

Utilizing ChatGPT in clinical research related to anesthesiology: a comprehensive review of opportunities and limitations

Sang-Wook Lee and Woo-Jong Choi

Department of Anesthesiology and Pain Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

Received May 2, 2023

Revised July 15, 2023

Accepted July 17, 2023

Corresponding author

Sang-Wook Lee, M.D., Ph.D.
Department of Anesthesiology and Pain Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea
Tel: 82-2-958-1783
Fax: 82-2-958-8580
E-mail: sangwooklee@amc.seoul.kr

Chat generative pre-trained transformer (ChatGPT) is a chatbot developed by OpenAI that answers questions in a human-like manner. ChatGPT is a GPT language model that understands and responds to natural language created using a transformer, which is a new artificial neural network algorithm first introduced by Google in 2017. ChatGPT can be used to identify research topics and proofread English writing and R scripts to improve work efficiency and optimize time. Attempts to actively utilize generative artificial intelligence (AI) are expected to continue in clinical settings. However, ChatGPT still has many limitations for widespread use in clinical research, owing to AI hallucination symptoms and its training data constraints. Researchers recommend avoiding scientific writing using ChatGPT in many traditional journals because of the current lack of originality guidelines and plagiarism of content generated by ChatGPT. Further regulations and discussions on these topics are expected in the future.

Keywords: Anesthesiology; Artificial intelligence; Hallucinations; Plagiarism; Transformer.

INTRODUCTION

A large language model is an artificial intelligence (AI) model that can understand and generate natural languages. Chat generative pre-trained transformer (ChatGPT) is a large language model developed by OpenAI, a non-profit startup founded by Elon Musk and Sam Altman. Microsoft invested approximately \$3 billion USD in the development and implementation of the current model. It is now being applied to Microsoft's search engine Bing [1]. ChatGPT was trained using a large number of documents available on the Internet, as well as various books and documents from various fields. Currently, publicly available large language models include OpenAI's ChatGPT and Google Bard.

ChatGPT 3.5 is generally trained on data until June 2021.

Performance is improved through reinforcement learning, which involves receiving feedback from interactions and conversations with humans [2].

One major difference in the training of ChatGPT is that it generates information by answering user questions instead of providing information based on a keyword-focused search, unlike conventional search engines, such as Google or Yahoo. ChatGPT, also known as generative AI, generates new information based on existing information, whereas conventional search engines simply retrieve existing information. In addition, conventional search engines do not "understand" questions or keywords from users and only provide information based on matching keywords, resulting in no interaction with the user. However, ChatGPT interacts with users through conversation, and the response can vary

depending on the direction or form of the user's question, making it very different from conventional search engines (Table 1).

BERT, or Bidirectional Encoder Representations from Transformers, is a language model developed by researchers at Google AI Language in 2018. BERT was trained on many documents and understands the context of sentences input by users and provides information that can be used for natural language processing (NLP) tasks by utilizing each word. It contains approximately 300 million parameters. Additionally, Turing, an AI verification algorithm developed by Microsoft, contains 17 billion parameters. However, ChatGPT 3.5 has over 1.7 trillion parameters, which is more than ten times that of Turing. The large number of parameters in ChatGPT enables it to produce dimensionally different responses from those of previous language models. Future studies should include additional parameters.

In this review article, the transformer model on which ChatGPT is based; its use in scientific writing, anesthesiology, and clinical research; and its limitations are discussed.

TRANSFORMER

The recurrent neural network (RNN) was first introduced in 1986 and was initially developed as a tool for learning time-series data [3]. RNNs have been widely used to process time-series data [4]. An RNN has a hidden state value that exists between the input and output, and it updates the output by learning the previous hidden state and the new input value at each time step to generate a new hidden state [4]. Long short-term memory (LSTM), one of the most prominent RNN algorithms, was introduced in 1997, and its sequence-to-sequence technique has become popular in fields such as translation [5]. LSTM is a representative RNN algorithm and was the most commonly used time-series data processing technique until the emergence of the transformer algorithm [5]. LSTM is used in time-series analysis to learn patterns of changes and predict trends. It can be ap-

plied to stock or weather forecasting, and such characteristics can be used to recognize speech, predict changes in speech over time, and improve speech recognition performance. The sequence-to-sequence model is often used to examine the translation process using an RNN. This model converts the given input words into vector format numerical data, which are then fed as inputs to each LSTM layer in the encoder. The encoder reads the input data and converts them into context-formatted data, which are then converted into embedded values of the target language using the LSTM layers in the decoder. Data may be lost during this process of sequentially reading input data in a time-sequence order and compressing it into a vector format, causing a bottleneck when processing a large amount of document data as input, leading to significantly decreased performance.

The convolutional neural network (CNN) was the dominant trend in deep learning until the emergence of AlphaGo in 2016. CNNs use a feature extraction method for convolution layers, which demonstrates excellent performance in processing dimensional image data [6]. CNN development has been a revolutionary turning point in image and video data processing, and CNN algorithms have been used in numerous medical imaging data processing applications. However, the CNN algorithm has limitations when processing time-series data other than image data.

"Attention Is All You Need" is a groundbreaking paper published by Google in 2017 that introduced a new neural network architecture called the transformer model for solving various NLP problems, including machine translation [7]. This paper announced the first appearance of the algorithm upon which ChatGPT is based. The transformer model was proposed to overcome the limitations of existing models based on RNNs. Its main features and advantages include the utilization of the self-attention mechanism to capture interactions between words within a sentence, allowing it to maintain the order information of the input data while managing long-range dependencies. The transformer simultaneously processes the input data, leading to faster

Table 1. Comparison of ChatGPT and Conventional Search Engines

Feature	ChatGPT	Conventional Search Engines
Technology utilized	AI-based language model (GPT)	Keyword-based search algorithms
Information generation	Generates human-like text responses	Provides a list of web pages
User interaction	Engages in a conversational manner	Requires specific keyword queries
Query & context understanding	A better comprehension of context and complex queries	Limited to simple keyword matching

AI: artificial intelligence, GPT: generative pre-trained transformer.

training and inference speeds, in contrast to the sequential computational structure of RNNs [7]. The transformer employs multihead attention, which is its core principle, enabling it to simultaneously learn multiple perspectives of the input data, thereby capturing various contextual information [7]. The attention method assigns weights to the input and output values and learns them. This is similar to the manner in which humans learn. Humans focus on a few critical pieces of information, rather than all the information, when acquiring information to understand the overall context and meaning. The attention module followed a similar learning method by assigning weights to important information during learning. The transformer algorithm processes data in parallel because this attention module is included in both the input and output interfaces. It also incorporates position encoding to add the position information of words, allowing the model to learn order information. Furthermore, the transformer model features an encoder-decoder structure with multiple layers in each encoder and decoder, which allows the model to learn more complex patterns. The encoder and decoder blocks are structured with a self-attention layer that demonstrates the relationships between words within the input sentence, and a fully connected feed-forward layer that is equally applied to all words [7]. In addition, the decoder block has an encoder-decoder attention layer between the two layers (Fig. 1). The transformer model

has been shown to solve various NLP problems efficiently and has become the foundation for representative models such as the GPT and BERT. Almost all language models have utilized the transformer model since Google proposed it in 2017 by stacking it as building blocks in various ways, making it a fundamental technique for text data processing that involves sentence structure and changing data over time. Since Google introduced the concept of the transformer algorithm in 2017, it has become a new trend in deep learning algorithms, and language models based on it, such as GPT and BERT, have been continuously introduced [7]. Before the emergence of the transformer algorithm, RNN-based algorithms evolved from serially processing input sequences to processing them in parallel, optimized for parallel computing hardware such as graphics processing units (GPUs), which further accelerated algorithm advancement. The characteristic of a GPU compared to a central processing unit (CPU) is that it processes data in parallel. GPUs are primarily used in computers to process graphics, facilitating the quick rendering of images and animations, especially in video games. However, the use of GPUs has been extended beyond graphics. They are also used in various fields, such as scientific computing, AI training, and virtual reality, owing to their ability to process large amounts of complex data quickly. For example, in the case of AI, training on large amounts of data can be significantly accelerated using GPUs,

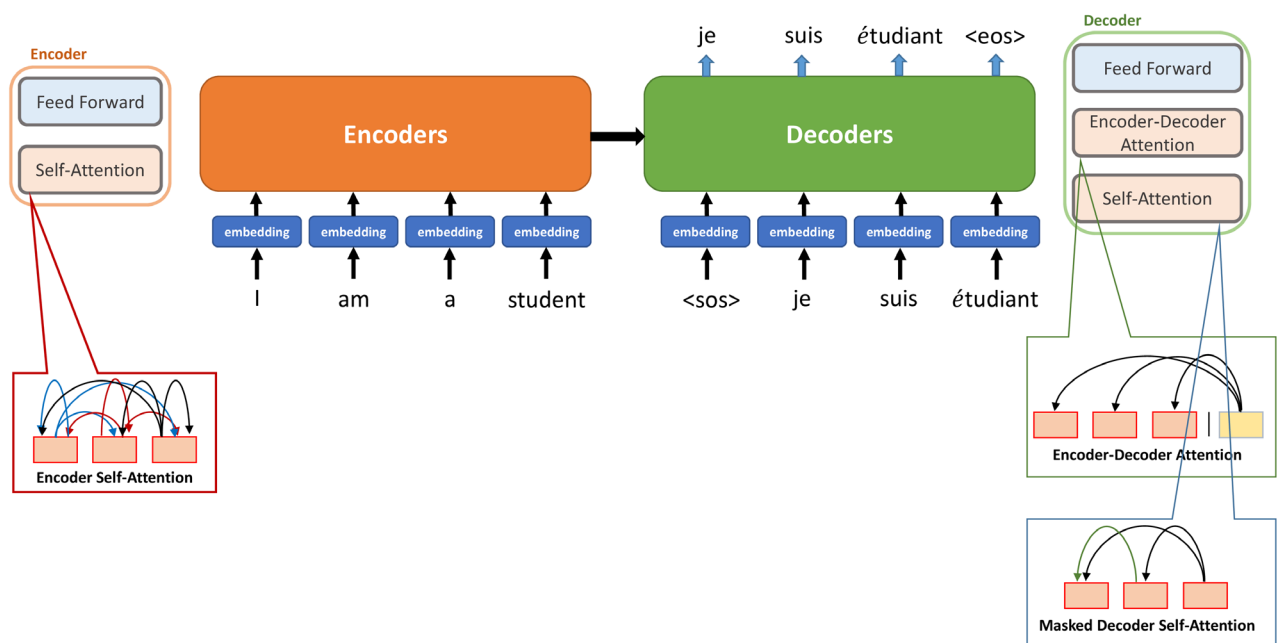


Fig. 1. Overview of the transformer architecture.

which is why many researchers have used them in AI research.

USING ChatGPT IN SCIENTIFIC WRITING

Utilizing ChatGPT, an AI chatbot, can greatly assist in writing scientific research papers through automatically generating drafts, summarizing papers, and providing translations [8-14]. In addition, it can write cover letters to send to editors, suggest titles for articles, and summarize the main content of a text to create abstracts [15]. This tool can streamline tasks and facilitate a smooth workflow for scientific writing.

ChatGPT can be used to review and summarize extensive text documents such as academic papers. Utilizing ChatGPT to summarize or condense lengthy papers can be very helpful. In addition, this feature can assist in writing the abstracts of research papers. The model can accept up to 25,000 words as input starting from GPT-4; therefore, summarizing the entire content of most papers should not be a significant challenge, considering that most papers contain approximately 5,000 words [16].

One of the most beneficial applications of ChatGPT is English proofreading. This demonstrates a high level of correction performance and explains the changes made. Hence, ChatGPT can be a useful tool for English proofreading and enhancing the quality of written content.

UTILIZING ChatGPT FOR CLINICAL RESEARCHERS RELATED TO ANESTHESIOLOGY

ChatGPT can be used to brainstorm research ideas, allowing users to discover and combine diverse subjects, ultimately leading to innovative research ideas. For example, ChatGPT can be asked to suggest relevant research topics when planning a clinical study related to robot-assisted surgery using remimazolam (Fig. 2). ChatGPT is an asset in various academic and professional settings. Furthermore, it can review and summarize extensive text documents such as academic papers, thereby streamlining the process of condensing content and writing abstracts.

ChatGPT can help create lecture materials and assist in drafting lecture outlines. For example, ChatGPT could efficiently organize the content if you were to request an outline for a lecture on anesthesia management for patients with hypertension, targeting anesthesiologists (Fig. 3). While ChatGPT can be very useful in this regard, more detailed in-

formation must be added when creating the actual lecture material.

Moreover, ChatGPT demonstrates excellent performance in writing R or Python code. Error messages are occasionally encountered when running R or Python code. Persistent errors that are difficult to resolve may occur even after multiple attempts. In such cases, ChatGPT can be a valuable resource for correcting and improving the code.

THE ROLE OF ChatGPT IN FUTURE CLINICAL FIELDS

ChatGPT not only helps with scientific writing and clinical research, but has recently been used to prepare regular discharge summaries and even directly generate medical records [17]. ChatGPT is believed to assist in automatically creating medical records by organizing patient test results generated in hospitals and documenting the patient's condition. ChatGPT may also contribute to future clinical decision-making based on this capability. However, the reliability of the information generated by ChatGPT needs to be further ensured before this becomes possible, because of the nature of medical practice dealing with lives.

LIMITATIONS OF ChatGPT AND ETHICAL CONCERNS

ChatGPT has many problems despite its differentiated performance compared to previous language models. The first is the AI hallucination phenomenon, in which the AI talks about things that are untrue as if they are true, which can make the generated information less reliable. The most critical limitation that makes it difficult to use ChatGPT in clinical research is the difficulty in finding references for the information generated by ChatGPT, and the fact that the references provided are often inaccurate. Additionally, the training data of ChatGPT 3.5 only extends up to June 2021, suggesting a lack of information on subsequent developments. However, this problem can be resolved by expanding the training dataset. Additionally, it can provide biased information based on training data. Reinforcing existing stereotypes is a risk, as evidenced by Google's initial introduction of the AI Interviews, which was later discontinued because of its tendency to be more blatantly discriminatory based on race and gender than humans. This indicates that AI can perpetuate existing discriminatory structures. Moreover, AI has no criteria for judging right or wrong when used

SA

I would like to conduct a clinical study related to remimazolam drugs in robotic surgery, so are there any topics you would like to recommend?



Remimazolam is an ultra-short-acting benzodiazepine used for sedation and anesthesia. It has potential applications in robotic surgery due to its rapid onset and offset, providing patients with a more controlled and shorter recovery period. Here are a few topics you might consider for your clinical study:

1. Comparison of remimazolam and other sedative agents: Compare the efficacy, safety, and recovery profiles of remimazolam with other commonly used sedative agents (e.g., propofol, midazolam) in robotic surgery.
2. Patient satisfaction and comfort: Assess the impact of remimazolam on patient satisfaction and comfort during robotic surgery, including anxiety, pain control, and overall experience.
3. Hemodynamic stability: Investigate the impact of remimazolam on hemodynamic stability in patients undergoing robotic surgery, focusing on blood pressure, heart rate, and other vital signs.
4. Anesthetic depth monitoring: Assess the accuracy and reliability of different anesthetic depth monitoring techniques, such as bispectral index (BIS) monitoring, in patients receiving remimazolam during robotic surgery.
5. Effects on postoperative outcomes: Examine the influence of remimazolam on postoperative outcomes, such as pain scores, postoperative nausea and vomiting, and the length of hospital stay, compared to other sedative agents.
6. Impact on surgical team performance: Investigate whether the use of remimazolam during robotic surgery affects the performance of the surgical team, including surgical time, precision, and overall surgical outcomes.
7. Remimazolam in specific patient populations: Explore the safety and efficacy of remimazolam in high-risk or special patient populations undergoing robotic surgery, such as elderly patients, patients with obesity, or patients with comorbidities.
8. Cost-effectiveness analysis: Evaluate the cost-effectiveness of using remimazolam in robotic surgery compared to other sedative agents, considering factors such as drug costs, recovery times, and patient satisfaction.

Remember to consult with experts in the field, review the existing literature, and ensure that your study design complies with ethical guidelines before commencing your research.

Fig. 2. Example of proposing a clinical study topic using ChatGPT. GPT: generative pre-trained transformer.

for criminal purposes, thereby raising ethical concerns. These are the fundamental limitations of generative AI. Human workers were employed to label answers for GPT 3.5 to calibrate the model to avoid politically sensitive responses. However, despite these efforts, ethical issues associated with AI continue to arise.

Concerning the authorship of text generated by ChatGPT, a question arises: Can ChatGPT be considered an author,

given that it created the content? Alternatively, should the user who directs the GPT to generate the text retain the copyright? This inquiry is relevant for discussing the implications of AI-generated content in scientific writing. In recent years, ChatGPT has been recognized as a contributing author to scientific research. Notably, several articles listing ChatGPT as authors have been registered and are now accessible through the Web of Science Core Collection [18-20].

SA I'd like to give a lecture on anesthesia management for hypertensive patients to anesthesiologists, so could you make a lecture about it?



Title: Anesthesia Management for Hypertensive Patients: Strategies and Challenges

Introduction:

1. Briefly introduce the topic and its relevance to anesthesiologists.
2. Explain the prevalence of hypertension and its potential complications during anesthesia.
3. Outline the objectives of the lecture: to review the pathophysiology of hypertension, discuss preoperative assessment and optimization, and explore anesthetic techniques for hypertensive patients.

I. Pathophysiology of Hypertension:

- A. Definition and classification of hypertension
- B. Factors contributing to hypertension (renal, vascular, and neural mechanisms)
- C. Complications of hypertension

1. Target organ damage (heart, brain, kidneys, and eyes)
2. Increased risk of perioperative cardiovascular events

II. Preoperative Assessment and Optimization:

- A. Thorough patient history and physical examination
- B. Laboratory and diagnostic tests (e.g., blood tests, ECG, echocardiography)
- C. Assessing blood pressure control

1. Optimal preoperative blood pressure targets
2. Antihypertensive medications: adjusting or continuing therapy
- D. Risk stratification and determining the need for further cardiac evaluation
- E. Patient education and addressing patient concerns

III. Anesthetic Management:

- A. Choice of anesthesia
 1. General anesthesia
 - a. Induction agents and their effects on hypertensive patients
 - b. Neuromuscular blocking agents
 - c. Maintenance and emergence considerations
 2. Regional anesthesia
 - a. Advantages and disadvantages
 - b. Techniques: spinal, epidural, and peripheral nerve blocks
- B. Intraoperative hemodynamic management
3. Monitoring: arterial line, central venous pressure, and other advanced monitoring techniques
4. Management of blood pressure fluctuations
 - a. Hypotension: vasopressors and fluid therapy
 - b. Hypertension: short-acting antihypertensive medications
- C. Postoperative care and pain management
5. Pain control strategies: multimodal analgesia and regional techniques
6. Monitoring for postoperative complications
7. Blood pressure management in the postoperative period

Conclusion:

1. Recap the key points of the lecture.
2. Emphasize the importance of a comprehensive approach to managing hypertensive patients undergoing anesthesia.
3. Encourage anesthesiologists to stay updated on new developments and guidelines in the field to optimize patient outcomes.

Of course, this is a suggested outline for the lecture, and you can adjust the topics and depth of content based on the audience's expertise and interests. Be sure to include relevant clinical case examples, recent research, and guidelines to illustrate your points and promote discussion among your peers.

Fig. 3. Example of developing a lecture outline using ChatGPT. GPT: generative pre-trained transformer.

However, at present, most traditional journals, along with most others, do not recognize AI systems such as ChatGPT as authors. Additionally, many traditional journals recommend that researchers refrain from using ChatGPT for scientific writing due to a lack of guidelines regarding plagiarism and originality of content generated by ChatGPT. The Editor-in-Chief of Science addressed this topic in an editorial emphasizing the importance of human authorship in scholarly publications [21]. In brief, the fundamental principle remains that the author of an academic paper should be a human, while AI systems such as ChatGPT can serve as invaluable tools for researchers to generate hypotheses, design experiments, and draw inspiration from results [21]. The ownership of the paper should belong to humans, as they are ultimately those utilizing AI systems such as ChatGPT and deriving the outcomes. ChatGPT is acknowledged as a tool; however, its authorship has not yet been recognized. The debate on AI authorship has elicited a variety of opinions. However, the editor suggests a more conservative approach to the authorship of ChatGPT, as loosening regulations is much easier than reversing them. Consequently, they suggest a high likelihood of changes in policies regarding AI authorship as technology advances.

Most journals do not recognize ChatGPT as an author. However, there seems to be no consensus among journals on obtaining ChatGPT's help in writing papers. Several journals have taken an open stance, requiring that the use of AI be mentioned in the acknowledgments or methods sections, whereas others outright prohibit its use [22,23].

CONCLUSION

In summary, the key points discussed reveal that generative AI, such as ChatGPT, is most effective when users ask precise questions because the response depends on question refinement. Notably, the original content should still be created by humans, and ChatGPT should only be used to provide an overall direction and not as a source of absolute truth. ChatGPT, developed using Google's 2017 transformer model, can be a helpful tool in various aspects of work, such as summarizing lengthy articles or assisting in paper reviews. However, its application in clinical research remains limited, and direct inclusion of the generated text in published papers is not advised. The ultimate goal of using ChatGPT is to facilitate brainstorming and idea generation, rather than to seek definitive answers. Individuals should tailor their approach when using ChatGPT to maximize their

work efficiency. AI can be beneficial for tasks such as reviewing long papers, writing R code, and proofreading English. ChatGPT is not perfect when used efficiently; however, it saves time and allows individuals to focus on more productive tasks. This makes ChatGPT a valuable tool for enhancing productivity and optimizing various aspects of work.

FUNDING

None.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

DATA AVAILABILITY STATEMENT

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

Conceptualization: Sang-Wook Lee, Woo-Jong Choi. Project administration: Sang-Wook Lee. Writing - original draft: Sang-Wook Lee. Writing - review & editing: Sang-Wook Lee, Woo-Jong Choi. Investigation: Sang-Wook Lee. Supervision: Sang-Wook Lee, Woo-Jong Choi.

ORCID

Sang-Wook Lee, <https://orcid.org/0000-0003-2967-9670>

Woo-Jong Choi, <https://orcid.org/0000-0002-5392-398X>

REFERENCES

1. Weise CMaK. Microsoft to Invest \$10 Billion in OpenAI, the Creator of ChatGPT. The New York Times [Internet]. 2023 [cited 2023 Jan 23]. Available from <https://www.nytimes.com/2023/01/23/business/microsoft-chatgpt-artificial-intelligence.html>.
2. Sheller MJ, Edwards B, Reina GA, Martin J, Pati S, Kotrotsou A, et al. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* 2020; 10: 12598.
3. Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. *Nature* 1986; 323: 533-6.

4. Sherstinsky A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Nature Physica D* 2003 doi: 10.1016/j.physd.2019.132306.
5. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9: 1735-80.
6. Valueva MV, Nagornov NM, Lyakhov PA, Valuev GV, Chervyakov NI. Application of the residue number system to reduce hardware costs of the convolutional neural network implementation. *Math Comput Simul* 2020; 177: 232-43.
7. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inf Process Syst* 2017; 30.
8. Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care* 2023; 27: 75.
9. Hutson M. Could AI help you to write your next paper? *Nature* 2022; 611: 192-3.
10. King MR. The future of AI in medicine: a perspective from a chatbot. *Ann Biomed Eng* 2023; 51: 291-5.
11. Hammad M. The impact of artificial intelligence (AI) programs on writing scientific research. *Ann Biomed Eng* 2023; 51: 459-60.
12. Hill-Yardin EL, Hutchinson MR, Laycock R, Spencer SJ. A Chat(GPT) about the future of scientific publishing. *Brain Behav Immun* 2023; 110: 152-4.
13. Gao CA HF, Markov NS, Dyer EC, Ramesh S, Luo Y, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *BioRxiv* 2022 doi: 10.1101/2022.12.23.521610.
14. Najafali D, Camacho JM, Reiche E, Galbraith L, Morrison SD, Dorafshar AH. Truth or lies? the pitfalls and limitations of ChatGPT in systematic review creation. *Aesthet Surg J* 2023; 43: NP654-NP655.
15. Else H. Abstracts written by ChatGPT fool scientists. *Nature* 2023; 613: 423.
16. Sharma U. OpenAI GPT-4: multimodal, new features, image input, how to use & more. Beebom. [Internet]. [2023 Apr 20]. Available from <https://beebom.com/openai-gpt-4-new-features-image-input-how-to-use/>.
17. Patel SB, Lam K. ChatGPT: the future of discharge summaries? *Lancet Digit Health* 2023; 5: e107-8.
18. O'Connor S. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? *Nurse Educ Pract* 2023; 66: 103537.
19. Curtis N; ChatGPT. To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. *Pediatr Infect Dis J* 2023; 42: 275.
20. Vrana J, Singh R. This Is Chatgpt; how may I help you? *Materials Evaluation* 2023; 81: 17-8.
21. Thorp HH. ChatGPT is fun, but not an author. *Science* 2023; 379: 313.
22. Flanagan A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "authors" and implications for the integrity of scientific publication and medical knowledge. *JAMA* 2023; 329: 637-9.
23. Stokel-Walker C. ChatGPT listed as author on research papers: many scientists disapprove. *Nature* 2023; 613: 620-1.