Shivam Chandra, BS[1]; Vineet Kumar, BS[2]; Lily Novack; Aditya Achyut; Ryan Hubbard, DO[1]
[1]A.T. Still School of Osteopathic Medicine Arizona
[2]Michigan State University, East Lansing ,Michigan

Background and Objective

- Patients increasingly use large language models (LLMs) to obtain health information outside clinical encounters.
- Prior studies show AI-generated health information often exceeds recommended patient reading levels.
- Patient education must be readable, understandable, and actionable to support safe care seeking.
- The impact of prompting LLMs to produce lower reading-level ophthalmic education remains unclear
- This study evaluated the readability and patient education quality of ChatGPT, Perplexity, and Google Gemini ophthalmic responses and tested whether simplified sixth- to eighth-grade prompting improves PEMAT actionability across models.

Methods and Timeline of Care

- Cross-sectional study using identical patient-oriented questions on diabetic retinopathy, macular degeneration, and retinal detachment entered into ChatGPT, Perplexity, and Google Gemini (free versions, logged out).
- Twenty-five responses per platform were collected and analyzed.
- Readability was assessed using Flesch Reading Ease and Flesch–Kincaid Grade Level, and patient education quality was evaluated using PEMAT understandability and actionability scores.
- Mean scores were compared using one-way ANOVA with Tukey post hoc testing, with statistical significance set at $p < 0.05$.

AI can simplify ophthalmic education, but it still falls short in guiding patients on what to do next.

Effect of Prompting on Readability

PEMAT Scores Across Models and Datasets
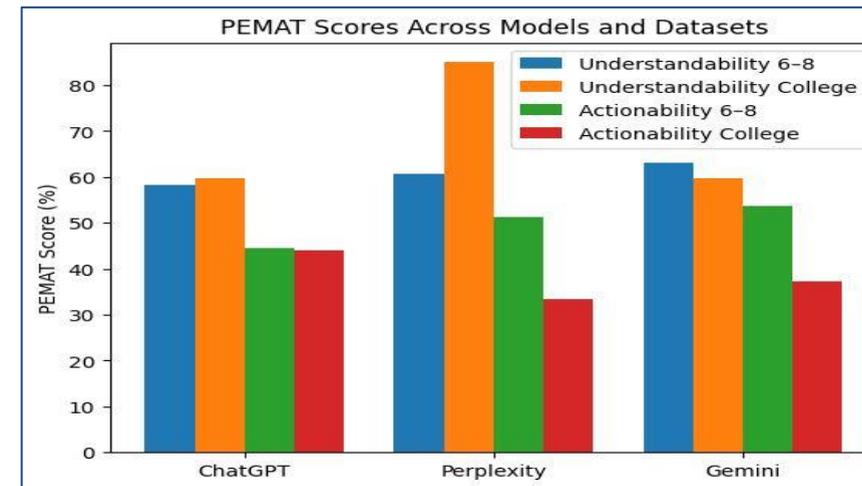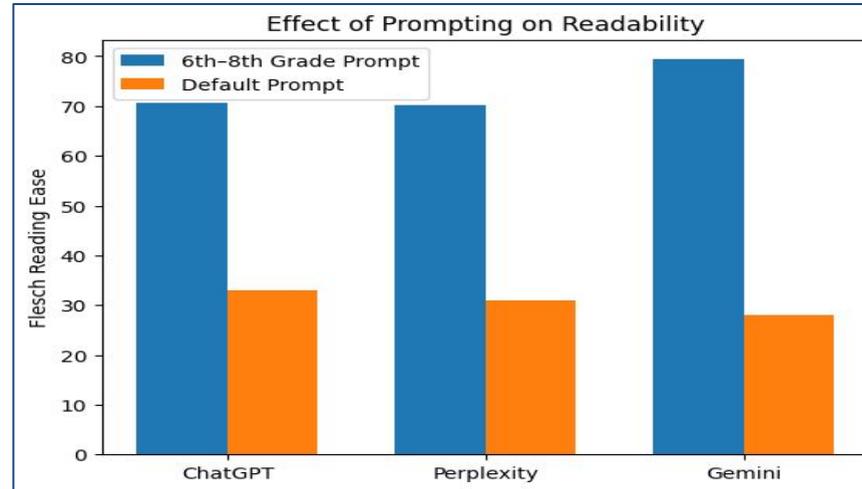
Results

- All models produced content above the recommended reading level, even after simplification.
- Lowering the reading level did not meaningfully improve understandability.
- Actionable guidance remained poor across platforms.
- Limited next-step guidance is concerning for urgent conditions such as glaucoma and retinal detachment.

Key Takeaways

- Prompting significantly improves readability
- Without prompting, outputs revert to college-level complexity
- Understandability varies across platforms
- Actionability remains limited across models
- Improved readability does not reliably improve actionable guidance

References

1. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J. 2019;6:94–98. doi: 10.7861/futurehosp.6-2-94. [DOI] [PMC free article] [PubMed] [Google Scholar]
2. Alhejaily AG. Artificial intelligence in healthcare (Review). Biomed Rep. 2024 Nov 12;22(1):11. doi: 10.3892/br.2024.1889. PMID: 39583770; PMCID: PMC11582508.
3. Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2:719–731. doi: 10.1038/s41551-018-0305-z.
4. Rajpurkar P, Chen E, Banerjee O, Topol EJ. AI in health and medicine. Nat Med. 2022;28:31–38. doi: 10.1038/s41591-021-01614-0.