

The reliability of freely accessible, baseline, general-purpose large language model generated patient information for frequently asked questions on liver disease: a preliminary cross-sectional study

Madunil A. Niriella , Pathum Premaratna, Mananjala Senanayake, Senerath Kodisinghe, Uditha Dassanayake, Anuradha Dassanayake, ...show all

Pages 437-442 | Received 13 Jun 2024, Accepted 21 Feb 2025, Accepted author version posted online: 22 Feb 2025, Published online: 27 Feb 2025

 Cite this article  <https://doi.org/10.1080/17474124.2025.2471874>

 Check for updates

Niriella, M. A., Premaratna, P., Senanayake, M., Kodisinghe, S., Dassanayake, U., Dassanayake, A., Ediriweera, D. S., & De Silva, H. J. (2025). The reliability of freely accessible, baseline, general-purpose large language model generated patient information for frequently asked questions on liver disease: a preliminary cross-sectional study. *Expert Review of Gastroenterology & Hepatology*, 19(4), 437–442.
<https://doi.org/10.1080/17474124.2025.2471874>

Background

We assessed the use of large language models (LLMs) like ChatGPT-3.5 and Gemini against human experts as sources of patient information.

Research design and methods

We compared the accuracy, completeness and quality of freely accessible, baseline, general-purpose LLM-generated responses to 20 frequently asked questions (FAQs) on liver disease, with those from two gastroenterologists, using the Kruskal–Wallis test. Three independent gastroenterologists blindly rated each response.

Results

The expert and AI-generated responses displayed high mean scores across all domains, with no statistical difference between the groups for accuracy [$H(2) = 0.421$, $p = 0.811$], completeness [$H(2) = 3.146$, $p = 0.207$], or quality [$H(2) = 3.350$, $p = 0.187$]. We found no statistical difference between rank totals in accuracy [$H(2) = 5.559$, $p = 0.062$], completeness [$H(2) = 0.104$, $p = 0.949$], or quality [$H(2) = 0.420$, $p = 0.810$] between the three raters (R1, R2, R3).

Conclusion

Our findings outline the potential of freely accessible, baseline, general-purpose LLMs in providing reliable answers to FAQs on liver disease.