GPT Family Validation in **Medical Education | Custom Tools**

Thank you for participating in this form and contributing to the review of LLMs in the education linked to medicine. This time, we experiment with ChatGPT relatives in performing tasks reflected from the downgrade results of this baseline in previous attempts (with *prompting*).

🎯 The ChatGPT environment has been extended with additional plugins making the model able to search information from external sources in the Internet, capturing up-to-date data. Here, we test this promising feature on the failures of native ChatGPT at this level using *Consensus* and *AI for Medical Students* customs GPTs.

📰 Our experiments encompass a sequence of **four** tasks in respect of learning Infective Endocarditis (IE):

1. Grounded Explanation
2. Material Update
3. Data Presentation
4. Multiple-Choice Question Solving Assistant

👍 The inquiry was organized to be swift and direct to the found ChatGPT's weak point! Please follow its outcomes in the received PDF sequentially and respond to the questions separately. Any other observations or suggestions are always welcome from you!

T1 - Grounded Explanation

Explain IE carefull align to the given student level, sustaining with references from accredited online sources.

How do you score the model IE explanations to medical students, considering it made its research on the Internet? Please mark from **1** (as **Unusable**) to **5** (as **Perfect**) when answering this question. *

	1	2	3	4	5
ToolsT1-o1	○	○	●	○	○
ToolsT1-o2	○	○	●	○	○
ToolsT1-o3	○	○	●	○	○

Which complexity level is more likely associated with those IE explanations? *

	Non-Medical Students	(Regular) Medical Students	Advanced Medical Students
ToolsT1-o1	●	○	○
ToolsT1-o2	○	●	○
ToolsT1-o3	○	●	○

Did you find any non-factual information or improper content you would like to report? If so, where? *

Sim. 1.01 alguma linguagem não medica clumps of bactéria and cells...
1.02 e 1.03 com várias referências que não existem!!!! Com informações sobre episimiologia Portugal e diagnóstico errada ou incompleta

Are references sustaining those explanations seamlessly (i.e., there is relevant content and no outdated information)? *

○ Yes

● No

Compared to *GPT-4o* ("Direct Inference for TEXT" task), how fluent and informative does this tool pass medical language? *

○ Better

● The Same

○ Worst

Do you have any extra observations or suggestions you want to provide? (Optional)

....

T2 - Material Update

Update old content from a given pedagogical material.

How do you agree with the updates the model give to *ChatGPT*-generated material? *

	Disagree	Agree
ToolsT2-o1	●	○
ToolsT2-o2	●	○

Did the model add non-factual information (including references) or inrelevant updates to the * *ChatGPT*-generated material? If so, which?

Outdated information. Ou seja o que acrescentou continua a ser desatualizado em particular no diagnóstico

Do you have any extra observations or suggestions you want to provide? (Optional)

T3 - Data Presentation

Create tables with revised data related to a given question.

How well does the model present information in tables for pedagogical purposes? Please * mark from **1** (as **Unusable**) to **5** (as **Perfect**) when answering this question.

	1	2	3	4	5
ToolsT3-o1	○	○	◉	○	○
ToolsT3-o2	○	○	◉	○	○

Did the model make non-factual information (including references) when creating a table? If * so, which?

Sim. A primeira mistura incidência com % de variação e as referências estão incompletas. Não referendo Journal e por isso não são passíveis de serem validadas. A segunda é um quadro melhor mas referências incompletas.

Would those tables fit into a medical theoretical slide presentation, for example? *

○ Both

○ Only the first one

◉ Only the second one

○ None of them

○ Outra: ....

Do you have any extra observations or suggestions you want to provide? (Optional)

....

Assist a medical student solving a MCQ without giving the solution right away. Note we opt to use*ChatGPT* pursuing the role of the student.

How well did the model assist the medical student in solving the given MCQ? Please mark from **1** (as **Unusable**) to **5** (as **Perfect**) to answer this question. *

○ 1

○ 2

◉ 3

○ 4

○ 5

Did you encounter any misleading information in that conversation between models? If so, which? *

Seria interessante . Mas têm erros. A primeira resposta do aluno por exemplo tem erros na escolha dos antibióticos por exemplo.

Do you see this tool applicable for education in assisting medical students solving questions and learning from that? *

⦿ Yes

◯ No

◯ Outra:

Do you have any extra observations or suggestions you want to provide? (Optional)

Google Formulários