

Perspective

Zhengwei Xie* and Gangqing Hu

The artificial intelligence pharma era after “Chat Generative Pre-trained Transformer”

<https://doi.org/10.1515/mr-2023-0023>

Received June 7, 2023; accepted June 27, 2023;

published online August 17, 2023

Abstract: The era of advanced artificial intelligence has arrived with the development of chatbots like ChatGPT (Chat Generative Pre-trained Transformer). As described by Ouyang et al. (2022), ChatGPT demonstrates an impressive ability to generate human-like responses and solve practical problems, surpassing original expectations for its capabilities. The rapid release and adoption of ChatGPT signals a new phase in AI development, powered by large language models that can be fine-tuned through human feedback. However, risks remain regarding how such powerful models may be misused. Further research is needed to ensure safe and ethical deployment of these transformative technologies.

Keywords: artificial intelligence; ChatGPT; pharma

The transformative era of artificial intelligence has arrived. Since its release in November 2022, ChatGPT [1] has surpassed initial projections, commanding widespread attention. The immense popularity of ChatGPT, and similar applications, can be attributed to their capacity for human-like dialogue and profound knowledge base, positioning them as invaluable tools for addressing complex issues, such as drug discovery in pharmacology.

The potency of ChatGPT derives from its high-quality training datasets, and a transformative neural network architecture, supplemented by an “attention mechanism” [2]. The success of this architecture may stem from the structure of its components—query (q), keys (k), and values (v)—which

facilitates efficient storage, comparison, and retrieval of diverse information types. The inclusion of residual layers potentially enhances model sensitivity and linear scalability. Stacking multiple nodes and layers vertically echoes the neural connectivity of the human brain, while the transformer structure enables weight sharing among different information components. Its internal hierarchy aids in restructuring information near the output, ensuring adherence to instruction requirements. All these facets of ChatGPT, along with some “black box” specifications, synergize to foster human-like intelligence in this large-scale model. The result is a multi-modal behemoth capable of various language tasks, such as translation, summarization, gaming, Q&A, proof-reading (as in this perspective), and brainstorming.

Thanks to ChatGPT’s robust natural language generation capabilities, it can be trained and fine-tuned using literature and knowledge pertaining to drug development. By introducing a curated list of query and answer pairs related to drug development and integrating citation information, researchers can leverage ChatGPT to swiftly retrieve relevant literature and patent information, uncover implicit knowledge, and gain research inspiration. This would accelerate preliminary research and foster the discovery of novel directions in drug development. Furthermore, ChatGPT can automatically generate biomedical experiment plans and protocols [3], easing the workload for researchers and expediting experiment design. Adopting an iterative approach, as introduced in bioinformatics education [4], researchers can describe the experiment’s purpose and key points to ChatGPT and refine the plan and operational details through subsequent dialogues. This iterative procedure, guided by critical and creative thinking, is expected to expedite experiment design, result generation, and verification.

When trained with synthetic pathway literature and descriptive text on molecular structure-activity relationships, ChatGPT can conjecture the synthesis route of molecules and suggest modifications to their chemical properties. Future iterations of chatbots like ChatGPT are anticipated to have a potent multi-modal comprehension capability, understanding and integrating various types of information, including text, images, and videos, to drive a wide array of applications in drug development.

*Corresponding author: **Zhengwei Xie**, Peking University International Cancer Institute and School of Basic Medicine, Peking University Health Science Center, Peking University, Beijing 100191, China; and Peking University – Yunnan Baiyao International Medical Research Center, Peking University Health Science Center, Peking University, Beijing 100191, China, E-mail: xiezhengwei@hsc.pku.edu.cn. <https://orcid.org/0000-0001-9572-878X>

Gangqing Hu, Department of Microbiology, Immunology, and Cell Biology, West Virginia University, Morgantown, WV, USA

In clinical settings, chatbots that prioritize privacy and security can analyze patient case information, medical images, surgical videos, and propose treatment options while also generating hypotheses for new drug targets [5].

However, ChatGPT does have limitations. It could not address questions about knowledge obtained after its last training period (September 2021) without the aid of plugins enabling real-time internet access, such as Bing. Though Bing can partially attribute credits to its responses, a comprehensive evaluation is yet to be carried out. From a historical standpoint, ChatGPT lacks an accurate understanding of past events, and in the context of drug discovery, it is unable to produce code to calculate the logP of numerous molecules [6]. A potential solution is the integration of results from smaller specialized models through ChatGPT. In this context, ChatGPT operates as an intelligent agent, summoning various sub-module software for drug design, analyzing the results, and making multiple attempts. If coding is required for this integration, ChatGPT could generate and execute the relevant code with proper instruction. Concurrently, drug design software would need a custom natural language framework for integration with ChatGPT to automate script writing.

We envisage an even more integrated and automated ChatGPT, possessing unprecedented capabilities. If combined with a domain-specific model targeting a particular disease, it could aid in project initiation with needs assessment, formulating drug development plans through literature review, and utilizing specialized smaller GPT models or traditional drug design software. The eventual goal is full automation of chemical synthesis in a chemistry laboratory, followed by relevant biological experiments in a molecular cell laboratory, and subsequent analysis of results. This would allow for a high degree of automation throughout the development process, from planning to execution.

Finally, ChatGPT's prowess in language translation is well suited for the translation work required for

drug applications. It also shines in assisting with medical record-keeping for clinical trials. It is conceivable that we are at the cusp of establishing an artificial intelligence doctor [7] such as a 'ChatDoctor', which could help in identifying rare diseases, and providing professional advice in prescribing medications and treatment plans.

Research funding: This work was supported by the National key R&D program of China (2018YFA0900200) and NSFC (31771519).

Author contributions: Original draft, Z. X; Review & Editing, G. H.

Conflict of interest: The authors declare that there is no conflict of interest.

References

1. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, et al. Training language models to follow instructions with human feedback; 2022. <https://arxiv.org/abs/2203.02155> [Accessed 8 Jun 2023].
2. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need; 2017. <https://arxiv.org/abs/1706.03762> [Accessed 8 Jun 2023].
3. Rehana H, Bengisu Çam N, Basmaci M, He Y, Özgür A, Hur J. Evaluation of GPT and BERT-based models on identifying protein-protein interactions in biomedical text; 2023. <https://arxiv.org/abs/2303.17728> [Accessed 8 Jun 2023].
4. Shue E, Liu L, Li B, Feng Z, Li X, Hu G. Empowering beginners in bioinformatics with ChatGPT. *bioRxiv* 2023. <https://doi.org/10.1101/2023.03.07.531414>.
5. Savage N. Drug discovery companies are customizing ChatGPT: Here's how. *Nat Biotechnol* 2023;41:585–6.
6. Castro Nascimento CM, Pimentel AS. Do large language models understand chemistry? A conversation with ChatGPT. *J Chem Inf Model* 2023;63:1649–55.
7. Li Y, Li Z, Zhang K, Dan R, Zhang Y. ChatDoctor: a medical chat model fine-tuned on LLaMA model using medical domain knowledge; 2023. <https://arxiv.org/abs/2303.14070>. [Accessed 8 Jun 2023].