Article published online: 2024-03-25

Response Generated by Large Language Models Depends on the Structure of the Prompt

Pradosh Kumar Sarangi[1] Himel Mondal[2]

[1] Department of Radiodiagnosis, All India Institute of Medical Sciences, Deoghar, Jharkhand, India
[2] Department of Physiology, All India Institute of Medical Sciences, Deoghar, Jharkhand, India

Address for correspondence Himel Mondal, MBBS, MD, Department of Physiology, All India Institute of Medical Sciences, Deoghar 814152, Jharkhand, India (e-mail: himelmkcg@gmail.com).

Indian J Radiol Imaging

We appreciate the opportunity to respond to the letter written in response to our published article titled "Assessing the capability of ChatGPT, Google Bard, and Microsoft Bing in solving radiology case vignettes."[1] We thank the authors for their interest in our work and thoughtful questions.

The use of large language models (LLMs) like ChatGPT, Google Bard, Microsoft Bing, etc., in radiology is indeed a burgeoning field, and we are pleased that our study has sparked further discussion. We would like to address the raised concerns and provide additional information to enhance the clarity of our methodology.

The authors asked us about the input prompts. We used the Fellowship of the Royal College of Radiologists Part 2A pattern questions directly as prompts. There was no prefix (e.g., role definition) or suffix (e.g., customizing response for reader) attached to the prompt. ►**Fig. 1** present such a question used as a prompt in ChatGPT (GPT3.5; free version).

We appreciate the authors' acknowledgment of the potential influence of prompt engineering on the performance of LLMs. Indeed, the intricacies of prompt design can change the response drastically. We summarized top ten tips in ►**Table 1** about designing prompts for better output

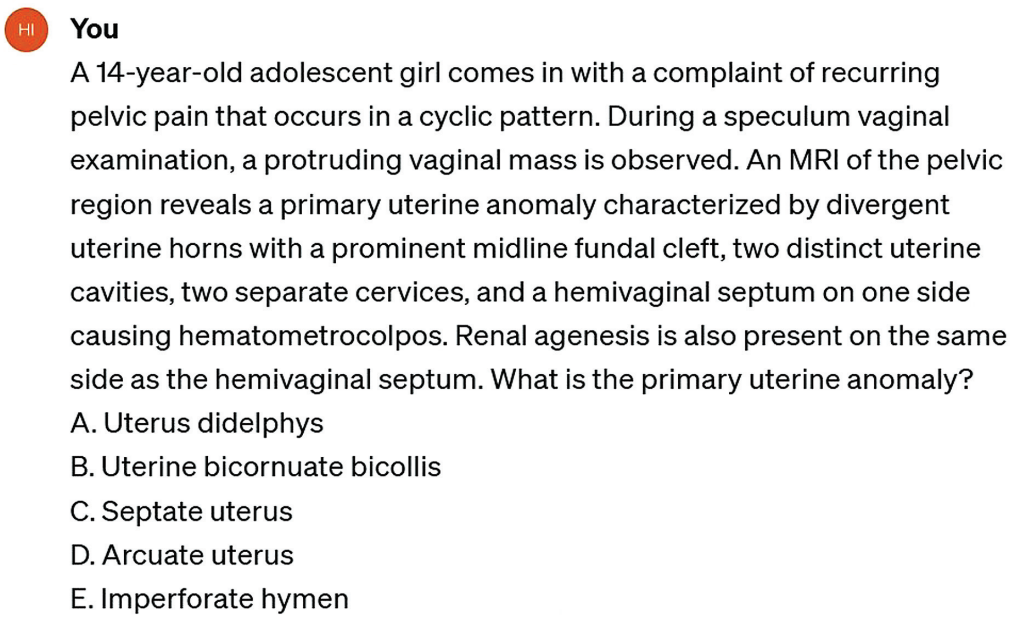

Fig. 1 Example of a prompt we used for getting response from ChatGPT3.5.

DOI https://doi.org/10.1055/s-0044-1782165.
ISSN 0971-3026.

© 2024. Indian Radiological Association. All rights reserved.
This is an open access article published by Thieme under the terms of the Creative Commons Attribution-NonDerivative-NonCommercial-License, permitting copying and reproduction so long as the original work is given appropriate credit. Contents may not be used for commercial purposes, or adapted, remixed, transformed or built upon. (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Thieme Medical and Scientific Publishers Pvt. Ltd., A-12, 2nd Floor, Sector 2, Noida-201301 UP, India

Table 1 Ten tips for using prompts for using large language models (LLMs)

Tip	Brief
1. Clearly define the objective	Clearly state the purpose of your inquiry or the type of information you are seeking
2. Ask to play a role	You can ask it to play a particular role also. For example, you can provide a prompt like "act like an academic writer"
3. Provide context and relevant information	Include background details that can help the model better understand your request. The more details are given to it, more specific would be the response
4. Use specific language	Incorporate precise and clear language to guide the model in generating responses. Although LLMs' comprehension is high, sometimes the models may generate undesired text due to ambiguity in input language
5. Experiment with formatting and structure	Explore different ways to structure your prompt for improved clarity and specificity
6. Provide example	An example of what answer you want helps the model to prepare better response
7. Divide task in smaller segments	Putting a complex task would make the response complex. Hence, large task can be divided to smaller fragments for a better output
8. Ask for references	Medical literature needs reference. Hence, the LLM may ask to write with reference
9. Chain of thoughts	A simple calculation can be mistaken by LLM. Hence, always ask to do a stepwise calculation or generate content step by step
10. Specify text volume	The length of text, sentence structure, and readability can be defined. For example, one can prompt to "generate content for a 6th standard student"

Note: A guide is also available from https://platform.openai.com/docs/guides/prompt-engineering/six-strategies-for-getting-better-results.

from the LLM. There are several websites that help training prompt engineering.[2,3]

The authors raise a valid question about any training we used for the chatbots. We confirm that the chatbots, including ChatGPT, Google Bard, and Microsoft Bing, were not trained by us. Instead, we utilized pretrained models from the respective platforms to ensure a fair evaluation of their out-of-the-box diagnostic capabilities. However, studies have reported that fine-tuned GPT 3.5 and GPT 4 models have shown better responses.[4]

We hope that these clarifications address the concerns raised by the authors. Thank you for providing this platform for academic discourse.

Funding
None.

Conflict of Interest
None declared.

References

1 Sarangi PK, Narayan RK, Mohakud S, Vats A, Sahani D, Mondal H. Assessing the capability of ChatGPT, Google Bard, and Microsoft Bing in solving radiology case vignettes. Indian J Radiol Imaging 2023 (e-pub ahead of print). Doi: 10.1055/s-0043-1777746
2 Meskó B. Prompt engineering as an important emerging skill for medical professionals: tutorial. J Med Internet Res 2023;25: e50638
3 Giray L. Prompt engineering with ChatGPT: a guide for academic writers. Ann Biomed Eng 2023;51(12):2629–2633
4 Gamble JL, Ferguson D, Yuen J, Sheikh A. Limitations of GPT-3.5 and GPT-4 in applying Fleischner Society Guidelines to incidental lung nodules. Can Assoc Radiol J 2023 (e-pub ahead of print). Doi: 10.1177/08465371231218250

 © 2024. Indian Radiological Association. All rights reserved.