Unveiling the Potential: ChatGPT's Impact on Vestibular Rehabilitation Education

- Trust, Learning, and Value

Yael Arbel P.T. [1], Yoav Gimmon P.T., Ph.D [2,3], Liora Shmueli, Ph.D.[1]

[1] Department of Management, Bar-Ilan University, Ramat-Gan, 52900, Israel.

[2] Department of Physical Therapy, Faculty of Social Welfare & Health Studies,

University of Haifa, Haifa, Israel

[3] Department of Otolaryngology-Head and Neck Surgery, Sheba Medical Center, Tel-

Hashomer, Israel

Corresponding Author: Liora Shmueli, Ph.D., Department of Management,

Bar-Ilan University, Ramat-Gan, 52900, Israel. Email: liora.shmueli@biu.ac.il

Abstract

15
Objective: To evaluate the accuracy, completeness, and explanations provided by

ChatGPT in response to multiple-choice questions related to vestibular rehabilitation.

Study Design: The study was conducted among 30 physical therapists professionals

experienced with vestibular rehabilitation and 30 physical therapy students. They were

asked to complete a Vestibular Knowledge Test consisting of 20 multiple-choice

questions categorized into three groups: (1) Clinical Knowledge, (2) Basic Clinical

Practice, and (3) Clinical Reasoning. Additionally, in May 2023, ChatGPT was tasked

with answering the same 20 VKT questions and providing rationales for its answers.

Three expert board-certified otoneurologists evaluated independently the accuracy of

each ChatGPT response on a 4-level scale.

Results: ChatGPT correctly answered 14 of the 20 multiple-choice questions (70%). It

excelled in Clinical Knowledge (100%) but struggled in Clinical Reasoning (50%).

According to three otoneurologic experts, ChatGPT's accuracy was "comprehensive" for

9 of the 20 questions (45%), while 5 (25%) were "completely incorrect". ChatGPT

provided "comprehensive" responses in 50% of Clinical Knowledge and Basic Clinical

Practice questions, but only 25% in Clinical Reasoning.

Conclusion: Caution is advised when using the current version of ChatGPT due to its

limited accuracy in clinical reasoning. While it provides accurate responses concerning

Clinical Knowledge, its reliance on web information may lead to inconsistencies.

Healthcare professionals should carefully formulate questions and be aware of the

potential influence of the online prevalence of information on ChatGPT's responses.

Combining clinical expertise and guidelines with ChatGPT can maximize benefits while

mitigating limitations.

Keywords: **ChatGPT, Accuracy, Vestibular Rehabilitation, Education**

Introduction

The integration of artificial intelligence (AI) technologies in healthcare has created a new era in patient care, marked by the exploration of the potential integration of AI chatbots, including ChatGPT (Generative Pre-trained Transformer), in medicine as a means to assist healthcare professionals. However, AI's use in vestibular rehabilitation (VR), which requires precise knowledge and expertise, has not yet been thoroughly investigated. Understanding the capabilities of ChatGPT and its like and their potential applications in VR is vital for advancing evidence-based practices and delivering accurate knowledge to patients. This study, therefore, sought to assess the accuracy and grading of the information provided by ChatGPT concerning specialized physical therapists (PTs) in vestibular rehabilitation.

VR is a specialized form of treatment for patients suffering from dizziness. This exercise-based treatment program is designed to promote vestibular adaptation and substitution. Since its advent in the 1940s[1], VR has undergone significant advancements and is now the recommended therapy for dizziness, vestibular dysfunction, and benign paroxysmal positional vertigo (BPPV).[2] The Barany Society, has amended its definitions for diagnosing vestibular disorders several times since 2009.[3–6] In addition, updated guidelines provide evidence-based recommendations on how to improve diagnostic accuracy and treatment efficiency.[2,7] For many clinicians, implementing these guidelines is challenging and requires an active effort of continuous ongoing learning to stay up-to-date. Research has shown that constant learning is essential to ensure consistent, evidence-based treatment and standards in VR.[8,9]

Effective use of AI technologies can assist in making knowledge accessible to

VR therapists. ChatGPT, an interactive chatbot powered by the GPT3.5 architecture

developed by OpenAI, is a new advanced Large Language Model (LLM) that has

demonstrated tremendous potential in accelerating learning and knowledge acquisition

in diverse medical fields.[10] Trained on a vast dataset from the Internet, ChatGPT excels

at generating human-like responses in conversations and prompts across multiple

languages and subject domains, making it a valuable tool for various applications.[11–13]

Recent studies have evaluated the accuracy level of the answers provided by

various version of ChatGPT to knowledge-based questions in the field of medicine, both

with respect to multiple-choice questions and to open-ended questions. When

answering multiple-choice questions related to well-defined, established medical

questionnaires, ChatGPT's accuracy rate was found to vary across different domains,

ranging from 42% in the field of ophthalmology, as measured by the OKAP

(Ophthalmic Knowledge Assessment Program) test, to 76% in cardiology, specifically in

the American Heart Association (AHA) and Advanced Cardiovascular Life Support

(ACLS) exams.[11,12]

The studies that assessed the accuracy and reproducibility of ChatGPT

responses to open-ended questions involved grading the responses by several clinical

domain experts using a scale that included four categories: (1) comprehensive, (2)

correct but inadequate, (3) some correct and some incorrect, and (4) completely

incorrect. For instance, a study on Bariatric Surgery[14] found that the model provided

"comprehensive" responses to 131 out of 151 questions, resulting in an accuracy rate of

86.8%. In a similar study regarding cirrhosis and hepatocellular carcinoma, an accuracy

rate of 76.9% was determined.[15]

An alternative approach implemented by Gilson to determine the degree of

accuracy on the AMBOSS Student Medical exam entailed evaluating the text output of

each ChatGPT response across 3 qualitative metrics: logical justification of the answer

selected, presence of information internal to the question, and presence of information

external to the question.[16]

In this study, we compared the performance of ChatGPT on these multiple-

choice questions with that of PTs who received training in VR and that of PT students.

Additionally, we assessed the accuracy and completeness of ChatGPT's answers and

explanations to a multiple-choice questionnaire we developed about VR.

Materials and Methods

Data source

VKT multiple choice Questionnaire

In May 2023, a Vestibular Knowledge Test (VKT) consisting of 20 multiple-choice

questions (one correct answer and three incorrect options (distractors)) was developed

(See supplemental Appendix 1). It was categorized into three groups: Clinical

Knowledge, Basic Clinical Practice, and Clinical Reasoning. The test was validated

through an anonymous online survey using Qualtrics.[17] The survey link was distributed

via WhatsApp, targeting two distinct groups of participants: 1) PTs who had previously

received training in VR, and 2) PT students from all years of the degree. In the survey,

before each question, the prompt "Please select the correct answer." appeared.

Out of the 60 respondents (30 PTs and 30 PT students) who completed the survey,

53% (n = 32) identified as female. The ages of the participants ranged from 21 to 51

years (M = 31.65, SD = 7.92). With respect to the 30 PTs, the median duration of work

as a PTs was 11.08 years (range <2–23 years), and the median time since their last

vestibular training was 2.44 years (range <0–12 years).

To assess the internal consistency of the questionnaire, reliability statistics were

calculated using Cronbach's alpha. For the PT students, the Cronbach's alpha value

was $\alpha = .37$, indicating relatively low internal consistency. In contrast, for the PT group,

the Cronbach's alpha value was $\alpha = .68$, suggesting a higher level of internal

consistency.

ChatGPT Response Generation

In order to generate responses, each VKT question was prompted to the ChatGPT (May 16th version) AI language model based on the GPT-3.5 architecture. The model is trained on data last updated in September 2021. The questions were entered separately using the same chat.[18]

Grading

Two grading methods were implemented: (1) Multiple-choice questionnaire grading: A score of 1 was assigned to each multiple-choice question answered correctly, and a grade of 0 to an incorrect answer, resulting in a maximum total score of 20 points. We chose this technique to simulate human test-taking. In order to assess the knowledge base of ChatGPT, we compared its performance on multiple-choice questions with that of PTs who had received training in VR and PT students.

(2) Question-response grading: A score was assigned to each explanation provided by ChatGPT, based on an independent assessment of its accuracy and compatibility with the answer provided. The review and grading of each response were independently performed by three board-certified otoneurologists who are experts in the vestibular field. The reviews were conducted based on evidence-based knowledge to determine the accuracy of the ChatGPT responses. The accuracy of each ChatGPT response was rated using the following scale.[14] 1. "Comprehensive": The answer is correct, and the explanation provided is accurate and comprehensive. 2. "Correct but inadequate": Both the answer and the explanation are correct but not satisfactory in terms of completeness. 3. "Mixed with correct and incorrect/outdated data": The answer is

incorrect, and the explanation is partially accurate. 4. "Completely incorrect": Both the

answer and the explanation are incorrect. The accuracy of each answer was

determined based on the median of the experts' answers.

The accuracy of the ChatGPT responses is displayed in Figure 2 according to three

VKT knowledge categories.

Statistical analyses

The grading was determined by comparing ChatGPT's answer to the correct

answer key. The model's grading was determined from a single run. The means and

standard deviation obtained on the 20 multiple-choice single-answer questions by the

PTs experienced in VR and PT students with no prior knowledge of vestibular

rehabilitation were compared to the values of the ChatGPT model using One Sample T-

Test (test of means) to determine if there is a significant difference between the

proportions of response points earned were calculated as a total grade for each domain

and reported as percentages.

The accuracy of the ChatGPT answers was rated by three experts and the score

was determined according to the median of the three given scores. A correlation test

was performed between each two experts using Spearman analysis.

Results

Overall Performance

A VKT consisting of 20 multiple-choice single-answer questions was completed by 30

PTs experienced in VR, 30 PT students and ChatGPT-3.5 (See supplemental Appendix

1). ChatGPT answered 70% of the questions correctly (14 out of 20), while the PTs

achieved a score of 76.3%, and the PT students a score of 40.5% (Figure 1). A

statistical analysis using t-tests revealed significant differences between the groups.

The PTs' performance was significantly better than that of ChatGPT (t = 2.46, p < .05),

with a large effect size (Cohen's d = 2.89). The PT students performed significantly

worse than ChatGPT (t = -13.24, p < .001), with a large effect size (Cohen's d = 2.44).

Figure 1: Performance of ChatGPT, the PTs and the PT Students on the VKT, stratified

by question type and topic.

187
188

Performance by Question Type

191   In terms of performance, ChatGPT excelled in Clinical Knowledge questions,

192   answering all 4 correctly (100%). However, in the category of Clinical Reasoning, it

193   managed to answer only 1 out of 4 correctly (25%), and in the Basic Clinical Practice

194   category, it provided the correct answer to only 8 of the 12 questions (66.66%).

Accuracy of the ChatGPT-generated explanations

The accuracy of ChatGPT's response was assessed and categorized as either

comprehensive, correct but inadequate, mixed (correct and incorrect/outdated data),

and completely wrong Based on the experts' evaluation of the ChatGPT answers, 9 of

the 20 answers were "comprehensive" (45%), 5 were "correct but inadequate" (25%), 1

was deemed "mixed with correct and incorrect responses" (5%) and the remaining 5

were considered "completely incorrect" (25%). When analyzed according to knowledge

category, it was found that ChatGPT provides "comprehensive" responses to 2 of the 4

(50%) question related to Clinical Knowledge, to 6 of the 12 (50%) questions concerning

"Basic Clinical Practice, and to only 1 of the 4 (25%) questions in the Clinical Reasoning

category (**Figure 2**). Figure 3 illustrates further examples of ChatGPT prompts by

category.

Figure 2: Accuracy of the responses generated by ChatGPT-3.5 to questions related to

VR, categorized by knowledge category

210
211

212
213
214 **Figure 3**: Examples of ChatGPT prompts for multiple-choice questions, and expert-
215 graded ChatGPT responses

Question	Which of the following examinations is used to diagnose the vestibular system?	When examining a dizzy patient, which of the following tests is different in terms of the meaning of the result that is obtained?	While performing oculomotor tests, you observe a vertical deviation during the Cover Uncover test, a changing-direction nystagmus when the gaze is held on both sides, and during the Head Impulse test - no noted vestibular impairment. Given these findings, what should you do?	In acute vestibular syndrome of the right ear, according to Alexander's law, what should we expect to see in the first few hours and days?
multiple-choice correct answer	Head Impulse test	VOR cancellation	Refer the patient urgently to the emergency room	Nystagmus increases when the gaze is diverted to the right side and nystagmus decreases when the gaze is diverted to the left
Chat GPT answers	CORRECT: Head Impulse test The Head Impulse test is used to assess the function of the vestibulo-ocular reflex (VOR), which is a crucial component of the vestibular system.	CORRECT: abnormal and indicates impairment of the VOR. This means that the eyes are unable to accurately maintain fixation on the target during head movements. It suggests a dysfunction within the vestibular system.	INCORRECT: Refer the patient to their General Practitioner doctor for further diagnosis. The presence of a vertical deviation during the Cover Uncover test suggests a potential ocular misalignment or extraocular muscle dysfunction.	In acute vestibular syndrome of the right ear, according to Alexander's law, we would expect to see the INCORRECT: Nystagmus increases when the gaze is directed forward, and nystagmus decreases when the gaze is diverted to the sides.
category	Clinical knowledge (100% correct)	Basic clinical practice (66.66% correct)	Clinical Reasoning (50% correct)	Basic clinical practice (66.66% correct)
Accuracy Grading of responses	comprehensive(45%)	Correct but inadequate (25%)	Mixed with correct and incorrect (5%)	completely incorrect (25%)

Discussion

In this study, we present a novel analysis of ChatGPT's performance in the domain of vestibular rehabilitation. We found that when answering multiple-choice questions, PTs perform significantly better than ChatGPT, whereas PT students fare much worse than ChatGPT. ChatGPT is accurate in answering Clinical Knowledge questions and in providing accurate explanations for the answers, but their accuracy starkly declines when it comes to Clinical Reasoning. Therefore, the way in which the questions for the chatGPT are formulated is crucial.

Our finding that ChatGPT achieved an accuracy of 70% on multiple-choice questions specifically focused on VR aligns with previous studies where ChatGPT's performance was near or passed the threshold of 60% accuracy on the United States Medical Licensing Exam (USMLE),[13] an accuracy of 76.3% (29/38) on two 38-question AHA ACLS exams,[12] and successfully cleared the Medical Physiology Examination of Phase I MBBS with distinction (>75% of marks)[19]. Likewise, in another such study involving quiz-style questions from the platform of the German Society of Oto-Rhino-Laryngology that subspecialties 2576 questions into 15 distinct otolaryngology, in the vestibular system category ChatGPT responded correctly to 63% of the questions (n=95 correct vs. n=57 false).[20]

The frequency of errors in ChatGPT's explanations in our study is particularly notable in the Best Clinical Practice topics. Indeed, errors often occur when addressing questions that involve clinical reasoning and descriptions, which require a deep understanding of medical contexts rather than solely relying on clinical knowledge. For example, when providing a description of a patient who exhibits a down-beating

nystagmus with a rightward rotation during Dix-Hallpike tests on both sides, the

ChatGPT response concluded that the "most appropriate treatment choice would be: A

positioning treatment for the right posterior canal by performing a Gufoni maneuver."

While ChatGPT did accurately diagnose the right ear, it was wrong when it came to the

identification of the canal and the appropriate treatment.

An interesting finding was revealed from ChatGPT's responses to post-treatment

restrictions regarding prohibitions and restrictions. In this context, ChatGPT's answers

were found to be incorrect and based on outdated approaches and previously published

articles[21] available on the Internet. For example, when asked what recommendations

should be given after a patient is successfully treated with BPPV concerning what

should he or she should avoid doing at home, The Chat incorrectly responded: "Avoid

lying on the side that was treated". ChatGPT's responses were influenced by the

abundance of information available online rather than by recent and reliable sources.[22]

ChatGPT, however, did provide accurate and current answers when the question was

rephrased: In response to, "After you successfully treat a patient diagnosed with BPPV,

what will be your post-treatment recommendations?", ChatGPT's response was,

"Return to normal activity; you must not refrain from movement or restrict the sleeping

position but should avoid climbing ladders or stools." Thus, careful consideration should

be given to the impact of the wording of the question and presentation scenarios, as it

could affect the quality of the ChatGPT response. Additionally, it is crucial to formulate

questions carefully and be aware that ChatGPT's responses may be influenced by the

prevalence of information on the web rather than being based on updated, current and

reliable sources. Of note, we ensured that no prompting or training was provided to the

AI, by entering each question separately using the same chat.

It is evident that the utilization of AI for enhancing clinical decision-making will

continue to expand. This growing trend highlights the necessity for effective

collaboration between medical professionals and technology developers. With the rapid

growth of medical knowledge, the integration of technologies like AI becomes crucial in

enabling healthcare professionals to effectively apply this knowledge in their practice.

Health care education emerges as a captivating domain to explore, given the vast

amount of information and diverse concepts that healthcare students are expected to

comprehend. [23] For example, in a recent editorial by Moons and Van Bulck (2023), they

highlight the potential of ChatGPT in cardiovascular nursing practice and research.

They further emphasize its ability to summarize large texts, facilitate the work of

researchers, and assist in data collection, making ChatGPT a potentially valuable tool in

health care practice. [24]

Our study, together with previous research, leads us to formulate the following

recommendations regarding the specific cases where ChatGPT can be utilized to

expedite the learning process in clinical knowledge. Instead of investing time on reading

and memorizing scientific literature updates and guidelines, ChatGPT can provide a

viable alternative. Indeed, a recent systematic review examined the potential

applications of Language Model models (LLMs) in healthcare education. The review

highlighted several benefits of ChatGPT in healthcare education, such as enhanced

personalized learning experiences and an emphasis on critical thinking and problem-

based learning.[25]

To determine to what extent ChatGPT could provide accurate responses and

reliable information, further research is needed to validate the efficacy of ChatGPT in

providing accurate answers in the medical domain, specifically in the context of VR.

Furthermore, studies should be conducted to determine if ChatGPT's performance can

be enhanced through techniques such as question repetition and integration of reliable

medical literature to ChatGPT. As emphasized by ChatGPT itself, the evaluation of its

output by professionals remains crucial in ensuring the accuracy and completeness of

the information provided. In our opinion, clinicians who wish to "consult" with ChatGPT

should simplify their questions and avoid clinical reasoning questions.

Limitations

With respect to the study's limitations, there is no common questionnaire that tests

knowledge and clinical reasoning in the field of vestibular rehabilitation. We compiled a

questionnaire and validated it. In this VKT questionnaire, there are only 20 multiple-

choice questions, which may have limited participants' ability to express nuanced

perspectives, potentially impeding the assessment of higher-order thinking skills.

Additionally, the survey focuses on three VR categories, which may overlook other

relevant knowledge categories. Furthermore, the survey was conducted among Hebrew

speakers and we translated into English, but as professional terminology, which is

primarily in English, was used in the Hebrew questionnaire, it may be that English

speakers might have performed better. Due to this discrepancy in linguistic proficiency,

ChatGPT may be biased and potentially discriminatory against participants.

Conclusions

The ChatGPT platform is an effective tool that can be used to obtain information and

answer questions related to a variety of fields, including VR. In spite of this, it is

important to recognize that ChatGPT has a number of limitations such as the potential

provision of false and inaccurate answers to questions in specific VR categories,

especially in clinical reasoning questions. The model is trained on an updated version of

ChatGPT from 2021 and is influenced by the abundance of information available online

rather than by recent and reliable sources. Medical professionals should therefore use

this tool carefully and be aware of its limitations.

Acknowledgments: We would like to express our gratitude to Prof. Avi Shupak, Dr.
Yahav Oron, Dr. Amit Wolfovitz, as well as the physical therapists and physiotherapy
students who participated in the study.
329

Funding: The authors received no financial support for the research, authorship, and/or

publication of this article.

Declaration of conflicting interests: The Authors declares that there is no conflict of

interest.

Ethics approval and consent to participate

The study was approved by the Ethics Committee for Non-clinical Studies at Bar Ilan-

University in Israel. The ethics form was signed by the committee head and the date of

approval was 21 May 2023.

CRediT authorship contribution statement

Yael Arbel, P.T: conceptualization, methodology, formal analysis resources, and writing the original draft.

Yoav Gimmon, P.T., Ph.D: methodology, formal analysis resources, writing - review & editing, supervision.

Liora Shmueli, Ph.D.: conceptualization, methodology, formal analysis resources, writing the original draft, writing - review & editing, supervision and project administration.

All authors approved the final draft.

Additional Contributions

Additional Information

This data was collected using the AI ChatGPT developed by OpenAI. The data was analyzed using SPSS version 28

Data Sharing Statement All the data generated or analyzed during this study is included in this published article [and its supplementary information files].

Consent for publication

Not applicable

Abbreviations

ChatGPT - Generative Pre-trained Transformer

VKT - Vestibular Knowledge Test

AI - artificial intelligence

VR - vestibular rehabilitation

PTs - Physical therapists

BPPV - benign paroxysmal positional vertigo

LLM - Large Language Model

AHA – merican Heart Association

ACLS – Advanced Cardiovascular Life Support

References

1. Cawthorne T. Vestibular Injuries. *Proceedings of the Royal Society of Medicine.* 1946;39(5):270-273. doi:10.1177/003591574603900522

2. Hall CD, Herdman SJ, Whitney SL, et al. Vestibular Rehabilitation for Peripheral Vestibular Hypofunction: An Updated Clinical Practice Guideline From the Academy of Neurologic Physical Therapy of the American Physical Therapy Association. *Journal of Neurologic Physical Therapy.* 2022;46(2):118-177. doi:10.1097/NPT.0000000000000382

3. Staab JP, Eckhardt-Henn A, Horii A, et al. Diagnostic criteria for persistent postural-perceptual dizziness (PPPD): Consensus document of the committee for the Classification of Vestibular Disorders of the Bárány Society. *VES.* 2017;27(4):191-208. doi:10.3233/VES-170622

4. Strupp M, Lopez-Escamez JA, Kim JS, et al. Vestibular paroxysmia: Diagnostic criteria. *VES.* 2017;26(5-6):409-415. doi:10.3233/VES-160589

5. Von Brevern M, Bertholon P, Brandt T, et al. Benign paroxysmal positional vertigo: Diagnostic criteria: Consensus document of the Committee for the Classification of Vestibular Disorders of the Bárány Society. *VES.* 2015;25(3,4):105-117. doi:10.3233/VES-150553

6. Bisdorff A, Von Brevern M, Lempert T, Newman-Toker DE. Classification of vestibular symptoms: Towards an international classification of vestibular disorders. *VES.* 2009;19(1-2):1-13. doi:10.3233/VES-2009-0343

7. Bhattacharyya N, Gubbels SP, Schwartz SR, et al. Clinical Practice Guideline: Benign Paroxysmal Positional Vertigo (Update). *Otolaryngol--head neck surg.* 2017;156(S3). doi:10.1177/0194599816689667

8. Male AJ, Ramdharry GM, Grant R, Davies RA, Beith ID. A survey of current management of Benign Paroxysmal Positional Vertigo (BPPV) by physiotherapists' interested in vestibular rehabilitation in the UK. *Physiotherapy.* 2019;105(3):307-314. doi:10.1016/j.physio.2018.08.007

9. Meldrum D, Burrows L, Cakrt O, et al. Vestibular rehabilitation in Europe: a survey of clinical and research practice. *J Neurol.* 2020;267(S1):24-35. doi:10.1007/s00415-020-10228-4

10. Liao W, Liu Z, Dai H, et al. Differentiate ChatGPT-generated and Human-written Medical Texts. Published online April 23, 2023. doi:10.48550/arXiv.2304.11567

11. Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the Performance of ChatGPT in Ophthalmology: An Analysis of its Successes and Shortcomings. *Ophthalmology Science.* Published online May 5, 2023:100324. doi:10.1016/j.xops.2023.100324

12. Fijačko N, Gosak L, Štiglic G, Picard CT, Douma MJ. Can ChatGPT pass the life
support exams without entering the American heart association course?
Resuscitation. 2023;185. doi:10.1016/j.resuscitation.2023.109732

13. Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE:
Potential for AI-assisted medical education using large language models. *PLOS*
Digital Health. 2023;2(2):e0000198. doi:10.1371/journal.pdig.0000198

14. Samaan JS, Yeo YH, Rajeev N, et al. Assessing the Accuracy of Responses by the
Language Model ChatGPT to Questions Regarding Bariatric Surgery. *OBES SURG*.
Published online April 27, 2023. doi:10.1007/s11695-023-06603-5

15. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in
answering questions regarding cirrhosis and hepatocellular carcinoma. Published
online February 8, 2023:2023.02.06.23285449. doi:10.1101/2023.02.06.23285449

16. Gilson A et al. JMIR Medical Education - How Does ChatGPT Perform on the United
States Medical Licensing Examination? The Implications of Large Language Models
for Medical Education and Knowledge Assessment. Published 2023. Accessed May
24, 2023. https://mededu.jmir.org/2023/1/e45312/

17. Qualtrics Survey Software. Published online 2023. https://www.qualtrics.com

18. OpenAI. ChatGPT (Version GPT-3.5). Published online 2021. https://openai.com

19. Subramani M, Jaleel I, Krishna Mohan S. Evaluating the performance of ChatGPT in
medical physiology university examination of phase I MBBS. *Advances in*
Physiology Education. 2023;47(2):270-271. doi:10.1152/advan.00036.2023

20. 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 Otorhinolaryngol*. Published
online June 7, 2023. doi:10.1007/s00405-023-08051-4

21. Herdman SJ. Treatment of Benign Paroxysmal Positional Vertigo. *Physical Therapy*.
1990;70(6):381-388. doi:10.1093/ptj/70.6.381

22. Bhattacharyya N, Gubbels SP, Schwartz SR, et al. Clinical Practice Guideline:
Benign Paroxysmal Positional Vertigo (Update). *Otolaryngol--head neck surg*.
2017;156(S3). doi:10.1177/0194599816689667

23. Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing Artificial
Intelligence Training in Medical Education. *JMIR Medical Education*.
2019;5(2):e16048. doi:10.2196/16048

24. Moons P, Van Bulck L. ChatGPT: can artificial intelligence language models be of
value for cardiovascular nurses and allied health professionals. *European Journal of*

Cardiovascular Nursing. Published online February 8, 2023:zvad022.
doi:10.1093/eurjcn/zvad022

25. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice:
Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare*.
2023;11(6):887. doi:10.3390/healthcare11060887

451
452