

Review of: "A Comparative Study of Large Language Models in Explaining Intrinsically Disordered Proteins"

Jean-Marc Lelièvre

Potential competing interests: No potential competing interests to declare.

This work first fulfills a lasting gap (despite the growing number of studies) between the large use of AI tools based on Large Language Models (LLM) by the public, including researchers and students, to get information rapidly, and the evaluation of the results generated by these tools. This latter is meaningful only if these results can be confronted with the specialists' knowledge of a given field.

Second, it is applied to an important emerging field. Even though it cannot be considered new anymore, it is still often ignored or misunderstood by reference textbooks and a large portion of biochemists and molecular and cell biologists. Key -if not all- biological functions cannot be approached by ignoring the role of intrinsically disordered proteins and protein regions (IDPs/IDPRs). It was therefore very interesting to analyze the responses brought by AI robots.

In this regard, this study is successful and provides useful guidelines for both teaching/learning about IDPs and the use of AI tools.

The performance of GPT-4 was certainly expected. On the opposite, its decreased performance when associated with the internet Bing Browser is more surprising and represents a challenge for LLM specialists. As mentioned by the authors, this result must be discussed along with the strengths and weaknesses of AI tools: however, it also illustrates how much these tools remain a "black box," leaving only the possibility to assess their results *a posteriori*.

Similarly, GPT-4 performed better than GPT-3.5 on 13 questions, but how GPT-3.5 can beat GPT-4 on 2 questions not only remains a mystery but highlights the necessity of ...human interactions (namely the necessity to keep human teachers and classes) to assess the results, and the stimulation of critical thinking, as mentioned by the authors.

Notwithstanding these criticisms, one can certainly share the optimistic view of the authors claimed in the Conclusions section.

Taken together, this kind of work highly deserves to be read and should inspire many readers to extend it.

Additional remarks :

- The supplementary file would be better referred to as Supplementary Figure S1 and associated with a legend.
- The organization of Table 4 might be improved to avoid orphan lines.
- In the Conclusions section, "its clear.." should be replaced by "It is..."

