Cardiology in the Young

cambridge.org/cty

Precision of artificial intelligence in paediatric cardiology multimodal image interpretation

Michael N. Gritti[1,2,3], Rahil Prajapati[3], Dolev Yissar[1] and Conall T. Morgan[1,2,3]

[1]Division of Cardiology, The Labatt Family Heart Centre, The Hospital for Sick Children, Toronto, Ontario, Canada; [2]Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada and [3]Temerty Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada

Original Article

Cite this article: Gritti MN, Prajapati R, Yissar D, and Morgan CT (2024). Precision of artificial intelligence in paediatric cardiology multimodal image interpretation. *Cardiology in the Young*, page 1 of 6. doi: 10.1017/S1047951124036035

Received: 12 June 2024
Revised: 18 September 2024
Accepted: 13 October 2024

Keywords:
Artificial intelligence; image interpretation; paediatric cardiology; chatGPT

Corresponding author:
Michael Gritti; Email: michael.gritti@sickkids.ca

Michael N. Gritti and Rahil Prajapati are co-first authorship.

Abstract

Multimodal imaging is crucial for diagnosis and treatment in paediatric cardiology. However, the proficiency of artificial intelligence chatbots, like ChatGPT-4, in interpreting these images has not been assessed. This cross-sectional study evaluates the precision of ChatGPT-4 in interpreting multimodal images for paediatric cardiology knowledge assessment, including echocardiograms, angiograms, X-rays, and electrocardiograms. One hundred multiple-choice questions with accompanying images from the textbook *Pediatric Cardiology Board Review* were randomly selected. The chatbot was prompted to answer these questions with and without the accompanying images. Statistical analysis was done using X^2, Fisher's exact, and McNemar tests. Results showed that ChatGPT-4 answered 41% of questions with images correctly, performing best on those with electrocardiograms (54%) and worst on those with angiograms (29%). Without the images, ChatGPT-4's performance was similar at 37% (difference = 4%, 95% confidence interval (CI) −9.4% to 17.2%, $p = 0.56$). The chatbot performed significantly better when provided the image of an electrocardiogram than without (difference = 18, 95% CI 4.0% to 31.9%, $p < 0.04$). In cases of incorrect answers, ChatGPT-4 was more inconsistent with an image than without (difference = 21%, 95% CI 3.5% to 36.9%, $p < 0.02$). In conclusion, ChatGPT-4 performed poorly in answering image-based multiple-choice questions in paediatric cardiology. Its accuracy in answering questions with images was similar to without, indicating limited multimodal image interpretation capabilities. Substantial training is required before clinical integration can be considered. Further research is needed to assess the clinical reasoning skills and progression of ChatGPT in paediatric cardiology for clinical and academic utility.

Introduction

The integration of artificial intelligence into our daily lives has marked a pivotal era of technological advancement. The development of artificial intelligence large language models has allowed for understanding of context, reason, and ultimately generating realistic conversation.[1] Large language model-based artificial intelligence assistants like Apple's Siri and Google's assistant have greatly improved our daily quality of life by helping us perform defined tasks in our daily lives.[2] Now, the introduction of ChatGPT-4, a novel large language model artificial intelligence released in 2023, has enhanced user interactions, accelerated workflows, and driven global innovation.[3]

Artificial intelligence chatbots, like ChatGPT-4, have shown great strides in medicine in a short period of time.[2] From performing literature searches and designing methodologies[2,4,5] to data analysis and article writing[6,7], researchers have found great success in utilising artificial intelligence as a tool in their research efforts. Some journals even published ChatGPT as an author or acknowledgement in their manuscripts.[8] These chatbots have not fallen short in clinical applications either, as they have been utilised to write medical notes[9,10], detect drug interactions[9], identify high-risk patients[9], overcome language barriers[10,11], and aid in patient education.[4,10,12] They have also shown success in passing the USMLE[13] and European Exam in Core Cardiology.[14] Recently, ChatGPT-4 has been used to interpret multimodal images in radiology and ophthalmology with some success.[4,10,15] These findings are promising as the use of artificial intelligence image interpretation can augment diagnostic accuracy and support clinicians in clinical decision-making.[7]

Our group's previous study found that ChatGPT's performance in text-based paediatric cardiology educational knowledge assessment is quickly advancing.[16] However, to our knowledge, the chatbot's proficiency in interpreting imaging in paediatric cardiology has not yet been assessed. Multimodal imaging, such as electrocardiogram, echocardiogram, angiogram, and X-ray, holds immense value in combination with clinical findings to allow for more accurate diagnoses and targeted interventions.[17] This study aims to evaluate the performance of

ChatGPT-4 in multimodal imaging interpretation in paediatric cardiology through single-best answer testing and compare its performance between different imaging modalities.

Methods

We used a dataset of image-based multiple-choice questions from Pediatric Cardiology Board Review by Eidem[18], a textbook resource for Paediatric Cardiology board certification examination preparation. Copyright permissions were obtained from the publisher to test artificial intelligence chatbots' ability to answer up to 100 questions. The default mode of ChatGPT-4, accessed through ChatGPT Plus, was utilised due to its ability to interpret multimodal imaging. Questions with accompanying images were first extracted and were screened by an independent reviewer for the exclusion criteria. Those that included multiple-choice answers in the image itself, as well as questions that were not specifically related to paediatric cardiology, for example, statistics, were excluded. The remaining dataset was further refined through random selection of 100 questions, as per the copyright agreement. The final dataset included questions from the following 10 paediatric cardiology topics: Cardiac Anatomy and Physiology, Congenital Cardiac Malformations, Diagnosis of Congenital Heart Disease, Cardiac Catheterization and Angiography, Non-invasive Cardiac Imaging, Electrophysiology Questions for Paediatrics, Cardiac Intensive Care and Heart Failure, Cardiac Pharmacology, and Surgical Palliation and Repair of Congenital Heart Disease. The accompanying images also varied, including echocardiograms, angiograms, X-rays, electrocardiograms, tables, and graphs. This study adhered to Strengthening the Reporting of Observational Studies in Epidemiology guidelines.

A new ChatGPT Plus account was utilised to ensure conversation history prior to the study's initiation did not affect the chatbot's answers. All questions and images were inputted in ChatGPT-4 exactly as presented in the textbook, without any alterations or preprocessing, from March 13, 2024 to March 25, 2024. Each image, with the same file name as described in the question (i.e. Figure 1), was attached to its corresponding question prompt through the attachment function on ChatGPT-4. Five answer choices were provided in each question, exactly as described in the textbook. Each question, along with its accompanying image, was entered as a separate new prompt, and previous dialogues were cleared to prevent any prior information from influencing the chatbot's responses. For reliability in the answer choice, this was repeated two more times per question, for a total of 3 samples per question. The chatbot was also given the same 100-question test without accompanying images to test for differences in responses. Responses were subsequently reviewed by members of our team to confirm the chatbot's accuracy in addressing the question. If ChatGPT-4 arrived at the correct answer across all three repeated inputs, the answer was scored as correct. Conversely, if the chatbot did not consistently arrive at the correct answer across the three repeated inputs, the answer was scored as incorrect. If ChatGPT 4.0 deemed that none, multiple, or all the answers were correct, when this was not one of the multiple-choice options, it was scored as incorrect. Responses were validated against the textbook's answer key. The chatbot's accuracy was reported as a proportion of correct responses, categorised by chapter or image type.

The primary outcome of this study was the accuracy of ChatGPT-4 in answering image-based multiple-choice questions, measured as a proportion of correct answers. Secondary outcomes included differences in accuracy between image types, paediatric cardiology topic, and when an image was provided compared to when one was not provided. We also assessed the inconsistency of incorrectly answered questions as a proportion of questions with varying answers across trials.

Various statistical tools were utilised for data analysis. X^2 was used to compare overall proportions of correct responses between questions with images and without images. Fisher's exact test was used to compare proportions of correct responses between groups, which had too small of a sample size to use X^2, such as when comparing groups stratified by chapter or image type. McNemar's test was used to complete a pairwise comparison of responses to questions when provided an image and when not provided an image. Statistical analyses were completed with an alpha value of 0.05, 95% confidence intervals, and two-tailed *p*-values.

Table 1. Number of correctly answered questions by ChatGPT-4 when provided the accompanying image, stratified by image type

Image type	Total Questions, n	Total Questions Correct, n (%)
Echocardiogram	22	11 (50)
Angiogram	17	5 (29)
X-ray	6	2 (33)
Electrocardiogram	39	21 (54)
Table	11	1 (9)
Graph	5	1 (20)
Total	100	41 (41)

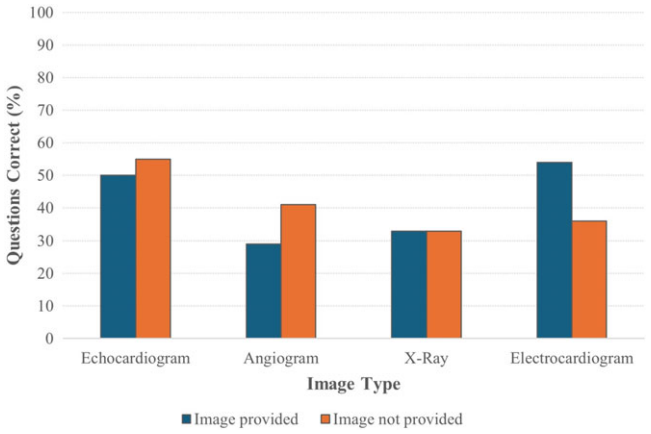

Figure 1. Proportion of questions correct by ChatGPT-4 with and without providing the accompanying image, stratified by image type of multimodal imaging typically performed in paediatric cardiology.

Results

ChatGPT-4 accuracy on questions with multimodal imaging

ChatGPT-4 was used to answer 100 multiple-choice questions with accompanying images from the *Paediatric Cardiology Board Review* textbook.[18] The chatbot answered 41 questions correctly (41%). Table 1 outlines questions correctly answered, sorted by image type.

Of questions with typical diagnostic imaging done in paediatric cardiology, such as an echocardiogram, angiogram, X-ray, and

https://doi.org/10.1017/S1047951124036035 Published online by Cambridge University Press

Table 2. Number of correctly answered questions by ChatGPT-4 with and without providing the accompanying image, stratified by chapter of the *Pediatric Cardiology Board Review* book[18]

Chapter topic	Total Questions, n	Total Questions with Image Correct, n (%)	Total Questions without Image Correct, n (%)
Cardiac Anatomy and Physiology	2	1 (50)	1 (50)
Congenital Cardiac Malformations	16	6 (38)	8 (50)
Diagnosis of Congenital Heart Disease	4	3 (75)	3 (75)
Cardiac Catheterization and Angiography	18	3 (17)	3 (17)
Non-invasive Cardiac Imaging	10	6 (60)	6 (60)
Electrophysiology	37	19 (51)	12 (32)
Outpatient Cardiology	1	0 (0)	0 (0)
Cardiac Intensive Care and Heart Failure	3	0 (0)	0 (0)
Cardiac Pharmacology	3	3 (100)	3 (100)
Surgical Palliation and Repair of Congenital Heart Disease	6	0 (0)	1 (17)
Total	100	41 (41)	37 (37)

electrocardiogram, 46% (39/84) were correctly answered. The chatbot performed best on questions with an electrocardiogram, correctly answering 54% (21/39) of questions, and poorest on questions with an angiogram, correctly answering 29% (5/17) of questions. Questions with a table or graph were typically answered poorly, with the chatbot achieving 9% (1/10) and 20% (1/5) correctly answered questions, respectively. When completing statistical analysis, no significant difference was found between groups.

When breaking down questions by chapter, ChatGPT-4 performed worst with questions on cardiac catheterization and angiography, only answering 17% (3/18) of questions correctly. This was significantly worse than its performance with questions on diagnosis of CHD (difference = 58%, 95% CI 12.5% to 100%, $p < 0.05$), non-invasive cardiac imaging (difference = 43%, 95% CI 8.4% to 78.2%, $p < 0.04$), and electrophysiology (difference = 35%, 95% CI 11.1% to 58.2%, $p < 0.02$), where the chatbot correctly answered 75% (3/4), 60% (6/10), and 51% (19/37) of questions, respectively. A complete breakdown can be found in Table 2.

Accuracy on questions without providing the accompanying image

ChatGPT-4 was also given the same 100 multiple-choice question test without the accompanying images to test for differences in responses. The chatbot answered 37 questions correctly (37%), which was not significantly different from when it was given the images correctly (difference = 4, 95% CI −9.4% to 17.2%, $p = 0.56$). Among questions with typical diagnostic imaging done in paediatric cardiology, performance was best when answering those with an echocardiogram, correctly answering 55% (12/22) of questions, and worst among those with an X-ray, correctly answering 33% (2/6) of questions. A complete breakdown can be found in Figure 1. There was no significant difference when completing statistical analysis comparing questions answered with and without images, stratified by image type.

Pairwise comparison

When completing a pairwise analysis of all questions answered, with the pairs being questions with and without accompanying images, no significant differences were found. However, a pairwise analysis with stratification by image type found that ChatGPT-4 performed significantly better when given the image of an electrocardiogram than without (difference = 18%, 95% CI 4.0% to 31.9%, $p < 0.04$).

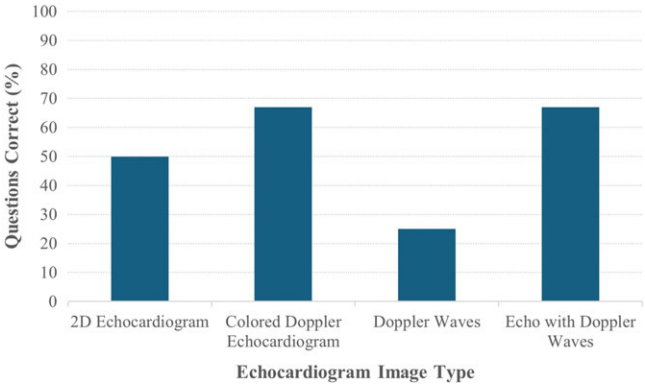

Figure 2. Proportion of questions correct by ChatGPT-4 when the provided the accompanying echocardiogram, stratified by echocardiogram image type.

Further stratifying the echocardiogram group

The echocardiogram group was further broken down by specific imaging modality and it was found that, when provided the image, ChatGPT-4 correctly answered 50% (6/12) of questions with a 2D echocardiogram, 67% (2/3) with a coloured doppler echocardiogram, 67% (2/3) with an echocardiogram with doppler waves, and 25% with doppler waves alone (Figure 2). When an image was not provided, the chatbot answered one more question correctly (3/3) in the echocardiogram with Doppler waves group.

Variation in responses

ChatGPT-4 often provided varied and inconsistent answers when the exact same question was prompted multiple times. Among the incorrectly answered questions, the chatbot offered significantly more inconsistent answers when an image was provided (53%)

https://doi.org/10.1017/S1047951124036035 Published online by Cambridge University Press

than when an image was not provided (difference = 21%, 95% CI 3.5% to 36.9%, $p < 0.02$).

Discussion

Overall findings

Image interpretation is a novel capability of artificial intelligence chatbots such as ChatGPT-4 that has yet to be explored in the context of paediatric cardiology. Our study found that ChatGPT-4 performed poorly in responding to image-based multiple-choice questions from a paediatric cardiology textbook, with an accuracy of 41% in the overall sample. Among imaging typically performed in the field, the chatbot performed best on questions with an electrocardiogram, and worst on those with an angiogram. The textbook from which questions for this study were extracted is typically used to prepare for the paediatric cardiology board examination in our country, which has a pass rate of 70%. Based on our findings, ChatGPT-4 would not pass this examination. By contrast, the chatbot has passed other image-based examinations such as the United States Medical Licensing Exam Step 1 and Step 2,[19] and American Heart Association Advanced Cardiac Life Support and Basic Life Support exams.[20,21] We believe this highlights the difficulty ChatGPT will have when dealing with increasingly complex medical problems.

Comparison to when images were not presented with the question

In comparison, when we presented the same questions without their accompanying images, the chatbot answered 37% of the questions correctly. Based on this comparison, it seems that the chatbot is determining its answer primarily based on the text rather than an interpretation of the image in combination with the provided text. This suggests that the chatbot may not be able to accurately interpret multimodal images and/or utilise its interpretation to arrive at logical conclusions in paediatric cardiology knowledge assessment. The chatbot's inability to consistently choose a single answer when prompted with the same question with its accompanying image multiple times further supports this point. Similar inconsistencies have been reported in the literature and are seen as a threat to the integration of artificial intelligence chatbots in clinical medicine.[22,23]

Interestingly, questions with electrocardiograms were more likely to be interpreted correctly by ChatGPT-4 when provided the image than not. One reason for this may be the high prevalence of electrocardiogram interpretation artificial intelligence models preceding the release of ChatGPT-4.[24] These are robust artificial intelligence that have been utilised and improved since the mid-1990s such that they can detect pathology with high accuracy. It is possible that ChatGPT-4 may have been trained on publicly available data from these artificial intelligence, thus allowing it to better interpret electrocardiograms. Electrocardiograms are also generally standardised thus making pattern recognition – the basis of machine learning algorithms – easier for artificial intelligence than echocardiograms, angiograms, or X-rays, which can vary due to anatomical variation and probe placement.[22]

Comparison to previous study

In a previous investigation of 88 text-based multiple-choice questions from the same textbook utilised in this study, we found that ChatGPT-4 correctly answered 66% of questions.[16] Compared to this prior investigation and other similar studies examining text-based questions,[14,25,26] the chatbot's performance on image-based questions in paediatric cardiology appears inferior. In fact, ChatGPT-4's performance with image-based questions was similar to that of ChatGPT-3.5, an older version of ChatGPT, which correctly answered 38% of paediatric cardiology-related text-based questions.[16] Given the chatbot's novel ability to provide answers to image-based questions, it is expected that with future versions of ChatGPT, a similar improvement in performance seen from ChatGPT-3.5 to ChatGPT-4 in text-based questions will be seen for image-based questions.

Comparison to other medical fields

ChatGPT-4's performance in clinical image analysis varies substantially in different medical specialties. Its performance in paediatric cardiology is similar to that in dermatology, where it was reported to be 36% accurate.[27] However, the chatbot is more accurate with other topics such as neuroradiology,[28] ophthalmology,[15] and pathology[29] which report a 50%, 65%, and 100% accuracy in image interpretation, respectively. Therefore, ChatGPT-4 seems to perform poorly at interpreting findings from paediatric cardiology imaging in comparison to most other medical specialties. In the ophthalmology study, they found that the chatbot performed poorer on topics like paediatric ophthalmology and neuro-ophthalmology.[15] In conjunction with our findings, this suggests that ChatGPT-4 may currently have limited image interpretation capacity in niche and highly subspecialized fields. This is further supported by a recent study which utilised ChatGPT's DALL · E 3 to illustrate CHDs with minimal success.[30] One explanation for this may be that niche subspecialties are underrepresented in literature, providing less publicly available data for artificial intelligence chatbots to train on, thus resulting in poorer performance. Furthermore, there are numerous imaging modalities utilised in paediatric cardiology, hence requiring an extensive database of images to be trained on. Based on our results, it can be hypothesised that the artificial intelligence model has not been trained on a sufficient database to correctly interpret all the imaging modalities. For ChatGPT to be clinically and academically useful in niche subspecialties like paediatric cardiology, it needs further training on a robust database.

Additionally, when clinicians come across a novel problem they have not previously encountered – that is to say, it is not in their 'database' of knowledge – they search for more information through numerous means such as academic literature and clinical guidelines before arriving at conclusions. ChatGPT-4 does not yet have this capability of self-identifying knowledge gaps, and instead tends to offer inaccurate but seemingly plausible explanations for its incorrect answers, a phenomenon common to artificial intelligence chatbots known as "hallucination."[9] This poses a threat to the clinical and academic integration of artificial intelligence chatbots as it requires the user to have sufficient knowledge and experience to differentiate between fact and fiction.[2,9] Therefore, its practical utility in settings where accuracy is crucial, such as a clinical tool, is currently unclear. A hope for future artificial intelligence chatbots is a feature that allows access to search the internet for relevant information to supplement its decision-making, much as a real clinician would. Although this relies on the gathering and interpretation of accurate and reliable information which poses another barrier, it would be a step forward towards a clinically useful and sentient artificial intelligence.

https://doi.org/10.1017/S1047951124036035 Published online by Cambridge University Press

ChatGPT-4 in the future

Nonetheless, ChatGPT-4's current performance in broader medical specialties and improvement over a short period of time provides promise for the future utility of artificial intelligence chatbots in clinical image interpretation. ChatGPT-4 was not specifically trained for healthcare and medical applications but is still performing well in many circumstances. Future chatbots that are designed for clinical purposes and trained on relevant data have the potential for substantial improvements not only in clinical image interpretation but also in other aspects of healthcare such as diagnostics and patient counselling.[2,9,31] This process could theoretically be accelerated by the incorporation of datasets from currently well-established artificial intelligence that uses alternative machine learning architectures like convolutional neural networks, such those for cardiac MRI,[32] echocardiograms,[32] electrocardiogram interpretation,[32,33] and brain tumour MRI analysis.[34] We believe this requires collaboration between academic groups and with industry to develop a robust artificial intelligence model that is accurate and useful.

Limitations

This study had several limitations. All the questions were extracted from a single textbook source, which limits our results' generalizability. Similarly, although the test was created through a random selection of questions in the textbook, it may not be fully representative of the breadth of knowledge in paediatric cardiology, further impeding its generalizability. Additionally, the textbook used in this study was not publicly accessible, while images used in comparable studies used primarily publicly available or licenced images, which may have been used in the training of ChatGPT-4. However, the training data used for ChatGPT has not been publicly disclosed. Ultimately, this confounds comparisons made in this study and portrays the chatbot to have seemingly inferior performance in paediatric cardiology in contrast to other medical specialties. In general, our findings are limited to ChatGPT-4 and are not generalisable to other artificial intelligence chatbots that may be designed for healthcare settings through training with healthcare-specific data. Our study also did not evaluate ChatGPT-4's performance against that of paediatric cardiologists, thereby offering a limited understanding of how artificial intelligence measures up to human expertise. Furthermore, although we made generalisations to the chatbot's ability to interpret images, its answer choices were confounded by text-based clinical information provided in the question. Similarly, all necessary clinical information necessary to answer the questions was provided, which may simulate a knowledge assessment setting (i.e. board examinations) but does not simulate a real-world clinical scenario in which a more nuanced approach may be required (i.e. gathering more information, ordering further diagnostics) to arrive at an informed answer. Lastly, this study examined ChatGPT-4's ability to answer single-answer multiple-choice questions with five answer choices, thus allowing for a 20% probability of arriving at the correct answer by chance alone. Although this pitfall was limited by repeated entries of the question, it is still a resultant non-zero probability. A possible next step could be to employ a short or long answer test format that addresses this issue while also providing an opportunity to judge the artificial intelligence's clinical reasoning skills. One study has suggested that this may paradoxically result in improved performance.[20]

Conclusion

In conclusion, ChatGPT-4 performed poorly when tasked with answering specialised, image-based medical questions regarding paediatric cardiology. By contrast, it has higher accuracy in answering solely text-based questions in paediatric cardiology and image-based questions in other medical specialties. ChatGPT-4 needs substantially more training with multimodal clinical imaging to be a reliable and accurate clinical tool. These improvements may be accelerated through collaboration within and between academia and industry. Future research will be necessary to further assess the clinical reasoning skills and progression of ChatGPT in paediatric cardiology to determine its clinical and academic utility.

Acknowledgements. None.

Financial support. This research received no specific grant from any funding agency, commercial, or not-for-profit sectors.

Competing interests. None.

References

1. Dave T, Athaluri SA, Singh S. ChatGPT in medicine: an overview of its applications, advantages, limitations, future prospects, and ethical considerations. Front Artif Intell 2023; 6: 1169595. DOI: 10.3389/frai.2023.1169595.
2. Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. New Engl J Med 2023; 388: 1201–1208. DOI: 10.1056/NEJMra2302038.
3. Sardana D, Fagan TR, Wright JT. ChatGPT: a disruptive innovation or disrupting innovation in academia? J Am Dent Assoc 2023; 154: 361–364.
4. Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications. Radiology 2024; 310: e232756. DOI: 10.1148/radiol.232756.
5. Ruksakulpiwat S, Kumar A, Ajibade A. Using chatGPT in medical research: current status and future directions. J Multidiscip Healthc 2023; 16: 1513–1520.
6. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to real abstracts with detectors and blinded human reviewers. NPJ Digit Med 2023; 6:75. DOI: 10.1038/s41746-023-00819-6.
7. Novak A, Rode F, Lisičić A, et al. The pulse of artificial intelligence in cardiology: a comprehensive evaluation of state-of-the-art large language models for potential use in clinical cardiology. medRxiv 2023; DOI: 10.1101/2023.08.08.23293689v1.
8. Flanagin A, Bibbins-Domingo K, Berkwits M, Christiansen SL. Nonhuman "Authors" and implications for the integrity of scientific publication and medical knowledge. JAMA 2023; 329: 637–639.
9. Lee P, Bubeck S, Benefits, Petro J. Limits, and risks of GPT-4 as an AI chatbot for medicine. New Engl J Med 2023; 388: 1233–1239. doi: 10.1056/NEJMsr2214184.
10. Deng J, Heybati K, Shammas-Toma M. When vision meets reality: exploring the clinical applicability of GPT-4 with vision. Clin Imaging 2024; 108: 110101.
11. Teixeira da Silva JA. Can chatGPT rescue or assist with language barriers in healthcare communication? Patient Educ Couns 2023; 115: 107940.
12. Kuckelman IJ, Yi PH, Bui M, Onuh I, Anderson JA, Ross AB. Assessing AI-powered patient education: a case study in radiology. Acad Radiol 2024; 31: 338–342.
13. Brin D, Sorin V, Vaid A, et al. Comparing chatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023; 13: 1–5. DOI: 10.1038/s41598-023-43436-9.
14. Skalidis I, Cagnina A, Luangphiphat W, et al. ChatGPT takes on the european exam in core cardiology: an artificial intelligence success story? Eur heart J Digit Health 2023; 4: 279–281.

15. Mihalache A, Huang RS, Popovic MM, et al. Accuracy of an artificial intelligence chatbot's interpretation of clinical ophthalmic images. JAMA Ophthalmol 2024; 142: 321.
16. Gritti MN, AlTurki H, Farid P, Morgan CT. Progression of an artificial intelligence chatbot (ChatGPT) for pediatric cardiology educational knowledge assessment. Pediatr Cardiol 2024; 45: 309–313. DOI: 10.1007/s00246-023-03385-6.
17. Opfer E, Shah S. Advances in pediatric cardiovascular imaging. Mo Med 2018; 115: 354–360.
18. Eidem B. Pediatric cardiology board review. 2nd ed. Philadelphia, PA, 2023.
19. Gilson A, Safranek CW, Huang T, et al. How does chatGPT perform on the United States medical licensing examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ 2023; 9: e45312.
20. Zhu L, Mou W, Yang T, Chen R. ChatGPT can pass the AHA exams: open-ended questions outperform multiple-choice format. Resuscitation 2023; 188: 109783.
21. King RC, Bharani V, Shah K, Yeo YH, Samaan JS. GPT-4V passes the BLS and ACLS examinations: An analysis of GPT-4V's image recognition capabilities. Resuscitation 2024; 195: 110106.
22. Lee KH, Lee RW. ChatGPT's accuracy on magnetic resonance imaging basics: characteristics and limitations depending on the question type. Diagnostics 2024; 14: 171.
23. Handa P, Chhabra D, Goel N, Krishnan S. Exploring the role of chatGPT in medical image analysis. Biomed Signal Process Control 2023; 86: 105292.
24. Martínez-Sellés M, Marina-Breysse M. Current and future use of artificial intelligence in electrocardiography. J Cardiovasc Dev Dis 2023; 10: 175.
25. Hoch CC, Wollenberg B, Lüers JC, et al. ChatGPT's quiz skills in different otolaryngology subspecialties: an analysis of 2576 single-choice and multiple-choice board certification preparation questions. Eur Arch Oto-RHINO-L 2023; 280: 4271–4278. DOI: 10.1007/s00405-023-08051-4.
26. Krusche M, Callhoff J, Knitza J, Ruffer N. Diagnostic accuracy of a large language model in rheumatology: comparison of physician and chatGPT-4. Rheumatol Int 2024; 44: 303–306. DOI: 10.1007/s00296-023-05464-6.
27. Shifai N, van Doorn R, Malvehy J, Sangers TE. Can chatGPT vision diagnose melanoma? An exploratory diagnostic accuracy study. J Am Acad Dermatol 2024; 90: 1057–1059.
28. Horiuchi D, Tatekawa H, Shimono T, et al. Accuracy of chatGPT generated diagnosis from patient's medical history and imaging findings in neuroradiology cases. Neuroradiology 2024; 66: 73–79.
29. Apornvirat S, Namboonlue C, Laohawetwanit T. Comparative analysis of chatGPT and bard in answering pathology examination questions requiring image interpretation. Am J Clin Pathol 2024; 162: 252–260. DOI: 10.1093/ajcp/aqae036.
30. Temsah MH, Alhuzaimi AN, Almansour M, et al. Art or artifact: evaluating the accuracy, appeal, and educational value of AI-generated imagery in DALL·E 3 for illustrating congenital heart diseases. J Med Syst 2024; 48: 54. DOI: 10.1007/s10916-024-02072-0.
31. McMahon CJ, Sendžikaitė S, Jegatheeswaran A, et al. Managing uncertainty in decision-making of common congenital cardiac defects. Cardiol Young 2022; 32: 1705–1717.
32. Sethi Y, Patel N, Kaka N, et al. Artificial intelligence in pediatric cardiology: a scoping review. J Clin Med 2022; 11: 7072.
33. Muzammil MA, Javid S, Afridi AK, et al. Artificial intelligence-enhanced electrocardiography for accurate diagnosis and management of cardiovascular diseases. J Electrocardiol 2024; 83: 30–40.
34. Pinto-Coelho L. How artificial intelligence is shaping medical imaging technology: a survey of innovations and applications. Bioengineering 2023; 10: 1435.

https://doi.org/10.1017/S1047951124036035 Published online by Cambridge University Press