Generative artificial intelligence in graduate medical education

Ravi Janumpally, Suparna Nanua, Andy Ngo and
Kenneth Youens*

Clinical Informatics Fellowship Program, Baylor Scott & White Health, Round Rock, TX, United States

Generative artificial intelligence (GenAI) is rapidly transforming various sectors, including healthcare and education. This paper explores the potential opportunities and risks of GenAI in graduate medical education (GME). We review the existing literature and provide commentary on how GenAI could impact GME, including five key areas of opportunity: electronic health record (EHR) workload reduction, clinical simulation, individualized education, research and analytics support, and clinical decision support. We then discuss significant risks, including inaccuracy and overreliance on AI-generated content, challenges to authenticity and academic integrity, potential biases in AI outputs, and privacy concerns. As GenAI technology matures, it will likely come to have an important role in the future of GME, but its integration should be guided by a thorough understanding of both its benefits and limitations.

1 Introduction

Generative artificial intelligence (GenAI) is a relatively new technology that uses advanced machine learning models to generate humanlike expression. Large language models (LLMs) like ChatGPT (OpenAI, San Francisco, United States) rely on a machine learning architecture called a "transformer." A key feature of transformers is their self-attention mechanism, which allows the model to assess the importance of words in a sequence relative to one another, enhancing its ability to understand context and, when trained on vast amounts of data, resulting in a remarkable ability to understand and generate humanlike text [1]. Such models excel at tasks like document summarization, sentiment analysis, question answering, text classification, translation, text generation, and as conversational chatbots. Related models called large vision models (LVMs), Vision-Language Models (VLMs), large multimodal models (LMMs), diffusion models, and generative adversarial networks (GANs) provide similar or overlapping functionality for image, audio, and video processing and generation. It is widely believed that GenAI will have far-reaching societal impact and will be incorporated into multiple aspects of our daily lives [2, 3]. GenAI has the potential to revolutionize multiple industries, with healthcare and education among the likely targets.

In healthcare, GenAI has shown promise in a broad range of applications such as clinical decision support, medical education, clinical documentation, research support, and as a communication tool [4]. GenAI models like ChatGPT, even without special fine-tuning for medical knowledge, achieve performance at or near the passing threshold on all three United States Medical Licensing Examination (USMLE) Step exams [5]. Studies evaluating performance on medical specialty board examination-or in-service examination-level questions have shown mixed results, but in some cases LLM performance has approached that of senior medical trainees [6–9]. GenAI-powered tools are deployed in production clinical

environments today, most notably in the patient care-adjacent domains of clinical documentation (10) and provider-patient communication, where they have shown promise in improving EHR-related provider inefficiency and burnout (11, 12).

In the medical educational setting, GenAI potentially offers multiple benefits such as easy personalization of learning experiences, simulation of real-world scenarios and patient interactions, and practicing communication skills (13). These potential gains are balanced by meaningful risks, such as the trustworthiness of AI-generated content, the deepening of socioeconomic inequalities, and challenges to academic integrity (14, 15).

Graduate medical education (GME) shares many characteristics with undergraduate medical education and with other types of healthcare education. As adult learners, medical trainees are theorized to learn best when self-motivated, self-directed, and engaged with task-centered, practical topics (16). Historically, medical education used time spent in the training environment as a proxy for learning success. More recently, there has been renewed interest in competency-based medical education (CBME), a paradigm that uses achievement of specific competencies rather that time spent (or other structural measures) as the key measure of learning success (17, 18). CBME serves as the foundation of the Accreditation Council for Graduate Medical Education (ACGME)'s accreditation model, and is the key theory underpinning the formative "Milestones" used by ACGME-accredited programs to assess trainee development and to improve education (19).

Having built a foundation in medical sciences and basic clinical skills in medical school, GME trainees spend little time in the classroom, with most of their learning occurring with real patients as they function as members of the healthcare team. A core tenet of GME is "graded authority and responsibility," where trainees progressively gain autonomy until they achieve the skills to practice independently. Additionally, trainees are expected to become "physician scholars"; participants in ACGME-accredited GME programs participate in scholarly pursuits like research, academic writing, quality improvement, and creation of educational curricula (20).

In this paper, we present concise summary of the existing literature (Table 1) and commentary on the potential opportunities and risks of GenAI in the GME setting.

2 Opportunities

2.1 EHR workload reduction

Given their long work hours and stressful work environment, GME trainees are particularly susceptible to burnout, with rates higher than their age-matched peers in non-medical careers and higher than early-career attending physicians (21). Burnout among the academic physicians who comprise most GME faculty also occurs, and may impact the quality of training they are able to deliver (22, 23). Thus, innovations that prevent overwork and burnout have the potential to benefit GME trainees and faculty.

One unintended consequence of the adoption of electronic health records (EHRs) has been a dramatic increase in time spent in documenting clinical encounters. Many physicians now spend as much time documenting in the EHR as they do in patient-facing activities (24). This documentation burden can result in medical

errors, threats to patient safety, poor quality documentation, and attrition, and is a major cause of physician burnout (25). Various strategies have been tried to reduce physician documentation burden, including medical scribes and various educational interventions, workflow improvements, and other strategies (26). Given its ability to summarize, translate and generate text, GenAI demonstrates clear potential as a technological aid to alleviate the burden of clinical documentation. The most notable current application is ambient listening tools that use GenAI to transcribe and analyze patient-doctor conversations, converting them into structured draft clinical notes that the physician would theoretically only then need to review for accuracy. Numerous organizations are piloting such technology as of the time of this writing (27), though the few results published so far about real-world performance have been mixed (10, 28, 29). Examples of other less commercially mature concepts for how GenAI could reduce clinical documentation burden include tools to improve medical coding accuracy (30), to generate clinical summary documentation like discharge summaries (31), and to draft GME faculty supervisory notes (32).

In addition to documenting clinical encounters, physicians (including GME trainees) spend large amounts of time in the EHR managing inbox messages, including patient messages, information about tests results, requests for refills, requests to sign clinical orders, and various administrative messages (33). As another major contributor to workload, EHR inbox management is also a cause of burnout (34, 35). This problem came to be of particular importance during the COVID-19 pandemic, where patient messaging increased by 157% compared to pre-pandemic levels (35). LLMs have shown the ability to draft high-quality, "empathetic" responses to patient questions (36). Early efforts to use LLMs for drafting replies to patient inbox messages have shown promising results, with multiple studies showing that LLMs can draft responses of good quality (37, 38) and at least one study showing good provider adoption with significant reductions in provider assessment of multiple burnout-related metrics (11). Multiple health information technology companies, including the largest United States EHR vendor, have already brought GenAI functionality for EHR inbox management to market (39–41).

2.2 Clinical simulation

Simulation-based medical education (SBME) has evolved significantly since the early use of mannequins for basic life support training 60 years ago, and simulation using high-fidelity mannequins and virtual and augmented reality tools are now a vital component of GME. There is a substantial body of evidence confirming the benefits of simulation-based training and the successful transfer of these skills to real patients (42, 43). Simulations are used both to educate and to assess performance in GME. For example, the American Board of Anesthesiology incorporates an Objective Structured Clinical Examination (OSCE) meant to assess communication and professionalism, as well as technical skills, into the board examination process for anesthesiology residents (44). Many of the current applications of SBME in GME are targeted at procedural skills like complex surgical techniques, bridging the gap for trainees' experiential

TABLE 1 Literature on GenAI in the GME setting.

Specialty	First author (Publication date)	Title (Citation)	Brief description
Administration	Mangold, S (2024)	Artificial Intelligence in Graduate Medical Education Applications (101)	Commentary on the use of GenAI in GME application materials.
Administration	Quinonez, S (2024)	ChatGPT and Artificial Intelligence in Graduate Medical Education Program Applications	Commentary on the use of GenAI in GME application materials.
Administration	Zumsteg, J (2023)	Will ChatGPT Match to Your Program? (97)	Commentary on the use of GenAI in GME application materials.
Anesthesiology	Sardesai, N (2023)	Utilizing Generative Conversational Artificial Intelligence to Create Simulated Patient Encounters: A Pilot Study for Anaesthesia Training (48)	Study using an LLM to simulate patient conversations for trainees regarding certain anesthesia procedures. The tool showed good accuracy in simulating patient responses and behavior.
Dermatology	Ayub, I (2023)	Exploring the Potential and Limitations of Chat Generative Pre-trained Transformer (ChatGPT) in Generating Board-Style Dermatology Questions: A Qualitative Analysis	Study using an LLM to generate board exam-style dermatology questions, showing poor performance of the model in generating accurate and appropriate questions.
Dermatology	Breslavets, M (2024)	Advancing dermatology education with AI-generated images.	Commentary with examples using a GAN to generate synthetic clinical images for dermatology training.
Dermatology	Lim, S (2024)	Exploring the Potential of DALL-E 2 in Pediatric Dermatology: A Critical Analysis	Study using a diffusion model to generate synthetic clinical images of dermatologic conditions for dermatology training, showing poor performance of the model for most tested conditions.
Emergency Medicine	Barak-Corren, Y (2024)	Harnessing the Power of Generative AI for Clinical Summaries: Perspectives From Emergency Physicians (32)	Study using an LLM to generate clinical supervisory notes, showing a significant reduction the in time and effort required to create notes, without any reduction in note quality on simpler notes.
Emergency Medicine	Webb, J (2023)	Proof of Concept: Using ChatGPT to Teach Emergency Physicians How to Break Bad News (49)	Proof-of-concept study using ChatGPT to roleplay breaking bad news to patients.
Neurosurgery	Arfaie, S (2024)	ChatGPT and Neurosurgical Education: A Crossroads of Innovation and Opportunity (90)	Review of the literature and summary of the uses of GenAI for educating neurosurgical trainees.
Neurosurgery	Bartoli, A (2024)	Probing Artificial Intelligence in Neurosurgical Training: ChatGPT Takes a Neurosurgical Resident's Written Exam (111)	Study evaluating the performance of an LLM in generating board examination-style questions, showing poor performance of the LLM in generating a small trial set of exam-quality questions.
Neurosurgery	McLean, A (2024)	Application of Transformer Architectures in Generative Video Modeling for Neurosurgical Education (112)	Detailed description of a planned study that would use a diffusion model to create synthetic neurosurgical training videos.
Ophthalmology	Sevgi, M (2024)	Medical Education with Large Language Models in Ophthalmology: Custom Instructions and Enhanced Retrieval Capabilities (113)	Description of tools using LLMs to teach clinical guidelines in ophthalmology and to summarize current ophthalmology research.
Orthopedic Surgery	DeCook, R (2024)	AI-Generated Graduate Medical Education Content for Total Joint Arthroplasty: Comparing ChatGPT Against Orthopedic Fellows (114)	Study using an LLM to generate educational summaries of total joint arthroplasty-related topics, showing that the LLM created better orthopedic training content than orthopedic fellows across several topics and domains.
Pathology	Cecchini, M (2024)	Harnessing the Power of Generative Artificial Intelligence in Pathology Education (18)	Review of the literature and summary of the uses of GenAI for educating pathology trainees.
Pediatrics	Ba, H (2024)	Enhancing Clinical Skills in Pediatric Trainees: A Comparative Study of ChatGPT-Assisted and Traditional Teaching Methods (115)	Study comparing LLM-assisted instruction with traditional instruction on pediatric clinical skill education, showing comparable or better performance of the LLM-assisted method.

(Continued)

TABLE 1 (Continued)

Specialty	First author (Publication date)	Title (Citation)	Brief description
Pediatrics	Suresh, S (2024)	Large Language Models in Pediatric Education: Current Uses and Future Potential (116)	Review of the literature and summary of the uses of GenAI for educating pediatrics trainees, showing that LLM-assisted instruction did not affect theoretical knowledge application but did enhance practical clinical skills.
Pediatrics	Waikel, R (2023)	Generative Methods for Pediatric Genetics Education (56)	Study using synthetic images of individuals with uncommon genetic conditions to train pediatric residents, showing that the synthetic images performed similarly but were slightly less helpful than real patient images.
Primary Care	Parente, D (2024)	Generative Artificial Intelligence and Large Language Models in Primary Care Medical Education (59)	Review of the literature and summary of the uses of GenAI for educating primary care trainees.
Radiology	Lyo, S (2024)	From Revisions to Insights: Converting Radiology Report Revisions into Actionable Educational Feedback Using Generative AI Models (60)	Study using an LLM to compare preliminary (trainee) and finalized radiology reports, identify discrepancies, and suggest review topics. The LLM consistently and accurately identified discrepancies and suggested relevant feedback.
Radiology	Meşe, I (2024)	Educating the Next Generation of Radiologists: A Comparative Report of ChatGPT and E-Learning Resources (117)	Review of the literature and summary of the uses of GenAI for educating radiology trainees.
Radiology	Mistry, N (2024)	Large Language Models as Tools to Generate Radiology Board-Style Multiple-Choice Questions (61)	Study using two LLMs to generate board exam-style radiology questions, demonstrating that one LLM generated questions of equivalent quality to real American College of Radiology in-service exam questions.
Surgery	Lia, H (2024)	Cross-Industry Thematic Analysis of Generative AI Best Practices: Applications and Implications for Surgical Education and Training (118)	Analysis of ethical considerations when integrating GenAI into surgical education, with example use cases.
Surgery	Sathe, T (2024)	How I GPT It: Development of Custom Artificial Intelligence (AI) Chatbots for Surgical Education (119)	Commentary on the use of GenAI chatbots for surgical education with description of several potential use cases.

Summary of existing literature of which we are aware focusing on GenAI in GME. The summary excludes papers focused on mainly on testing LLM performance on medical knowledge tasks, papers on non-GME-specific clinical, educational or academic applications of GenAI, and papers about artificial intelligence in general.

learning on invasive, uncommon, or high-acuity procedures (45). The integration of artificial intelligence into clinical simulations would theoretically allow for the customization of scenarios based on a trainee's skill level and performance data, providing a personalized learning experience and potentially opening the door to new types of patient simulation (43). Accordingly, there has been interest in using conversational GenAI to simulate patient encounters to practice cognitive and communication skills, though this application is more often focused on undergraduate medical education (15, 33, 46–49).

Among the most interesting potential applications of GenAI in GME is the concept of using synthetic data as training material for visual diagnosis. GANs and diffusion models have shown promise in generating realistic images of pathology findings (50, 51), skin lesions (52–54), chest X-rays (55), genetic syndromes (56), and ophthalmological conditions (57). The synthetic data approach may ultimately address important limitations in image-based training data

sets, such as underrepresentation of certain patient demographics and adequate demonstration of rare findings.

2.3 Individual education

Individualized tutoring produces better academic outcomes than learning in a traditional classroom setting (58). Skilled teachers can guide learners at different levels through complex topics, offering tailored and accessible explanations. One-on-one tutoring delivered by humans is costly, and skilled teachers are not available everywhere, but GenAI tools may have some of the same benefits at a fraction of the cost. LLMs show promise as a tool for explaining challenging concepts to graduate medical trainees in a manner tailored to the learner's level (18), and LLMs could be configured to act as personalized tutors (59). In one study, an LLM successfully reviewed trainee-generated radiology reports and generated relevant

educational feedback, a concept which could be extended to other types of clinical documentation (60).

GME trainees preparing for board examinations often use question banks to study, and GME programs use board-exam style questions to assess trainee progress. Question generation can be a costly and labor-intensive proposition (61). Authors report mixed success with using LLMs to generate board exam-style questions (61, 62), but as the technology matures, it seems likely that LLMs will be used by trainees and educators alike to create high-quality self-directed study materials and test questions.

2.4 Research and analytics support

LLMs are powerful tools for academic research and writing, and can assist in idea generation, processing complex background information, and proposing testable hypotheses (63, 64). LLMs readily summarize complex academic papers and draft academic text, abilities that can accelerate academic productivity (65). When paired with reliable academic databases and search engines and/or when fine-tuned with specific knowledge, LLMs do a serviceable job of conducting literature reviews (66, 67), synthesizing findings from existing literature, and drafting new scientific text with accurate literature citations (68). LLMs have great utility in assisting non-native English speakers with academic writing, representing a cost-effective and always-available alternative to commercial editing and proofreading services or to searching for native English-speaking collaborators (69).

Among the competencies listed in the ACGME's Common Program Requirements is the ability to "systematically analyze practice using quality improvement (QI) methods" (20). GME trainees are required to participate in QI projects, which are typically require quantitative data analysis. Trainees are often underrepresented in organizational quality improvement activities, with one potential reason being the substantial time and effort needed for data collection and analysis (70). LLMs have some ability to facilitate straightforward data analysis and can generate serviceable code for statistical and programming tasks (71). LLMs are also adept at natural language processing tasks like extracting structured data from unstructured medical text (72).

2.5 Clinical decision support

Computer-based clinical decision support (CDS) systems are among the most effective tools for guiding good clinical decision-making (73). For GME trainees, CDS that provides authoritative, evidence-based guidance has both great practical clinical and educational utility (73). CDS that delivers evidence-based clinical guidance based on relevant patient data is a required feature for EHR systems certified by the United States government (74). A widely accepted CDS framework explains that CDS should be delivered according to the "five rights": the right information, to the right person, in the right format, through the right channel, at the right time (75). Most current CDS consists of rule-based expert systems that display alerts to providers. While such systems are effective, rule-based alerts often suffer from practical problems such as a lack of specificity, poor timing, and incomplete characterization of clinical context (76).

The potential for intelligent, interactive, authoritative, LLMs to serve as always-available clinical consultants and educators has generated compelling speculation (77). LLMs can provide context-sensitive and specific guidance incorporating clinical context and patient data, they can be accessed through readily available communication channels, and--in contrast to rule-based alerts--they are interactive. However, studies done to evaluate the potential of LLMs for clinical decision support in various clinical contexts (78–83) have shown mixed results so far, with limitations in their ability to handle nuanced judgment and highly specialized decision-making. Thus, while GenAI for CDS is an area of great potential and ways to improve performance are under development, GME faculty and trainees cannot yet rely on LLMs to directly guide clinical care.

3 Risks

Despite its recent public availability, GenAI use is widespread and continues to grow quickly in both business and personal contexts. ChatGPT has the fastest-growing user base of any consumer web application in history (84), and a McKinsey & Company survey in early 2024 reported that 65% of businesses are regularly using generative AI, a rate nearly twice the year before (85). In another McKinsey report, more than 70% of healthcare leaders say they are using or pursuing GenAI technologies in their organizations (86). This explosive growth will undoubtedly have many benefits, but there are there are practical risks associated with GenAI that should temper optimism. Below we summarize the principal known risks as applicable in the GME setting:

3.1 Inaccuracy and overreliance

In essence, LLMs are statistical models that predict the most likely continuation of a given input sequence, based on their training data. Sometimes this approach results in plausible sounding but factually incorrect outputs, a phenomenon called "hallucination." This problem can be especially difficult when dealing with topics requiring nuanced understanding of context or specialized knowledge, conditions very common in healthcare and specialized academic settings. For example, a biomedical researcher recently reported a cautionary tale in which ChatGPT generated incorrect information about tick ecology, complete with an entirely fake but plausible-appearing source document citation (87). In clinical settings, LLMs have been found to occasionally add fabricated information to clinical documentation (88) and to provide incorrect clinical recommendations (89).

In GME, trainees learn in a real clinical environment where accuracy and context are critical. There is a risk that overreliance on LLMs can result in an incomplete or incorrect understanding of complex topics, contributing to a poor educational outcomes, loss of critical thinking skills, and/or to suboptimal care and patient harm (15, 90, 91). Techniques like retrieval augmented generation, fine-tuning and prompt engineering show promise in reducing or eliminating the problems of inaccuracy and hallucination (92–94), but at present, reliance on GenAI as a source of factual information in any important clinical or academic context is risky. In our view, assertions made by GenAI should be validated by the user to avoid

misinformation, and GME trainees should not use GenAI to directly guide patient care decisions outside of a controlled research context. GenAI users should be aware of automation bias, a cognitive bias in which people tend to excessively trust automated systems (95).

3.2 Authenticity and integrity

In reviewing applications for GME positions, personal statements are one of the most important elements that program directors review (96), especially in modern era where in-person residency and fellowship interviews are less common. Personal statements allow program directors to assess an applicant's interest in their program and the clarity, organization and effectiveness of their written communication (97). There have long been concerns about plagiarism in personal statements (98, 99), and these concerns are magnified by GenAI tools that can readily produce writing that is clear, well-structured and compelling but that lacks an applicant's unique voice, style and values (100, 101). Similarly, through letters of recommendation (LORs), faculty advocate for applicants by highlighting qualities observed in longitudinal relationships; using GenAI to draft LORs may have benefits but raises similar concerns about authenticity (102). GenAI-written content can be difficult to detect, even with software assistance (103). Some authors recommend that program draft policies for the use of GenAI in personal statements and LORs, with a common recommendation being that the use of GenAI be disclosed by the writer (97, 101).

As noted above, GME trainees are also expected to participate in research, academic writing, quality improvement summaries, creation of educational curricula, and similar academic activities. There are currently no consensus standards for using GenAI in academic medicine, but a recent review synthesized existing papers into a proposed set of guidelines, paraphrased as: (1) LLMs should not be cited as coauthors in academic works, (2) LLMs should not be used to produce the entirety of manuscript text, (3) authors should have an understanding of how LLMs work, (4) humans are accountable for content created by the LLM, and (5) use of an LLM should be clearly acknowledged in any resulting manuscripts (104).

3.3 Bias

GenAI tools are typically trained on huge corpora of data from the internet such as informational web sites, public forums, books, research literature, and other digitized media. Given the uncontrolled nature of the training data, it is unsurprising that they can exhibit social bias and stereotypes in their output (105). If unmanaged, these biases have the potential to reinforce detrimental beliefs and behaviors (106). In healthcare, GenAI may overrepresent, underrepresent or mis-characterize certain groups of people or certain medical conditions (18).

3.4 Privacy and security

GenAI is computationally intensive and expensive to operate. Thus, many resource-limited healthcare organizations or individual physicians may rely on third-party, external GenAI tools (107). Given the great utility of GenAI, knowledge workers may be sorely tempted to upload confidential information, despite significant risks (108). In healthcare, such risks are legal as well as ethical in nature, and transgressions can have implications for professional development.

4 Conclusion

Though the timeline is uncertain, GenAI technology will continue to advance. There is little question that GenAI will come to have a key role in the medical education landscape. We are optimistic about the potential of GenAI to enhance GME for both learners and educators, but enthusiasm should be tempered by a realistic understanding of the risks and limitations of this technology. We believe specific education on artificial intelligence should be included in medical curricula, and that research should continue on the risks and benefits of artificial intelligence as a tool for medical education (109, 110).

Author contributions

RJ: Writing – original draft. SN: Writing – original draft. AN: Writing – original draft. KY: Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Funding

Conflict of interest

Authors RJ, SN, AN and KY were employed by company Baylor Scott & White Health.

All authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Generative AI was used in the creation of this manuscript. GPT 4o (version 2024-08-06, OpenAI) was used to refine each individual contributor's section(s) of the manuscript draft into a more cohesive writing style.

Publisher's note

References

1. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv; (2023). Available from: http://arxiv.org/abs/1706.03762 (Accessed November 4, 2024).

2. Ooi KB, Tan GWH, Al-Emran M, Al-Sharafi MA, Capatina A, Chakraborty A, et al. The potential of generative artificial intelligence across disciplines: perspectives and future directions. *J Comput Inf Syst* (2023) 1–32. doi: 10.1080/08874417.2023.2261010

3. Brynjolfsson E, Li D, Raymond L. Generative AI at work. Cambridge, MA: National Bureau of Economic Research (2023). w31161 p.

4. Moulaei K, Yadegari A, Baharestani M, Farzanbakhsh S, Sabet B, Reza AM. Generative artificial intelligence in healthcare: a scoping review on benefits, challenges and applications. *Int J Med Inform*. (2024) 188:105474. doi: 10.1016/j.ijmedinf.2024.105474

5. Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. *PLOS Digit Health*. (2023) 2:e0000198. doi: 10.1371/journal.pdig.0000198

6. Kung JE, Marshall C, Gauthier C, Gonzalez TA, Jackson JB. Evaluating ChatGPT performance on the Orthopaedic in-training examination. JB JS Open. *Access*. (2023) 8:e23.00056. doi: 10.2106/JBJS.OA.23.00056

7. Lum ZC. Can artificial intelligence pass the American Board of Orthopaedic Surgery Examination? Orthopaedic residents versus ChatGPT. *Clin Orthop Relat Res*. (2023) 481:1623–30. doi: 10.1097/CORR.0000000000002704

8. Cheong RCT, Pang KP, Unadkat S, Mcneillis V, Williamson A, Joseph J, et al. Performance of artificial intelligence chatbots in sleep medicine certification board exams: ChatGPT versus Google bard. *Eur Arch Otorrinolaringol*. (2024) 281:2137–43. doi: 10.1007/s00405-023-08381-3

9. Khan AA, Yunus R, Sohail M, Rehman TA, Saeed S, Bu Y, et al. Artificial intelligence for anesthesiology board-style examination questions: role of large language models. *J Cardiothorac Vasc Anesth*. (2024) 38:1251–9. doi: 10.1053/j.jvca.2024.01.032

10. Liu TL, Hetherington TC, Stephens C, McWilliams A, Dharod A, Carroll T, et al. AI-powered clinical documentation and clinicians' electronic health record experience: a nonrandomized clinical trial. *JAMA Netw Open*. (2024) 7:e2432460. doi: 10.1001/jamanetworkopen.2024.32460

11. Garcia P, Ma SP, Shah S, Smith M, Jeong Y, Devon-Sand A, et al. Artificial intelligence-generated draft replies to patient inbox messages. *JAMA Netw Open*. (2024) 7:e243201. doi: 10.1001/jamanetworkopen.2024.3201

12. Small WR, Wiesenfeld B, Brandfield-Harvey B, Jonassen Z, Mandal S, Stevens ER, et al. Large language model-based responses to patients' in-basket messages. *JAMA Netw Open*. (2024) 7:e2422399. doi: 10.1001/jamanetworkopen.2024.22399

13. Boscardin CK, Gin B, Golde PB, Hauer KE. ChatGPT and generative artificial intelligence for medical Education: potential impact and opportunity. *Acad Med*. (2024) 99:22–7. doi: 10.1097/ACM.0000000000005439

14. Bhardwaj P, Bookey L, Ibironke J, Kelly N, Sevik IS. A Meta-analysis of the economic, social, legal, and cultural impacts of widespread adoption of large language models such as ChatGPT|OxJournal. (2023). Available from: https://www.oxjournal.org/economic-social-legal-cultural-impacts-large-language-models/ (Accessed November 4, 2024).

15. Knopp MI, Warm EJ, Weber D, Kelleher M, Kinnear B, Schumacher DJ, et al. AI-enabled medical Education: threads of change, promising futures, and risky realities across four potential future worlds. *JMIR Med Educ*. (2023) 9:e50373. doi: 10.2196/50373

16. Knowles MS III, Swanson RA. The adult learner: The definitive classic in adult Education and human resource development. *7th* ed. London New York: Butterworth-Heinemann (2011). 424 p.

17. Carraccio C, Englander R, Van Melle E, Ten Cate O, Lockyer J, Chan MK, et al. Advancing competency-based medical Education: a charter for clinician-educators. *Acad Med*. (2016) 91:645–9. doi: 10.1097/ACM.0000000000001048

18. Cecchini MJ, Borowitz MJ, Glassy EF, Gullapalli RR, Hart SN, Hassell LA, et al. Harnessing the power of generative artificial intelligence in pathology Education. *Arch Pathol Lab Med*. (2024). doi: 10.5858/arpa.2024-0187-RA

19. Edgar L, McLean S, Hogan S, Hamstra S, Holmboe E. The milestones guidebook. Accreditation Council for Graduate Medical Education. (2020).

20. Accreditation Council for Graduate Medical Education. ACGME common program requirements (residency). (2023). Available from: https://www.acgme.org/globalassets/pfassets/programrequirements/cprresidency_2023.pdf (Accessed November 5, 2024).

21. Dyrbye LN, West CP, Satele D, Boone S, Tan L, Sloan J, et al. Burnout among U.S. medical students, residents, and early career physicians relative to the general U.S. population. *Acad Med*. (2014) 89:443–51. doi: 10.1097/ACM.0000000000000134

22. Shah DT, Williams VN, Thorndyke LE, Marsh EE, Sonnino RE, Block SM, et al. Restoring faculty vitality in academic medicine when burnout threatens. *Academic Med: J Assoc American Medical Colleges*. (2018) 93:979–84. doi: 10.1097/ACM.0000000000002013

23. Nassar AK, Waheed A, Tuma F. Academic clinicians' workload challenges and burnout analysis. *Cureus*. (2019) 11:e6108. doi: 10.7759/cureus.6108

24. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med*. (2016) 165:753–60. doi: 10.7326/M16-0961

25. Moy AJ, Schwartz JM, Chen R, Sadri S, Lucas E, Cato KD, et al. Measurement of clinical documentation burden among physicians and nurses using electronic health records: a scoping review. *J Am Med Inform Assoc*. (2021) 28:998–1008. doi: 10.1093/jamia/ocaa325

26. Sloss EA, Abdul S, Aboagyewah MA, Beebe A, Kendle K, Marshall K, et al. Toward alleviating clinician documentation burden: a scoping review of burden reduction efforts. *Appl Clin Inform*. (2024) 15:446–55. doi: 10.1055/s-0044-1787007

27. Blum K. Association of Health Care Journalists. (2024). All ears: What to know about ambient clinical listening. Available from: https://healthjournalism.org/blog/2024/03/all-ears-what-to-know-about-ambient-clinical-listening/ (Accessed November 7, 2024).

28. Bundy H, Gerhart J, Baek S, Connor CD, Isreal M, Dharod A, et al. Can the administrative loads of physicians be alleviated by AI-facilitated clinical documentation? *J Gen Intern Med*. (2024) 39:2995–3000. doi: 10.1007/s11606-024-08870-z

29. Haberle T, Cleveland C, Snow GL, Barber C, Stookey N, Thornock C, et al. The impact of nuance DAX ambient listening AI documentation: a cohort study. *J Am Med Inform Assoc*. (2024) 31:975–9. doi: 10.1093/jamia/ocae022

30. Abdelgadir Y, Thongprayoon C, Miao J, Suppadungsuk S, Pham JH, Mao MA, et al. AI integration in nephrology: evaluating ChatGPT for accurate ICD-10 documentation and coding. *Front Artif Intell*. (2024) 7:1457586. doi: 10.3389/frai.2024.1457586

31. Falis M, Gema AP, Dong H, Daines L, Basetti S, Holder M, et al. Can GPT-3.5 generate and code discharge summaries? *J Am Med Inform Assoc*. (2024) 31:2284–93. doi: 10.1093/jamia/ocae132

32. Barak-Corren Y, Wolf R, Rozenblum R, Creedon JK, Lipsett SC, Lyons TW, et al. Harnessing the power of generative AI for clinical summaries: perspectives from emergency physicians. *Ann Emerg Med*. (2024) 84:128–38. doi: 10.1016/j.annemergmed.2024.01.039

33. Akbar F, Mark G, Warton EM, Reed ME, Prausnitz S, East JA, et al. Physicians' electronic inbox work patterns and factors associated with high inbox work duration. *J Am Med Inform Assoc*. (2021) 28:923–30. doi: 10.1093/jamia/ocaa229

34. Tai-Seale M, Dillon EC, Yang Y, Nordgren R, Steinberg RL, Nauenberg T, et al. Physicians' well-being linked to in-basket messages generated by algorithms in electronic health records. *Health Aff (Millwood)*. (2019) 38:1073–8. doi: 10.1377/hlthaff.2018.05509

35. Holmgren AJ, Downing NL, Tang M, Sharp C, Longhurst C, Huckman RS. Assessing the impact of the COVID-19 pandemic on clinician ambulatory electronic health record use. *J Am Med Inform Assoc*. (2022) 29:453–60. doi: 10.1093/jamia/ocab268

36. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence Chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. (2023) 183:589–96. doi: 10.1001/jamainternmed.2023.1838

37. Scott M, Muncey W, Seranio N, Belladelli F, Del Giudice F, Li S, et al. Assessing artificial intelligence-generated responses to urology patient in-basket messages. *Urol Pract*. (2024) 11:793–8. doi: 10.1097/UPJ.0000000000000637

38. Liu S, McCoy AB, Wright AP, Carew B, Genkins JZ, Huang SS, et al. Leveraging large language models for generating responses to patient messages—a subjective analysis. *J American Medical Info Assoc: JAMIA*. (2024) 31:1367–79. doi: 10.1093/jamia/ocae052

39. Droxi. Droxi digital health. Available from: https://www.droxi.ai (Accessed November 7, 2024).

40. Epic. Epic and Microsoft Bring GPT-4 to EHRs. (2023). Available from: https://www.epic.com/epic/post/epic-and-microsoft-bring-gpt-4-to-ehrs/ (Accessed November 7, 2024).

41. Affineon. The AI inbox that saves provider time. Available from: https://www.affineon.com/ (Accessed November 7, 2024).

42. Elendu C, Amaechi DC, Okatta AU, Amaechi EC, Elendu TC, Ezeh CP, et al. The impact of simulation-based training in medical education: a review. *Medicine (Baltimore)*. (2024) 103:e38813. doi: 10.1097/MD.0000000000038813

43. Komasawa N, Yokohira M. Simulation-based Education in the artificial intelligence era. Cureus. (2023); Available from: https://www.cureus.com/articles/161951-simulation-based-education-in-the-artificial-intelligence-era (Accessed November 7, 2024).

44. Rothkrug A, Mahboobi SK. Simulation training and skill assessment in anesthesiology In: StatPearls. FL: StatPearls Publishing (2024)

45. Kothari LG, Shah K, Barach P. Simulation based medical education in graduate medical education training and assessment programs. *Prog Pediatr Cardiol*. (2017) 44:33–42. doi: 10.1016/j.ppedcard.2017.02.001

46. Holderried F, Stegemann-Philipps C, Herschbach L, Moldt JA, Nevins A, Griewatz J, et al. A generative Pretrained transformer (GPT)-powered Chatbot as a simulated patient to practice history taking: prospective. *Mixed Methods Study JMIR Med Educ.* (2024) 10:e53961. doi: 10.2196/53961

47. Borg A, Jobs B, Huss V, Gentline C, Espinosa F, Ruiz M, et al. Enhancing clinical reasoning skills for medical students: a qualitative comparison of LLM-powered social robotic versus computer-based virtual patients within rheumatology. *Rheumatol Int.* (2024) 44:3041–51. doi: 10.1007/s00296-024-05731-0

48. Sardesai N, Russo P, Martin J, Sardesai A. Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training. *Postgrad Med J.* (2024) 100:237–41. doi: 10.1093/postmj/qgad137

49. Webb JJ. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. *Cureus.* (2023) 15:e38755. doi: 10.7759/cureus.38755

50. Mahmood F, Borders D, Chen RJ, Mckay GN, Salimian KJ, Baras A, et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans Med Imaging.* (2020) 39:3257–67. doi: 10.1109/TMI.2019.2927182

51. Zargari A, Topacio BR, Mashhadi N, Shariati SA. Enhanced cell segmentation with limited training datasets using cycle generative adversarial networks. *iScience.* (2024) 27:00962–3. doi: 10.1016/j.isci.2024.109740

52. Ghorbani A, Natarajan V, Coz D, Liu Y. DermGAN: synthetic generation of clinical skin images with pathology. arXiv; (2019). Available from: http://arxiv.org/abs/1911.08716 (Accessed November 8, 2024).

53. Breslavets M, Breslavets D, Lapa T. Advancing dermatology education with AI-generated images. *DOJ.* (2024) 30. doi: 10.5070/D330163299

54. Lim S, Kooper-Johnson S, Chau CA, Robinson S, Cobos G. Exploring the potential of DALL-E 2 in pediatric dermatology: a critical analysis. *Cureus.* (2024) 16:e67752. doi: 10.7759/cureus.67752

55. Waheed A, Goyal M, Gupta D, Khanna A, Al-Turjman F, Pinheiro PR. CovidGAN: data augmentation using auxiliary classifier GAN for improved Covid-19 detection. *IEEE Access.* (2020) 8:91916–23. doi: 10.1109/ACCESS.2020.2994762

56. Waikel RL, Othman AA, Patel T, Hanchard SL, Hu P, Tekendo-Ngongang C, et al. Generative methods for pediatric genetics Education. *med Rxiv.* (2023) 2023:23293506. doi: 10.1101/2023.08.01.23293506

57. Sonmez SC, Sevgi M, Antaki F, Huemer J, Keane PA. Generative artificial intelligence in ophthalmology: current innovations, future applications and challenges. *Br J Ophthalmol.* (2024) 108:1335–40. doi: 10.1136/bjo-2024-325458

58. Bloom BS. The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educ Res.* (1984) 13:4–16. doi: 10.3102/0013189X013006004

59. Parente DJ. Generative artificial intelligence and large language models in primary care medical Education. *Fam Med.* (2024) 56:534–40. doi: 10.22454/FamMed.2024.775525

60. Lyo S, Mohan S, Hassankhani A, Noor A, Dako F, Cook T. From revisions to insights: converting radiology report revisions into actionable educational feedback using generative AI models. *J Digit Imaging Inform med.* (2024):1–15. doi: 10.1007/s10278-024-01233-4

61. Mistry NP, Saeed H, Rafique S, Le T, Obaid H, Adams SJ. Large language models as tools to generate radiology board-style multiple-choice questions. *Acad Radiol.* (2024) 31:3872–8. doi: 10.1016/j.acra.2024.06.046

62. Ayub I, Hamann D, Hamann CR, Davis MJ. Exploring the potential and limitations of chat generative pre-trained transformer (ChatGPT) in generating board-style dermatology questions: a qualitative analysis. *Cureus.* (2023) 15:e43717. doi: 10.7759/cureus.43717

63. Girotra K, Meincke L, Terwiesch C, Ulrich K. Ideas are dimes a dozen: large language models for idea generation in innovation. *SSRN Electron J.* (2023). doi: 10.2139/ssrn.4526071

64. Park YJ, Kaplan D, Ren Z, Hsu CW, Li C, Xu H, et al. Can ChatGPT be used to generate scientific hypotheses? *J Mater.* (2024) 10:578–84. doi: 10.1016/j.jmat.2023.08.007

65. Rahman M, Terano HJ, Rahman M, Salamzadeh A. ChatGPT and Academic Research: A Review and Recommendations Based on Practical Examples. *J Educ Manag Develop Stud.* (2023) 3:1–12. doi: 10.52631/jemds.v3i1.175

66. Agarwal S, Laradji IH, Charlin L, Pal C. Lit LLM: a toolkit for scientific Literature Review. arXiv; (2024). Available from: http://arxiv.org/abs/2402.01788 (Accessed November 8, 2024).

67. Guo E, Gupta M, Deng J, Park YJ, Paget M, Naugler C. Automated paper screening for clinical reviews using large language models: data analysis study. *J Med Internet Res.* (2024) 26:e48996. doi: 10.2196/48996

68. Susnjak T, Hwang P, Reyes NH, Barczak ALC, McIntosh TR, Ranathunga S. Automating research synthesis with domain-specific large language model Fine-tuning. arXiv; (2024). Available from: http://arxiv.org/abs/2404.08680 (Accessed November 8, 2024).

69. Hwang SI, Lim JS, Lee RW, Matsui Y, Iguchi T, Hiraki T, et al. Is ChatGPT a "fire of Prometheus" for non-native English-speaking researchers in academic writing? *Korean J Radiol.* (2023) 24:952–9. doi: 10.3348/kjr.2023.0773

70. Jones JH, Fleming N. Quality improvement projects and anesthesiology graduate medical Education: a systematic review. Cureus. (2024); Available from: https://www.cureus.com/articles/243594-quality-improvement-projects-and-anesthesiology-graduate-medical-education-a-systematic-review (Accessed November 8, 2024).

71. Nejjar M, Zacharias L, Stiehle F, Weber I. LLMs for science: Usage for code generation and data analysis. ICSSP Special Issue in Journal of Software Evolution and Process); In Print (2023).

72. Liu Z, Zhong T, Li Y, Zhang Y, Pan Y, Zhao Z, et al. Evaluating large language models for radiology natural language processing. arXiv; (2023). Available from: http://arxiv.org/abs/2307.13693 (Accessed November 9, 2024).

73. Education M, Systems D-S. Medical Education and decision-support Systems. *AMA J Ethics.* (2011) 13:156–60. doi: 10.1001/virtualmentor.2011.13.3.medu1-1103

74. Clinical decision support (CDS). HealthIT.gov. (2024). Available from: https://www.healthit.gov/test-method/clinical-decision-support-cds (Accessed November 9, 2024).

75. Sirajuddin AM, Osheroff JA, Sittig DF, Chuo J, Velasco F, Collins DA. Implementation pearls from a new guidebook on improving medication use and outcomes with clinical decision support: effective CDS is essential for addressing healthcare performance improvement imperatives. *J Healthc Inf Manag.* (2009) 23:38–45.

76. Liu S, Wright AP, Patterson BL, Wanderer JP, Turer RW, Nelson SD, et al. Assessing the value of ChatGPT for clinical decision support optimization. *Health Informatics.* (2023). doi: 10.1101/2023.02.21.23286254

77. Lee P, Goldberg C, Kohane I. The AI revolution in medicine: GPT-4 and beyond. Erscheinungsort nicht ermittelbar: Pearson Education (2023). 282 p.

78. Ahmed W, Saturno M, Rajjoub R, Duey AH, Zaidat B, Hoang T, et al. ChatGPT versus NASS clinical guidelines for degenerative spondylolisthesis: a comparative analysis. *Eur Spine J.* (2024) 33:4182–203. doi: 10.1007/s00586-024-08198-6

79. Nietsch KS, Shrestha N, Mazudie Ndjonko LC, Ahmed W, Mejia MR, Zaidat B, et al. Can large language models (LLMs) predict the appropriate treatment of acute hip fractures in older adults? Comparing appropriate use criteria with recommendations from ChatGPT. *J Am Acad Orthop Surg Glob Res Rev.* (2024) 8:e24.00206. doi: 10.5435/JAAOSGlobal-D-24-00206

80. Sandmann S, Riepenhausen S, Plagwitz L, Varghese J. Systematic analysis of ChatGPT, Google search and llama 2 for clinical decision support tasks. *Nat Commun.* (2024) 15:2050. doi: 10.1038/s41467-024-46411-8

81. Kao HJ, Chien TW, Wang WC, Chou W, Chow JC. Assessing ChatGPT's capacity for clinical decision support in pediatrics: a comparative study with pediatricians using KIDMAP of Rasch analysis. *Medicine (Baltimore).* (2023) 102:e34068. doi: 10.1097/MD.0000000000034068

82. Lahat A, Sharif K, Zoabi N, Shneor Patt Y, Sharif Y, Fisher L, et al. Assessing generative Pretrained transformers (GPT) in clinical decision-making: comparative analysis of GPT-3.5 and GPT-4. *J Med Internet Res.* (2024) 26:e54571. doi: 10.2196/54571

83. Jo E, Song S, Kim JH, Lim S, Kim JH, Cha JJ, et al. Assessing GPT-4's performance in delivering medical advice: comparative analysis with human experts. *JMIR Med Educ.* (2024) 10:e51282. doi: 10.2196/51282

84. Hu K, Hu K. ChatGPT sets record for fastest-growing user base - analyst note. Reuters. (2023); Available from: https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/ (Accessed November 8, 2024).

85. State of AI. Exhibit 11. (2024). Available from: http://ceros.mckinsey.com/stateofai2024-ex11 (Accessed November 8, 2024).

86. McKinsey. The future of generative AI in healthcare. (2024). Available from: https://www.mckinsey.com/industries/healthcare/our-insights/generative-ai-in-healthcare-adoption-trends-and-whats-next (Accessed November 8, 2024).

87. Goddard J. Hallucinations in ChatGPT: a cautionary tale for biomedical researchers. *Am J Med.* (2023) 136:1059–60. doi: 10.1016/j.amjmed.2023.06.012

88. Williams CYK, Bains J, Tang T, Patel K, Lucas AN, Chen F, et al. Evaluating large language models for drafting emergency department discharge summaries. *med Rxiv.* (2024):24305088. doi: 10.1101/2024.04.03.24305088

89. Williams CYK, Miao BY, Kornblith AE, Butte AJ. Evaluating the use of large language models to provide clinical recommendations in the emergency department. *Nat Commun.* (2024) 15:8236. doi: 10.1038/s41467-024-52415-1

90. Arfaie S, Sadegh Mashayekhi M, Mofatteh M, Ma C, Ruan R, MacLean MA, et al. ChatGPT and neurosurgical education: a crossroads of innovation and opportunity. *J Clin Neurosci.* (2024) 129:110815. doi: 10.1016/j.jocn.2024.110815

91. Ahmad O, Maliha H, Ahmed I. AI syndrome: an intellectual asset for students or a progressive cognitive decline. *Asian J Psychiatr.* (2024) 94:103969. doi: 10.1016/j.ajp.2024.103969

92. Gao Y, Xiong Y, Gao X, Jia K, Pan J, Bi Y, et al. Retrieval-augmented generation for large language models: a survey. *arXiv.* (2024). doi: 10.48550/arXiv.2312.10997

93. Parthasarathy VB, Zafar A, Khan A, Shahid A. The ultimate guide to Fine-tuning LLMs from basics to breakthroughs: an exhaustive review of technologies, research, best practices, Applied Research Challenges and Opportunities. *arXiv.* (2024). doi: 10.48550/arXiv.2408.13296

94. Barkley L, Merwe B. Investigating the role of prompting and external tools in hallucination rates of large language models. *arXiv*. (2024). doi: 10.48550/arXiv.2410.19385

95. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. *J American Medical Info Assoc: JAMIA*. (2011) 19:121–7. doi: 10.1136/amiajnl-2011-000089

96. Johnstone RE, Vallejo MC, Zakowski M. Improving residency applicant personal statements by decreasing hired contractor involvement. *J Grad Med Educ*. (2022) 14:526–8. doi: 10.4300/JGME-D-22-00226.1

97. Zumsteg JM, Junn C. Will ChatGPT match to your program? *Am J Phys Med Rehabil*. (2023) 102:545–7. doi: 10.1097/PHM.0000000000002238

98. Parks LJ, Sizemore DC, Johnstone RE. Plagiarism in personal statements of anesthesiology residency applicants. *A&A Practice*. (2016) 6:103. doi: 10.1213/XAA.0000000000000202

99. Segal S, Gelfand BJ, Hurwitz S, Berkowitz L, Ashley SW, Nadel ES, et al. Plagiarism in residency application essays. *Ann Intern Med*. (2010) 153:112–20. doi: 10.7326/0003-4819-153-2-201007200-00007

100. Quinonez SC, Stewart DA, Banovic N. ChatGPT and artificial intelligence in graduate medical Education program applications. *J Grad Med Educ*. (2024) 16:391–4. doi: 10.4300/JGME-D-23-00823.1

101. Mangold S, Ream M. Artificial intelligence in graduate medical Education applications. *J Grad Med Educ*. (2024) 16:115–8. doi: 10.4300/JGME-D-23-00510.1

102. Leung TI, Sagar A, Shroff S, Henry TL. Can AI mitigate Bias in writing letters of recommendation? *JMIR Medical Educ*. (2023) 9:e51494. doi: 10.2196/51494

103. Weber-Wulff D, Anohina-Naumeca A, Bjelobaba S, Foltýnek T, Guerrero-Dib J, Popoola O, et al. Testing of detection tools for AI-generated text. *arXiv*. (2023). doi: 10.48550/arXiv.2306.15666

104. Kim JK, Chua M, Rickard M, Lorenzo A. ChatGPT and large language model (LLM) chatbots: the current state of acceptability and a proposal for guidelines on utilization in academic medicine. *J Pediatr Urol*. (2023) 19:598–604. doi: 10.1016/j.jpurol.2023.05.018

105. Open AI, Achiam J, Adler S, Agarwal S, Ahmad L, Akkaya I, et al. GPT-4 technical report. *arXiv*. (2024). doi: 10.48550/arXiv.2303.08774

106. Zhou M, Abhishek V, Derdenger T, Kim J, Srinivasan K. Bias in generative AI. *arXiv*. (2024). doi: 10.48550/arXiv.2403.02726

107. Templin T, Perez MW, Sylvia S, Leek J, Sinnott-Armstrong N. Addressing 6 challenges in generative AI for digital health: a scoping review. *PLOS Digit Health*. (2024) 3:e0000503. doi: 10.1371/journal.pdig.0000503

108. Cyberhaven. 11% of data employees paste into ChatGPT is confidential. (2023). Available from: https://www.cyberhaven.com/blog/4-2-of-workers-have-pasted-company-data-into-chatgpt (Accessed October 21, 2024).

109. Russell RG, Lovett Novak L, Patel M, Garvey KV, Craig KJT, Jackson GP, et al. Competencies for the use of artificial intelligence-based tools by health care professionals. *Acad Med*. (2023) 98:348–56. doi: 10.1097/ACM.0000000000004963

110. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME guide no. 84. *Med Teach*. (2024) 46:446–70. doi: 10.1080/0142159X.2024.2314198

111. Bartoli A, May AT, Al-Awadhi A, Schaller K. Probing artificial intelligence in neurosurgical training: ChatGPT takes a neurosurgical residents written exam. *Brain and Spine*. (2024) 4:102715. doi: 10.1016/j.bas.2023.102715

112. Lawson McLean A, Gutiérrez PF. Application of transformer architectures in generative video modeling for neurosurgical education. *Int J CARS*. (2024). doi: 10.1007/s11548-024-03266-0

113. Sevgi M, Antaki F, Keane PA. Medical education with large language models in ophthalmology: custom instructions and enhanced retrieval capabilities. *Br J Ophthalmol*. (2024) 108:1354–61. doi: 10.1136/bjo-2023-325046

114. DeCook R, Muffly BT, Mahmood S, Holland CT, Ayeni AM, Ast MP, et al. AI-generated graduate medical Education content for Total joint arthroplasty: comparing ChatGPT against Orthopaedic fellows. *Arthroplast Today*. (2024) 27:101412. doi: 10.1016/j.artd.2024.101412

115. Ba H, Zhang L, Yi Z. Enhancing clinical skills in pediatric trainees: a comparative study of ChatGPT-assisted and traditional teaching methods. *BMC Med Educ*. (2024) 24:558. doi: 10.1186/s12909-024-05565-1

116. Suresh S, Misra SM. Large language models in pediatric Education: current uses and future potential. *Pediatrics*. (2024) 154:e2023064683. doi: 10.1542/peds.2023-064683

117. Meşe İ, Taşlıçay CA, Kuzan BN, Kuzan TY, Sivrioğlu AK. Educating the next generation of radiologists: a comparative report of ChatGPT and e-learning resources. *Diagn Interv Radiol*. (2024) 30:163–74. doi: 10.4274/dir.2023.232496

118. Lia H, Atkinson AG, Navarro SM. Cross-industry thematic analysis of generative AI best practices: applications and implications for surgical education and training. *Global Surg Educ*. (2024) 3:61. doi: 10.1007/s44186-024-00263-4

119. Sathe TS, Roshal J, Naaseh A, L'Huillier JC, Navarro SM, Silvestri C. How I GPT it: development of custom artificial intelligence (AI) Chatbots for surgical Education. *J Surg Educ*. (2024) 81:772–5. doi: 10.1016/j.jsurg.2024.03.004