#3952

Accuracy and Readability of ChatGPT-4o Generated Responses to queries on Age-related Macular Degeneration for Patient Education and Counselling

[1]Nabin Paudel, PhD, [1]Ellen Margaret Moran, PhD, [2]Subash Bhatta, MD, [1]Avril Daly, MA

[1]Retina International, Dublin, Ireland

[2]Pacific Eye Institute, Suva, Fiji

Retina International

PURPOSE:

- Poor clinical-patient communication remains a major challenge in Age-related Macular Degeneration (AMD) care.(1,2)
- Large Language Models (LLMs) are emerging as a potential solution for patient education and counselling.
- Before these tools are implemented in practice it is necessary to assess their suitability in terms of the accuracy and readability of their outputs.
- This study aims to assess ChatGPT-4o's (the latest version the popular LLM from OpenAI) performance in addressing AMD-related patient queries.

METHODS:

Compilation

60 FAQs
- Definition
- Diagnosis
- Risk Factors
- Treatment/prognosis

AI response

- Questions submitted to ChatGPT-4o
- ChatGPT instructed to act as a retina specialist with expertise in simpilfying complex medical content

Expert Validation

Independent retina specialist:
- Accuracy of responses
- Accurate, incomplete or inaccurate
- Quality and appropriateness of references

Readability Assessment

Readable
Readability Metrics
- Flesch Kincaid Grade
- Flesch Reading Ease Score
- Gunning Fog Score
- Coleman Liau Index
- Simple Measure of Gobbledygook (SMOG)

RESULTS:

Accuracy

85% accurate,
13.5% incomplete and
1.6% as inaccurate

References

Total references - 65
26% accurate
22% broken or misdirected links
28% information not easily accessible
25% incomplete

Readability

Flesch Kincaid Grade - 11.5±0.2 (11th Grade)
Flesch Reading Ease Score - 41.4±1.3 (Difficult)
Gunning Fog Score - 12.9±0.2 (College freshman)
Coleman Liau Index - 15.3±0.4 (College +2 years)
SMOG index - 13.7±0.2 (8th Grade)

With the exception of the SMOG index these scores indicate that the answers provided would require college/university level education to understand.

American Medical Association and National Institute of Health recommend patient education materials to be between 6-8th grade reading level. (3,4)

CONCLUSIONS:

- ChatGPT-4o provided accurate response to most of the queries posed.
- Although ChatGPT did not hallucinate any references in our study, many references were broken, were not easily accessible and were incomplete.
- ChatGPT generated responses in AMD may be too difficult for patients and public to understand.
- Optimizing prompts and training the model in ophthalmology specific materials may enhance its performance.
- Further enhancements are required before ChatGPT can be confidently used by clinicians, patients and the public.

REFERENCES:

1. Scheffer M, Menting J, Rausch-Koster P, van Nispen R, van Dulmen S. Looking beyond the eyes of the patient: The importance of effective communication in the treatment of age-related macular degeneration. Acta Ophthalmologica. 2025 Mar;103(2):205-14.
2. Scheffer M, Menting J, Roodbeen R, van Dulmen S, van Hecke M, Schlingemann R, van Nispen R, Boeije H. Patients' and health professionals' views on shared decision-making in age-related macular degeneration care: a qualitative study. Ophthalmic and Physiological Optics. 2022 Sep;42(5):1015-22.
3. Weis BD. Health literacy: a manual for clinicians. American Medical Association Foundation and American Medical Association; 2003.
4. National Institute of Health. How to write easy-to-read health materials.

AUTHOR :



Author email: nabin.paudel@retina-international.org, Retina International, Dublin, Ireland