Role of ChatGPT in predicting protein–protein interactions

Artificial intelligence (AI), particularly natural language processing (NLP) technologies, have made significant strides in recent years. One of the most impactful developments has been ChatGPT, an advanced language model built on the GPT-4 architecture. We report here the role of ChatGPT in predicting protein–protein interactions (PPI), highlighting its contributions, limitations and potential ethical considerations[1]. We present some ways in which ChatGPT can contribute to predicting PPI.

ChatGPT can efficiently scan, analyse and summarize relevant research articles, reviews and other resources, enabling researchers to stay up-to-date on the latest advancements in PPI prediction methodologies, datasets and tools. Providing quick access to relevant information can save researchers time and help them identify novel techniques, data sources, and insights that can enhance their PPI prediction efforts. Processing vast amounts of data and identifying patterns within the available information can generate hypotheses about potential PPIs or suggest new approaches to predicting interactions[2]. Researchers can then investigate these hypotheses or consider the suggested approaches to improve their predictive models. Such AI tools can help researchers analyse and interpret data related to PPIs. For example, it can aid in the preliminary interpretation of the results from machine learning models or network-based analyses. This assistance can help researchers focus on more complex tasks and reduce the time spent on data analysis. Researchers often use various programming languages and software tools in the PPI prediction process. This AI technology can provide suggestions for code snippets or assist with troubleshooting issues related to software or programming languages[3]. This support can save researchers time and improve the efficiency of their work. It can also facilitate communication among researchers working on PPI prediction, helping them discuss their findings, challenges and ideas more effectively. Bridging language barriers and providing accurate translations helps foster international collaboration and the exchange of ideas in the field of PPI prediction. Further, it may act as an educational resource for researchers, students or other professionals interested in PPI prediction by providing explanations of complex concepts, methodologies and algorithms, helping individuals build a strong foundation in the field. While ChatGPT can be a valuable tool in assisting researchers with predicting PPI, it is important to recognize its limitations and the key challenges associated with its use in this context.

Some of these limitations and challenges include[4,5]:

(i) *Lack of domain-specific knowledge:* ChatGPT is a general-purpose language model and may not possess the same domain-specific knowledge as a specialized bioinformatics tool or expert in the field. This limitation could result in less accurate or less insightful information being provided.

(ii) *Inherent biases and errors:* An AI model learns from the data it is trained on, which can contain biases or inaccuracies. These issues may be reflected in the generated responses, leading to misleading or incorrect information. Researchers must be cautious when relying on its output and cross-reference it with reliable sources.

(iii) *Inability to perform specialized bioinformatics tasks:* It is not specifically designed for bioinformatics tasks, such as sequence alignment, protein–structure prediction, or machine learning model development[6]. Therefore, it cannot replace specialized bioinformatics tools or algorithms for PPI prediction.

(iv) *Limited understanding of biological context:* While it can provide information and generate hypotheses, it may not fully grasp the biological context or underlying mechanisms of PPIs. As a result, some of its suggestions might not be biologically relevant or feasible.

(v) *Overreliance on AI:* Relying too heavily on this tool could hinder critical thinking and reduce researchers' ability to independently assess and analyse data. Researchers need to balance leveraging AI tools like ChatGPT and using their expertise in the field.

(vi) *Intellectual property and authorship concerns:* The use of ChatGPT in generating ideas, hypotheses, or even manuscript drafts raises questions about intellectual property and authorship. Ensuring fair recognition of contributions and establishing clear guidelines for attribution is essential when using AI-generated content.

(vii) *Data privacy and security:* Using ChatGPT to analyse sensitive or proprietary data could pose risks to data privacy and security. Researchers must ensure that they follow proper data handling procedures and adhere to the guidelines set forth by their institutions or funding agencies.

It is important to note that while it can be a valuable tool in predicting PPI, it should not replace the expertise of bioinformatics researchers or the specialized tools and techniques specifically designed for PPI prediction tasks[7]. Additionally, it is crucial to be aware of the limitations of ChatGPT, such as potential biases in its training data and the possibility of generating inaccurate information. Researchers can enhance their PPI prediction efforts and accelerate scientific progress by considering these factors and using them as complementary tools. However, to overcome the above-mentioned limitations and challenges, researchers should use it as a complementary tool alongside their expertise and specialized bioinformatics software.

1. Wang, D. Q., Feng, L. Y., Ye, J. G., Zou, J. G. and Zheng, Y. F., *MedComm – Fut. Med.*, 2023, **2**(2), e43.
2. Majumder, P., In Statistical Modelling and Machine Learning Principles for Bioinformatics Techniques, Tools and Applications (eds Srinivas, K. G. *et al.*), Springer, Singapore, 2020, 1st edn, pp. 119–133.
3. Majumder, P., Ray, P. P., Ghosh, S. and Dey, S. K., *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2019, **17**(4), 1364–1371.
4. Agathokleous, E., Saitanis, C. J., Fang, C. and Yu, Z., *Sci. Total Environ.*, 2023, **25**, 888, 164154.
5. Cheng, K., Li, Z., He, Y., Guo, Q., Lu, Y., Gu, S. and Wu, H., *Ann. Biomed. Eng.*, 2023, **19**, 1–6.
6. Xue, V. W., Lei, P. and Cho, W. C., *Clin. Transl. Med.*, 2023, **13**(3).
7. Tong, Y. and Zhang, L., *Synth. Syst. Biotechnol.*, 2023, **8**(2), 220.

PARTHA PRATIM RAY[1,*]
POULAMI MAJUMDER[2]

[1]Sikkim University,
6th Mile, PO Tadong,
Gangtok 737 102, India
[2]Maulana Abul Kalam Azad University of Technology,
BF Block, Sector 1, Bidhannagar,
Kolkata 700 064, India
*e-mail: ppray@cus.ac.in