Orbital
The electronic journal of chemistry

the electronic journal of chemistry

Technical Note | http://dx.doi.org/10.17807/orbital.v16i4.21129

Investigating the Prospects of ChatGPT in Training Medicinal Chemists and the Development of Novel Drugs

Michell de Oliveira Almeida *, Artur Caminero Gomes Soares *, Gustavo Henrique Marques de Sousa *, Witor Ribeiro Ferraz , and Gustavo Henrique Goulart Trossini **

This scientific article delves into the advantages, insights, and limitations of ChatGPT in various scientific domains. Alongside other large language models, this tool exhibits the potential to directly or indirectly assist in a range of scientific areas including Computer Science, Chemistry, Biology/Bioinformatics, and Medicine. Some of the functionalities of ChatGPT include text translation, code improvement, data visualization, and database cleaning. The model can aid in writing and translating scientific articles from mostly any language. In the field of chemoinformatics and computational chemistry, ChatGPT can provide code examples and assist in code development, by evaluating and enhancing code readability and project documentation. Furthermore, it can assist in the database cleaning process and create customized functions for performing specialized tasks. However, ChatGPT does possess some limitations, such as frequent occurrences of artificial hallucinations (a response generated by AI that comprises erroneous or misleading information presented as true), the inability to process multimodal information, and the potential for biases in its training datasets. Therefore, caution must be exercised when incorporating these technologies, considering their social impact and implications for the job market. Acknowledging limitations is crucial when using these tools. With careful and proper use, they can aid the scientific process with the potential to speed up drug discovery.

Graphical abstract

Keywords

Natural Language Processing
Medicinal Chemistry
Drug design and development
Chemoinformatic
Computational chemistry

Article history

Received 14 May 2014
Revised 02 Aug 2024
Accepted 02 Aug 2024
Available online 29 Dec 2024

Handling Editor: Jamal Rafique

1. Introduction

LITEC - Laboratório de Integração entre Técnicas Experimentais e Computacionais no planejamento de fármacos, School of Pharmaceutical Sciences, Pharmacy Department, University of São Paulo, Brazil. *These authors contributed equally to the execution of this work. **Corresponding author. E-mail: **trossini@usp.br**

Natural Language Processing (NLP)-based technologies, often known as chatbots, have radically revolutionized several areas of society, including the social, economic, and scientific spheres, by providing a diverse set of functions. These potent instruments can be utilized in a wide range of activities, including text translation, user interface development, and journalistic and scientific writing. Among their many functions, such tools can be extremely useful for improving the in-silico phase in the discovery of new bioactive compounds, a critical stage in the lengthy drug development process that is becoming increasingly important in Medicinal Chemistry (MC) [1]. Medicinal chemistry is a multidisciplinary science that examines medications using a variety of scientific disciplines, including organic chemistry, biochemistry, physics, pharmacology, computer science, and molecular biology [2]. From this science, it is possible to discover new medicinal compounds and one example of a widely used study is drug design. Drug design seeks to find new drugs according to a biological target and one of the most widely used ways of doing this is through computational tools [3]. It is therefore very important to use as many tools as possible, such as chatbots. The chatbot "ELIZA" is credited with pioneering NLP-based approaches in the mid-1960s [4]. This technology has improved over time, culminating in the 1995 release of the A.L.I.C.E (Artificial Linguistic Internet Computer Entity) programmer [5], which could simulate a conversation in a more rudimentary way. In recent years, artificial intelligence models have advanced. Currently, the influence of these instruments is undeniable, with major benefits for education, research, and the labor market [6–8]. Language modeling (LM) is fundamental to machine language intelligence. The objective is to forecast the probabilities of future (or absent) tokens in sequences of words. Researchers have observed that scaling pre-trained language models (PLM) boosts their ability in future tasks. Studies have exploited this limit by training larger and larger PLMs, such as GPT-3 (175B parameters) and PaLM (540B parameters) [9, 10]. Although scaling is based purely on the size of the model, these enormous PLMs behave differently from smaller PLMs and display surprise aptitude in accomplishing complicated tasks. For example, GPT-3 can handle few-shot tasks by context learning, while GPT-2 cannot [11]. The research community uses the term large language models (LLMs) to refer to these massive PLMs, which are getting more and more attention. In this setting of perpetual change, the great growth of LLMs and generative models has produced a plethora of potential, stimulating significant interest in scientific applications. Pharmaceutical companies such as BioXcel Therapeutics consider LMMs for studies of possible drug candidates [12]. Among them, certain fields of research stand out, such as Computer research [13], Chemistry [14], Biology/Bioinformatics [15], and Medicine [16], where the potential benefits of the direct or indirect use of these tools have been widely recognized, as illustrated in Figure 1.

A specific Large Language Model (LLM) has become quite popular since its launch on November 30, 2022, as Figure 1 illustrates: the Chat GPT (Generative Pre-trained Transformer) created by OpenAI [17]. Other comparable technologies like Google Bard [18] and Bing AI [19] have also emerged in tandem with its widespread adoption, underscoring the development and variety of these technical solutions. Technically speaking, programs such as ChatGPT use architectures that make use of Transformers [20], which are particular neural network models used in Natural Language Processing (NLP) to maximize textual data processing efficiency through parallelization and textual context consideration, resulting in notable advances in LLM development [21]. To enable more seamless user interactions, these models are trained with a lot of parameters. For example, GPT-4 uses more than 1 trillion parameters in its most recent platform version, while GPT-3 used 175 billion parameters in its predecessor. It is significant to note that comprehensive information regarding the textual datasets utilized to train these models has not been made public. The developers have defended this omission by citing the need to remain competitive in the current market [22, 23]. Prior to getting too enthusiastic about this kind of technology and thinking about potential uses, it is important to point out that LLM, and ChatGPT in particular, have a number of restrictions that need to be taken into account, especially in a scientific setting. When asked where the response originated, the model frequently reports experiencing hallucinations, which might undermine the validity of the generated responses [24]. For example, it may quote completely fictitious sources. Moreover, multimodal data, including audio, video, and image processing—all frequent results of computer studies commonly used in machine learning—cannot be processed by the present versions. The bias in the datasets used for model training, which can lead to discriminating or biased results, is another important consideration. Furthermore, LLMs have performed inadequately in terms of conforming to human values and viewpoints. For instance, the model considers all nouns, prepositions, and numerals similarly, but people are far more concerned about inaccurate age statements than they are about a preposition being misplaced [25], not to mention the possibility of the model being used to spread false information. Finally, it's critical to take into account how using these tools may affect society [6, 23, 26]. When integrating these technologies into their everyday work, chemists across all specialties should proceed with prudence and avoid expecting them to be the one and only answer to every problem.

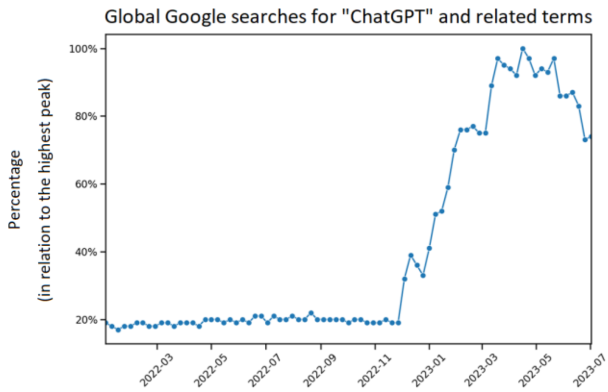

Fig. 1. Searches by users on the Google search engine conducted globally using the terms 'ChatGPT' or 'Chat GPT' or 'GPT' or 'Chat.' The data were obtained using the Google Trends tool.

Following this concise introduction, we possess sufficient data to delineate the essence of ChatGPT, encompassing its proficiency and limitations. In the future, we will examine the practical benefits offered by this instrument in the creation and analysis of new medications. We will demonstrate its importance in different stages of the process of discovering physiologically active compounds and in teaching aspiring medicinal chemists. An example of chatbot application in medicinal chemistry and drug development is the work by Wang and collaborators who utilized ChatGPT to research the development of anticocaine addiction medicines [27]. In this paper, the authors employed ChatGPT plugins to conduct research into the treatment of cocaine addiction. The report also employs ChatGPT to analyze similarity indices for compounds effective against cocaine transporters. Another intriguing use of this work was to employ ChatGPT in virtual screening, where they examined whether drug candidates have favorable ADMET values. Therefore, our objective is to offer illustrations and perspectives that showcase how ChatGPT can accelerate certain and vital phases in the exploration of potentially beneficial therapeutic structures, as well as enhance the skills and knowledge of students and enthusiasts in the field, such is the usage of chatbots for language translation, data cleaning and improved data visualization.

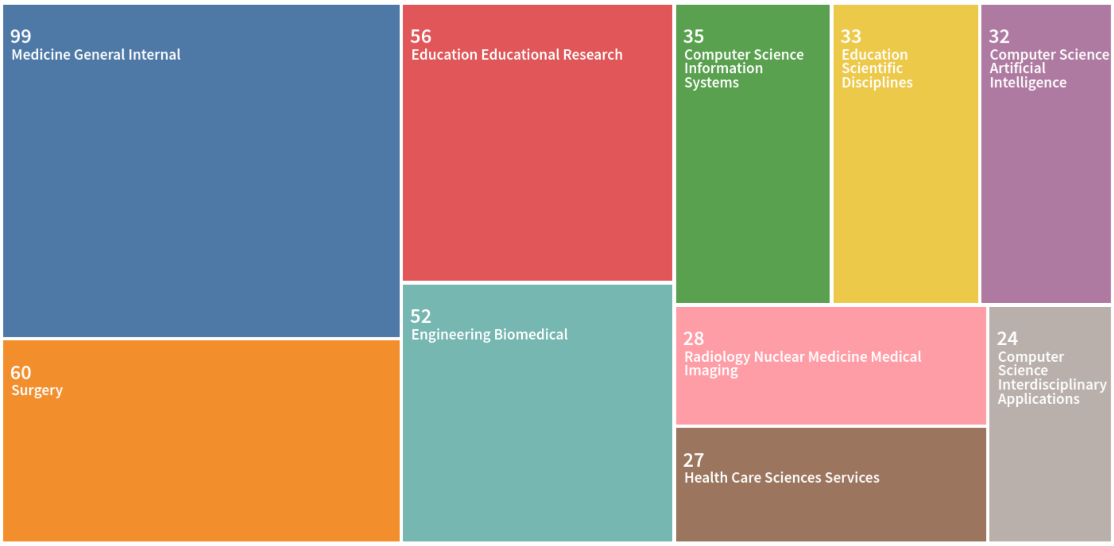

Fig. 2. Treemap of articles published on the topic of ChatGPT, divided by publication category in the Web of Science.

2. Translation

A commonly utilized functionality of ChatGPT is to aid in translation and text editing, notably for those writing in the English language. In this perspective, the tool bears tremendous potential to aid medicinal chemists not born in English speaking nations, given that scientific studies are generally grounded in English. Therefore, ChatGPT is particularly valuable for researchers and students to acquaint themselves with the common jargon of their study areas, enabling them to develop content that helps them comprehend the contexts in which these terms are employed. Another feature of major utility in this functionality is that, in addition to conducting direct translations, the chatbot allows interactions in various languages [28, 29]. Furthermore, the model exhibits a high potential to summarize material, able to take an English scientific paper as input and provide a cohesive summary of the most significant elements in different languages. However, due to the lack of internet connection and the model being trained on data available only up to 2021, the system cannot retrieve a scientific publication based on its Digital Object Identifier (DOI) or title. Therefore, texts must be supplied directly for the model to process them. It is crucial to note that ChatGPT has a character limit, which may require breaking the material into smaller portions [24].

3. Code Improvement

Modern drug planning procedures are becoming more diverse and interdisciplinary, integrating many computational tools into their workflow. These tools are typically advanced, such as chemometric studies, and are commonly found in software without a graphical interface. This frequently demands programming knowledge, posing a significant challenge for those wishing to enter this field [30]. Given that drug planning is, as mentioned, interdisciplinary and involves individuals with diverse educational backgrounds, learning programming can be a complex task for some. In this context, the use of ChatGPT can offer valuable assistance in this process [31, 32].

Regarding the production of custom code, ChatGPT can offer support to enhance its readability. The model is capable of performing tasks such as including descriptive comments throughout the code, creating docstrings (which are texts that describe the functioning of functions), renaming variables (for greater clarity), and code refactoring. Additionally, the LLM can contribute to the overall project documentation by generating a "readme" file when appropriate contexts are provided [30, 32].

Another interesting application of ChatGPT in this regard is its support in unit testing, which is the process of software testing aiming at confirming the functionality of individual components. An example of a suitable request would be: "Develop a Python unit test to evaluate the following function, which converts molecules in SMILES (Simplified Molecular-Input Line-Entry System) format to the MOL format." It is worth mentioning that the tests created by the model may provide limited examples, and it can be necessary to manually incorporate more test situations [30, 31]. These computational methods can also help the advancement of drug discovery and expand our comprehension of the chemistry and biology of diseases. ChatGPT can be employed to facilitate drug discovery research. One such example is the utilization of ChatGPT to provide input files for programs like Gaussian and AutoDock. ChatGPT can be utilized for PDB searches, however it does not supply solutions to intricate questions within these searches. Additionally, it is worth stressing the application of ChatGPT in medical and clinical education. In this regard, the research done by Sallam and colleagues indicates the benefits of utilizing ChatGPT in the field of education and medicine [33, 34]. These advantages fundamentally involve raising the capacity for individualized learning, enhancing clinical reasoning skills, and facilitating comprehension of tough medical issues. Furthermore, the authors stress the limits of using ChatGPT in the sectors of medical and education, including privacy concerns, the potential for biased and erroneous information, and the risk of weakening communication and critical thinking capacities. In order to tackle these restrictions, Sallam and his colleagues advocated three approaches to minimize the issues: adding human oversight, implementing privacy and security standards, and stimulating critical assessment of the generated content [34].

4. Improvement in data Visualization

The processes of medicinal chemistry related to the search for bioactive molecules often involve steps of molecular modeling or chemometric analyses, generating a vast profusion of data, with their organization and visualization considered a daunting task for scientists in the field. However, in the last two decades, there has been a technological advent of a series of tools that can assist researchers in this complex mission, including recently the ChatGPT [31]. Data visualization plays a crucial role in conveying information to third parties, as well as in the in-depth exploration and understanding of the information set. Within this context, ChatGPT provides valuable assistance, becoming a study companion in suggesting various types of graphs, as well as in characterizing their advantages and disadvantages [35]. However, at the time of crafting this material, the model is not yet capable of receiving images as input, but developers have announced the future implementation of this functionality, which will be extremely useful in aiding data visualization [17].

Additionally, if we provide information about the columns to ChatGPT and explain the dataset, it can offer suggestions on the types of charts that can be created, assessing which columns can be plotted. It is important to note that the results provided by the model are preliminary and require refinement, as they do not cover all possible evaluations that can be performed for a given dataset but constitute a good starting point [26, 35]. An example of input for the described situation could be: "I have a database with drugs withdrawn from the market. It has the following columns: drug name, approval year, molecular mass, logP, year of market withdrawal. Give me suggestions on which data I can visualize, and what charts I can use to visualize them." Furthermore, if we present our code used for generating the chart to ChatGPT, it can make suggestions to improve it, such as changing colors to make the diagram more accessible to people with color blindness or enhancing overall legibility [29]. The feature for creating figures using ChatGPT was recently implemented and is a promising application for generating and evaluating figures in bioinformatics and medication planning. Even while this application offers a rapid way to visualize a figure related to a given piece of data, in order to properly analyze the image generated, specific knowledge in the subject is required, as the chatbot is not yet fully trained for this purpose. In addition, when a chatbot performs a task outside its scope, it can apply certain "hallucinations" to the analysis. A systematic review is incredibly necessary to grasp the best ways of using a chatbot to generate figures linked to medication design, as well as a critical analysis for a better scientific interpretation of the data generated by the chatbot.

5. Data Cleaning

Until the mid-2010s, the chemical universe was expected to comprise roughly 10^{60} chemical substances known or produced by humans. This quantity includes inorganic compounds, natural products, synthetic molecules, among others [36, 37]. In other words, there is a huge chemical area to be explored by medicinal chemistry in the quest for bioactive compounds. Obviously, this translates into a tremendous amount of molecular data to be used in drug planning, available, for instance, on platforms like as PubChem, ZINC, and CHEMBL [38–40]. However, before employing this data, various processes of cleaning and curation are essential. In this approach, ChatGPT can become a valuable ally [41].

As an example, if the database is smaller than the 4000-character limit of ChatGPT, it is possible to directly input the content of a .CSV (Comma Separated Values) file, indicating to the model that it is a database in this format. This way, the model can interpret the data and perform actions on it, such as converting the words in a column to lowercase or extracting a regular text pattern. However, the use of this method is not recommended due to the possibility of model hallucinations and its low practicality [42].

Another way to utilize ChatGPT in database filtering or cleaning activities is by asking the LLM to create functions that perform the necessary tasks for data cleaning. In these cases, it can be helpful to provide ChatGPT with a portion of the dataset so that it can understand the column names and data types present [43]. Additionally, describing the dataset is an important step that can significantly aid the process. An example input in this case could be "I have a dataset of biological activities for different molecules. It has the following columns: molecule_name, smiles, administration_route, and biological_activity. Create a Python function that creates a column called 'active_molecule,' whose value is 'True' if the biological activity is greater than 10μM and 'False' if it is less."

For simple tasks, as exemplified earlier, ChatGPT performs well. However, in more specific cases, careful analysis by the user is necessary. For instance, the authors requested ChatGPT to create a function that retrieves all FDA-approved drugs from the ChEMBL database. The model's response provided only 20 (arbitrary) molecules as a result. Furthermore, when obtaining a database from ChEMBL, it was observed to have a column of molecular structures, where each data point is a dictionary with various formats, such as SMILES, MOL, MOL2, among others. Subsequently asking ChatGPT to create a function capable of interpreting this dictionary and extracting only the structure in SMILES format, the model was unable to fulfill the task. On the other hand, the model succeeded in creating a function that interpreted a SMILES column and removed salts when present. Thus, ChatGPT can be extremely useful in data cleaning, but caution and careful consideration are necessary when using its responses.

6. Assistance in the Use of Modeling Tools

Command-line scripts and shell scripts are commonly used in the field of computational chemistry and quantum mechanics [41, 44, 45]. Therefore, basic tasks like merging multiple text files into one, detecting text patterns using tools like grep or awk, and applying regular expressions to files are frequently performed by medicinal chemists, particularly those engaged in chemometric analyses. These chemists can greatly benefit from the enhancement of ChatGPT [26, 32].

As an initial illustration, we shall employ the previously mentioned capability of ChatGPT to solicit particular regular expressions using natural language. For example, suppose the medicinal chemist encounters the following issue: they own a file in the .smi format (which is a text file where each line contains a SMILES) and desire to identify all SMILES that have one or more fragments. It would be feasible to question ChatGPT about the position of all SMILES that contain the character "." character, which is used to split fragments, and save them sequentially in a different file for more careful curation. Additionally, conversions of command-line functionality to scripts in other languages, such as Python or Perl, can be implemented fairly easily [32]. An exemplary instance would be the subsequent query: "What is the procedure for merging all files within a directory into a solitary file using the shell? And how can I accomplish the same action with Python?".

The proficiency to manage diverse file formats is crucial in the repertoire of a medicinal chemist. Nevertheless, the amalgamation or alteration of structural files, such as .pdb, .mol, or .sdf, entails a more intricate procedure. The outcomes of our attempts to utilize bash scripts in combination with ChatGPT for the purpose of converting several .mol files into a unified SDF file were unsatisfactory. Hence, we strongly underline the critical necessity of considering more appropriate instruments for executing this particular duty.

It is also important to emphasize that there are other examples of prompts that play a crucial role not only in performing daily tasks but also in the learning process of specific tools used in quantum mechanics (QM), and ChatGPT can be a great learning tool when combined with reading the official manuals of these software [45]. Some of these examples may include: "What are the meanings of each parameter in the .mdp file for a simulation using GROMACS?" or even "Which formats are supported by Gaussian? Generate a script that can convert a .mol file to a format suitable with Gaussian." In this context, the employment of tools for routine automation, such as molecular docking of numerous ligands to a specific target, can be directly influenced by the application of these prompts. However, by adopting common language, even users with low knowledge of computer languages like Python can make requests through prompts [45], for example: "Develop a Python script that enables the docking of various structures using AutoDock Vina." It is vital to note that while these tools may provide improvement regarding pose parameters and energy ranges, it is imperative for the user to have prior technical knowledge and check the official library documentation to ensure proper and efficient use [26, 45].

As a final example and to highlight the use of the LLM with the official documentation of software, we will consider a widely recognized and used tool in the routine of many in medicinal chemistry, OpenBabel. OpenBabel is capable of converting over 110 types of data relevant to the chemical structure of molecules and macromolecules. The obstacle to using such a tool generally derives from the fact that its handling is mostly through the command line. Although the package manual provides clear instructions on how to convert a text file with SMILES into a series of .mol files in batch mode using OpenBabel, ChatGPT was unable to produce the desired command-line results even after trying different prompts. LLM technologies, like ChatGPT, require more training when utilized in the field of medicinal chemistry, as these models lack human-like reasoning capabilities. However, in the future, the implementation of LLM models in medicinal chemistry has the potential to overcome scientific obstacles in the identification of novel molecules. Chatbots might assist in molecular design, synthesis planning, and experiment validation, thereby streamlining the discovery process [46] .

7. Conclusions

Here, we have explored the advantages, insights, and limitations of some possible applications of ChatGPT in an extremely interdisciplinary field, drug planning. This tool, along with other large language models, has the potential to directly or indirectly assist in a range of scientific areas such as Computer Science, Chemistry, Biology, Bioinformatics, and Medicine.

Among some of ChatGPT's functionalities are text translation, code improvement, data visualization, and database cleaning. The model can aid in the writing and translation of scientific articles from English to several languages (or vice versa) and can generate coherent summaries of the relevant points in these articles.

Regarding chemoinformatics and computational chemistry routines, the tool can provide code examples, assist in the development of users with beginner programming levels, evaluate codes written by other professionals for improved readability, and document programming projects. In the context of data visualization, ChatGPT can offer suggestions for the types of charts to create, along with direct improvements in visibility. Additionally, the LLM can assist in the database cleaning process and in the creation of specific/custom functions to perform particular tasks.

However, ChatGPT also has several limitations, including frequent hallucinations, an inability to process multimodal information, and biases present in the training datasets. Moreover, LLMs have shown unsatisfactory performance in aligning with human values and the potential to spread misinformation.

In conclusion, it is important to consider the limitations of ChatGPT and approach its use with caution. With proper and conscientious application, LLMs can play a relevant role in assisting various stages of the scientific process, with the potential to accelerate drug planning and development. Furthermore, the use of chatbots has the potential to assist medicinal chemists in various areas, including the creation of novel molecules and the strategic planning and execution of synthesis.

Author Contributions

M.O.A., A.C.G.S., G.H.M.S. and W.R.F. collected data and wrote the manuscript. M.O.A., A.C.G.S., G.H.M.S. and G.H.G.T. tested and validated the information. A.C.G.S., G.H.M.S. and G.H.G.T. idealized the project.

References and Notes

1. Bruno, A.; Constantino, G.; Sartori, L.; Radi, M. *Curr. Med. Chem.* **2019**, *26*, 3838. **[Crossref]**
2. Guido, R. V. C.; Andricopulo, A. D.; Oliva, G. *Estud. Av.* **2010**, *70*, 81. **[Crossref]**
3. Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. *Pharmacol. Rev.* **2014**, *66*, 334. **[Crossref]**
4. Weizenbaum, J. *Commun. ACM* **1966**, *9*, 36. **[Crossref]**

4. Wallace, R. S. In: Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer. Epstein, R; Roberts, G; Beber, G., eds. Dordrecht: Springer, 2009, chapter 13.
6. Owens, B. *Nature* **2023**, *615*, 20. [Crossref]
7. Teubner, T.; Flath, C. M.; Weinhardt, C.; van der Aalst, W.; Hinz, O. *Bus. Inf. Syst. Eng.* **2023**, *65*, 95. [Crossref]
8. Thorp, H. H. *Science* **2023**, *379*, 313. [Crossref]
9. Naveed, H.; Khan, A. U.; Qiu, S.; *et al. arXiv:2307.06435v10 [cs.CL]* **2024**. [Crossref]
10. Raiaan, M. A. K.; Mukta, Md. S. H.; Fatema, K.; *et al*. *IEEE Access*. **2024**, *12*, 26839. [Crossref]
11. Li, T.; Shetty, S.; Kamath, A.; *et al. Npj Digit. Med.* **2024**, *7*, 40. [Crossref]
12. Chakraborty, C.; Bhattacharya, M.; Lee, S-S. *Mol. Ther. - Nucleic Acids* **2023**, *33*, 866. [Crossref]
13. Xu, F. F.; Alon, U.; Neubig, G.; Hellendoorn, V. J. A systematic evaluation of large language models of code. In Proceedings of the 6th ACM SIGPLAN International Symposium on Machine Programming, Association for Computing Machinery: San Diego, CA, USA, 2022, 1.
14. White, A. D.; Hocky, G. M.; Gandhi, H. A.; *et al. Digit. Discov.* **2023**, *2*, 368. [Crossref]
15. Lubiana, T.; Lopes, R.; Medeiros, P.; *et al*. *PLoS Comput Biol.* **2023**, *19*, e1011319. [Crossref]
16. Lobentanzer, S.; Saez-Rodriguez, J. *arXiv:2307.06435v10 [cs.CL]* **2024**. [Crossref]
17. Available from: https://openai.com/chatgpt Acess April 2024.
18. Available from: https://bard.google.com Access April 2024.
19. Available from: https://www.microsoft.com/en-us/bing?form=MA13FJ Access April 2024.
20. Dufter, P.; Schmitt, M.; Schütze, H. *Comput. Linguist* **2022**, *48*, 733. [Crossref]
21. Sejnowski, T. J. *Neural Comput.* **2023**, *35*, 309. [Crossref]
22. GPT-4 Technical Report. Available from: http://arxiv.org/abs/2303.08774 Access April 2024.
23. Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. Available from: http://arxiv.org/abs/2102.02503. Access April 2024.
24. Zheng, H.; Zhan, H. *Am. J. Med.* **2023**, *136*, 725. [Crossref]
25. Krügel, S.; Ostermaier, A.; Uhl, M. *Sci. Rep.* **2023**, *13*, 4569. [Crossref]
26. Ray, P. P. *Internet Things Cyber-Phys. Syst.* **2023**, *3*, 121. [Crossref]
27. Wang, R.; Feng, H.; Wei, G-W. *J. Chem. Inf. Model.* **2023**, *63*, 7189. [Crossref]
28. Castro Nascimento, C. M.; Pimentel, A. S. *J. Chem. Inf. Model.* **2023**, *63*, 1649. [Crossref]
29. Humphry, T.; Fuller, A. L. *J. Chem. Educ.* **2023**, *100*, 1434. [Crossref]
30. Surameery, N. M. S.; Shakor, M. Y. *Int. J. Inf. Technol. Comput. Eng.* **2023**, *3*, 17. [Crossref]
31. Blanco-González, A.; Cabezón, A.; Seco-González, A.; *et al. Pharmaceuticals* **2023**, *16*, 891. [Crossref]
32. Chen, E.; Huang, R.; Chen, H. S.; Tseng, Y. H.; Li, L. Y. GPTutor: A ChatGPT-Powered Programming Tool for Code Explanation. In: Wang, N., Rebolledo-Mendez, G., Dimitrova, V., Matsuda, N., Santos, O.C., eds. Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners, Doctoral Consortium and Blue Sky. AIED 2023. Communications in Computer and Information Science, vol. 1831. Springer, Cham. [Crossref]
33. Sallam, M.; Salim, N. A.; Al-Tammemi, A. B.; *et al. Cureus* **2023**, *15*, e35029. [Crossref]
34. Sallam, M.; Salim, N.; Barakat, M.; Al-Tammemi, A. B. *Narra J.* **2023**, *3*, e103. [Crossref]
35. Hassani, H.; Silva, E. S. *Cogn. Comput.* **2023**, *7*, 62. [Crossref]
36. Reymond, J-L. *Acc. Chem. Res.* **2015**, *48*, 722. [Crossref]
37. Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J-L. *J. Chem. Inf. Model.* **2012**, *52*, 2864. [Crossref]
38. Banegas-Luna, A-J.; Cerón-Carrasco, J. P.; Pérez-Sánchez, H. *Future Med. Chem.* **2018**, *10*, 2641. [Crossref]
39. Irwin, J. J.; Tang, K. G.; Young, J.; *et al. J. Chem. Inf. Model.* **2020**, *60*, 6065. [Crossref]
40. Kim, S.; Thiessen, P. A.; Bolton, E. E.; *et al. Nucleic Acids Res.* **2016**, *44*, D1202. [Crossref]
41. Yoshimori, A.; Chen, H.; Bajorath, J. *Future Med. Chem.* **2023**, *15*, 119. [Crossref]
42. Porter, L.; Daniel Zingaro. Learn Ai-Assisted Python Programming: With Github Copilot and Chatgpt, 1th ed. United States: Manning Publications, 2024.
43. Ma, L.; Thakurdesai, N.;Chen, J.; Xu, J.; Korpeoglu, E.; Kumar, S.; Achan, K. In LLMs with User-defined Prompts as Generic Data Operators for Reliable Data Processing, 2023 IEEE International Conference on Big Data (BigData), 15-18 December 2023, 3144-3148. [Crossref]
44. Mente, S.; Kuhn, M. *Curr. Top. Med. Chem.* **2012**, *12*, 1957. [Crossref]
45. Rahman, M. M.; Watanobe, Y. *Appl. Sci.* **2023**, *13*, 5783. [Crossref]
46. Language models in molecular discovery. Available from: http://arxiv.org/abs/2309.16235. Access April 2024.

How to cite this article

Almeida, M. O.; Soares, A. C. G.; de Souza, G. H. M.; Ferraz, W. R.; Trossini, G. H. G. *Orbital: Electron. J. Chem.* **2024**, *16*, 319. DOI: http://dx.doi.org/10.17807/orbital.v16i4.21129