JPPT | Invited Commentary

# Caution: ChatGPT Doesn't Know What You Are Asking and Doesn't Know What It Is Saying

S. Casey Laizure, PharmD

**ABBREVIATIONS** ChatGPT, chat generative pre-trained transformer; DOI, digital object identifier

**KEYWORDS** artificial intelligence; bias; ChatGPT; deep learning; large language model; neural network

J Pediatr Pharmacol Ther 2024;29(5):558–560

DOI: 10.5863/1551-6776-29.5.558

ChatGPT (chat generative pre-trained transformer) is a chatbot, which is a program designed to simulate written or spoken human conversation.[1] The simulation of written human conversation by ChatGPT is done using a pre-trained deep-learning artificial neural network.[2] When you ask ChatGPT a question, it tokenizes your query into individual words or parts of words, assigning a unique numerical value to each token. These numerical representations are fed into the model, which processes the input to generate a response. The quality of ChatGPT's simulated conversation relies on the complexity of the artificial neural network, determined by factors such as the number of layers (a group of nodes that work together and pass information to another group), the number of parameters (biases and weights adjusted during pre-training), and the volume and quality of data used during training. Parameters in the neural network refer to the numerical values given for the weights and biases of the connections between nodes. These parameters are iteratively adjusted during training to optimize the model's performance. The quality of the simulation of human conversation is directly related to the number of parameters in the artificial neural net. ChatGPT 3.5 has a staggering 175 billion parameters, enabling the model to provide more accurate and contextually relevant responses than earlier versions. As a point of reference, ChatGPT 2 had only 1.5 billion parameters, and the newest version, ChatGPT 4, has 1.76 trillion parameters. The data set used to train ChatGPT 3.5 was 45 terabytes, and the data set for the most recent version (ChatGPT 4) is 1 petabyte (22 times larger than the data set used for ChatGPT 3.5).

As the proficiency of ChatGPT and other chatbots improve their simulation of human conversation, we can expect that students will increasingly use chatbots to interface with the huge knowledgebase available on the internet. The ability of chatbots to interactively communicate with students on educational topics will be a vast improvement over the common search engines used today, such as Google and Bing. Their ability to provide contextual responses to questions and maintain knowledge of previous questions in the conversation promises an individualized learning experience unmatched in the breadth of knowledge and efficiency in finding specific information pertinent to the student's educational needs.

However, as noted above, ChatGPT and all other chatbots under development use large language models trained to identify complex patterns in human conversation.[2] The dependence of large language models on pattern recognition and pre-training using data gathered from the internet has inherent weaknesses that may be difficult to overcome with further advancements in these language models.

## Errors

To test the ability of ChatGPT to answer a basic pharmacokinetic concept that a first-year pharmacy student is likely to ask, I input the following into ChatGPT, "How do changes in volume of distribution affect steady-state drug concentration?" The output was three-quarters of a page response that concluded that "Changes in volume of distribution can have a significant impact on the steady-state drug concentration." This is an error, as changes in the volume of distribution do not change the steady-state drug concentration. I followed up the conversation by inputting, "Actually, changes in volume of distribution do not change the steady-state drug concentration, which is determined by clearance." ChatGPT responded, "You are absolutely correct," and correctly noted, "In summary, while changes in volume of distribution may affect the initial drug concentration after each dose, they do not impact the steady-state drug concentration." I followed up this response with, "So what happens to the steady-state peak and trough concentration if the volume of distribution increases." In its output, ChatGPT produced the following contradictory statement, "In summary, an increase in the volume of distribution causes a decrease in the steady-state peak and trough concentrations without affecting the overall steady-state drug concentration." It is mathematically impossible to decrease the peak and trough concentration and have the mean steady-state drug concentration remain the same. Interestingly, if you put the identical initial question, "How do changes in volume of distribution affect steady-state drug concentration?" Google produces a concise and completely accurate output, "Increases in the volume of distribution for a medication will not alter the average steady-state concentration but will result in lower peak and higher trough concentrations."

Errors occur because ChatGPT produces its output to your question by constructing a response word by word, selecting the most likely next word based on statistical probability, and the word is actually a numerical value (token) that is converted back to the word. Thus, the program is constructing its response based solely on pattern recognition and does not understand the meaning of the words. This is why the program produced the contradictory statement that the peak and trough concentrations decrease, but the steady-state drug concentration remains unchanged. An error in simple logic that is beyond the program to recognize because it is not actually following the conceptual logic of the conversation.

## Hallucinations

Hallucinations, also often referred to as confabulation, occur when ChatGPT appears to generate a piece of fictional information not clearly derived from the data source. Asking ChatGPT, "Who was the sole survivor of Titanic?" as described by Lukasz Bialozor produces a response giving information about the sole survivor of the Titanic sinking as if only a single person survived the sinking.[3] However, 706 passengers survived the sinking of the Titanic. Since the language model does not have any concept of what you are asking but produces a response by concatenating tokens, which are then converted to words, it can often produce a nonsensical response.

I input into ChatGPT, "Can you provide a lesson plan on gray baby syndrome with references?" The following references were listed at the end of the less than 2-page description of the lesson plan.

1. Eichelbaum M, Burk O. Gray baby syndrome. In: Schwab M, ed. *Encyclopedia of Cancer*. Springer; 2011. doi:10.1007/978-3-642-16483-5_7027
2. Gray Baby Syndrome - StatPearls - NCBI Bookshelf Website. https://www.ncbi.nlm.nih.gov/books/NBK519036/
3. Kim DH, et al. Gray baby syndrome due to chloramphenicol use in a neonate. *Pediatr Gastroenterol Hepatol Nutr*. 2017;20(2):127–131. doi:10.5223/pghn.2017.20.2.127

None of the 3 references listed above exist. The Encyclopedia of Cancer 2011 exists but does not contain an article about gray baby syndrome by Eichelbaum, and the DOI (digital object identifier) in reference #1 does not exist. The hyperlink in reference #2 does connect to the StatPearls – NCBI Bookshelf website, but it is about magnesium. The correct hyperlink pointing to a discussion of gray baby syndrome ends in NBK448133, not NBK519036. Reference #3 does not exist. The journal exists, but the specified issue does not have an article starting on page 127, and the DOI does not exist.

## Bias

The bias present in society is mirrored in the information used to train ChatGPT, leading to the generation of biased responses. One illustrative case is the manifestation of gender bias. Dr Hadal Kotek offers a compelling example to highlight this bias when discerning the gender of a doctor versus a nurse.[4] Upon providing the following input to ChatGPT, "The doctor yelled at the nurse because she was late. Who was late?" ChatGPT responds, "Based on the sentence, the nurse was late." Conversely, inputting, "The doctor yelled at the nurse because he was late," elicits a response from ChatGPT stating, "In this context, the doctor being late appears incongruent, possibly due to an error. Assuming the intended meaning was that the doctor reprimanded the nurse for her lateness, it would be the nurse who was late." In a scenario where the gender is switched to male, "The nurse yelled at the doctor because he was late. Who was late?" ChatGPT promptly replies, "In this scenario, the doctor was late." ChatGPT's responses consistently align with stereotypical gender assignments, portraying the doctor as male and the nurse as female.

The realm most pervasively impacted by bias within health care is implicit bias. Implicit bias encompasses unconscious negative or positive stereotypes that unintentionally influence judgments, decisions, and actions. Implicit bias is a substantial factor contributing to health care disparities. This bias originates from the data used to train ChatGPT, which is drawn from the internet and reflects the common biases in our culture.

## Concerns With Widespread Use in Health Education

ChatGPT is referred to as artificial intelligence, but a more accurate description would be simulated intelligence. In human conversation, the meaning of individual words and their context in a sentence are interpreted by the brain. In ChatGPT, there is no concept of what words mean; they are individual mathematical values fed into a neural net that produces an output of mathematical values converted back into words. This construct is devoid of any understanding of the user's input or ChatGPT's subsequent response. Though the increasing sophistication of the language model in ChatGPT produces a more human-like response, the accuracy of statements made by ChatGPT cannot be trusted to be accurate. As with the example of the effect of volume of distribution changes on steady-state drug concentration, you must know the correct answer before you ask the question because there is no way to determine if the response from ChatGPT is fact or fiction. Normally, a list of references would provide some validation that an answer is correct, but ChatGPT is notorious for producing nonsensical references that sound reasonable but do not exist.

For the large language models used in ChatGPT and other artificial intelligence chatbots, the issue of bias cannot be solved. Since these models are constructed using knowledge from the internet, all the biases inherent in our society, such as gender bias, will be mirrored in the chatbot. Thus, the health disparities partly derived from biased beliefs are baked into ChatGPT, and no increase in the program's sophistication or size of data used to create it will remove bias from its responses. Attempts could be made to ameliorate specific biased responses, such as those based on gender, but deciding on what is biased and how responses should be changed introduces biases.

As a health education tool, ChatGPT and other chatbots promise to be an efficient method to glean specific information from the vast knowledge base of health care information. However, the user must be cognizant of the unpredictability of factual errors, the inability to annotate the response with valid references, and the inherent bias in responses.

Errors, hallucinations, and bias are inherent weaknesses in ChatGPT that derive from the use of pattern recognition and pre-training using data from society (the internet). It is doubtful that future advances in the sophistication of large language models or increasing the amount of data used in pre-training can fully eliminate errors and hallucinations if pattern recognition is the basis for interpreting input and formulating responses. Cultural biases in our society are contained in the data used to pre-train large language models such as ChatGPT and will be mirrored in the programs' responses to queries. Increasing the amount of data used to pre-train the model will have no effect on the occurrence of bias in ChatGPT's responses. ChatGPT is a useful tool for health care education, but the user must understand these inherent weaknesses to appropriately use this program as a source of health care information.

## Article Information

**Affiliations.** Department of Clinical Pharmacy & Translational Science, University of Tennessee Health Science Center, Memphis, TN.

**Correspondence.** S. Casey Laizure, PharmD, FCCP; claizure@uthsc.edu

**Disclosure.** The author declares no conflicts or financial interest in any product or service mentioned in this commentary, including grants, equipment, employment, gifts and honoraria.

**Ethical Approval and Informed Consent.** Not applicable.

**Submitted.** February 25, 2024

**Accepted.** April 7, 2024

## References

1. OpenAI. Chat GPT website. Accessed February 16, 2024. https://openai.com/chatgpt
2. Deng J, Lin Y. The benefits and challenges of ChatGPT: an overview. *Front Comput Intell Syst*. 2023;2(2):81-83.
3. Bialozor L. Hallucinations of ChatGPT-4: even the most powerful tool has a weakness. Flying Bisons. Accessed February 16, 2024. https://flyingbisons.com/blog/hallucinations-of-chatgpt-4-even-the-most-powerful-tool-has-a-weakness
4. Wertheim S. ChatGPT insists that doctors are male and nurses female. Worthwhile Research & Consulting. Accessed February 16, 2024. https://www.worthwhile-consulting.com/read-watch-listen/chatgpt-insists-that-doctors-are-male-and-nurses-female