BMJ Health & Care Informatics

Unleashing the potential of AI: a deeper dive into GPT prompts for medical research

Dorian Garin [1,2]

To cite: Garin D. Unleashing the potential of AI: a deeper dive into GPT prompts for medical research. *BMJ Health Care Inform* 2023;**30**:e100857. doi:10.1136/bmjhci-2023-100857

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/bmjhci-2023-100857).

Accepted 15 August 2023

[1]Internal Medicine, HFR - Hôpital Cantonal, Fribourg, Switzerland
[2]Faculty of Medicine, UNIGE - Université de Genève, Geneva, Switzerland

Correspondence to
Dr Dorian Garin;
dorian.garin@icloud.com

I read the article by Haemmerli *et al* on the performance of ChatGPT-3.5 in generating treatment recommendations for central nervous system (CNS) tumours, which were then evaluated by tumour board (TB) experts. While the study did illuminate promising aspects of the Artificial Intelligence (AI) model, the design of the prompt used to interact with ChatGPT warrants further consideration.

In the study, the prompt employed was a brief patient history, followed by two questions, which appears to have limited the model's performance. As a sophisticated large language model (LLM), GPT-3.5 relies heavily on the context and specificity of the provided prompt.[1 2] Based on cited literature, an alternative prompt structure could have included context, specific intent, a question and an expected response format. Moreover, pretraining the LLM with examples of the expected answer significantly improves the quality of the answer.[2 3] Finally, the introduction of GPT-4 in early March 2023 has shown considerable improvement in understanding and generating responses when compared with ChatGPT-3.5.[4 5]

Table 1 ChatGPT-3.5 from the paper's online supplemental material S1 (6), chaptGPT-3.5 and GPT-4 adapted output

ChatGPT-3.5 with simple prompt (from the study[6])	ChatGPT-3.5 with pretraining and structured prompt	GPT-4 with pretraining and structured prompt
Patient 9		
'(…)the patient has a malignant brain tumour known as an astrocytoma. These tumours are typically treated with a combination of surgery, radiation therapy and chemotherapy.(…)'	'Oncological diagnosis: Glioblastoma (WHO grade IV) right precentral lesion(…) Proposed treatment: Palliative care, supportive management. Discussion: Considering the patient's age, comorbidities and postoperative complications, the panel recommends a palliative approach.(…)'	'Oncological diagnosis: Glioblastoma (WHO grade IV) Right precentral Lobe(…) Proposed treatment: Due to the poor general condition and advanced age of the patient,(…)focus should be on symptomatic relief, quality of life, and palliative care. Discussion: (…)The tumour board suggests a supportive and palliative care approach(…). Regular follow-ups with palliative care and neurology teams should be established(…).'
Patient 10		
'(…)the patient likely has a grade III or IV astrocytoma. The best adjuvant treatment for this type of tumour would likely be a combination of radiation therapy and chemotherapy with temozolomide It is also important to note that the patient has IDH-1 wild type, which is associated with poor outcome and less response to chemotherapy and radiation therapy.'	'Oncological diagnosis: Glioblastoma (WHO grade IV) left frontotemporal lesion Proposed treatment: Radiotherapy and chemotherapy. Discussion: In the case of a gross total resection of a left frontotemporal glioblastoma, the neurotumour board recommends adjuvant treatment with radiotherapy and chemotherapy. (…)Close monitoring and follow-up will be essential to assess treatment response(…)'	'Oncological diagnosis: Glioblastoma (WHO grade IV) left frontotemporal lobe proposed treatment: (…)Standard adjuvant radiotherapy and chemotherapy using temozolomide according to the STUPP protocol.(…) Discussion: (…)an unmethylated MGMT promoter suggests a less favourable response to temozolomide chemotherapy, however, given the patient's overall good health status, adjuvant radiochemotherapy remains the standard of care.(…)'

MGMT, O6-methylguanine-DNA methyltransferase; IDH-1, isocitrate dehydrogenase 1; STUPP protocol, temozolomide + radiotherapy according to Roger Stupp's regimen

With the application of these techniques, researchers could have guided the predictive capabilities of the LLM to generate more relevant and contextually nuanced responses. This could have particularly helped in areas where the model underperformed, such as precision in glioma subtypes and considerations of patient functional status.

As an illustration, both ChatGPT-3.5 and GPT-4 were pretrained with eight examples (patients 1–8, patient history followed by TB response) from online supplemental material of the study. A more context-specific prompt was then used with the history of patients 9 and 10. Table 1 displays main output obtained using this technique, revealing enhanced precision in oncological diagnosis, treatment discussions and patient functional status from ChatGPT-3.5 compared with what was presented in the paper. GPT-4 seemed to align even more closely with the board's opinion, which was defined as the gold standard. Full discussion with the chatbot is available in online supplemental material 1.

It is critical to acknowledge that the efficiency of LLMs applications heavily depends on the prompt used and the quality of the data given. Future research needs to employ a refined, context-driven approach in interacting with these models and the development and sharing of prompt engineering techniques should continue to be prioritised.

In conclusion, the exploration of LLM in CNS oncology research is commendable, but it is essential to optimise the methodology to fully unlock the true potential of AI tools in such a complex and challenging clinical landscape.

Acknowledgements The author would like to thank Dr. Marie-Claude Blatter, PhD, for raising interest in the use of LLMs in medicine, and GPT-4 for writing and proofreading assistance.

Contributors DG created the concept of the letter, reviewed the literature and wrote the manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iD

Dorian Garin http://orcid.org/0000-0002-3872-3496

REFERENCES

1 Gordijn B, ten HH. Chatgpt: evolution or revolution? *MED Health Care Philos* 2023;26:1–2.
2 Lyu Q, Tan J, Zapadka ME, *et al*. Translating radiology reports into plain language using chatgpt and GPT-4 with prompt learning: results, limitations, and potential. *Vis Comput Ind Biomed Art* 2023;6:9.
3 Liu J, Shen D, Zhang Y, *et al*. What makes good in-context examples for GPT-3? [Internet]. arXiv. 2021. Available: http://arxiv.org/abs/2101.06804 [Accessed 15 Jul 2023].
4 Nori H, King N, McKinney SM, *et al*. Capabilities of GPT-4 on medical challenge problems. 2023. Available: http://arxiv.org/abs/2303.13375 [Accessed 15 Jul 2023].
5 Gpt-4. Available: https://openai.com/research/gpt-4 [Accessed 15 Jul 2023].
6 Haemmerli J, Sveikata L, Nouri A, *et al*. Chatgpt in glioma adjuvant therapy decision making: ready to assume the role of a doctor in the tumour board *BMJ Health Care Inform* 2023;30:e100775.