OPEN ACCESS
ISSN: 2835-6764
Lin AJ and Srinivasan DK (2024) 516
DOI: 10.47991/2835-6764/AJSER-163

Mini-review
American Journal of Science Education Research

ChatGPT in Anatomy Teaching and Learning: A Review

Ao Jie Lin[a] and Dinesh Kumar Srinivasan[*]

[a]NUS High School of Mathematics and Science, Singapore – 129957, Singapore
[b]Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore, Singapore – 117594, Singapore

[*]Corresponding author: Dinesh Kumar Srinivasan, Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore-117594, Singapore. Email address: dineshkumar@nus.edu.sg (D.K. Srinivasan).

Citation: Lin AJ and Srinivasan DK (2024) ChatGPT in Anatomy Teaching and Learning: A Review. American J Sci Edu Re: AJSER-163.

Received Date: 24 January, 2024; **Accepted Date:** 02 February, 2024; **Published Date:** 09 February, 2024

Abstract

ChatGPT is the first public generative artificial intelligence (AI) model pre-trained to generate human-like text responses. Owing to its comprehensive reservoir of knowledge in various fields, researchers have been exploring its potential applications in medical education, such as in the Human Anatomy discipline. This article aims to provide a summary of the existing literature on ChatGPT in anatomy education (AE) and research (AR), aggregating them to condense their discoveries in their research. In this review, we introduce the current information for the mechanism of ChatGPT revealed to the public, summarise the advantages and disadvantages of ChatGPT and attempt to evaluate its practicality in AE and AR. Further, we identify ChatGPT's potential implications in AE and AR along with areas of improvement to analyse the feasibility of deploying ChatGPT in AE and AR.

Keywords: *ChatGPT, artificial intelligence, anatomy, education, research.*

1. Introduction

The use of Artificial Intelligence (AI) in medicine has long been prevailing, dating back to as early as 1976 [1]. Yet the feasibility of AI in medical education came only with the introduction of ChatGPT, the first publicly accessible and user-friendly AI model [2]. Equipped with the Generative Pre-Trained Transformer (GPT) model, the first autoregressive generative model based on the Transformer architecture [3], the free GPT-3 excels in zero- or few-shot scenarios [4] while GPT-4 takes it one step further by taking images inputs and generating them as outputs [5]. Additionally, unlike previous AI models, which can only take in specific inputs and cannot answer in the context of the prompt, ChatGPT possesses the capability to interpret the context of the question and answer accordingly [6], producing human-like and nuanced essay outputs [7]. Thus, with the plethora of information in ChatGPT's databases, as well as its ability to use past inputs to adjust and correct itself at the individual's discretion, ChatGPT has revealed the possibility of a resourceful tool in education capable of filling gaps where human educators could not cover or replacing them in areas when students require immediate and individualized factual responses. This article thus explores the potential of ChatGPT in anatomy education (AE) and anatomy research (AR), a branch of medical education with a focus on the composition and structure of the human body. The ChatGPT versions which are in use, current literature on the applications of ChatGPT AE and AR, and advantages and limitations of ChatGPT in anatomy education, are reiterated in this review. Owing to the specificity of the research question targeting only AE, and not medical education, few publications on the topic are discoverable despite nearly a year of advent since the AI's public release on 30th January 2023 and the widespread attention ChatGPT received in the field of medical education.

2. Various ChatGPT versions

ChatGPT, which was launched on 30th November 2022, had reached 100 million users by January 2023 [8], making this platform the fastest-growing consumer application of all time. The chatbot was based on the GPT-3.5 model upon its release and updated to GPT-4 in March 2023 [9] on the paid version. The variations of GPT models, comparing their training processes, architecture, and model size are presented below.

2.1. GPT-3.5

GPT-3.5 is based on GPT-3 released in 2020, pre-trained with 45TB of human text data including books, blogs and other online content, up to September 2021 and could take up to 1.3B GPT model, fewer than the 175B GPT-3 model, with an emphasis on conforming to human ethics by detection and exclusion of hate speech and other derogatory languages and fine-tuning data collected from human inputs [11], for deep learning using a subfield of AI known as Reinforcement Learning from Human Preferences (RLHF). This enables a more human-like response with greater details and specificity A 12-billion-parameter version of GPT-3 was also used in Dall-E, an AI image generator capable of image outputs from user text prompts.

2.2. GPT-4

GPT-4, released in 2023 on the paid version of ChatGPT is updated with data up to the date of its release. However, compared to its predecessor, where all technical information has been disclosed, OpenAI has revealed little information on the training processes, dataset construction and other details

considering the growing competition in generative large-scale language models. The greatest visible difference between GPT-3 and GPT-4 lies in the modality of inputs and outputs the model could handle, with GPT-4 able to accept both text and image prompts and output likewise, parallel to the text-only GPT-3, which would prove GPT-4 advantageous in AE, given the large number of images required for the students to fully comprehend the anatomical structure and their relative positions.

2.3. Modification of ChatGPTs

Relative to GPT-3's full disclosure of its base architecture, little is known about GPT-4, though with similar architecture, with only these key differences noted in ChatGPT:

a. Accessibility

The most prominent difference out of the box between the two ChatGPT versions would be their availability. While ChatGPT-3.5 is free for consumer usage, ChatGPT-4 can only be accessed with a monthly subscription. Given that large-scale application is required for implementation in AE, most of the papers reviewed thus focus on ChatGPT-3.5.

b. Model Size

GPT-4 is 10 times more advanced than GPT-3.5. The difference indirectly signifies the amplified performance and accuracy of GPT-4, allowing GPT-4 to process more complex and nuanced language prompts.

c. Context Window Length

The difference in context window length indicates the number of previous words the chatbot refers to before producing an output. This allows the chatbot to understand the context of the conversation without reiteration. GPT-4 has a maximum window length of 24000 words, eight times higher than GPT-3.5

d. *Training Process*

While GPT-3.5 uses LRHF to simulate human-like responses, GPT-4 introduced a rule-based reward model in addition to LRHF, which increases the reliability and safety of the response, reducing hallucination cases dramatically as compared to GPT-3.5.

3. ChatGPT-based Publications in AE and AR

As ChatGPT is a nascent technology, there are few papers available on ChatGPT and AE/AR as opposed to ChatGPT and medical education, since AE is only a subset of medical education.

3.1. ChatGPT and AE

Mogali tested ChatGPT on its ability to answer questions on contextual knowledge, tutor individuals and produce MCQs [12]. The author set off open-ended questions, tasked the chatbot to simulate a question-and-answer session and to mass produce AE MCQs respectively to gauge the chatbot's proficiencies. The study discovered the chatbot was able to handle all these tasks up to a certain extent, excelling particularly in the simulation of question-and-answer sessions, providing elaborate and extensive feedback to the inputted answer. However, the chatbot was not as concise in its response to the anatomy question, with omissions and mistakes in minor parts of its response and was unable to process the prompt to generate 100 MCQs. The author concluded that ChatGPT would need further development before deploying in AE. Overall, the study highlighted that ChatGPT can currently be used only as a tool to supplement AE outside the classroom setting to provide a prompt response to any queries.

ChatGPT was pitted in an AE test consisting of 40 MCQs against students in the Faculty of Health Sciences at a state university in Turkey who took a 4-week long course on a certain anatomy topic [13]. The study showed that ChatGPT answered 67.5% of the questions correctly in a zero-shot setting, higher than 83% of the 38 students participating in the study. Furthermore, the answering techniques used by ChatGPT were significantly superior to the students, being able to provide explanations for their answers in the MCQ. This aids the researchers in gauging the proficiency of ChatGPT in the paper, pinpointing the areas of mistakes of ChatGPT for correction and modification. The authors expect ChatGPT to improve performance through deep learning and hope to conduct similar tests for different educational levels. The research performed has shown the proficiency of ChatGPT in anatomy via a side-by-side comparison with students of undergraduate level, hence determining the possibility for ChatGPT to be used in AE such education levels.

Totlis et al. (2023) randomly selected 18 questions they formulated after obtaining a subscription to ChatGPT-4 [14], and evaluated the output's accuracy, relevance, and comprehensiveness. In its responses, ChatGPT acknowledges its inability to provide consistent accurate responses and is meant to be used as a supplement to AE. Evidence of this response was displayed when omissions and mistakes in its answers to anatomical features asked were detected. The authors propose ChatGPT as a potential interactive learning tool providing organized and precise summaries of anatomical terminology in bullet points, and suggest that ChatGPT helps answering anatomical queries, but requires specific and nuanced prompts on the users' end to obtain desired results. ChatGPT was also found to be unable to provide reliable sources for its outputs and to output images crucial for anatomical studies. The study thus calls for further understanding of ChatGPT to provide guidelines for optimal utilization of ChatGPT in AE.

Chheang et al. (2023) combined ChatGPT with virtual reality (VR) in a VR-based anatomy education system [15]. In this setting, ChatGPT was integrated into a virtual assistant to generate responses to the user's questions asked not in text but through real-time speech-to-text. The added interactions of the virtual assistant were found to offer an engaging, immersive, and interactive learning experience, enhancing the learning process and motivation. Moreover, the authors state that conventional virtual assistants could only run pre-programmed scenarios with little or no adaptability, thus appearing to be rigid and may not meet individual learning needs and queries of the students. ChatGPT's ability to query a vast database of information and provide comprehensive information and resources according to the student's needs, further enhances the student's VR experience and adapts the user with personalized support. The authors also suggest further developing and

integrating visual input and output to introduce greater flexibility to amplify the already phenomenal capabilities of ChatGPT in VR AE.

Ilgaz et al. (2023) weighed ChatGPT-3.5 against Google Bard in attempting the last five years of the Medical Specialty Examination questions in both English and Turkish [16] and instructed both AI platforms to construct questions of similar, harder, and easier difficulties. Both models answered around half of the questions correctly, though they mentioned that ChatGPT-4 had a higher success rate. ChatGPT-3.5 also produced errors in its answer scheme provided for the questions it generated, but the greater capability of ChatGPT-4 may resolve such issues. The authors noted that ChatGPT has the capability to improve its accuracy and ability in answering and generating AE questions, proving it likely possible for AE in the future.

Kundakcı reviewed in his paper the advantages and limitations of ChatGPT in AE [17]. ChatGPT provides equal opportunities to students in every country and university, being accessible for free to all with an internet connection. It has the potential to increase student participation in lessons as ChatGPT can answer to a group of students simultaneously and without pause. Being an AI model, it boasts 24/7 availability and students would feel more comfortable making mistakes as they are not conversing with humans. Lastly, ChatGPT can produce innovative and practical solutions if students provide information from an accurate and reliable database. The article hence emphasizes the benefits offered by ChatGPT and urges further development into it.

3.2. ChatGPT and AR

Ilgaz et al. (2023) weighed ChatGPT-3.5 against Google Bard in producing an academic paper on the anatomy of the facial nerve [16]. ChatGPT produced a basic and less detailed article without any referencing, compared to previous papers on such subjects. However, the authors noted that ChatGPT's article structure was more in line with academic norms on such subjects, with sections divided clearly and accurately.

Kundakcı scrutinized the deployment of ChatGPT in AR by its credibility and ability to answer to comments posed [17]. The author warned that ChatGPT, with all its learning models, does not have the conscientious conviction of a judge or the capability of a doctor caring for the patients. It would also be unable to respond to letters to the referee comments if it was a co-author for a paper. The author concludes therefore ChatGPT can only be used as a complementary tool with clear awareness of its risks, including the lack of credibility and lack of empathy.

4. Advantages and Limitations of ChatGPT and AE/AR

To evaluate ChatGPT's potential in AE, researchers first pitted it in medical exams against students [13, 16] and tested its reservoir of anatomical knowledge [12]. ChatGPT showed impressive capabilities in passing the exam with higher-than-average results in zero-shot [13, 16] and could provide elaborate responses to anatomical feature questions [12]. Another aspect scrutinized was ChatGPT's ability to host interactive learning sessions with students to provide a personalized learning experience [12, 15]. ChatGPT displayed highly human-like behavior in its outputs in a tutor-student simulation [12] and utilization of ChatGPT's generative language models in virtual assistants provided highly immersive and flexible study sessions [15]. Lastly, ChatGPT's ability to mass produce AE questions to lessen the workload of human educators was put to the test. It has been reported that ChatGPT churns out viable AE questions with comprehensible working and solutions and was able to alter the difficulty of the questions when prompted [16]. However, ChatGPT possesses stark limitations in AE hindering it from being deployed. Its contextual knowledge in anatomy had many generalizations and false information with risks of spreading misinformation to students [12, 14, 16]. This is particularly detrimental in AE as medical misinformation may produce dire consequences when students graduate and become practitioners of the misinformation if undiscovered. ChatGPT can also never fully replicate the characteristics of human educators in emotional and motivational support for the students [12]. It is also currently unfeasible to mass produce AE questions due to the task's resource intensiveness and may include questions produced. Ultimately, the flaws in ChatGPT require extensive fine-tuning by professionals to ensure safe usage [17].

The use of ChatGPT in AR was assessed through its ability to generate academic papers [16] and an evaluation of the feasibility of ChatGPT as an author in publicized papers [17]. However, contrary to the promising capabilities ChatGPT had for AE, its usefulness is lacklustre at best. The papers it generated for an academic topic were basic and lacked any form of references to external sources to support its arguments [16]. When asked for literature on anatomical subjects, the literature it provided was also undiscoverable [12]. The lack of credibility of ChatGPT's information and its inability to access online unless one pays a premium for its monthly subscription hence prevents ChatGPT from being a reliable source of information or producing any academic literature [9]. In addition, ChatGPT is unable to respond to any questions directed to it in its literature due to its offline nature. Hence, unlike in AE, ChatGPT in AR still has a long way to go before it may be practical.

5. Conclusion

As the first consumer-accessible generative AI model capable of responding to inputs in a human-like manner, ChatGPT showed much promise as papers with extensive research on its use in various fields exploded with its release. This review thus provides insight into the ongoing study of ChatGPT's usage in AE and AR, specific fields with little literature compared to the more generalized medical education and research given the short timespan from its release. Overall, this paper serves to provide an informative guide to educators and professionals in anatomy seeking to add ChatGPT to their reservoir of teaching resources, or to students considering making use of ChatGPT to aid their learning.

Acknowledgement
We would like to thank Professor Bay Boon Huat at the Department of Anatomy, Yong Loo Lin School of Medicine, National University of Singapore for commenting on earlier version of this paper. The authors would like to acknowledge financial support for the publication of this paper from the Research Department at NUS High School of Mathematics and Science, Singapore

Conflict of Interest: The authors declare no conflicts of interest.

Author contributions: Conceptualization, D.K.S.; writing—original draft preparation, A.J.L.; writing—review and editing, D.K.S.; supervision, D.K.S. Both authors have read and agreed to the published version of the manuscript.

Funding support: Not Available.

References

1. Weiss S, Kulikowski CA, Safir A. (1978). Glaucoma consultation by computer. Comput Biol Med; 8(1):25-40. https://doi.org/10.1016/0010-4825(78)90011-2.
2. How ChatGPT managed to grow faster than TikTok or Instagram. TIME Blog. [Accessed Jan 21, 2024]. Available online: https://time.com/6253615/chatgpt-fastest-growing/.
3. Radford A, Narasimhan K, Salimans T, and Sutskever I. (2018) Improving language understanding with unsupervised learning. Technical Report, OpenAI Blog. [Accessed Jan 21, 2024]. Available online https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf
4. Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, et al. (2020). Language models are few-shot learners. 2020. Advances in Neural Information Processing Systems 33. [Accessed Jan 21, 2024]. Available online: https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf
5. Image inputs for ChatGPT. OpenAI Blog. [Accessed Jan 21, 2024]. Available online: https://help.openai.com/en/articles/8400551-image-inputs-for-chatgpt-faq
6. Teo S. (2022). ChatGPT: The end of online exam integrity? [Accessed Jan 21, 2024]. Available online: https://arxiv.org/abs/2212.09292.
7. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2018). Language models are unsupervised multitask learners. [Accessed Jan 21, 2024]. Available online: https://life-extension.github.io/2020/05/27/GPT%E6%8A%80%E6%9C%AF%E5%88%9D%E6%8E%A2/language-models.pdf
8. Shewale R (2024). ChatGPT statistics - user demographics. Demandsage Blog. [Accessed Jan 21, 2024]. Available at: https://www.demandsage.com/chatgpt-statistics/.
9. Natalie (2023). ChatGPT - release notes. OpenAI Blog. [Accessed Jan 21, 2024]. Available at: https://help.openai.com/en/articles/6825453-chatgpt-release-notes.
10. GPT-3 vs GPT-3.5: Key Differences and Applications. Iffort.Blog. [Accessed Jan 21, 2024]. Available at: https://www.iffort.com/blog/2023/03/31/gpt-3-vs-gpt-3-5/
11. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL et al. (2022). Training language models to follow instructions with human feedback. arXiv. [Accessed Jan 21, 2024]. Available at: https://arxiv.org/pdf/2203.02155.pdf
12. Mogali SR. Initial impressions of ChatGPT for anatomy education (2023). Anat Sci Educ. https://doi.org/10.1002/ase.2261.
13. Talan T and Kalinkara Y (2023). The role of artificial intelligence in higher education: ChatGPT assessment for anatomy course. International Journal of Management Information Systems and Computer Science, 7(1): 33-40. Available at: https://dergipark.org.tr/en/download/article-file/2927015.
14. Totlis T, Natsis K, Filos D, Ediaroglou V, Mantzou N et al (2023). The potential role of ChatGPT and artificial intelligence in anatomy education: a conversation with ChatGPT. Surg Radiol Anat 45, 1321–1329 (2023). https://doi.org/10.1007/s00276-023-03229-1.
15. Chheang V, Márquez-Hernández R, Patel M, Rajasekaran D, Sharmin S, et al. (2023). Towards anatomy education with generative AI-based virtual assistants in immersive virtual reality environments. arXiv. [Accessed Jan 21, 2024]. Available at: https://arxiv.org/pdf/2306.17278.pdf.
16. Ilgaz HB, and Çelik Z (2023) The significance of artificial intelligence platforms in anatomy education: An experience with ChatGPT and Google Bard. Cureus 15(9): e45301. https://doi.org/10.7759/cureus.45301.
17. Kundakcı, YE (2023). ChatGPT's capabilities for use in anatomy education and anatomy research. European Journal of Therapeutics. https://doi.org/10.58600/eurjther1842.