



University of Southern Denmark

ChatGPT: First glance from a perspective of clinical pharmacology

Krumborg, Julie Rudbech; Mikkelsen, Nicolaj; Damkier, Per; Ennis, Zandra Nymand; Henriksen, Daniel Pilsgaard; Lillevang-Johansen, Mads; Pedersen, Sidsel Arnspang; Bergmann, Troels K

Published in:
Basic & Clinical Pharmacology & Toxicology

DOI:
10.1111/bcpt.13879

Publication date:
2023

Document version:
Accepted manuscript

Citation for published version (APA):
Krumborg, J. R., Mikkelsen, N., Damkier, P., Ennis, Z. N., Henriksen, D. P., Lillevang-Johansen, M., Pedersen, S. A., & Bergmann, T. K. (2023). ChatGPT: First glance from a perspective of clinical pharmacology. *Basic & Clinical Pharmacology & Toxicology*, 133(1), 3-5. <https://doi.org/10.1111/bcpt.13879>

Go to publication entry in University of Southern Denmark's Research Portal

Terms of use

This work is brought to you by the University of Southern Denmark.
Unless otherwise specified it has been shared according to the terms for self-archiving.
If no other license is stated, these terms apply:

- You may download this work for personal use only.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying this open access version

If you believe that this document breaches copyright please contact us providing details and we will investigate your claim.
Please direct all enquiries to puresupport@bib.sdu.dk

Krumborg Julie Rudbech (Orcid ID: 0009-0000-5549-6590)
Mikkelsen Nicolaj (Orcid ID: 0009-0008-6020-0844)
Damkier Per (Orcid ID: 0000-0003-0591-7187)
Henriksen Daniel Pilsgaard (Orcid ID: 0000-0003-1303-6195)
Pedersen Sidsel Arnspang (Orcid ID: 0000-0001-7457-9092)
Bergmann Troels K (Orcid ID: 0000-0001-8313-0721)

ChatGPT: First Glance from a Perspective of Clinical Pharmacology

Authors: Julie Rudbech Krumborg^{1,2}, Nicolaj Mikkelsen^{1,2}, Per Damkier^{1,2}, Zandra Nymand Ennis^{1,2}, Daniel Pilsgaard Henriksen^{1,2}, Mads Lillevang-Johansen^{1,2}, Sidsel Arnspang Pedersen^{1,2}, Troels K Bergmann^{1,3}

1. Department of Clinical Pharmacology, Odense University Hospital, Denmark
2. Department of Clinical Research, Faculty of Health Science, University of Southern Denmark, Denmark
3. Department of Regional Health Research, University of Southern Denmark, Esbjerg, Denmark

Corresponding author: Julie Rudbech Krumborg: Julie.Rudbech.Krumborg@rsyd.dk

Keywords: Artificial intelligence, Clinical Pharmacology, Drug Information Services, evidence-based medicine, Medical Writing

Conflict of interest statement: The authors state no conflicts of interest

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: [10.1111/bcpt.13879](https://doi.org/10.1111/bcpt.13879)

A white rabbit runs by

ChatGPT, an artificial intelligence (AI) chatbot released in November 2022, has caused controversy within the world of academia¹. The number of articles on the subject continues to increase almost exponentially. A general theme is: To what extent can an AI platform write a reliable scientific text?

In our department – Clinical Pharmacology at Odense University Hospital – we decided to approach the question in a practical manner by testing ChatGPT in a routine task. In our Drug Information Centre (DIC), we answer clinical, drug-related questions from healthcare professionals.

The foundation of our answers is evidence-based medicine². We follow a standardized procedure to ensure scientific quality and reproducibility: From a focused clinical question, we perform a thorough literature search and write a mini-review based on the available evidence with an accompanying interpretation of the clinical relevance. The work is reviewed by the department academic staff on a bi-weekly conference before the report is sent to the inquiring healthcare professional.

We dedicate a significant portion of our work-day to writing and reviewing evidence. The emergence of ChatGPT made us wonder if it could serve as a valid review- or writing assistant – or even replace us. Since our product (the mini-review) is purely text, we can directly compare our answers to ChatGPT's text-based output.

When we asked ChatGPT to write a scientific review on a pharmacological topic, our initial discovery was, that although the answer on surface-level looked convincing, some of the provided references were utter nonsense. In other words: when prompted to write a scientific review, ChatGPT will provide a scientific-looking article including a list of references with an obscure mix of relevant and real articles, real but irrelevant articles, and articles that are complete bogus.

This finding is echoed in the rapidly exploding literature about ChatGPT's capability to produce reliable scientific texts^{3,4}. Likewise, it is openly stated by the creators of ChatGPT that ChatGPT "*sometimes writes plausible-sounding but incorrect answers*"⁵.

Intrigued by this, we decided follow the white rabbit...

So, what exactly is ChatGPT?

ChatGPT is short for Chat Generative Pre-trained Transformer. It is an AI "chatbot"-software program, created by the San Francisco-based tech-company OpenAI¹. The program can generate human-like text responses to any question one can think of. ChatGPT does not only summarize existing text, but also creates

completely new content. In March 2023, a major update to ChatGPT was released (ChatGPT 4.0), which promised improved reasoning and conciseness, and a 40% increased likelihood of producing factual responses⁶.

At its core, an AI is an algorithm, but, unlike static mathematical models, it uses machine learning to adjust its parameters and improve performance over time. To do so, the program needs “training”. ChatGPT was trained by being exposed to a huge amount of text from a database of internet pages. The program then iteratively adjusted its parameters based on human feedback, which gradually made it more capable of generating human-like text. As this whole process took place offline, ChatGPT does not “search the internet in real-time”. It will not consider anything published that was not included in its training material, and it is not getting trained by the input from the current users. However, it does store the content of the current chat and previous “conversations” with the same user, which is a powerful feature because it allows the bot to retain the context. OpenAI states that the conversations the users have with ChatGPT may be reviewed and used for training purposes later, and warns about not sharing sensitive information⁷.

How do I use ChatGPT?

To access ChatGPT, go to chat.openai.com and create a user by providing an e-mail and a phone number for verification. Currently, you can access the GPT-3 version for free and the newer version for a subscription fee. After login, the ChatGPT interface looks like an internet search engine start-page, and it starts replying as soon as you hit enter after writing something. The difference is that instead of getting a long list of web pages with content related to your question, ChatGPT generates a complete answer in a format that is relevant to the question. We have included a link to a video that illustrates the process (supplement 1).

Down the rabbit hole – clinical pharmacology versus ChatGPT

We asked both versions of ChatGPT a question on the safety of methylphenidate in pregnancy and its possible adverse effects on female fertility. We then compared the answers to our DIC mini-review (full answers in supplements 2, 3 and 4).

The overall structure of our answer and that of ChatGPT 3 was almost identical; both provide background information on methylphenidate, followed by a section on female fertility and pregnancy safety and finally a conclusion and list of references. ChatGPT 4.0 provided a more abstract-like summary, and did not include background information about methylphenidate. Some sentences were strikingly similar:

DIC: “(...) some of the included studies found delayed vaginal opening and abnormal estrous cycling in rats given methylphenidate. However, this data cannot be extrapolated to humans.”

ChatGPT 3: “A study conducted on rats showed that methylphenidate treatment reduced ovarian weight and impaired ovulation (Dudley et al., 2009). However, extrapolating these findings to humans is challenging and requires further investigation.”

ChatGPT 4.0: “One study conducted on rats reported that high doses of methylphenidate led to a decrease in ovarian weight and follicle number, suggesting potential adverse effects on fertility (Motabar et al., 2016).”

However, it is important to note that this study used doses that were much higher than those typically prescribed for humans.”

A thorough check of the references provided by both versions of ChatGPT showed that some of them are hallucinations, i.e. completely made-up; while the journals and authors exist, there are no papers with the specific titles in any database we could find. A couple of the references are real though, and some even quite relevant. However, in its current version, ChatGPT does not provide an estimate on the legitimacy of the references, and therefore, you will need to cross-check every reference to separate the wheat from the chaff, which is tedious at best and unfeasible at worst.

To make the confusion complete, ChatGPT also reports plausible, but made-up results: The chatbot replied to our question by presenting specific but spurious odds ratios including a meaningful 95% confidence interval, and referred to a cohort study as being a meta-analysis. Also, the conclusion made by ChatGPT was not the conclusion made in the actual study⁸.

The conclusions from the DIC and ChatGPT are overall comparable. To our relief, admittedly, we found that as clinical pharmacologists, we did a better job at translating the findings to clinical practice. We also included information on breastfeeding, which often turns out relevant in the context of a pregnancy-related question. The AI can answer the question at hand, but we are able to anticipate other needs from the inquiring healthcare professional.

If it grins like a cat, disappears like a cat, and speaks riddles like a cat

Before writing this essay, it was a common belief in our department that ChatGPT was looking at the knowledge currently available and considering “all of the internet”. However, ChatGPT is a text-based AI. It is not a tool of knowledge but a tool of text. On the one hand, in the specific context of writing in the field of clinical pharmacology, our initial finding is that ChatGPT is unreliable because it generates nonsensical references and postulates, and it is impossible to reverse engineer an answer – the very essence of scientific method. On the other hand, ChatGPT is currently leading a revolution in how we use AI technology, and its impact within healthcare is yet to be determined.

According to Gartner’s hype cycle⁹, when a new technology is released, in the first wave of hype, the possibilities appears limitless. The media talks nonstop about how AI and robots will disrupt entire professions. After a while, when the new technology has been tested on real-life applications, its underlying issues and limitations are discovered, and the hype falls drastically. In the plateau phase, the technology is implemented, often in a much narrower frame of use than we first dreamed of.

After our preliminary explorations, we are not convinced that ChatGPT is going to replace us, but we do believe it is likely that clinical pharmacologists who speak ChatGPT fluently will have a substantial advantage. AI, as a whole, has strong potential as a tool in clinical pharmacology – for instance as a writing support tool, as an assistant data analyst in pharmacogenomics or as an automated answer-generator for TDM results. It should not, though, act as a substitute for the extensive literature review, wealth of cross-references and skeptical mind of a clinical pharmacologist.

As clinical pharmacologists, we are cautiously optimistic and enthusiastically curious about the future of AI within our field, and we encourage our peers worldwide to share their experiences. AI has immense potential to change medical writing, drug information services, clinical pharmacology, and the whole medical profession in fundamental ways and we are hopeful it is for the better.

References

1. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature*. 2023;613(7945):620-621. doi:10.1038/d41586-023-00107-z
2. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet*. 2017;390(10092):415-423. doi:10.1016/S0140-6736(16)31592-6
3. Lee P, Bubeck S, Petro J. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *N Engl J Med*. 2023;388(13):1233-1239. doi:10.1056/NEJMSr2214184
4. Homolak J. Opportunities and risks of ChatGPT in medicine, science, and academic publishing: a modern Promethean dilemma. *Croat Med J*. 2023;64(1):1-3.

5. ChatGPT: Optimizing Language Models for Dialogue. Available from: <https://openai.com/blog/chatgpt/>. Accessed: March 7, 2023.
6. ChatGPT: 2023. <https://openai.com/product/gpt-4>, Accessed: March 21, 2023.
7. ChatGPT General FAQ. Available from: <https://openai.com/blog/chatgpt/>. Accessed: March 8, 2023.
8. Cooper WO, Habel LA, Sox CM, et al. ADHD drugs and serious cardiovascular events in children and young adults. *N Engl J Med*. 2011;365(20):1896-1904. doi:10.1056/NEJMoa1110212
9. Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Med*. 2019;17(1):143. doi:10.1186/s12916-019-1382-x