

Residents, Faculty, and Artificial Intelligence: Brave New World or Utopia?

Gail M. Sullivan, MD, MPH
 Deborah Simpson , PhD
 Lalena M. Yarris , MD, MCR
 Anthony R. Artino Jr , PhD

Since the late 2022 release of OpenAI's ChatGPT, an open-source natural language processing tool that can generate human-like intellectual content and responses, journal editors, including those at the *Journal of Graduate Medical Education* (JGME), have intensified debates around fair, appropriate use of this technology for submitted manuscripts.¹⁻⁴ We have recently expanded JGME's overall policy of transparency to include artificial intelligence (AI): "Authors must communicate any use of artificial intelligence technology and similar tools, whether for writing assistance, storage, interpretation of qualitative research data, statistical analysis of numerical data, creation of visual imagery, or other uses."⁵ In addition, listed JGME authors must be human. Why? Only humans can be accountable for all aspects of the work, a critical authorship criterion.⁶ Beyond these considerations of authorship and transparency, we are curious about potential AI uses in graduate medical education (GME) for both the process and content of training.⁷⁻⁹

In 2020, we asked 40 thought leaders in medical education what they believed would be the most significant transformation in GME over the next decade: AI was a major theme.¹⁰ Three years later, AI use in daily life is everywhere, from cell phones and credit checks to weather and climate predictions. AI examples include natural language processing, speech recognition, and creating written works (See if you can tell the difference between human- and AI-generated text in BOX 1). The volume of health information has increased beyond the capacity of humans to sift and digest: in 1950 medical information doubled in 50 years; by 2010, it took 3.5 years, and the estimate for 2020 was 73 days.¹¹ However, only a few undergraduate medical schools have added AI experiences, and the Accreditation Council for Graduate Medical Education Milestones do not include AI competencies.¹¹ Currently AI is being introduced into learning and clinical diagnosis support, automated assessments of clinical skills, setting

educational goals, and designing curriculum and performance assessments.^{12,13}

Scholars have hailed the potential benefits of AI, such as improving manuscripts for authors writing in a second language, saving researchers time when creating research documents or conducting literature searches, and freeing trainees from memorization to focus on reasoning, counseling, and shared decision-making with patients.^{11,14} Other scholars are concerned that AI tools in GME present serious risks. These include loss of trainee and patient anonymity in confidential information, failure to develop key competencies as those tasks are outsourced to AI tools, erroneous evidence and assessment summaries, and "paper mills," or fabrication of fake research to pad resumes and grant applications.¹⁵⁻¹⁸ As AI is wholly dependent upon the quality and included biases of the available training data, it may generate inaccurate syntheses, which could be missed by inexperienced trainees and beleaguered faculty. In addition, AI could negatively affect trainees' ability to develop mental models or learn key foundational skills, such as clinical reasoning in treating complex patients. Many patients, with their unique values, social supports, finances, and medical histories, may be hard to fit into AI's derived diagnoses and management plans. In 2023, the optimal interactions between physicians and AI, patients and AI, and trainees and AI are unknown, yet graduating residents and fellows will increasingly manage these interactions and confront numerous ethical sequelae.

Medical education is notorious for diving headlong into the next innovation, often with scanty evidence. But as Dr Rachel Ellaway points out, "the genie is out of the bottle."³ We need research that examines the best uses of AI in GME. This research should include exploration of both the use of AI during training and a greater understanding of the ethical, beneficial uses of AI in medical education research. Which research tasks remain uniquely human? Which AI educational functions require minimal human oversight? Will AI replace some physician skills, rendering them obsolete and no longer taught in GME? Will AI technology transform

BOX 1 Chatbot vs Human: Can You Identify Which? (See end of this article for the answer.)**Briefly describe the differences between formative and summative feedback in residency education**

#1

Formative feedback and summative feedback are 2 distinct types of assessment in residency education.

Formative feedback is designed to provide ongoing support and guidance to learners during their educational journey. It focuses on identifying strengths and areas for improvement, with the goal of helping residents develop and grow. Formative feedback is usually given throughout the learning process and is more subjective and descriptive in nature. It emphasizes constructive criticism, suggestions for improvement, and actionable recommendations. The primary purpose of formative feedback is to enhance learning and facilitate skill development.

On the other hand, summative feedback is given at the end of a specific period or task, such as an examination, rotation, or residency program. It aims to evaluate the overall performance and achievement of learners against predetermined standards or objectives. Summative feedback tends to be more objective and evaluative in nature, focusing on measuring competency or proficiency levels. It often involves grading or rating systems and provides a comprehensive assessment of a resident's abilities and readiness for advancement or certification.

In summary, formative feedback is ongoing, developmental, and serves to guide and support learners, whereas summative feedback is final, evaluative, and provides an overall judgment of performance. Both types of feedback are valuable in residency education as they help foster growth, identify areas of improvement, and measure competence and progress.

#2

Formative feedback in graduate medical education (GME) occurs throughout clinical and other rotations, often in the form of immediate feedback regarding performance. For example, after a difficult patient interview or procedure, the attending physician may ask the resident what they think were aspects that went well, or less well. The attending physician can corroborate, enhance, or correct the resident's self-evaluation, while also setting a few goals for future encounters. Formative feedback also occurs at scheduled times, such as a mid-rotation meeting, with attention to specific competencies assigned to the current experience or assignments. Thus, formative feedback provides ongoing information to the resident to improve performance as well as reinforce strong performances.

Summative feedback is associated with end-of-rotation assessments or biannual milestones assessments, akin to "grades" in school. Most assessment experts now consider that all assessments should primarily align with formative, rather than summative feedback approaches, to be more effective and support a growth mindset in GME trainees.

cognitive skills, similar to robotics in surgery? We invite reports of how GME trainees, programs, faculty, and institutions are learning about and using AI tools (BOX 2).

BOX 2 Potential Research Topics for AI in Graduate Medical Education

- Curriculum design, implementation, and evaluation
- Analysis of faculty clinical teaching performance to provide feedback
- Formative feedback on communication and interpersonal skills, based on AI analysis of videos of learner interactions with patients or colleagues
- Analysis of EHR data to understand trainee clinical diagnostic reasoning
- Visual summary representation of resident assessments—text, audio, image, video—for an assessment portfolio, for competency committee decisions and individual learning goals
- Creating faculty avatars for coaching, advising, or assessment of learners linked to specific competencies and performance standards
- Resident AI technology skills that increase time with patients without lowering patient and resident satisfaction, patient quality of care, and resident learning
- Resident and faculty skills in detecting and avoiding chatbots or other AI-generated "hallucinations" such as fake citations, resident history and physical examinations, and other materials
- Interactive AI for simulations
- Faculty and trainee skills to efficiently verify evidence compiled by AI technology
- Effects of AI technology on trust in resident-patient relations
- Effects of AI on professional identity formation in GME trainees

Abbreviations: AI, artificial intelligence; EHR, electronic health record; GME, graduate medical education.

Of substantial concern is whether for-profit entities, which have contributed to the growing disparities in US health care, will encourage problematic AI use in medical education and patient care. We prefer that ethics and evidence lead the way. The US health care system is increasingly a "flailing medical industry shaped by profit rather than ethics."^{19,20} Thus we need AI tools that can enhance health equity. AI is hyped as a way to generate faster, more accurate diagnoses; reduce physician error; reduce or eliminate repetitive, mundane physician tasks such as the electronic health record (EHR); and reduce the high cost of health care.¹¹ Some of us are wary of these claims. They are reminiscent of those touted for EHRs, now shown to produce increased physician documentation time and burnout, especially for the endangered species known as primary care physicians.²¹

What do most GME educators want from AI? Our guess is they desire more time for teaching and tools that will help more learners reach "mastery" competency levels. They also want their trainees well

prepared for future AI uses, in a medical utopia rather than a dystopian *Brave New World*.²²

References

- Committee on Publication Ethics. Authorship and AI tools: COPE position statement. Accessed June 26, 2023. <https://publicationethics.org/cope-position-statements/ai-author>
- DeVilbiss MB. Artificial intelligence tools in scholarly publishing: guidance for Academic Medicine authors [published online ahead of print April 28, 2023]. *Acad Med*. doi:10.1097/ACM.0000000000005261
- Ellaway RH, Tolsgaard M. Artificial scholarship: LLMs in health professions education research [published online ahead of print June 19, 2023]. *Adv Health Sci Educ Theory Pract*. doi:10.1007/s10459-023-10257-4
- Flanagan A, Bibbins-Domingo K, Berkswits M, et al. Nonhuman “authors” and implications for the integrity of scientific publication and medical knowledge. *JAMA*. 2023;329(8):637-639. doi:10.1001/jama.2023.1344
- Journal of Graduate Medical Education. Instructions for Authors. Accessed July 28, 2023. <https://meridian.allenpress.com/jgme/pages/Instructions-for-Authors>
- International Committee of Medical Journal Editors. Defining the role of authors and contributors. Accessed July 10, 2023. <https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>
- Shoja MM, Van de Ridder JMM, Rajput V. The emerging role of generative artificial intelligence in medical education, research, and practice. *Cureus*. 2023;15(6):e40883. doi:10.7759/cureus.40883
- Dwivedi YK, Kshetri N, Hughes L, et al. “So what if ChatGPT wrote it?” Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int J Info Manag*. 2023;71:102642. doi:10.1016/j.ijinfomgt.2023.102642
- Lee J, Wu AS, Li D, Kulasegaram KM. Artificial intelligence in undergraduate medical education: a scoping review. *Acad Med*. 2021;96(suppl 11):62-70. doi:10.1097/ACM.0000000000004291
- Simpson D, Sullivan GM, Artino Jr AR, Deiorio NM, Yarris LM. Envisioning graduate medical education in 2030. *J Grad Med Educ*. 2020;12(3):235-240. doi:10.4300/JGME-D-20-00292.1
- Paranjape K, Schinkel M, Panday RN, Car J, Nanayakkara P. Introducing artificial intelligence training in medical education. *JMIR Med Educ*. 2019;5(2):e16048. doi:10.2196/16048
- Tolsgaard MG, Pusic MV, Sebok-Syer SS, et al. The fundamentals of artificial intelligence in medical education research: AMEE guide no. 156. *Med Teach*. 2023;45(6):565-573. doi:10.1080/0142159X.2023.2180340
- Latif E, Mai G, Nyaaba M, et al. Artificial general intelligence (AGI) for education. *arXiv*. Published May 15, 2023. Accessed June 26, 2023. <https://arxiv.org/abs/2304.12479>
- Lee P, Bubeck S, Petro J. Benefit, limits, and risks of CPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023;388(13):1233-1239. doi:10.1056/NEJMSr2214184
- Masters K. Ethical use of artificial intelligence in health professions education: AMEE guide no. 158. *Med Teach*. 2023;45(6):574-584. doi:10.1080/0142159X.2023.2186203
- Schinkel M, Paranjape K, Nanayakkara P. Written by humans or artificial intelligence? That is the question. *Ann Intern Med*. 2023;176(4):572-573. doi:10.7326/M23-0154
- Committee on Publication Ethics. Paper mills: research report from COPE & STM. Published June 2022. Accessed June 26, 2023. <https://publicationethics.org/sites/default/files/paper-mills-cope-stm-research-report.pdf>
- Marks M, Haupt CE. AI chatbots, health privacy, and challenges to HIPAA compliance [published online ahead of print July 6, 2023]. *JAMA*. doi:10.1001/jama.2023.9458
- Reinhart E. The political education of US physicians. *JAMA Netw Open*. 2023;6(6):e2320447. doi:10.1001/jamanetworkopen.2023.20447
- Berwick DM. Salve lucrum: the existential threat of greed in US health care. *JAMA*. 2023;329(8):629-630. doi:10.1001/jama.2023.0846
- Budd J. Burnout related to electronic record use in primary care. *J Prim Care Commun Health*. 2023;14:21501319231166921. doi:10.1177/21501319231166921
- Huxley A. *Brave New World*. Harper Perennial, 2006.



Gail M. Sullivan, MD, MPH, is Editor-in-Chief, *Journal of Graduate Medical Education* (JGME), and Associate Director for Education, Center on Aging, and Professor of Medicine, University of Connecticut Health Center; **Deborah Simpson, PhD**, is Deputy Editor, JGME, and Director of Education, Academic Affairs at Advocate Aurora Health, and Clinical Adjunct Professor of Family & Community Medicine, Medical College of Wisconsin, University of Wisconsin School of Medicine and Public Health; **Lalena M. Yarris, MD, MCR**, is Deputy Editor, JGME, and Professor of Emergency Medicine, Oregon Health & Science University; and **Anthony R. Artino Jr, PhD**, is Deputy Editor, JGME, and Professor and Associate Dean for Evaluation and Educational Research, The George Washington University School of Medicine and Health Sciences.

Corresponding author: Gail M. Sullivan, MD, MPH, University of Connecticut Health Center, gsullivan@uchc.edu

Answer to box 1: #1: Chatbot; #2: Human, JGME editor