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

Dorian Garin

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. 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. 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.

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

<table>
<thead>
<tr>
<th>Patient 9</th>
<th>ChatGPT-3.5 with simple prompt (from the study)</th>
<th>ChatGPT-3.5 with pretraining and structured prompt</th>
<th>GPT-4 with pretraining and structured prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘(...)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(...)’</td>
<td>‘Oncological diagnosis: Glioblastoma (WHO grade IV) right precentral lesion(...)’</td>
<td>‘Oncological diagnosis: Glioblastoma (WHO grade IV) Right precentral Lobel(...)’</td>
<td>‘Oncological diagnosis: Glioblastoma (WHO grade IV) Right precentral Lobel(...)’</td>
</tr>
<tr>
<td></td>
<td>Proposed treatment: Palliative care, supportive management. Discussion: Considering the patient’s age, comorbidities and postoperative complications, the panel recommends a palliative approach(...)’</td>
<td>Proposed treatment: Palliative care, supportive management. Discussion: 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(...)’</td>
<td>Proposed treatment: Palliative care, supportive management. Discussion: 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(...)’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient 10</th>
<th>ChatGPT-3.5 with simple prompt (from the study)</th>
<th>ChatGPT-3.5 with pretraining and structured prompt</th>
<th>GPT-4 with pretraining and structured prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>‘(...)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.’</td>
<td>‘Oncological diagnosis: Glioblastoma (WHO grade IV) left frontotemporal lesion’</td>
<td>‘Oncological diagnosis: Glioblastoma (WHO grade IV) left frontotemporal lesion’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>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(...)’</td>
<td>Proposed treatment: Radiotherapy and chemotherapy. Discussion: (...)Standard adjuvant radiotherapy and chemotherapy using temozolomide according to the STUPP protocol(...)’</td>
<td>Proposed treatment: Radiotherapy and chemotherapy. Discussion: (...)Standard adjuvant radiotherapy and chemotherapy using temozolomide according to the STUPP protocol(...)’</td>
</tr>
</tbody>
</table>

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.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

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

REFERENCES
1 Goré Di, ten HH. Chatgpt: evolution or revolution? MED Health Care Philos 2023;26:1–2.