

A Framework for Artificial Intelligence in Medical Education: Could I, Would I, Should I?

Rick Peter Fritz Wolthusen¹, MD, MPP

Mountasir El-tohami², MD

Reuben Abraham Hendler³, MD

Paul Allen Riordan⁴, MD

Joseph Brian Stoklosa, MD

Medical Education and Artificial Intelligence

Artificial intelligence (AI) is increasingly influencing medical education, with rapid developments in tools that promise to enhance how physicians teach, learn, and assess competency. While the broader literature on AI in medical education is growing, most studies focus on specific AI functions, like rapidly training novices in bronchoscopy,¹ and not on generative AI (GAI)—text-based models like ChatGPT.

The Association of American Medical Colleges has offered ChatGPT-generated, then human-edited principles for responsible use of AI in medical education.² Initial work in undergraduate medical education has led to proposed AI competencies.³ However, graduate medical education (GME) lags in pragmatic frameworks and strategies for their application. Among the existing GME-relevant literature, much of the work is theoretical and descriptive, only outlining capabilities or raising concerns. A 2024 review in *Frontiers in Medicine*, for example, provides summaries of the use of GAI in GME settings and discusses opportunities, such as prompt engineering, and risks, such as automation bias, with minimal discussion of how GME educators can practically implement GAI tools in real-world settings.⁴ If practical tips are provided, they are general in nature without actionable steps to guide day-to-day teaching.⁵

This Perspectives article offers a practical, theory-informed, decision-making tool—the “Could I, Would I, Should I?” framework—to help health professions education leaders and educators fill this gap by engaging proactively with GAI tools like ChatGPT.

Could I Use It?

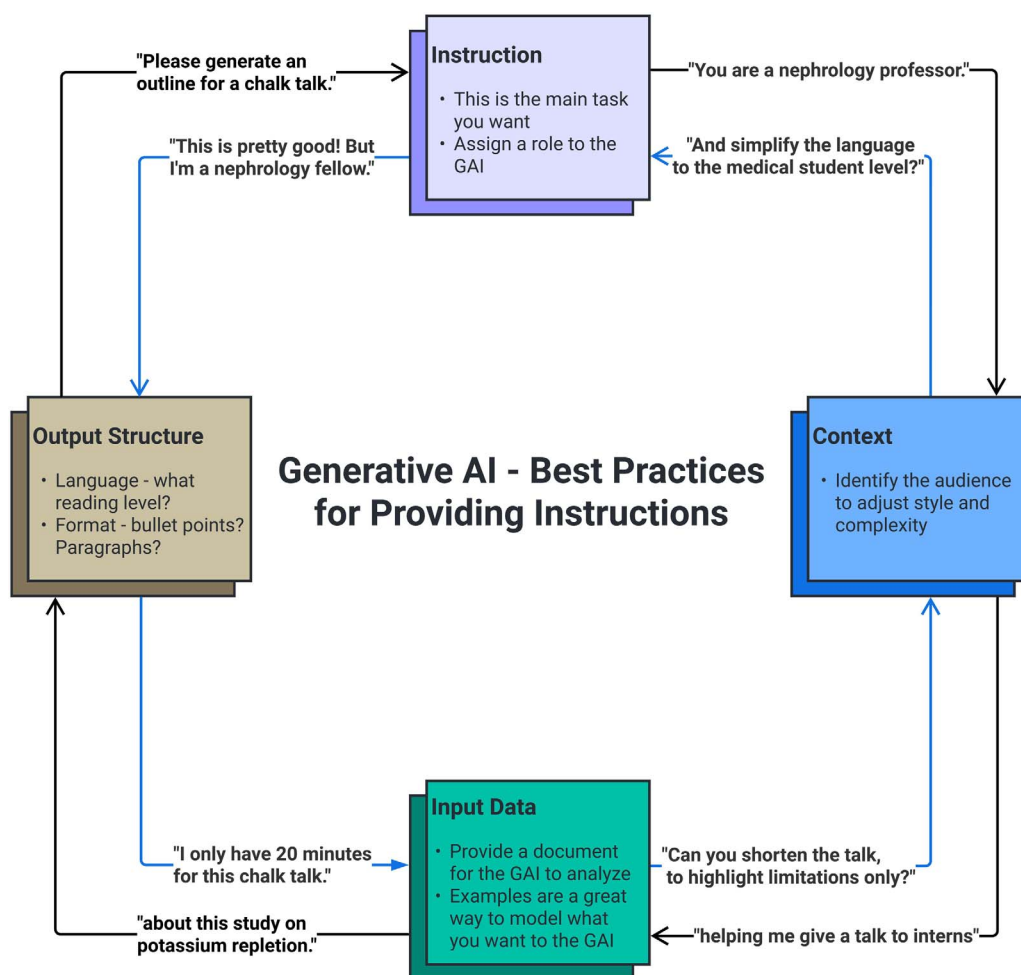
GAI has rapidly attracted many users because it is powerful, accessible, and easy to use. Through natural language processing, GAI has learned to communicate

in human languages and can readily translate between these languages, assisting educators in engaging more diverse learners. Unfortunately, as GAI works by predicting a likely response based on probabilistic reasoning about data it has previously encountered, the content it produces can be vague, generic, and inaccurate. Therefore, at least for the time being, GAI may be most useful in the beginning stages of brainstorming and middle stages of refining educational work, rather than producing a final product. For example, GAI can initially serve as a standardized “patient” for practicing communication skills by offering generic, typical responses. In later editing stages, GAI can rapidly analyze a practice transcript to provide coaching and feedback. GAI is designed for dialogue, so iterative back-and-forth exchanges tend to elicit more satisfying results than a single request. If unsure how to write an effective prompt, you can simply ask the GAI to guide you (and, for reasons that remain unclear, you may get better results if you ask nicely). For effective prompting, we often utilize the user-friendly ICIO framework: Instruction, Context, Input Data, and Output Structure. In the FIGURE, we provide details about the framework and display an example.

Would I Use It?

One of the challenges facing learners in medicine is how to organize and integrate facts and ideas into frameworks that form the building blocks of expertise in medicine. Cognitive psychologists have studied this learning process for many years, and 5 core learning strategies have been identified as critical for mastery of a topic.⁶ These strategies are (1) spaced retrieval, (2) elaboration, (3) reflection, (4) interleaving, and (5) generation, which form the mnemonic RE-RIG (see TABLE). Each of these strategies is a highly active process that requires the learner to engage with the material. Additionally, to advance from expertise to mastery of a clinical topic, clinicians require repeated exposure to multiple patient

DOI: <http://dx.doi.org/10.4300/JGME-D-25-00009.1>

**FIGURE****Best Practices for Providing Instructions**

Abbreviations: AI, artificial intelligence; GAI, generative artificial intelligence.

Note: Follow the outer arrows for an initial set of instructions that may be given to generate a talk on repleting potassium to medical interns. Imagine that you receive a good output but want to do some refining. Follow the inner arrows for additional instructions that may be used to refine the output that was provided. This iterative process will lead to a chalk talk that can then be further refined by the medical educator.

cases with feedback on one's performance, so that the pathophysiological knowledge becomes embedded within the mind as "illness scripts."⁷ These scripts are effortlessly retrieved, allowing for highly accurate diagnoses within minutes of working with patients. Unfortunately, GAI can remove much of the mental effort of learning by taking a learner's input and producing a pleasing summary of complex ideas for memorization. This bypasses the struggle required by these learning strategies and may lead to a regression to the mean, both elevating struggling learners by providing them with average answers to clinical problems and lowering top performers by limiting the creative insights produced by mastery of a topic. To counter this trend, as educators, we will need to teach our learners how to use GAI in a manner that allows these cognitive strategies to flourish

so that expertise can grow. Educators may benefit the most from GAI if they understand it as a "copilot" that helps provide feedback so that they can challenge and refine their ideas.⁸ In terms of "Would I use it?" we believe that the use of AI in medical education is most effective when combined with evidence-based effective learning strategies.

Should I Use It?

Numerous ethical concerns have been associated with the development, implementation, and use of AI. Well-established bioethical principles like non-maleficence, beneficence, justice, and fidelity⁹ provide a familiar structure, conceptual clarity, and proven applicability across similar contexts for organizing key ethical concerns.

TABLE

How a Learner Can Use ChatGPT as a Copilot and Leverage its Power Synergistically With Cognitive Learning Strategies

Cognitive Learning Strategy	Description	Example of How to Leverage ChatGPT With Learning Strategy
Spaced retrieval	Spaced retrieval is the practice of recalling information over time to help information encode in long-term memory.	One can use ChatGPT to create flashcards of material for later testing.
Elaboration	Elaboration is the process of connecting new information to information already known. It is the process of adding the details, nuances, and refinements (ie, the branches, leaves, and flowers) to the trunk of a framework.	Once one has elaborated and filled in details to a medical question, one could turn to ChatGPT to ask it to critique one's thoughts for missing details and perspectives. For a clinical case, ChatGPT may help to hierarchically arrange a differential diagnosis for the presentation, helping the learner elaborate on diagnoses that are less likely but not considered by the learner.
Reflection	Reflection is the process of reviewing and analyzing recently learned material to prompt deeper understanding of the material.	One can use ChatGPT to create reflective questions for material that one has been learning. One can then type those answers back into ChatGPT to see if ChatGPT agrees with or disagrees with the analysis. This dialogue with ChatGPT can lead to deeper understanding.
Interleaving	Interleaving is the process where one learns better by studying 2 different topics simultaneously rather than focusing on a single topic.	One can use ChatGPT as an interlocutor to compare and contrast different diagnoses that are only loosely related. For example, in psychiatry, one could interleave studying on borderline personality disorder with bipolar disorder and then debate with ChatGPT about the similarities and differences of the 2 diagnoses.
Generation	Generation is the process of forcing oneself to create sentences, thoughts, and solutions to questions about a topic.	This is one of the hardest challenges, as ChatGPT effortlessly creates intelligent-sounding language from the simplest prompts. It will require discipline to first generate one's own thoughts and then run them by ChatGPT as an "editor" or "sounding board" if one wants to avoid regression to the mean.

Do No Harm—Non-Malefeasance

Because GAI is designed more to sound correct than to be correct, AI can convincingly spread misinformation. Educators should avoid instructing learners to "find the answer" using GAI and should instead teach learners to critically appraise AI-generated content.

Beneficence

AI has the potential to improve GME's efficiency to the detriment of its quality, by providing an alluring but inferior substitute for thoughtful teaching and active learning. In each use case, educators should ask themselves: Will AI take this lesson further, or shortcut to a less desirable destination?

Justice

GAI tends to "learn" the biases patterned in its (opaque) training data inputs and then perpetuate those biases in its outputs. Educators should affirmatively monitor GAI outputs for bias and encourage learners to consider how bias could impact GAI's reliability and social impact.

Fidelity

AI systems risk compromising our patients' health information confidentiality when AI companies control data fed into their systems. Educators should warn trainees against prompting unsecured AI systems with protected information.

In summary, clinical educators should use a bioethics framework to thoughtfully consider any use of GAI and its potential to help and harm trainees, patients, and society to optimize benefits while mitigating risk.

Conclusion

AI is here to stay and will substantially impact medical education. While tools like ChatGPT can help medical educators generate large amounts of structured content quickly if prompted effectively, their predictive nature can also bypass essential learning processes and raise ethical concerns, making the role of educators more critical than ever. We propose the "Could I, Would I, Should I?" framework and associated strategies to help GME leaders make thoughtful,

context-specific decisions about when and how to use GAI. To move to deliberate integration, we recommend that GME curricula support AI literacy. These curricula should include structured prompting strategies such as ICIO; practical use that reinforces critical thinking, reflection, and deep learning; and ethical considerations. Educators should link emerging GAI competencies to established GME core competencies and offer adaptable, program-specific frameworks. Additionally, we encourage GME stakeholders to consult the American Association of Directors of Psychiatric Residency Training AI in Psychiatric Education Taskforce Report,¹⁰ which outlines potential AI applications to support informed, context-driven decisions. By deliberately shaping AI's use, GME programs can ensure that GAI strengthens—rather than undermines—the mission to train skillful, ethical, and reflective physicians.

References

1. Cold KM, Xie S, Nielsen AO, Clementsen PF, Konge L. Artificial intelligence improves novices' bronchoscopy performance: a randomized controlled trial in a simulated setting. *Chest*. 2024;165(2):405-413. doi:10.1016/j.chest.2023.08.015
2. Association of American Medical Colleges. Principles for the use of artificial intelligence in medical education. Accessed May 20, 2025. <https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use>
3. Lee YM, Kim S, Lee YH, et al. Defining medical AI competencies for medical school graduates: outcomes of a Delphi survey and medical student/educator questionnaire of South Korean medical schools. *Acad Med*. 2024;99(5):524-533. doi:10.1097/ACM.0000000000005618
4. Janumpally R, Nanua S, Ngo A, Youens K. Generative artificial intelligence in graduate medical education. *Front Med (Lausanne)*. 2025;11:1525604. doi:10.3389/fmed.2024.1525604
5. Buckley PJ. Practical tips for enhancing academic skills with generative artificial intelligence tools. *Acad Psychiatry*. 2025;49(1):40-43. doi:10.1007/s40596-024-02055-w
6. Brown PC, Roediger HL III, McDaniel MA. *Make It Stick: The Science of Successful Learning*. The Belknap Press of Harvard University Press; 2014.
7. Norman GR, Grierson LE, Sherbino J, Hamstra SJ, Schmidt HG, Mamede S. Chapter 19: expertise in medicine and surgery. In: Ericsson KA, Hoffman RR, Kozbelt A, Williams AM, eds. *The Cambridge Handbook of Expertise and Expert Performance*. Cambridge University Press; 2018:331-355.
8. Mollick E. *Co-Intelligence: Living and Working With AI*. Penguin Publishing Group; 2018.
9. Beauchamp TL, Childress JF. *Principles of Biomedical Ethics*. 8th ed. Oxford University Press; 2019.
10. American Association of Directors of Psychiatric Residency Training. Artificial intelligence in psychiatric education: a report from the AADPRT AI Task Force. Accessed May 20, 2025. https://www.aadprt.org/application/files/1717/4343/2312/AADPRT_AI_Task_Force_Report_F_small.pdf



Rick Peter Fritz Wolthusen, MD, MPP*, is a Medical Instructor in Psychiatry, Harvard Medical School, Boston, McLean Hospital, Belmont, Massachusetts, USA; **Mountasir El-tohami, MD***, is an Attending Psychiatrist, New York City Health + Hospitals/King's County, Brooklyn, New York, USA; **Reuben Abraham Hendler, MD**, is a Medical Instructor in Psychiatry, Harvard Medical School, Boston, Massachusetts, USA; **Paul Allen Riordan, MD**, is a Medical Assistant Professor, Department of Psychiatry and Behavioral Sciences, Duke University Hospital, Durham, North Carolina, USA; and **Joseph Brian Stoklosa, MD**, is an Assistant Professor in Psychiatry, Harvard Medical School, Boston, Massachusetts, USA.

* Denotes co-first authors.

Corresponding author: Rick Peter Fritz Wolthusen, MD, MPP, McLean Hospital, Belmont, Massachusetts, USA, rwolthusen@mgh.harvard.edu