



## Review Article

# ChatGPT and large language model (LLM) chatbots: The current state of acceptability and a proposal for guidelines on utilization in academic medicine



<sup>a</sup>Division of Urology,  
Department of Surgery, The  
Hospital for Sick Children,  
Toronto, Canada

<sup>b</sup>Division of Urology,  
Department of Surgery,  
University of Toronto, Toronto,  
Canada

<sup>c</sup>Institute of Urology, St. Luke's  
Medical Center, Quezon City,  
Philippines

\* Correspondence to: Division  
of Urology, Hospital for Sick  
Children, 555 University Ave  
Toronto, ON M5G 1X8, Canada.  
[jjk.kim@mail.utoronto.ca](mailto:jjk.kim@mail.utoronto.ca)  
(J.K. Kim)

### Keywords

Large language model;  
ChatGPT; Generative pre-  
trained transformer; Artificial  
intelligence

Received 1 March 2023

Revised 14 May 2023

Accepted 27 May 2023

Available online 2 June 2023

Jin K. Kim <sup>a,b,\*</sup>, Michael Chua <sup>a,b,c</sup>, Mandy Rickard <sup>a</sup>,  
Armando Lorenzo <sup>a,b</sup>

## Summary

### Introduction

There is currently no clear consensus on the standards for using large language models such as ChatGPT in academic medicine. Hence, we performed a scoping review of available literature to understand the current state of LLM use in medicine and to provide a guideline for future utilization in academia.

### Materials and methods

A scoping review of the literature was performed through a Medline search on February 16, 2023 using a combination of keywords including artificial intelligence, machine learning, natural language processing, generative pre-trained transformer, ChatGPT, and large language model. There were no restrictions to language or date of publication. Records not pertaining to LLMs were excluded. Records pertaining to LLM ChatBots and ChatGPT were identified and evaluated separately. Among the records pertaining to LLM ChatBots and ChatGPT, those that suggest recommendations for ChatGPT use in academia were utilized to create guideline statements for ChatGPT and LLM use in academic medicine.

## Introduction

ChatGPT is a large language model (LLM) developed and recently released for public use by OpenAI (<https://chat.openai.com/>). It belongs to the GPT-3.5 series, which is the latest iteration of the Generative Pre-trained Transformer (GPT) model developed by the company [1]. This revolutionary model employs deep learning algorithms to predict the next word in a given sequence based on a large corpus of texts.

## Results

A total of 87 records were identified. 30 records were not pertaining to large language models and were excluded. 54 records underwent a full-text review for evaluation. There were 33 records related to LLM ChatBots or ChatGPT.

## Discussion

From assessing these texts, five guideline statements for LLM use was developed: (1) ChatGPT/LLM cannot be cited as an author in scientific manuscripts; (2) If use of ChatGPT/LLM are considered for use in academic work, author(s) should have at least a basic understanding of what ChatGPT/LLM is; (3) Do not use ChatGPT/LLM to produce entirety of text in manuscripts; humans must be held accountable for use of ChatGPT/LLM and contents created by ChatGPT/LLM should be meticulously verified by humans; (4) ChatGPT/LLMs may be used for editing and refining of text; (5) Any use of ChatGPT/LLM should be transparent and should be clearly outlined in scientific manuscripts and acknowledged.

## Conclusion

Future authors should remain mindful of the potential impact their academic work may have on healthcare and continue to uphold the highest ethical standards and integrity when utilizing ChatGPT/LLM.

ChatGPT, in particular, has been fine-tuned as a ChatBot and is capable of engaging in dialogues with humans and answering queries. Trained on a vast body of texts with over 300 billion words, and supplemented with human feedback, ChatGPT can generate reasonable answers to various prompts [2].

Other publicly available LLMs include Google's Language Model for Dialogue Applications (LaMDA), which has garnered attention for producing what are believed by some to be

**Table 1** Records pertaining to ChatBot LLMs (or LLMs utilized in similar capacity as a ChatBot) including ChatGPT.

Author	Type of Article	Type of LLM model	Topic evaluated	Provides support for guideline statement(s)
Discusses authorship for LLMs				
Anonymous (2023)	Editorial	ChatGPT	Authorship for LLMs including ChatGPT	1, 5
Huh S (2023)	Editorial	ChatGPT	Authorship for LLMs including ChatGPT	1, 3, 4, 5
Park SH (2023)	Editorial	ChatGPT	Authorship for LLMs including ChatGPT	1
Looi MK (2023)	News Article	ChatGPT	Authorship for LLMs including ChatGPT	1
Anonymous (2023)	Editorial	ChatGPT	Authorship for LLMs including ChatGPT	1
Yeo-Teh NSL (2023)	Letter to Editor	ChatGPT	Authorship for NLPs including ChatGPT	1
Discusses general issues surrounding LLMs in academia				
Gordjin B (2023)	Editorial	ChatGPT	General issues surrounding ChatGPT use in academia	3
Shen Y (2023)	Editorial	ChatGPT	General issues surrounding ChatGPT use in academia	3
Stokel-Walker C (2023)	News Article	ChatGPT	General issues surrounding ChatGPT use in academia	N/A
Nigel C (2023)	Opinion/Perspectives	ChatGPT	General issues surrounding ChatGPT use in academia	3
Stokel-Walker C (2023)	News Article	ChatGPT	General issues surrounding ChatGPT use in academia and suggestions for appropriate utilization of ChatGPT in academia	1, 3
Donato H (2023)	Editorial	ChatGPT	General issues surrounding ChatGPT use in academia including plagiarism	1, 3
Chatterjee J (2023)	Opinion/Perspectives	ChatGPT	General issues surrounding ChatGPT use in academia, promotes academics to play a role in navigating issues such as plagiarism with ChatGPT	2, 3
Van Dis EAM (2023)	Comment	ChatGPT	General Issues surrounding ChatGPT use in academia	2,5
Discusses LLM use in medical writing				
Kitamura F (2023)	Editorial	ChatGPT	The role of ChatGPT in medical writing and cautions surrounding ChatGPT use	1, 3
D'Amico RS (2023)	Editorial	ChatGPT	The role of ChatGPT in medical writing and cautions surrounding ChatGPT use	3
Biswas (2023)	Opinion/Perspectives	ChatGPT	The role of ChatGPT in medical writing and cautions surrounding ChatGPT use	2, 3, 4, 5
Cahan P (2023)	Editorial	ChatGPT	Use of ChatGPT for scientific writing (in computational biology)	3
Else H (2023)	News Article	ChatGPT	ChatGPT generated abstracts and peer-review process	3, 5
Stokel-Walker C (2022)	News Article	ChatGPT	ChatGPT's ability to write essays and outlook on adoption of ChatGPT technology	3
Evaluates LLM function (accuracy, use in programming, education, clinical context, etc.)				
Huh S (2023)	Original research	ChatGPT	Evaluates ChatGPT's performance on parasitology exam designed for Korean medical students	N/A
Klang E (2023)	Research Letter	ChatGPT	Evaluates ChatGPT's ability to answer three queries related to thrombosis and hemostasis	N/A
Tsigaris P (2023)	Letter to Editor	ChatGPT	Evaluates ChatGPT's ability to provide accurate estimates of surface area of lakes	N/A
Fijacko N (2023)	Letter to Editor	ChatGPT	Evaluates ChatGPT's performance for American Heart Association Basic Life Support (BLS) and Advanced Cardiovascular Life Support (ACLS) exams	N/A

(continued on next page)

Table 1 (continued)

Author	Type of Article	Type of LLM model	Topic evaluated	Provides support for guideline statement(s)
Lahat A (2023)	Letter to Editor	ChatGPT	Explores potential for LLMs to provide more accurate and efficient care, especially in underserved communities	N/A
Holzinger A (2023)	Editorial	ChatGPT and other AI	Explores potential of using artificial intelligence in biotechnology and bioinformatics	N/A
Moons P (2023)	Editorial	ChatGPT	Explores role of ChatGPT for cardiovascular nurses and allied healthcare professionals	N/A
Korngiebel DM (2021)	Comment	GPT-3	Explores the application of GPT-3 in healthcare	N/A
Castelvecchi D (2022)	News Article	ChatGPT, AlphaCode	Role of ChatGPT in programming and coding; outlook of programming professions with availability of ChatGPT	N/A
Patel SB (2023)	Comment	ChatGPT	The role of ChatGPT in automation of medical texts such as discharge summaries	N/A
Mogali SR (2023)	Letter to Editor	ChatGPT	Use of ChatGPT in anatomy education	N/A

\*ChatGPT has been listed as a co-author.

"sentient" answers [3]. As more LLMs become available, there is both excitement and concern over their use, including in academia. The potential for indiscriminate use, plagiarism and inaccuracies in generative artificial intelligence have been discussed, but there is currently no clear consensus on the standards for using these tools in academic medicine [4]. Hence, herein we present the output of a scoping review of available literature on the topic of 'LLM use in medicine' to understand the current state of LLM use in preparation of academic work to provide a framework for future utilization in manuscript preparation. In addition, we also explore the current use of LLM in clinical medicine to understand the capability of LLM beyond preparation of scientific work.

## Methods

A literature search was performed one MEDLINE on February 16, 2023 using combinations of keywords including "artificial intelligence", "machine learning", "natural language processing", "generative pre-trained transformer", "ChatGPT", and "large language model" (Search strategy in Appendix A). There were no restrictions to language or date of publication. Records not pertaining to LLMs were excluded. Records specifically addressing LLM ChatBots and ChatGPT were identified and evaluated separately. Among the records pertaining to LLM ChatBots and ChatGPT, those that suggest recommendations for ChatGPT use in academia were utilized to create guideline statements for ChatGPT and LLM use in medicine.

## Results

A total of 87 records were identified. 30 records were not pertaining to large language models and were excluded. 54 records underwent a full-text review for evaluation.

### Summary of current state of LLM chatbots in medicine

There were 33 records related to LLM ChatBots or ChatGPT. Among these, 4 were research letters or Letters to Editor evaluating ChatGPT performance and the remainder were editorials, perspectives, opinion, comments, and news articles that highlight issues, concerns, or promises of LLM in academic medicine (Table 1). Twenty-one records evaluated LLMs, not necessarily finetuned as ChatBots, and LLMs were being used in variety of areas including: molecular drug design, length of hospital stay predictions, creation of discharge summaries and clinical notes (Table 2).

### Summary guideline for ChatGPT/LLM use in academia based on available literature

1. ChatGPT or LLMs should not be cited as an author in scientific manuscripts

The academic community has reached a general consensus that ChatGPT should not be cited as an author in medical literature [4–13]. Although the introduction of large language models (LLMs) to the public is a recent development,

**Table 2** Records pertaining to use of LLM in academic research beyond the use as a ChatBot.

Li J (2023)	Journal Article	GPT-2	Evaluating clinical notes produced by GPT-2
Schimalzle R (2022)	Journal Article	GPT-2	Use of GPT-2 as a message engine in era of social media
Liang K (2021)	Journal Article	BERT-GPT, Transformer, Dialogue-GPT	Additional medical knowledge assisted mechanism for generative models may improve model outputs for medical conversational tasks
Bagal V (2021)	Journal Article	MolGPT	Drug design using LLM models
Danilov (2022)	Journal Article	Ru-GPT-3	Evaluation of a GPT-3 model trained in Russian language in anticipating length of stay in hospital for patients
Serna Garcia G (2022)	Journal Article	Genomic Metadata Integration (GeMI) based on GPT-2	Evaluation of GeMI's ability to contribute to genomics research
Drori I (2022)	Journal Article	GPT-3	Evaluates GPT-3 language model and how it can be improved to solve university-level mathematics course questions at human level
Heilbron M (2022)	Journal Article	GPT-2	Evaluates the role of human prediction in language processing through GPT-2
Lee J (2022)	Journal Article	GPT-2	Utilization of GPT-2 based Transformer model as an effort to characterize human motion
Sap M (2022)	Journal Article	GPT-3	Quantifies the narrative flow of imagined stories by GPT-3 versus autobiographical stories from humans
Luo R (2022)	Journal Article	BioGPT	Evaluate BioGPT on six biomedical natural language processing tasks
Russo AG (2022)	Journal Article	GPT-2	Explores neural activity in human listeners with deep learning via natural language processing of narrative text
Sezgin E (2022)	Journal Article	GPT-3	Evaluates the utility of GPT-3 in current US healthcare system
Sirrianni J (2022)	Journal Article	GPT-2, GPT-Neo	Evaluates the role of GPT-2 and GPT-Neo in medical text prediction
Yang X (2022)	Journal Article	Gator-Tron	Evaluates the role of Gator-Tron for clinical NLP tasks
Michelmann S (2023)	Journal Article	GPT-3	Utilization of GPT-3 to segment narrative events
Levine DM (2023)	Journal Article	GPT-3	Evaluates GPT-3's ability to appropriately diagnose and triage patients
Pivdori M (2023)	Journal Article	Manubot/GPT-3	Utilization of GPT-3 for manuscript preparation and revision of academic work
Madani A (2023)	Journal Article	ProGen	Evaluation of ProGen for generation of protein sequences and prediction of protein function
Sejnowski TJ (2023)	Journal Article	GPT-3, LaMDA	Explores language abilities in LLMs and interactions between LLMs and humans

some publications have already listed ChatGPT as an author [14,15]. To address this concern, many journals, including Science [8], Nature (and Springer Journals) [3,4], Lancet (and Elsevier Journals) [9], Journal of Education Evaluation for Health [13], and The Korean Journal of Radiology [12], have made explicit efforts to clarify that ChatGPT should not be considered an author. Currently, LLMs, including ChatGPT, lack the capacity to assume accountability for the scientific literature they contribute to, and humans must be responsible for verifying any information generated by ChatGPT and other LLMs prior to publication.

2. If use of ChatGPT or LLMs are considered for use in academic work, author(s) should have at least a basic understanding of what ChatGPT or LLM is.

Just as a surgeon needs to be knowledgeable about the tools they use in the operating room, or an internist about the drugs they prescribe, authors of academic work should be well-versed in the LLM they are using. For instance, it is important to understand that ChatGPT is a model fine-tuned from GPT-3.5, and completed training in early 2022, which means it may not have access to the latest data in science and medicine, where new trials and information are emerging at an astonishing pace. There are also clear limitations to the ChatGPT model, as outlined by OpenAI, including the possibility of incorrect or nonsensical answers, or biases in the training data [1]. Moreover, as the technology matures, and new versions or iterations of different models become available, the specific capabilities and characteristics of the employed LLM should be understood by the author.

Like traditional statistical methods, artificial intelligence models are not immune to the adage of “garbage in, garbage out”, albeit that in the case of these models, the amount of data (including inaccurate or outdated information) is substantially large. The accuracy of generated texts depends on the initial training, and potential biases, inaccuracies, or misconceptions in the data used to train the model may all contribute to erroneous or nonsensical answers. LLMs are trained on an unfiltered large body of text that does not necessarily have a “source of truth,” which means the data they use may be biased, potentially containing harmful associations including racism and sexism. Consequently, texts generated by AI can propagate such biases [6,16,17]. In light of this, BigScience Large Open-access Multilingual Language Model (BLOOM), an LLM developed by BigScience, pioneered a collaborative effort to create a model that was trained on multicultural and multilingual text which was screened by developers to reduce overrepresentation of pornography, which can lead to sexist associations in the model [18].

As the number of LLMs available for use continues to increase, it is crucial to understand the source of their training and development, as well as any potential conflicts of interest among technological giants. This knowledge will enable authors to better identify the strengths and limitations of the LLMs they encounter.

3. Do not use ChatGPT or LLMs to produce the entirety of text in manuscripts; humans must be held accountable for use of ChatGPT or LLMs and contents created by

ChatGPT or LLMs should be meticulously edited and verified by humans.

Although ChatGPT and other LLMs can generate an initial draft of text, it is crucial for human authors to use this solely as a backbone and ensure its contents are verified and edited to meet the high standards of scientific integrity required in the academic community. While these tools may expedite the writing process, relying solely on them for manuscript production raises a myriad of ethical concerns, such as intellectual property and plagiarism, as well as potential issues with scientific validity.

When ChatGPT is scrutinized in its level of knowledge and ability to answer questions in medical fields, it can perform reasonably well, including its feat of achieving similar score as a third-year medical student for a medical examination question set [19]. However, it generally is subpar to a human expert [20]. For example, it performed worse for parasitology examination compared to Korean medical students [21], did not meet passing thresholds for life support examinations [22], provides inaccurate anatomical information [23], and left out significant details when asked about certain hematology topics [24]. It also demonstrated inaccuracy in providing factual accuracy in metric conversion [25]. Moreover, its answers can involve fabrication of citations and it may not synthesize medical knowledge appropriately, possibly a result of ‘AI hallucination’ where there is no true internal understanding of the language, and LLMs produce plausible sounding text based on the training data it has been exposed to [9,10,26]. As an example in pediatric urology, specifically, when ChatGPT was asked about appropriate management for a non-palpable undescended testis, it suggested a pediatric urology referral and the high likelihood of requiring imaging such as ultrasound or magnetic resonance imaging to locate the testis prior to surgical planning. However, we know from well-established guidelines from multiple governing bodies in pediatric urology that imaging should not be utilized for evaluation of non-palpable testis, but rather examination under anesthesia and diagnostic laparoscopy (Supplementary Figure 1) [27].

There have been concerns about the factual inaccuracies of ChatGPT’s generated texts. Many, including OpenAI, have warned against relying exclusively on generative AI models, as these models are becoming increasingly popular in the public sector [28]. Some generative AI models, such as Stable Diffusion and Midjourney, are under scrutiny for their training data involving original artists and photography agencies [6]. However, there are currently no laws that specifically govern the use of AI in medical literature, and it may take some time until the legal status of using AI is established in the medical community worldwide [16]. While there have been attempts to detect AI-related plagiarism, conventional plagiarism detectors do not perform well against ChatGPT-generated texts, and they can even fool blinded human reviewers [29]. Therefore, as LLMs become more prevalent in academic medicine, it is increasingly important to maintain the accuracy and integrity of the academic field, particularly for work that will directly affect patient care.

In addition to factual inaccuracies and potential plagiarism, the same reason that authors should have a

basic understanding of what LLMs they are utilizing applies for why LLMs should not be used for generation of entirety of texts — potential for model bias and “garbage in, garbage out” exists whenever a LLM is being used. Hence, texts generated by LLMs should be scrutinized by the authors prior to distribution to the scientific community.

4. ChatGPT or LLMs may be used for editing and refining of text.

Despite the previously mentioned concerns and limitations, there are several roles for ChatGPT and LLMs in the generation, editing and refinement of academic text, with the potential to expedite the production of high-quality scholarly work. ChatGPT can revise academic manuscripts and provide stylistic improvements based on the body of text provided to it. Like previously existing and recently developed tools for manuscript editing and preparation, ChatGPT/LLMs can certainly serve as an adjunct to the workflow if the author(s) understand their capabilities and limitations [16].

While some academics argue that large commercial LLMs have no place in academic medicine due to the significant amount of energy required to train and maintain the model, the risk of incurring bias, and potential to offload the delivery of thought -which is a critical skill for academics- others believe they can be beneficial when used appropriately [5,6].

5. Any use of ChatGPT or LLMs should be transparent and should be clearly outlined in scientific manuscripts and acknowledged.

As it becomes increasingly difficult to distinguish between human-written and LLM-generated text, those who use LLMs to prepare their academic work should prioritize full transparency regarding their use of these tools [17]. This includes specifying the exact LLM used and the extent of its use, such as for text generation, stylistic formatting, and revisions. Additionally, authors may opt to include the exact outputs of the LLM in supplementary files submitted alongside the manuscript. These details should be clearly stated in the appropriate sections of the academic work, including the acknowledgements section and any other relevant areas of the text. Journals, depending on the extent of LLM that was used, may also consider requesting prompts and outputs from ChatGPT or other LLM that may have significantly influenced the contents of any submitted academic work.

## Discussion

The emergence of Language Models, including ChatGPT, is already reshaping the landscape of academic medicine. Since its release on November 30, 2022, numerous publications have evaluated the use of ChatGPT [1]. Even before the advent of ChatGPT, efforts were already underway to utilize Language Models in various aspects of science. Although the potential of ChatGPT has been extensively discussed in editorials, concerns regarding its use have also been prominent. This scoping review evaluates the current acceptance of LLMs in academic medicine and highlights

potential concerns with the use of ChatGPT and other LLMs. Based on discussions from available literature, we have identified five general recommendations that can serve as guidelines for authors who may use ChatGPT and other Language Models in the future.

It is worth noting that when asked about the current guidelines for using ChatGPT or other large language models (LLMs), ChatGPT generated an answer that admits the absence of official guideline as of September 2021. However, it does provide recommendations like those in our manuscript (Supplementary Figure 2). When prompted about the potential impact of ChatGPT on academic medicine, ChatGPT suggests that it could improve diagnosis and treatment recommendations, enhance research capabilities, and improve communication with patients (Supplementary Figure 3).

While it is still premature to rely solely on ChatGPT for these tasks given the current literature showing that ChatGPT is inferior to medical professionals for reaching a diagnosis and may struggle with relatively simple tasks like metric conversions, there is undoubtedly room for improvement in how we use ChatGPT and other LLMs. As we gain a better understanding of how these models respond to prompts, researchers will likely develop more sophisticated ways to design their prompts in the future. Moreover, as we explore the capabilities of ChatGPT and other LLMs in the future, we will more clearly understand the limitations of this technology and how it fits into our academic workflow. Undoubtedly, with the pace of advances in the field, these recommendations may need to be revised regularly. However, as with the advent of computers in the late 20th century, artificial intelligence is set to bring significant changes to how humans interact with the world, and academia must learn to navigate, coexist and utilize these powerful tools in the coming years. Efforts to block or dissuade the use of this technology are bound to be unsuccessful, and potentially detrimental.

To our knowledge, this is the first scoping review to evaluate the role and potential utility of ChatGPT and other LLMs in academic medicine. We are optimistic that our recommendations for guidelines for ChatGPT/LLM use will provide a valuable framework for future academic productivity, helping ensure that ethical and integrity standards are maintained. As more LLMs are developed in the future, we anticipate the guidelines proposed in this scoping review will become increasingly relevant for potential users of these tools.

## Conclusion

We present a summary of the current state of acceptability for ChatGPT and other LLM’s chatbots in academia, along with a concise recommendation for guideline regarding their use. In the coming years, ChatGPT and other LLM chatbots are expected to become ubiquitous tools. While they hold tremendous promise, authors should remain mindful of the potential impact their academic work may have on healthcare and continue to be vigilant and uphold the highest ethical standards and integrity.

## Funding

No funding has been secured for this study.



## Conflicts of interest

The authors do not have any conflict of interest to disclose.

## Acknowledgements

ChatGPT (OpenAI) was used for stylistic formatting and revision of this manuscript. ChatGPT was also prompted to provide answers for specific prompts regarding use of ChatGPT in academic medicine (detailed in discussion).

## References

- [1] ChatGPT. Optimizing Language models for dialogue. OpenAI. <https://openai.com/blog/chatgpt/>. [Accessed 16 February 2023].
- [2] Davey M. GPT-3: the AI language tool that could change how we talk to computers. BBC Science Focus. September 7, 2020. <https://www.sciencefocus.com/future-technology/gpt-3/>. [Accessed 16 February 2023].
- [3] Vallance C. Google engineer says Lamda AI system may have its own feelings. BBC News. June 13, 2022. <https://www.bbc.com/news/technology-61784011>. [Accessed 16 February 2023].
- [4] Anonymous. Tools such as ChatGPT threaten transparent science; here are our ground rules for their use. *Nature* 2023; 613(7945):612. <https://doi.org/10.1038/d41586-023-00191-1>. 10.1038/d41586-023-00191-1.
- [5] Stokel-Walker C. AI bot ChatGPT writes smart essays - should professors worry? *Nature* 2022. <https://doi.org/10.1038/d41586-022-04397-7>. 10.1038/d41586-022-04397-7.
- [6] Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. *Nature* 2023;614(7947):214–6. <https://doi.org/10.1038/d41586-023-00340-6>. 10.1038/d41586-023-00340-6.
- [7] Looi MK. Sixty seconds on . . . ChatGPT. *BMJ* 2023;380:205. <https://doi.org/10.1136/bmj.p205>. 10.1136/bmj.p.205.
- [8] Thorp HH. ChatGPT is fun, but not an author. *Science* 2023; 379(6630):313. <https://doi.org/10.1126/science.adg7879>. 10.1126/science.adg7879.
- [9] Liebreiz M, Schleifer R, Buadze A, Bhugra D, Smith A. Generating scholarly content with ChatGPT: ethical challenges for medical publishing. *Lancet Digit Health* 2023. [https://doi.org/10.1016/S2589-7500\(23\)00019-5](https://doi.org/10.1016/S2589-7500(23)00019-5). 10.1016/S2589-7500(23)00019-5.
- [10] Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large Language Models are double-edged swords. *Radiology* 2023;230163. <https://doi.org/10.1148/radiol.230163>.
- [11] Yeo-Teh NSL, Tang BL. Letter to editor: NLP systems such as ChatGPT cannot be listed as an author because these cannot fulfill widely adopted authorship criteria. *Account Res* 2023; 1–3. <https://doi.org/10.1080/08989621.2023.2177160>. 10.1080/08989621.2023.2177160.
- [12] Park SH. Authorship policy of the Korean journal of Radiology regarding artificial intelligence large Language Models such as ChatGPT. *Korean J Radiol* 2023. <https://doi.org/10.3348/kjr.2023.0112>. 10.3348/kjr.2023.0112.
- [13] Huh S. Issues in the 3rd year of the COVID-19 pandemic, including computer-based testing, study design, ChatGPT, journal metrics, and appreciation to reviewers. *J. educational eval. health prof.* 2023;20:5. <https://doi.org/10.3352/-jeehp.2023.20.5>. 10.3352/-jeehp.2023.20.5.
- [14] O'Connor S, ChatGPT. Open artificial intelligence platforms in nursing education: tools for academic progress or abuse? *Nurse Educ Pract* 2023 Jan;66:103537. <https://doi.org/10.1016/j.nepr.2022.103537>.
- [15] Curtis N, ChatGPT. To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. *Pediatr Infect Dis J* 2023. <https://doi.org/10.1097/INF.0000000000003852>. 10.1097/INF.0000000000003852.
- [16] Biswas S. ChatGPT and the future of medical writing. *Radiology* 2023 Feb 2:223312. <https://doi.org/10.1148/radiol.223312>. Epub ahead of print. PMID: 36728748.
- [17] van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature* 2023;614(7947):224–6. <https://doi.org/10.1038/d41586-023-00288-7>. 10.1038/d41586-023-00288-7.
- [18] Gibney E. Open-source language AI challenges big tech's models. *Nature* 2022 Jun;606(7916):850–1. <https://doi.org/10.1038/d41586-022-01705-z>. PMID: 35732717.
- [19] Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, et al. How does ChatGPT perform on the United States medical licensing examination? The implications of large Language Models for medical education and knowledge assessment. *JMIR Med Educ* 2023 Feb 8;9:e45312. <https://doi.org/10.2196/45312>. PMID: 36753318.
- [20] Binz M, Schulz E. Using cognitive psychology to understand GPT-3. *Proc Natl Acad Sci U S A* 2023;120(6):e2218523120. <https://doi.org/10.1073/pnas.2218523120>. 10.1073/pnas.2218523120.
- [21] Huh S. Are ChatGPT's knowledge and interpretation ability comparable to those of medical students in Korea for taking a parasitology examination?: a descriptive study. *J. educational eval. health prof.* 2023;20:1. <https://doi.org/10.3352/-jeehp.2023.20.1>. 10.3352/-jeehp.2023.20.1.
- [22] Fijacko N, Gosak L, Stiglic G, Picard CT, John Douma M. Can ChatGPT pass the life support exams without entering the American heart association course? *Resuscitation* 2023;109732. <https://doi.org/10.1016/j.resuscitation.2023.109732>. 10.1016/j.resuscitation.2023.109732.
- [23] Mogali SR. Initial impressions of ChatGPT for anatomy education. *Anat Sci Educ* 2023 Feb 7. <https://doi.org/10.1002/ase.2261>.
- [24] Klang E, Levy-Mendelovich S. Evaluation of OpenAI's large language model as a new tool for writing papers in the field of thrombosis and hemostasis. *J Thromb Haemostasis* 2023. <https://doi.org/10.1016/j.jth.2023.01.011>. 10.1016/j.jth.2023.01.011.
- [25] Tsigaris P, Teixeira da Silva JA. Can ChatGPT be trusted to provide reliable estimates? *Account Res* 2023. <https://doi.org/10.1080/08989621.2023.2179919>. 10.1080/08989621.2023.2179919.
- [26] Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. *ACM Comput Surv* 2022. <https://doi.org/10.1145/3571730>.
- [27] Kim JK, Chua ME, Ming JM, Santos JD, Zani-Ruttenstock E, Marson A, et al. A critical review of recent clinical practice guidelines on management of cryptorchidism. *J Pediatr Surg* 2018 Oct;53(10):2041–7. <https://doi.org/10.1016/j.jpedsurg.2017.11.050>. Epub 2017 Nov 24. PMID: 29269095.
- [28] Gordijn B, Have HT. ChatGPT: evolution or revolution? *Med Health Care Philos* 2023. <https://doi.org/10.1007/s11019-023-10136-0>. 10.1007/s11019-023-10136-0.
- [29] Gao CA, Howard FM, Nikolay S, Dyer EC, Ramesh S, Yuan L, 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. <https://doi.org/10.1101/2022.12.23.521610>.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpuro.2023.05.018>.