

# Application of Artificial Intelligence Generated Content in Medical Examinations

Rui Li<sup>1</sup>, Tong Wu<sup>2-4</sup>

<sup>1</sup>Emergency Department, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China; <sup>2</sup>National Clinical Research Center for Obstetrical and Gynecological Diseases, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China; <sup>3</sup>Key Laboratory of Cancer Invasion and Metastasis, Ministry of Education, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China; <sup>4</sup>Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, People's Republic of China

Correspondence: Tong Wu, Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, No. 1095, Jiefang Avenue, Wuhan, 430030, People's Republic of China, Email [tongwu66@tjh.tjmu.edu.cn](mailto:tongwu66@tjh.tjmu.edu.cn)

**Abstract:** As the rapid development of large language model, artificial intelligence generated content (AIGC) presents novel opportunities for constructing medical examination questions. However, it is unclear about the way of effectively utilizing AIGC for designing medical questions. AIGC is characterized by its rapid response capabilities and high efficiency, as well as good performance in mimicking clinical realities. In this study, we revealed the limitations inherent in paper-based examinations, and provided a streamlined instruction for generating questions using AIGC, with a particular focus on multiple-choice questions, case study questions, and video questions. Manual review remains necessary to ensure the accuracy and quality of the generated content. Future development will be benefited from technologies like retrieval augmented generation, multi-agent system, and video generation technology. As AIGC continues to evolve, it is anticipated to bring transformative changes to medical examinations, enhancing the quality of examination preparation, and contributing to the effective cultivation of medical students.

**Keywords:** artificial intelligence generated content, medical education, multiple-choice question, large language model

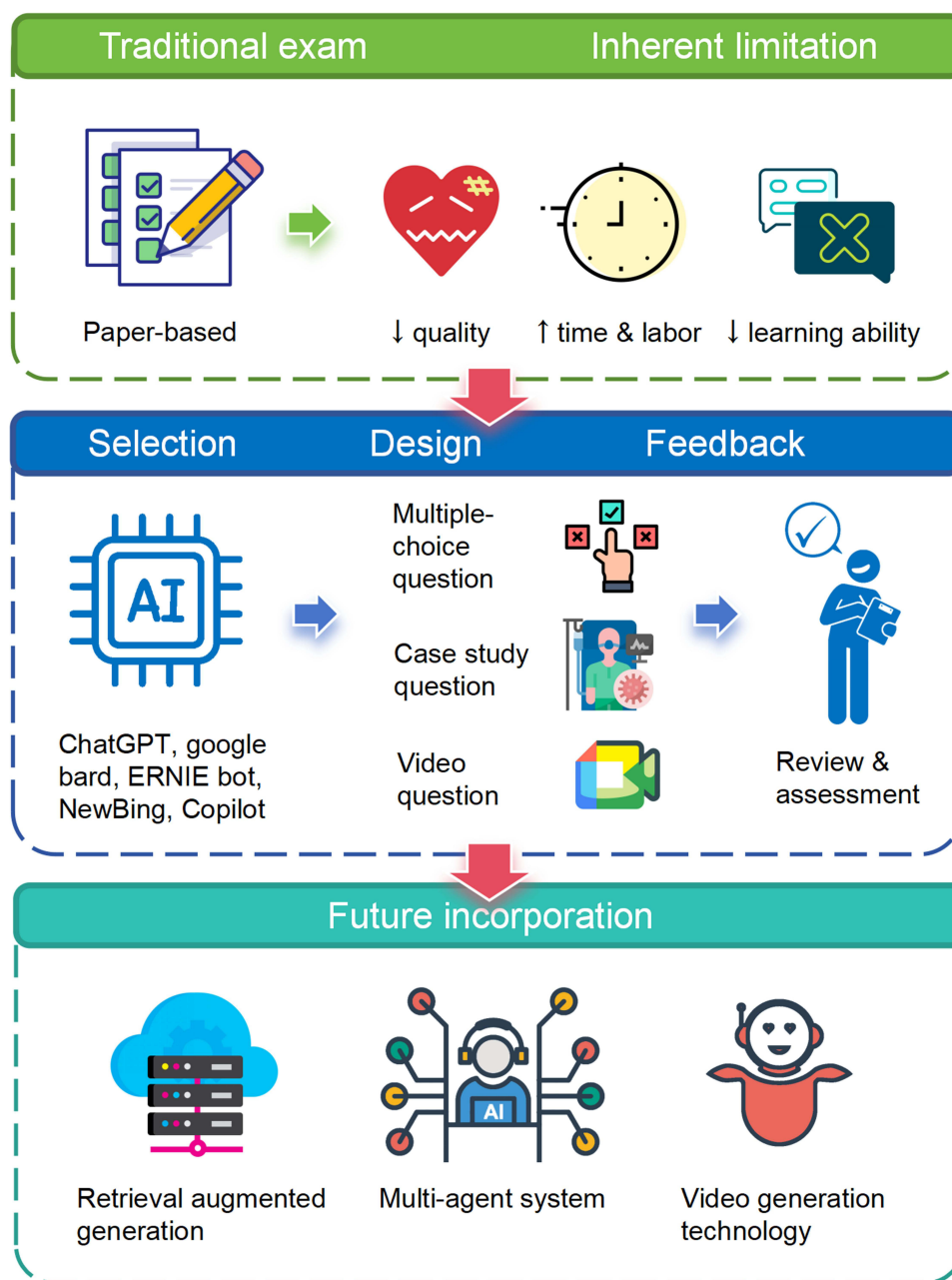
## Background

Artificial intelligence generated content (AIGC) refers to the utilization of artificial intelligence to produce text, image, audio, and video. It is greatly promoted by the release of Chat Generative Pre-trained Transformer (ChatGPT), and causes an immediate change in medical education.<sup>1</sup> AIGC offers educators opportunities to innovate teaching methodologies, enhance learning experiences, and improve educational effectiveness.<sup>1</sup> As an example, Fuller et al analyzed the course evaluation comments through instructors and ChatGPT, and found a high agreement between these two groups, so that ChatGPT could ease the burden of instructors during teaching feedback.<sup>2</sup> Additionally, integrating AI tools like ChatGPT into learning practices has been shown to assist learners in conducting literature reviews, addressing personalized inquiries, and promoting autonomous learning. As a result, students become more adept at navigating resources and synthesizing information.<sup>3</sup> This self-directed learning model is essential in today's fast-paced educational environment, where learners are encouraged to develop critical thinking and problem-solving skills. However, despite these advancements, the potential of AIGC in the development of medical questions remains underrecognized.

Examinations serve as an essential tool for assessing and promoting learning outcomes, as well as provide students with a structured means of revision and self-assessment. Repeated testing through quizzes significantly enhances material retention in subjects such as anatomy, where students have demonstrated improved scores after multiple assessments.<sup>4</sup> Most medical schools employ question banks to generate streamlined examination papers. It has become a prevalent practice, primarily due to the ability of question banks to enhance the efficiency and effectiveness of examination preparation and assessment. Question banks can serve as a repository of multiple-choice questions (MCQs) that can be utilized to create tailored assessments, while also ensure that the content aligns with the curriculum.<sup>5,6</sup> Despite their

advantages, challenges such as inconsistent question quality and inadequate real-world scenarios persist. In this context, AIGC offers several benefits, including low cost, high volume, and novelty. For instance, Klang et al leveraged GPT-4 to generate a substantial set of 210 medical MCQs, demonstrating the model's potential in educational contexts. Remarkably, the study reported only one error in the entire set of questions, underscoring the impressive accuracy and reliability of GPT-4. This finding suggests that AI is highly efficient in generating high-quality medical educational materials, where precise knowledge assessment is crucial for training future healthcare professionals.<sup>7</sup>

This study explores the application of AIGC in the development of medical examination questions (Figure 1). It begins by discussing the limitations of traditional paper-based examination methods, highlighting the need for innovation in assessment techniques. We then propose practical strategies for designing medical exams using AIGC, supported by illustrative examples. Additionally, the study examines the future potential of AIGC in conjunction with other emerging



**Figure 1** Flowchart of artificial intelligence generated content in medical examinations.

technologies, anticipating a transformative impact on medical examinations. As AIGC evolves, it is expected to significantly enhance the quality of examination preparation and contribute to the effective training of medical students.

## Issues in Traditional Medical Examination Questions

### Inconsistent Quality of Questions

Constructing high-quality examination questions is a highly specialized process that requires specific skills, which even experienced medical educators may lack. The National Board of Medical Examiners' Item-Writing Guide provides essential frameworks to assist educators in this endeavor.<sup>8</sup> When designing exams, it is crucial for educators to focus on the key knowledge points outlined in the teaching syllabus, adhere to specific principles in question formulation, and ensure that answer options reflect appropriate levels of difficulty. Meticulously constructed questions exhibit psychometric properties that enhance learners' performance.<sup>9</sup> Conversely, poorly designed MCQs reduce pass rates, impede clinical reasoning skills, and misinterpret test scores.<sup>10</sup> As a whole, although question banks can cover a wide range of topics, they often fall short in addressing the nuances of complex subjects that require critical thinking and the application of knowledge.

### Difficulties in Developing and Maintaining Question Banks

As revealed by Bharat et al that developing a question bank for rheumatology took four years compromising of a rigorous 12-step process.<sup>11</sup> This process is not only time-consuming but also labor-intensive and requires substantial resources. Most importantly, given the rapid evolution of clinical medicine, the questions in exam banks may quickly become outdated, necessitating annual revisions to reflect the latest advancements. Failure to update these questions results in students receiving inaccurate information, potentially leading to incorrect clinical management decisions. Consequently, it is imperative that new questions are continually developed and rigorously assessed to ensure the accuracy.<sup>12</sup> This iterative process also requires collaborations between experienced educators and professional software engineers.<sup>11</sup> Although commercial question banks are available, they often lack alignment with specific teaching content and their quality remains uncertain. These commercial resources are typically designed to prepare students for board certification examinations, making them less suitable for college-level graduation assessments.

### Declined Learning Ability

The increasing reliance on question banks has led to significant challenges in maintaining the integrity and effectiveness of the assessment process. Students may become overly familiar with question types and content, and even cause leakage and omission of questions. When students have access to previous exam questions or practice materials, it can undermine the integrity of the assessment process. This leakage not only diminishes the value of the evaluation but also fosters an environment where students may prioritize rote memorization over genuine understanding and critical thinking.<sup>13,14</sup> In addition, the prevalent use of multiple-choice questions (MCQs) or fill-in-the-blank formats encourages students to focus on short-term recall rather than long-term comprehension. This concern is echoed in discussions about the effectiveness of different question types, where non-compound questions tend to yield more reliable assessments than compound questions, as the latter can confuse respondents.<sup>10</sup> In the long term, such declines in educational rigor are likely to exacerbate disparities and uneven competences in educational quality among different universities.

Taken together, ensuring a fair and effective evaluation of medical students' competence necessitates the quality of examination questions, the construction of question banks. By addressing these challenges, we can not only enhance the assessment process but also contribute to the overall improvement of healthcare standards and uphold the professionalism of medical students. It is essential to emphasize the importance of flexibility and innovation in medical examinations.

## Construction of Medical AIGC Questions

### Step I Select an Appropriate AIGC Model

The quality of AIGC questions is significantly influenced by the type of large language models (LLMs) employed in their creation. Recent studies have highlighted the varying capabilities of different LLMs.<sup>15,16</sup> Rohaid et al conducted a study

on the accuracy of three LLMs in neurosurgery examinations, revealing that GPT-4 achieved a high score (82.6%), which surpassed that of GPT-3.5 (62.4%) and Google Bard (44.2%).<sup>17</sup> These discrepancies may be attributed to different data quality, model architecture, and technical version. Notably, GPT-4 relies on pre-existing training data and lacks web crawling capabilities. In contrast, Google Bard can access and incorporate real-time information from the internet. GPT-4 demonstrated a 20% improvement in three United States Medical Licensing examinations.<sup>18</sup> These variabilities underscore the necessity for careful selection and fine-tuning of LLMs to enhance their effectiveness in generating AIGC questions.

Step 2 Specific Prompts to Construct Questions

While AI can streamline the question creation process, the validity of these AI-generated questions varies significantly based on the specificity and clarity of the prompts used by educators.<sup>19</sup> This underscores the necessity for medical educators to engage in effective prompt engineering to maximize the potential of AI tools. By refining and enhancing the instructional prompts, it becomes feasible to achieve more accurate and scientific output. Fundamental prompts include delineating the knowledge areas and assessment criteria explicitly, thereby enabling LLMs to generate medical examination questions, answers and explanations. This approach minimizes the risk of question leakage, and facilitates students’ foundational understanding of specific concepts. Examination questions can further be derived from existing question banks. This method optimizes the use of existing resources, and keeps consistency in question styles. It also saves time and efforts in selecting relevant knowledge points.

Multiple-Choice Questions

MCQs are the predominant questions in medical examinations, and encompass a wide array of learning objectives. Poorly constructed questions can lead to misleading assessments of student knowledge and skills. High-quality MCQs are particularly effective in assessing students’ abilities to apply knowledge, interpret information, and synthesize concepts.<sup>20</sup> AIGC further enhances the efficiency and clarity of MCQ creation. Medical educators can leverage AIGC to formulate more intricate prompts, such as generating plausible yet incorrect distractors, or crafting questions that mirror real-life clinical scenarios, or increasing the difficulty. Table 1 presents the original options as simple and clear, allowing students to directly compare them to find answers. In contrast, the new test questions include options with

Table 1 Examples of AIGC in Multiple-Choice Questions

Prompt	Response
Write a multiple-choice question based on “Knowledge Point: Nursing of Postpartum Women – Breastfeeding”.	Question: What is the recommended initial feeding method for newborns according to the guidelines for postpartum nursing and breastfeeding? A. Formula feeding immediately after birth B. Exclusive breastfeeding within the first hour of life C. Offering a sugar water solution to soothe the baby D. Waiting for the mother’s milk to come in before starting breastfeeding Correct Answer: B.
Increase the difficulty of the above questions.	Question: In the context of postpartum nursing and breastfeeding, which of the following statements best aligns with the evidence-based guidelines for promoting optimal maternal and infant health outcomes? A. Colostrum, the first milk produced after birth, is nutritionally insufficient for the newborn’s needs and should be supplemented with formula. B. Mothers should wait until they feel the “let-down” reflex before attempting to breastfeed, as this indicates milk availability. C. The use of nipple shields is universally recommended to assist with latch issues and promote successful breastfeeding. D. Skin-to-skin contact immediately after birth, followed by early and frequent breastfeeding attempts. Correct Answer: D.

extensive modifications and descriptions, making them appear more familiar and authentic while also imposing greater demands on students. These innovative strategies can improve the accuracy and complexity of MCQs, as well as differentiate students across different competency levels.<sup>21</sup> Moreover, the use of AIGC in MCQ creation can lead to a more standardized approach, ensuring that questions are not only relevant but also aligned with learning objectives. This standardization can help mitigate biases that may arise from human-generated questions, as AIGC can be programmed to adhere to specific guidelines and criteria. For instance, studies have indicated that negatively-marked MCQ assessments, which reward partial knowledge, do not cause gender bias, and can enhance the performance and satisfaction of students.<sup>22</sup>

### Case Study Questions

The incorporation of AIGC into case study questions enhances the richness of situational descriptions, making a more realistic learning environment and better reflect the actual clinical practice. By this way, AIGC can facilitate deeper learning experiences and improve the assessment of higher-order cognitive skills among students. It aligns with the findings of various studies that emphasize the importance of context and complexity in educational assessments.<sup>23</sup> Medical students need to decipher clues and formulate innovative responses. Moreover, the use of AIGC in crafting case-based questions can draw from a wide array of data sources, providing a more comprehensive view of patient scenarios. This is particularly relevant in medical education, where understanding the interplay of various factors, such as patient history, social context, ethical dimensions, and clinical guidelines, is crucial for effective decision-making.<sup>24</sup> Consequently, such an alteration encourages medical students to engage in critical thinking. It aligns with the growing recognition of the importance of holistic approaches in healthcare education, and understanding the patient's narrative is as vital as the clinical facts.

### Video Question

By utilizing text driven image generation technology, students have access to highly simulated medical imaging questions without contact with real patients. This not only enriches students' learning experience but also protects patients' personal privacy. Moreover, it alleviates the shortage and inequality of medical teaching resources and multimedia resources. The development of denoising diffusion probabilistic models has enabled the stable generation of high-quality medical images. Image generation tools, such as Imagen and latent diffusion models, utilize text prompts to provide fine-grained guidance during the image generation process.<sup>25,26</sup> This capability holds significant promise for the formulation of questions related to medical imaging. Recently, Xu et al introduced a model named MedSyn, which generates high-resolution, anatomy-aware CT images based on user-input text prompts.<sup>27</sup> These images effectively retain intricate details of the lung's airway, vessels, and lobular structure while addressing privacy concerns. Furthermore, MedSyn demonstrated superior performance compared to the most advanced models based on GAN and diffusion techniques. Therefore, it can be used to help students grasp a deeper understanding of the anatomical structure and pathological manifestations of the lungs.

Traditional skill examinations often face limitations related to venues, equipment, and personnel. AIGC can generate simulated examination questions, students are required to complete tasks that involve moving objects and drawing diagrams.<sup>28</sup> This innovation significantly broadens the range of examination formats, and raises the standards for candidates' foundational knowledge and judgment abilities.<sup>29</sup>

## Step 3 Assess and Review the Outcomes

Although AIGC has shown great potential in generating high-quality medical questions, ensuring the accuracy and reliability of its output remains a key challenge. A study that examined the performance of ChatGPT on medical examinations found that the accuracy and applicability of AIGC varied based on the complexity of the questions posed.<sup>30</sup> Most open-access LLMs lack specific medical training. Consequently, they may lead to misdiagnoses or misinterpretations, particularly in contexts involving age and gender. For example, a 35-year-old female presenting with irregular menstruation might be incorrectly diagnosed with menopausal syndrome. A male patient with abdominal pain could receive an option for ectopic pregnancy among the choices. A study compared the responses of GPT-3.0 to those of



medical consultants on 41 case-based questions, and found the medical inadequacy and unconciseness of GPT-3.0.<sup>31</sup> This deficiency may be partly attributed to the model's inability to recognize non-verbal cues. Therefore, routine assessment and review of AIGC questions before formal examinations is essential to ensure the scientific validity, clarity, and feasibility. Teachers must evaluate the alignment of these questions with course objectives, and exclude any inaccurate or misleading contents. Furthermore, it is crucial to analyze the difficulty, discrimination, and reliability of the questions based on the performance of students, facilitating the optimization of teaching research and practice.

Some other AI tools can also be utilized for error recognition and correction. One method is to compare the generated information with a structured medical knowledge base, such as knowledge graphs, to find errors or inconsistencies.<sup>32</sup> Another approach is to encourage the medical expert systems to review and edit the AIGC content. Medical expert systems combine the judgement from human experts and AI systems, which can significantly improve the accuracy and reliability of AIGC applications in medical education.<sup>33</sup> As a result, the application of AIGC provides a more comprehensive and authentic assessment of students' clinical abilities and literacy. This approach contributes to their holistic development and better prepares them for professional practice.

## Future Prospects

### Retrieval Augmented Generation

Retrieval augmented generation (RAG) technology, which integrates retrieval and generation functions by introducing an external knowledge base, significantly improves the quality and accuracy of AIGC. This benefit is particularly useful in the medical field, where accurate information is critical for patient care and decision-making. One notable application of RAG technology is in the development of advanced chatbots for healthcare, such as the GastroBot. The system demonstrated a remarkable improvement in response accuracy and contextual relevance, achieving a context recall rate of 95% and a faithfulness to the source of 93.73%.<sup>34</sup> Similarly in emergency medical triage, RAG significantly outperformed traditional methods, achieving a correct triage rate of 70% compared to lower rates from human practitioners.<sup>35</sup> When designing questions, educators can leverage RAG technology to accurately retrieve relevant case information, treatment methods, and the latest research progress from a wide range of medical literature and databases, thereby deepening the comprehensive understanding of students concerning the knowledge point.

### Multi-Agent System

Multi-agent systems (MAS) have emerged as a powerful framework for simulating collaboration and communication in complex environments. By modeling various agents, such as doctors, nurses, and patients, these systems can effectively replicate the dynamics of real-world interactions, thereby enhancing our understanding of cooperative behaviors and improving healthcare delivery. In a microworld experiment, it was found that participants who interacted with a highly cooperative agent were more likely to engage in effective interactions and resource sharing compared to those who worked with a less cooperative agent. This finding underscores the potential of MAS to enhance resilience in healthcare systems by fostering better collaboration among human and automated agents.<sup>36</sup> MASs have demonstrated unique advantages in the representation and inference of medical knowledge. Students can better understand the interrelationships between medical concepts and enhance their abilities in comprehensive analysis and problem-solving. When used in medical exams, students may be required to effectively respond to the condition of simulated patients, communicate and collaborate effectively with other intelligent agents (nurse, attending physician, and others), and develop reasonable treatment plans. This teaching method not only enhances students' clinical practice ability, but also strengthens their teamwork and communication skills.

### Video Generation Technology

Video generation technology (VGT) represented by Sora provides more intuitive and vivid teaching resources for medical education by generating high-quality medical teaching videos. This technology presents unintelligible medical concepts and complex physiological processes in a dynamic and visual form, thereby reducing the difficulty of understanding and enhancing the learning interest. In anatomy teaching, VGT can generate 3D animated videos that showcase

various organs and structures of the human body in detail. Instructional videos can serve as effective teaching tools, especially in remote education settings, where traditional methods may fall short.<sup>37</sup> By watching videos, students can gain a clearer understanding of the internal structure and physiological functions of the human body. The application of cognitive load theory (CTL) in designing these educational technologies can further optimize their effectiveness. By structuring video content to align with CLT principles, educators can minimize extraneous cognitive load and enhance intrinsic cognitive load, leading to better learning outcomes. A scoping review highlighted that CLT-based interventions in medical education often yield positive results, suggesting that thoughtful integration of technology can foster a more productive learning environment.<sup>38</sup> When used in designing medical exams, Sora can be used to generate error example videos with incorrect operation sequences or violated medical regulations. These negative materials add value to deepen the judgement of the potential harm of erroneous behavior and the importance of correct practices through comparing correct operating procedures, thereby avoiding similar errors in practice, and promote clinical decision-making abilities.

## Conclusions

The integration of AI technology into medical education represents a significant trend for its future development. Against the limitations brought about by the use of question banks, AIGC in medical examinations is an innovative approach that enhances medical students' foundational knowledge, clinical understanding, and practical skills. This study offers a practical solution in designing MCQs, case-based questions, and Video question. It is anticipated that future development will incorporate with technologies like RAG, multi-agent system, and VGT. By leveraging these tools, educators can create more intuitive and vivid questions that not only enhance student learning but also improve patient education and engagement. The ongoing exploration of innovative educational technologies will undoubtedly continue to shape the future of medical training and practice.

## Abbreviations

AIGC, artificial intelligence generated content; ChatGPT, Chat Generative Pre-trained Transformer; CLT, cognitive load theory; MAS, multi-agent systems; MCQ, multiple-choice questions; REG, retrieval augmented generation; VGT, video generation technology.

## Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

## Funding

This work is financially supported by the grants from the Teaching and Research Project of the Second Clinical College of Tongji Medical College, Huazhong University of Science and Technology (TJXYJ2023016; TJSZ2024016).

## Disclosure

The authors declare that they have no competing interests in this work.

## References

1. Chen X, Hu Z, Wang C. Empowering education development through AIGC: a systematic literature review. *Educat Inform Technol*. 2024;29(13):17485–17537. doi:10.1007/s10639-024-12549-7
2. A. Fuller K, Morbitzer KA, Zeeman JM, M. Persky A, C. Savage A, McLaughlin JE. Exploring the use of ChatGPT to analyze student course evaluation comments. *BMC Med Educ*. 2024;24(1):423. doi:10.1186/s12909-024-05316-2
3. Eysenbach G. The role of ChatGPT, generative language models, and artificial intelligence in medical education: a conversation with ChatGPT and a call for papers. *JMIR Med Educ*. 2023;9:e46885. doi:10.2196/46885
4. Logan JM, Thompson AJ, Marshak DW. Testing to enhance retention in human anatomy. *Anatomical Sci Educ*. 2011;4(5):243–248. doi:10.1002/ase.250

5. Fisher J, Leahy D, Lim JJ, Astles E, Salvatore J, Thomson R. Question banks: credit? Or debit? A qualitative exploration of their use among medical students. *BMC Med Educ.* 2024;24(1):569. doi:10.1186/s12909-024-05517-9
6. Xu L, Jiang Z, Cai F, Ouyang J, Liu H, Cai T. Optimizing a national examination for medical undergraduates via modern automated test assembly approaches. *BMC Med Educ.* 2024;24(1):919. doi:10.1186/s12909-024-05905-1
7. Klang E, Portugez S, Gross R, et al. Advantages and pitfalls in utilizing artificial intelligence for crafting medical examinations: a medical education pilot study with GPT-4. *BMC Med Educ.* 2023;23(1). doi:10.1186/s12909-023-04752-w
8. Mahoney MT, Linkowski LC, Wu TC, et al. Exploring radiation oncology representation on the national board of medical examiners (NBME) official practice material for the undergraduate United States national standardized medical board examinations. *Int J Radiat Oncol Biol Phys.* 2023;116(3):E7–E8. doi:10.1016/j.ijrobp.2023.03.015
9. DeSantis M, McKean TA. Efficient validation of teaching and learning using multiple-choice exams. *Adv Physiol Educ.* 2003;27(1):3–14. doi:10.1152/advan.00016.2001
10. Mackillop L, Parker-Swift J, Crossley J. Getting the questions right: non-compound questions are more reliable than compound questions on matched multi-source feedback instruments. *Medical Educ.* 2011;45(8):843–848. doi:10.1111/j.1365-2923.2011.03996.x
11. Kumar B, Suneja M, Swee ML. Development and test-item analysis of a freely available 1900-item question bank for rheumatology trainees. *Cureus J Med Sci.* 2021;13(9).
12. Alade KH, Marin JR, Constantine E, et al. Development of a novel pediatric point-of-care ultrasound question bank using a modified Delphi process. *Aem Educ Training.* 2021;5(4). doi:10.1002/aet2.10651
13. Baños JH, Pepin ME, Van Wagoner N. Class-wide access to a commercial step 1 question bank during preclinical organ-based modules: a pilot project. *Academic Med.* 2018;93(3):486–490. doi:10.1097/ACM.0000000000001861
14. White LJ, McGowan HW, McDonald AC. The effect of content delivery style on student performance in anatomy. *Anatomical Sci Educ.* 2019;12(1):43–51. doi:10.1002/ase.1787
15. Mehta S, Mehta N, Benjamin J, Mehta S, MacNeill H, Masters K. Embracing the illusion of explanatory depth: a strategic framework for using iterative prompting for integrating large language models in healthcare education. *Med Teach.* 2024;1–4. doi:10.1080/0142159X.2024.2418937
16. De Busser B, Roth L, De Loof H. The role of large language models in self-care: a study and benchmark on medicines and supplement guidance accuracy. *Int J Clin Pharmacy.* 2024. doi:10.1007/s11096-024-01839-2
17. Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT, GPT-4, and google bard on a neurosurgery oral boards preparation question bank. *Neurosurgery.* 2023;93(5):1090–1098. doi:10.1227/neu.0000000000002551
18. Guerra GA, Hofmann H, Sobhani S, et al. GPT-4 artificial intelligence model outperforms ChatGPT, medical students, and neurosurgery residents on neurosurgery written board-like questions. *World Neurosurg.* 2023;179:E160–E165. doi:10.1016/j.wneu.2023.08.042
19. Kiyak YS, Emekli E. ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: a literature review. *Postgraduate Med J.* 2024;100(1189):858–865. doi:10.1093/postmj/qgae065
20. Rashwan NI, Aref SR, Nayel OA, Rizk MH. Postexamination item analysis of undergraduate pediatric multiple-choice questions exam: implications for developing a validated question Bank. *BMC Med Educ.* 2024;24(1). doi:10.1186/s12909-024-05153-3
21. Collins J. Education techniques for lifelong learning - Writing multiple-choice questions for continuing medical education activities and self-assessment modules. *Radiographics.* 2006;26(2):543–U516. doi:10.1148/rg.262055145
22. Bond AE, Bodger O, Skibinski DO, et al. Negatively-marked MCQ assessments that reward partial knowledge do not introduce gender bias yet increase student performance and satisfaction and reduce anxiety. *PLoS One.* 2013;8(2):e55956. doi:10.1371/journal.pone.0055956
23. Sun Y, Li X, Liu H, et al. The effectiveness of using situational awareness and case-based seminars in a comprehensive nursing skill practice course for undergraduate nursing students: a quasi-experimental study. *BMC Med Educ.* 2024;24(1):118. doi:10.1186/s12909-024-05104-y
24. Zhu Z, Ying Y, Zhu J, Wu H. ChatGPT's potential role in non-English-speaking outpatient clinic settings. *Digital Health.* 2023;9. doi:10.1177/20552076231184091
25. Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding. Paper presented at: 36th Conference on Neural Information Processing Systems (NeurIPS); 2022; Electr Network.
26. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. Paper presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022; New Orleans, LA.
27. Xu Y, Sun L, Peng W, et al. MedSyn: text-guided anatomy-aware synthesis of high-fidelity 3-D CT images. *IEEE Transactions Med Imag.* 2024;43(10):3648–3660. doi:10.1109/TMI.2024.3415032
28. Karakus A, Senyer N. The preparedness level of final year medical students for an adequate medical approach to emergency cases: computer-based medical education in emergency medicine. *Int J Emerg Med.* 2014;7(1):3. doi:10.1186/1865-1380-7-3
29. Bloom TJ, Rich WD, Olson SM, Adams ML. Perceptions and performance using computer-based testing: one institution's experience. *Curr Pharm Teach Learn.* 2018;10(2):235–242. doi:10.1016/j.cptl.2017.10.015
30. Thirunavukarasu AJ, Hassan R, Mahmood S, et al. Trialling a large language model (ChatGPT) in general practice with the applied knowledge test: observational study demonstrating opportunities and limitations in primary care. *JMIR Med Educ.* 2023;9:e46599. doi:10.2196/46599
31. Buhr CR, Smith H, Huppertz T, et al. ChatGPT versus consultants: blinded evaluation on answering otorhinolaryngology case-based questions. *JMIR Med Educ.* 2023;9:e49183. doi:10.2196/49183
32. Wu JT, Shenoy ES, Carey EP, Alterovitz G, Kim MJ, Branch-Elliman W. ChatGPT: increasing accessibility for natural language processing in healthcare quality measurement. *Infect Control Hosp Epidemiol.* 2024;45(1):9–10. doi:10.1017/ice.2023.236
33. Saibene A, Assale M, Giltri M. Expert systems: definitions, advantages and issues in medical field applications. *Expert Syst Appl.* 2021;177:114900. doi:10.1016/j.eswa.2021.114900
34. Zhou Q, Liu C, Duan Y, et al. GastroBot: a Chinese gastrointestinal disease chatbot based on the retrieval-augmented generation. *Front Med.* 2024;11:1392555. doi:10.3389/fmed.2024.1392555
35. Yazaki M, Maki S, Furuya T, et al. Emergency patient triage improvement through a retrieval-augmented generation enhanced large-scale language model. *Prehospital Emerg Care.* 2024. 1–7. doi:10.1080/10903127.2024.2374400
36. Chiou EK, Lee JD. Cooperation in human-agent systems to support resilience: a microworld experiment. *Human Factors.* 2016;58(6):846–863. doi:10.1177/0018720816649094



37. Garbin CAS, Pacheco Filho AC, Garbin AJI, Pacheco K. Instructional video as a teaching/learning tool in times of remote education: a viable alternative. *J Dental Educ.* 2021;85(Suppl 3):2034–2035. doi:10.1002/jdd.12536
38. Hochstrasser K, Stoddard HA. Use of cognitive load theory to deploy instructional technology for undergraduate medical education: a scoping review. *Med Sci Educator.* 2022;32(2):553–559. doi:10.1007/s40670-021-01499-1

#### Advances in Medical Education and Practice

**Dovepress**  
Taylor & Francis Group

#### Publish your work in this journal

Advances in Medical Education and Practice is an international, peer-reviewed, open access journal that aims to present and publish research on Medical Education covering medical, dental, nursing and allied health care professional education. The journal covers undergraduate education, postgraduate training and continuing medical education including emerging trends and innovative models linking education, research, and health care services. The manuscript management system is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <http://www.dovepress.com/advances-in-medical-education-and-practice-journal>