GPT Family Validation in **Medical Education | Error Detection & Correction**

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*).

🎯 Considered "best" *prompts*, they can still lead to mistakes - since no safety guarantee was found. As a second step, we test if *GPT-4o* can instead detect those errors and misconception information in wrongly *ChatGPT*-generated material.

❓ Given the uncertainty in obtaining poor results in **MCQ generation**, we ask the model to review a given **MCQ** created by *ChatGPT* alongside the Infective Endocarditis (IE) **topic** and **student level**, from which the question originated.

👍 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!

Do you agree with the model MCQ reviews about IE in assessing medical students? *

	Disagree	Agree
EDC-o1	○	●
EDC-o2	○	●
EDC-o3	●	○
EDC-o4	○	●
EDC-o5	○	●
EDC-o6	○	●
EDC-o7	○	●
EDC-o8	○	●

For those reviews where you disagreed with *GPT-4o*'s opinion, indicate which form parts were * wrongly argued for evaluating medical students.

Na questão 3 falhou a alínea d). Mas de resto esteve bem.

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

Google Formulários