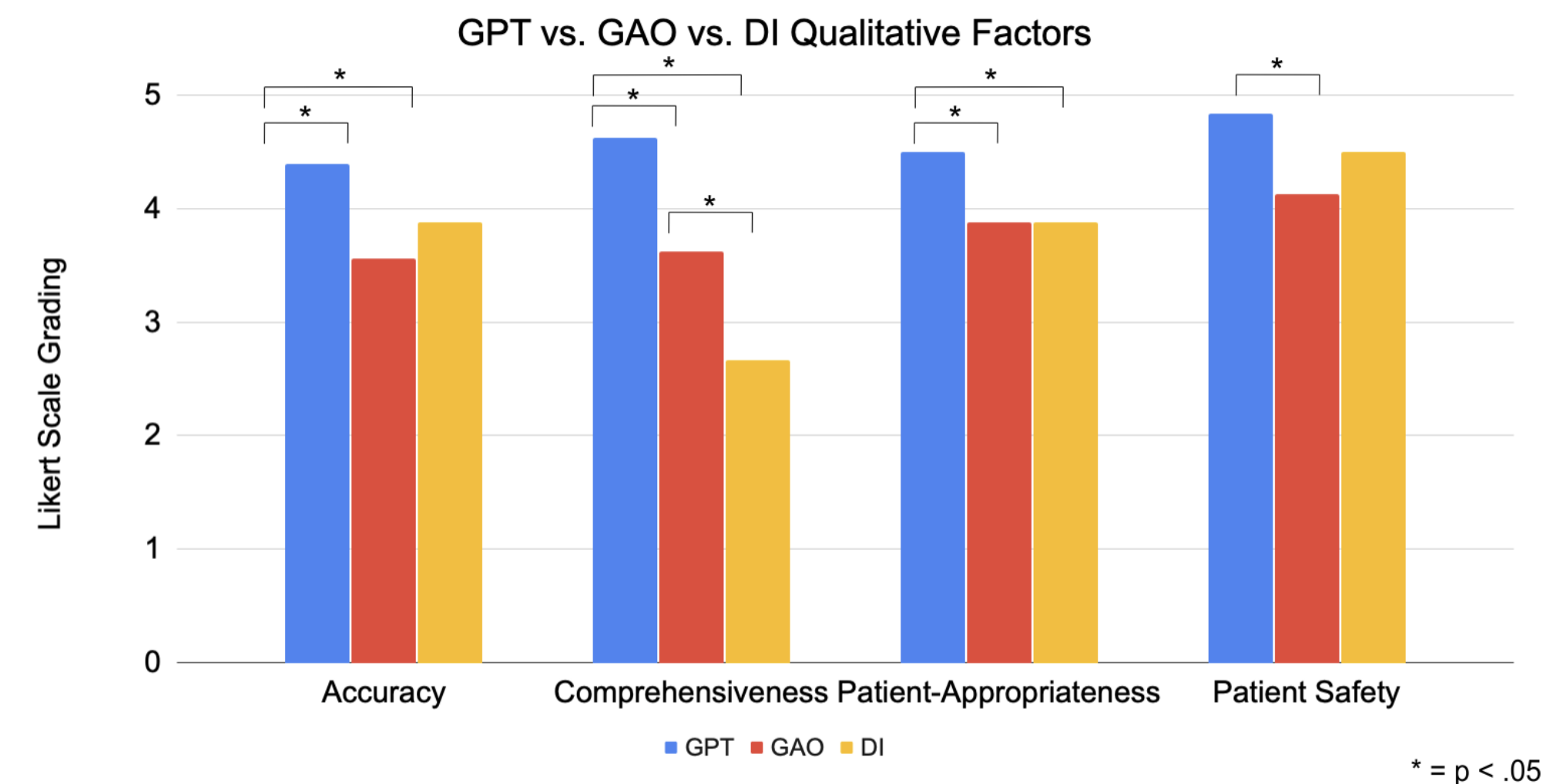
RESULTS

Figure 1: Average physician assessment of ChatGPT vs. GAO vs. DI responses.
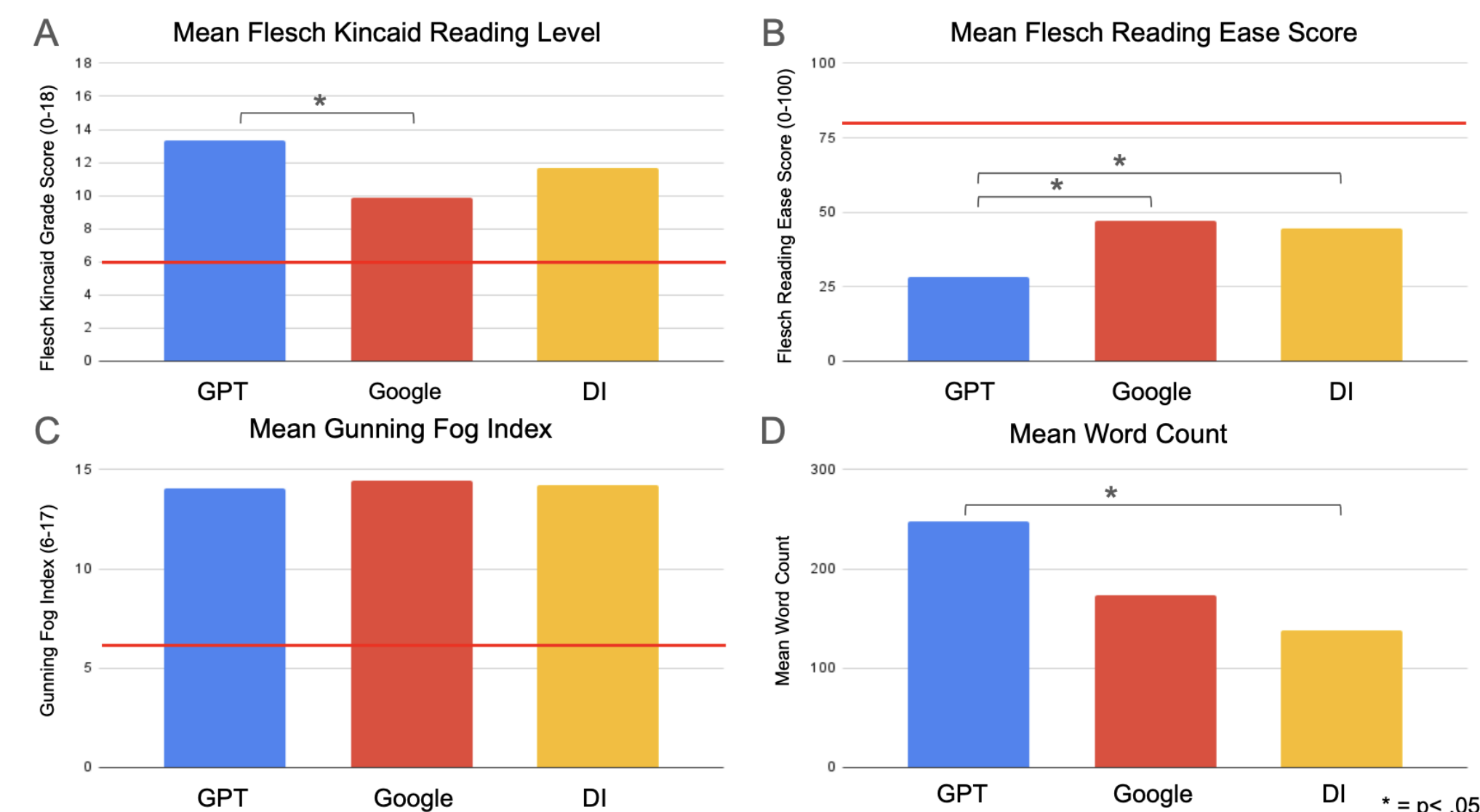

Figure 2: Readability scores by model using FKGL, FRE, GFI, and word count. Red line indicates AMA readability recommendations. Scales: FKGL (0–18), FRES (0–100), GFI (6–17).

RESULTS (contd.)

- ChatGPT performed significantly better than GAO and DI with regards to comprehensiveness, accuracy, appropriateness, and was an overall safer AI-resource to use than GAO.
- Interrater reliability was 0.8.
- All resources were above recommended reading grade level.

DISCUSSION

- AI models prove to be a formidable alternative resource for clinical information and can be responsibly integrated into patient education.
- AI models may require a higher-than-recommended health literacy for full use.
- When AI is properly supervised by board-certified professionals, it can improve medical knowledge and bridge communication gaps between patient and provider.

REFERENCES

[1] Carnino, J. M., Pellegrini, W. R., Willis, M., Cohen, M. B., Paz-Lansberg, M., Davis, E. M., Grillone, G. A. & Levi, J. R. (2024). Assessing ChatGPT's responses to otolaryngology patient questions. Annals of Otology, Rhinology & Laryngology, 133(7), 658-664.

[2] Hoch, C. C., Wollenberg, B., Lüers, J. C., Knoedler, S., Knoedler, L., Frank, K., ... & Alfertshofer, M. (2023). ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. European Archives of Oto-Rhino-Laryngology, 280(9), 4271-4278.

[3] Maksimoski, M., Noble, A. R., & Smith, D. F. (2024). Does Chat GPT Answer Otolaryngology Questions Accurately?. The Laryngoscope, 134(9), 4011-4015.