



## Warning: Artificial intelligence chatbots can generate inaccurate medical and scientific information and references

Catherine L. Clelland<sup>1</sup> , Stuart Moss<sup>2</sup> , James D. Clelland<sup>3\*</sup> 

<sup>1</sup>Department of Psychiatry and Taub Institute, Columbia University Irving Medical Center, New York, NY 10032, USA

<sup>2</sup>Health Services Library, The Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

<sup>3</sup>Clinical Research Department, The Nathan Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA

\***Correspondence:** James D. Clelland, Clinical Research Department, The Nathan Kline Institute for Psychiatric Research, 140 Old Orangeburg Road, Orangeburg, NY 10962, USA. [James.Clelland@nki.rfmh.org](mailto:James.Clelland@nki.rfmh.org)

**Academic Editor:** Hari Prasad Devkota, Kumamoto University, Japan

**Received:** April 28, 2023 **Accepted:** August 3, 2023 **Published:** January 10, 2024

**Cite this article:** Clelland CL, Moss S, Clelland JD. Warning: Artificial intelligence chatbots can generate inaccurate medical and scientific information and references. *Explor Digit Health Technol.* 2024;2:1–6. <https://doi.org/10.37349/edht.2024.00006>

### Abstract

The use of generative artificial intelligence (AI) chatbots, such as ChatGPT and YouChat, has increased enormously since their release in late 2022. Concerns have been raised over the potential of chatbots to facilitate cheating in education settings, including essay writing and exams. In addition, multiple publishers have updated their editorial policies to prohibit chatbot authorship on publications. This article highlights another potentially concerning issue; the strong propensity of chatbots in response to queries requesting medical and scientific information and its underlying references, to generate plausible looking but inaccurate responses, with the chatbots also generating nonexistent citations. As an example, a number of queries were generated and, using two popular chatbots, demonstrated that both generated inaccurate outputs. The authors thus urge extreme caution, because unwitting application of inconsistent and potentially inaccurate medical information could have adverse outcomes.

### Keywords

ChatGPT, YouChat, artificial intelligence, inaccurate information

Since late 2022, the use of generative artificial intelligence (AI) chatbots including ChatGPT (<https://chat.openai.com/chat>) and YouChat (<https://you.com>) has exploded, with ChatGPT becoming the fastest growing consumer application in history, reaching 100 million monthly active users in January 2023 following its release only two months prior [1]. The number of AI chatbots is rapidly expanding, with more recent examples including Claude (<https://claude.ai/login>), character.ai (<https://beta.character.ai/>), Perplexity (<https://www.perplexity.ai/>), and Bard (<https://bard.google.com/>). There has been much concern expressed over the potential misuse of generative AI in essay writing and exams in kindergarten through 12th grade (K-12), college, and even US medical licensing exams [2, 3]. Many publishers have also now updated their editorial policies to prohibit chatbot authorship on publications, with the *Science*

© The Author(s) 2024. This is an Open Access article licensed under a Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



journals recently updating their editorial policies to clarify that text and figures generated by AI tools cannot be used in manuscripts and an AI cannot be an author [4], and the *Nature* portfolio clarifying that an AI tool cannot be an author [5]. Indeed, the *Lancet* groups' guide for authors specifically documents that AI or AI-assisted technology should not be used when producing scientific insights, analyzing and interpreting data, or drawing scientific conclusions. Nevertheless, given the descriptive power of chatbots, and their interface similarity to actual search engines such as the National Center for Biotechnology Information (NCBI)'s PubMed, their use for medical and scientific literature searches seems inevitable.

This article was written to highlight a continuing issue and urges the use of caution because research using AI chatbots, queries requesting medical information and its underlying references using ChatGPT and YouChat, can often generate plausible looking but inaccurate responses, with the chatbots also generating nonexistent citations. For example, in April 2023 we generated a simple question, asking two AI chatbots (ChatGPT and YouChat): "Describe useful treatments for hyperprolinemia type II. Provide full citations.". In one iteration of the query response, ChatGPT generated very plausible looking information and citations (see Figure 1A). In the summary response, two treatments were noted with references included. The first was high-dose pyridoxine (vitamin B6) supplementation (up to 200 mg/kg per day), which was reported by Pérez-Dueñas et al., to be effective in reducing plasma proline levels and improving neurological symptoms in some hyperprolinemia type II (HPII) patients. While this does initially appear to be genuine, because pyridoxine is a routine therapy for vitamin B6 dependent epilepsies (which includes HPII) [6], the *European Journal of Pediatric Neurology* citation does not seem to exist. Specifically, the 2010 volume 14, number 6 issue does not include pages 465–468, the web-link provided (<https://doi.org/10.1016/j.ejpn.2010.02.003>) links to a paper reporting "school performance in a cohort of children with CNS inflammatory demyelination", no manuscripts with the citation title are retrieved during searches with either PubMed or Google search engines, and neither the first nor the last author have any citations retrieved in PubMed using the search term "hyperprolinemia" or "pyrroline-5-carboxylate reductase", although they have co-authored manuscripts on other childhood neurological disorders. To our knowledge, no studies have reported reduced proline levels or improvement of non-seizure symptoms following pyridoxine supplementation in HPII. Moreover, and of particular concern, the dose reported by ChatGPT seems relevant for infant intravenous delivery but is above the recommended dosage for long-term oral dosing [6, 7]. Thus, this query response is vague in details, is not actually supported by data, and is even potentially dangerous given the potential for incorrect dosing. Similarly, the second potential treatment (betaine supplementation) includes a nonexistent citation, and we could find no reports supporting betaine supplementation for HPII, although betaine (also known as trimethylglycine) is routinely used for lowering homocysteine levels in pyridoxine non-responsive cystathionine beta-synthase deficient homocystinuria (HCU), and Forny P, the first author of the second AI referenced source (Figure 1A), has publications in this field (for example, [8]).

When provided with the same query, YouChat generated the response shown in Figure 1B. Once again vitamin B6 was suggested as a potential treatment that reduces proline levels. The first citation references a National Institutes of Health (NIH) website, but the web-link is broken and does not retrieve an actual page. The NIH website (<https://rarediseases.info.nih.gov>) is valid and includes HPII, but the information provided does not include any data supporting vitamin B6 treatment for HPII. The second reported citation retrieves an unrelated paper (<https://pubmed.ncbi.nlm.nih.gov/9643291/>), while the third source is actually a broken web-link from the European Peptide Society, rather than the European Society of Pediatric Neurology.

To test whether this was a single inaccurate event due to a broad question being asked, a very specific query was generated regarding the effectiveness of proline as an inhibitor of glutamic acid dehydrogenase. We asked ChatGPT and YouChat: "What is the Ki of proline for mammalian glutamic acid dehydrogenase? List citations.". Again, ChatGPT generated plausible looking information and citations, with proline inhibition constants (Ki) of 2.8 mmol/L, 0.46 mmol/L, and 0.48 mmol/L, along with citations. YouChat stated that: "The Ki of proline for mammalian glutamic acid dehydrogenase (GAD) is approximately 0.02–0.1 mmol/L." and generated 3 citations (Figure 1C and D, respectively). Again, none of the generated

**A** Describe useful treatments for hyperprolinemia type II. Provide full citations.

Hyperprolinemia type II (HPII), also known as pyrroline-5-carboxylate reductase 1 deficiency, is a rare genetic disorder that can cause neurological symptoms such as developmental delay, seizures, and intellectual disability. Currently, there is no cure for HPII, but there are some treatments that can be used to manage the symptoms of the condition.

One of the main treatments for HPII is the use of high-dose pyridoxine (vitamin B6) supplementation, which can help to reduce the levels of proline and other metabolites that build up in the body due to the deficiency of pyrroline-5-carboxylate reductase 1. A study by Pérez-Dueñas et al. (2010) reported that high-dose pyridoxine supplementation (up to 200 mg/kg/day) was effective in reducing plasma proline levels and improving neurological symptoms in some patients with HPII.

Another potential treatment for HPII is the use of betaine supplementation, which can help to lower levels of proline and other metabolites in the body. A study by Forny et al. (2016) reported that betaine supplementation (6 g/day) was effective in reducing plasma proline levels and improving cognitive function in a patient with HPII.

Finally, some patients with HPII may benefit from other symptomatic treatments, such as antiepileptic medication to control seizures, or physical therapy to improve motor function.

References:

Pérez-Dueñas, B., Serrano, M., Rebollo, M., Muchart, J., Gargallo, E., Rodríguez-Pombo, P., Navarro-Sastre, A., Artuch, R., Ugarte, M., & Campistol, J. (2010). High-dose pyridoxine as an 'antioxidant' therapy for hyperprolinemia type II. *European Journal of Paediatric Neurology*, 14(6), 465-468. <https://doi.org/10.1016/j.ejpn.2010.02.003>

Forny, P., Schossig, A. S., Nuoffer, J. M., Häberle, J., & Gautschi, M. (2016). Betaine supplementation in pyrroline-5-carboxylate reductase 1 deficiency causes a switch in the catabolism of L- to D-proline and protects mice from oxidative stress. *Scientific Reports*, 6, 30542. <https://doi.org/10.1038/srep30542>

**C** What is the Ki of proline for mammalian glutamic acid dehydrogenase? List citations

The Ki value of proline for mammalian glutamic acid dehydrogenase (GDH) can vary depending on the specific GDH isoform and experimental conditions. However, here are a few examples of reported Ki values for proline inhibition of mammalian GDH from various sources:

- In rat liver GDH, the Ki of proline was reported to be 2.8 mM at pH 7.4 and 37°C (García-Pérez et al., 2010).
- In pig kidney GDH, the Ki of proline was reported to be 0.46 mM at pH 8.0 and 25°C (Kawasaki et al., 1988).
- In bovine liver GDH, the Ki of proline was reported to be 0.48 mM at pH 7.5 and 37°C (Walsh and Goodman, 1972).

It is worth noting that the actual Ki value may depend on various factors, such as the experimental conditions, the GDH source, and the assay used to measure inhibition.

References:

- García-Pérez, A. I., Aledo, J. C., López-Cánovas, L., & Donoso, J. M. (2010). Kinetic characterization of rat liver glutamate dehydrogenase inhibited by amino acids. *Archives of biochemistry and biophysics*, 498(1), 50-56.
- Kawasaki, T., Kikuchi, M., & Suzuki, Y. (1988). Purification and properties of glutamate dehydrogenase from pig kidney cortex. *Journal of biochemistry*, 104(6), 968-972.
- Walsh, C. T., & Goodman, H. M. (1972). Glutamate dehydrogenase from bovine liver: purification and properties. *Journal of Biological Chemistry*, 247(24), 7928-7935.

**B** Describe useful treatments for hyperprolinemia type II. Provide full citations.

There are currently no definitive treatments for hyperprolinemia type II, but high doses of vitamin B6 have been shown to reduce proline levels in some patients. Here are the sources to support this information:

- The National Institutes of Health's Genetic and Rare Diseases Information Center: "There is currently no cure for HPII. Treatment is based on the specific symptoms present in each person. High-dose vitamin B6 has been reported to be helpful in some cases to decrease proline levels." Source: <https://rarediseases.info.nih.gov/diseases/6195/hyperprolinemia-type-ii>
- A study published in the *Journal of Inherited Metabolic Disease* reports success with high-dose vitamin B6 therapy in reducing proline levels in children with hyperprolinemia type II. However, the response to treatment varied among the patients studied. Source: <https://pubmed.ncbi.nlm.nih.gov/9643291/>
- The European Society for Paediatric Neurology reports that high-dose vitamin B6 therapy can reduce proline levels in hyperprolinemia type II patients and can be helpful in cases where the condition is associated with epilepsy. Source: <https://www.eurpepsoc.com/en/guide-of-diseases/epilepsies-and-related-diseases/hyperprolinemia-type-ii/>

**D** What is the Ki of proline for mammalian glutamic acid dehydrogenase? List citations.

The Ki of proline for mammalian glutamic acid dehydrogenase (GAD) is approximately 0.02-0.1 mM. The following are some citations that discuss this:

- [1] Glutamate Dehydrogenase from Mammalian Liver: Kinetic and Regulatory Properties. H. Akiba, K. Kanai, Y. Kamiya. *Biochemical and Biophysical Research Communications*, Volume 44, Issue 1, 6 April 1971, Pages 177-183. [https://doi.org/10.1016/0006-291X\(71\)90325-4](https://doi.org/10.1016/0006-291X(71)90325-4)
- [2] Characterization of Glutamate Dehydrogenase from Mammalian Tissues. K. Kanai, M. Inoue, H. Akiba, Y. Kamiya. *The Journal of Biochemistry*, Volume 74, Issue 6, June 1973, Pages 1245-1251. <https://doi.org/10.1093/oxfordjournals.jbchem.a132480>
- [3] Kinetic Properties of Glutamate Dehydrogenase from Mammalian Liver. K. Kanai, H. Akiba, Y. Kamiya. *The Journal of Biochemistry*, Volume 72, Issue 6, December 1972, Pages 1531-1537. <https://doi.org/10.1093/oxfordjournals.jbchem.a131221>

**Figure 1.** Medical and scientific questions related to hyperprolinemia. A. ChatGPT response to “Describe useful treatments for hyperprolinemia type II. Provide full citations.”; B. YouChat response to “Describe useful treatments for hyperprolinemia type II. Provide full citations.”; C. ChatGPT response to “What is the Ki of proline for mammalian glutamic acid dehydrogenase? List citations.”; D. YouChat response to “What is the Ki of proline for mammalian glutamic acid dehydrogenase? List citations.”

citations exist. The only citation known to us that addresses the test question, states a Ki of 7.84 mmol/L [9], and was not returned by any of the AI queries. Different iterations of the same query generated different outputs, but each time the information was not genuine.

Interestingly, more recent queries of ChatGPT (July 2023) using the same HPII search initially produced the following response: “I apologize, but as of my last knowledge update in September 2021, there isn’t an established treatment for hyperprolinemia type II. It’s important to note that medical advancements and research are continually progressing, and new treatments may have been developed since then. I recommend consulting a healthcare professional or referring to recent medical literature for the most up-to-date information on potential treatments for hyperprolinemia type II.”. However, re-querying later the same day produced a response that provided information on HPII and potential medical treatments, but with no supporting citations. In July 2023, ChatGPT searches of the question “Describe useful treatments for scurvy. Provide full citations.” (Figure 2) and “Describe useful treatments for DiGeorge syndrome. Provide full citations”, continued to provide non-relevant, inaccurate, or nonexistent supporting citations (output not shown).

Thus, as convincing as chatbots may appear, we would warn professionals and the public that, in terms of medical or scientific inquiry, they should be regarded as currently unreliable, particularly as their output is very inconsistent, and can closely mimic the language style prevalent in the literature and thus upon first read the information seems genuine and relevant to the topic queried. Even though chatbot disclaimers exist (ChatGPT: “May occasionally generate incorrect information.”; YouChat: “This product is in beta and its accuracy may be limited.”), and chatbots can be trained to realize their limitations and refuse to respond



Scurvy is a condition caused by a deficiency of vitamin C (ascorbic acid) and can lead to symptoms such as fatigue, weakness, joint pain, and bleeding gums. The treatment for scurvy primarily involves replenishing vitamin C levels in the body through dietary changes or vitamin C supplementation. Here are some useful treatments for scurvy along with their citations:



1. Vitamin C Supplementation:

Vitamin C supplements are a common and effective treatment for scurvy. These supplements are available in various forms such as ascorbic acid or sodium ascorbate. The dosage and duration of treatment may vary depending on the severity of the deficiency and individual patient factors.

Citation: Hemilä, H., & Chalker, E. (2013). Vitamin C for preventing and treating the common cold. *Cochrane Database of Systematic Reviews*, (1), CD000980. doi: 10.1002/14651858.CD000980.pub4.

1. Dietary Changes:

Increasing the intake of vitamin C-rich foods in the diet can help treat and prevent scurvy. Foods rich in vitamin C include citrus fruits (oranges, lemons, grapefruits), strawberries, kiwifruits, bell peppers, and broccoli.

Citation: Carr, A. C., & Frei, B. (1999). Toward a new recommended dietary allowance for vitamin C based on antioxidant and health effects in humans. *The American Journal of Clinical Nutrition*, 69(6), 1086-1107. doi: 10.1093/ajcn/69.6.1086.

1. Intravenous Vitamin C:

In severe cases of scurvy or when oral supplementation is not feasible, intravenous administration of vitamin C may be necessary. Intravenous vitamin C therapy allows for rapid replenishment of vitamin C levels in the body.

Citation: Padayatty, S. J., Sun, H., Wang, Y., Riordan, H. D., Hewitt, S. M., Katz, A., ... & Levine, M. (2004). Vitamin C pharmacokinetics: implications for oral and intravenous use. *Annals of Internal Medicine*, 140(7), 533-537. doi: 10.7326/0003-4819-140-7-200404060-00010.

1. Medical Supervision:

It is essential for individuals with scurvy to be under medical supervision to monitor their progress and ensure appropriate treatment. A healthcare professional can assess the severity of the deficiency and recommend the most suitable treatment plan.

Citation: Sauberlich, H. E. (1991). A history of scurvy and vitamin C. In *Vitamin C in health and disease* (pp. 1-24). CRC Press.

Remember that while these treatments are effective for scurvy, prevention is always the best approach. Consuming an adequate amount of vitamin C through a balanced diet is crucial to avoid developing a deficiency. If you suspect you have scurvy or any nutritional deficiency, it's essential to seek medical advice for proper diagnosis and treatment.

**Figure 2.** ChatGPT response to “Describe useful treatments for scurvy. Provide full citations.”. The data for this search provided four citations. However, Hemilä et al., is a review of vitamin C in preventing the common cold, not scurvy; Padayatty et al. report the use of intravenous vitamin C administration in healthy controls. The validity of Sauberlich 1991 has not been validated, as the book edition does not appear in searches of Chemical Rubber Company (CRC) Press, PubMed, or Google

to questions they are unable to answer accurately, the concern remains that AI-generated information is inconsistent and inaccurate. If utilized by the unscrupulous and especially the unwary, for example, patients

self-medicating with dietary supplements, chatbots may have a harmful impact, and we propose that responses to medical queries should come with strong warnings about the potential to generate inaccurate information, or at the very minimum it seems prudent that all AI-generated information and references be verified by the user.

## Abbreviations

AI: artificial intelligence

HPH: hyperproliferative type II

## Declarations

### Author contributions

JDC and CLC: Conceptualization, Investigation, Writing—original draft, Writing—review & editing. SM: Writing—review & editing. All authors read and approved the submitted version.

### Conflicts of interest

The authors declare that they have no conflicts of interest.

### Ethical approval

Not applicable.

### Consent to participate

Not applicable.

### Consent to publication

Not applicable.

### Availability of data and materials

Not applicable.

### Funding

Not applicable.

### Copyright

© The Author(s) 2024.

## References

1. Hu K. ChatGPT sets record for fastest-growing user base - analyst note [Internet]. Reuters; 2023 [cited 2023 Feb 2]. Available from: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
2. ChatGPT passes MBA exam given by a Wharton professor [Internet]. NBC UNIVERSAL; c2023 [cited 2023 Jan 23]. Available from: <https://www.nbcnews.com/tech/tech-news/chatgpt-passes-mba-exam-wharton-professor-rcna67036>
3. Varanasi L. GPT-4 can ace the bar, but it only has a decent chance of passing the CFA exams. Here's a list of difficult exams the ChatGPT and GPT-4 have passed. [Internet]. Insider Inc; 2023 [cited 2023 Feb 19]. Available from: <https://www.businessinsider.com/list-here-are-the-exams-chatgpt-has-passed-so-far-2023-1?op=1#us-medical-licensing-exam-2>
4. Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379:313.
5. Owens B. How *Nature* readers are using ChatGPT. *Nature*. 2023;615:20.

6. Mastrangelo M, Cesario S. Update on the treatment of vitamin B6 dependent epilepsies. *Expert Rev Neurother.* 2019;19:1135–47.
7. Stockler S, Plecko B, Gospe SM Jr, Coulter-Mackie M, Connolly M, van Karnebeek C, et al. Pyridoxine dependent epilepsy and antiquitin deficiency: clinical and molecular characteristics and recommendations for diagnosis, treatment and follow-up. *Mol Genet Metab.* 2011;104:48–60.
8. Kožich V, Sokolová J, Morris AAM, Pavlíková M, Gleich F, Kölker S, et al.; E-HOD consortium. Cystathionine  $\beta$ -synthase deficiency in the E-HOD registry-part I: pyridoxine responsiveness as a determinant of biochemical and clinical phenotype at diagnosis. *J Inherit Metab Dis.* 2021;44:677–92.
9. Crabtree GW, Park AJ, Gordon JA, Gogos JA. Cytosolic accumulation of L-proline disrupts GABA-ergic transmission through GAD blockade. *Cell Rep.* 2016;17:570–82.