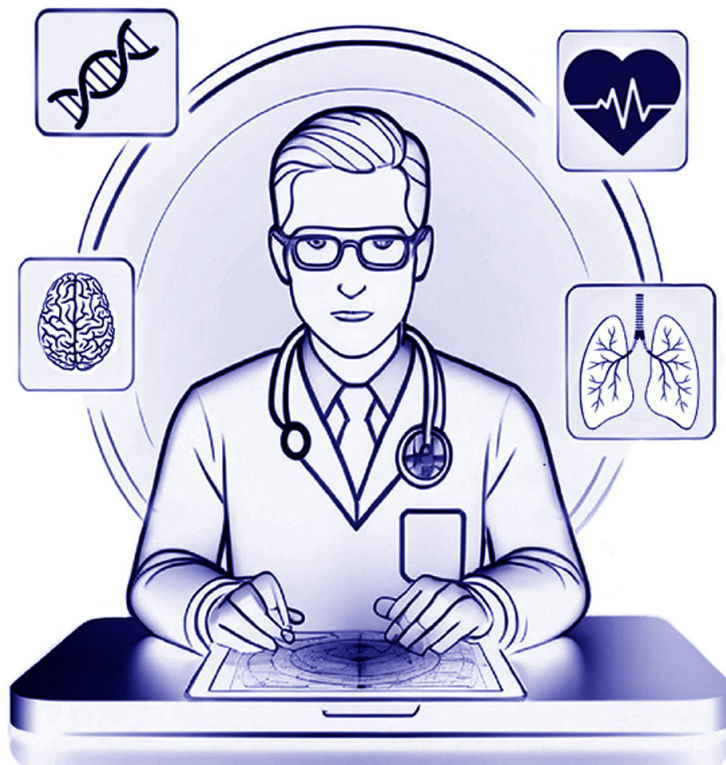




Shapiro Institute

BETH ISRAEL DEACONESS MEDICAL CENTER
HARVARD MEDICAL SCHOOL

Artificial Intelligence in Medical Education



Newsletter
October 2024

THE PEARLS AND PERILS OF AI IN MEDICAL EDUCATION



Dear friends,

Artificial intelligence, or more specifically generative artificial intelligence (GAI) using large language models (LLM), has taken the world by storm in the past several years. Is it one more step towards ultimate efficiency, which will free us from the drudgery at work, another technology that will replace jobs (remember those people who were once called secretaries?), or actually challenge humans for control of the universe (see: the computer HAL in the movie 2001)? While the answers to these questions will not be known for certain for many years, there are clearly uses for GAI in place now and growing rapidly. Medical educators need to consider how GAI can be incorporated into their work and how to prepare students and residents for the potential uses (and abuses) of GAI at the bedside in clinical medicine.

In the classroom or clinical teaching environment, GAI can help us create teaching cases with relative ease. Multiple choice questions are also products that ChatGPT can do within seconds and provide an answer and explanation as well. Assessments of open-ended questions, that might require hours of faculty time to score, are potentially feasible after “training” the GAI on your ideal answer. Studies are appearing in which students use GAI to do self-study by asking for review questions

(and associated answers) on a particular topic. If homework is given or a paper required for a course, however, can we be sure the student is doing their own work or merely submitting a computer generated manuscript? What is the ethical use of GAI in the academic setting? If the student becomes dependent on GAI, what learning is occurring? And how accurate is GAI? There are multiple studies in the literature looking at the performance of GAI on prep tests for standardized exams for licensing or credentialing. While the accuracy of GAI is typically about 60-70%, and improving rapidly, it does make mistakes. How does one discern whether the answer provided is correct or not if you cannot reason on your own?

In the clinical environment, GAI has been used to provide differential diagnoses in morbidity and mortality conferences with some dexterity. But we must remember that, when used for medical decision making, GAI will be subject to biases found on the internet and materials on which it has been trained and has been found to create answers, with great authority, that are incorrect; these have been referred to as “hallucinations.” Since GAI is essentially using pattern recognition to do its job, it is great at chess but not as great at clinical medicine; the former has been described by David Epstein in his book *Range* as a “kind” environment with well delineated rules and rapid feedback on whether a decision was correct or not and the latter is a “wicked” environment in which signs and symptoms are complex and feedback on whether you made the correct decision can be greatly delayed. Thus, proceeding with GAI in clinical decision-making must be done with caution and checked with the clinician’s foundational knowledge of medicine and ability to reason through a problem that they may never have seen before.

The future is now, or at least rapidly arriving.
Proceed with your eye’s wide open!

- Richard M. Schwartzstein, MD

Dr. Adam Rodman is general internist at Beth Israel Deaconess Medical Center and an instructor at Harvard Medical School. In 2024 he was named Director of AI Programs at the Shapiro Institute.

Since the release of ChatGPT in November 2022, the medical education field has been grappling with the impact of large language models (LLMs) and generative artificial intelligence (GAI). These models excel in areas once thought inviolable—acing licensing exams, communicating empathetically with patients, making diagnoses, serving as scribes, and even taking patient histories. However, they also generate false information (hallucinations), encode human biases, and raise privacy concerns, particularly in healthcare. Despite claims that AI will soon replace human clinicians, these technologies have had limited influence on daily practice.

With this in mind, I'm honored to be the first Director of AI Programs at the Carl J. Shapiro Institute for Education and Research. History shows that technology has always disrupted how physicians think, from Laennec's stethoscope, which birthed modern diagnostic medicine, to the problem-oriented medical record, which reshaped clinical practice. I believe LLMs will be as impactful as these innovations. As Foucault described, we may be witnessing an epistemic shift, though predicting the full impact of AI is difficult as we build the future in real-time. Nonetheless, in this role, I am guided by several practical principles:

1. Meaningfully evaluate new technologies. I am not only interested in process evaluations but fundamentally how AI technologies change how people think. These are challenging questions to answer and require a multidisciplinary approach. But the team at Shapiro is up for the challenge!

2. There's no clean divide between education and clinical care. An AI scribe that listens in to our conversations, a chatbot that elicits patient preferences about their care, or a summarization tool that helps write discharge information are not typically thought of as educational interventions. But each of them has important impacts on how we teach and think about our patients. The "hidden curriculum" around how we interact with the EHR has far more

impact on learners than a noon lecture, but it's not something that is traditionally studied.

3. Focus on faculty development. Just like the digital natives of a generation ago, incoming medical students are increasingly GAI natives, having used these technologies throughout undergrad. Faculty have a lot to learn from our trainees, but there are also considerable skill gaps that we will need to bridge.

4. Serving as a learning lab. GAI tools are different from other digital interventions that have come before, even other machine learning risk prediction tools, because of how they interact with human cognition. I truly believe that medical education research, which already focuses on the evaluation of how people think, is poised to play a major part in the rollout of GAI technologies. I hope that BIDMC will play a major role as a learning lab going forward.

5. The importance of data for education. The future of medical education is going to be multidisciplinary, involving not only physicians, but also social scientists, psychologists, computer scientists, and data scientists. Through capacity building and grant funding, I hope to develop a data infrastructure that will allow for true precision medical education.

I honestly don't know how this situation will play out – many questions about the future of our field are out of our hands, dependent on technology companies, chip manufacturers, and government regulators. And medicine is notoriously slow to adopt new technologies – we're one of the only fields that still uses pagers and fax machines. But I also truly believe that all of us are living through the beginning of a tremendous shift in the nature of medicine – one that I am honored and excited to take part in!

As part of this new role, please feel free to reach out to discuss any study ideas, plan faculty development sessions or just brainstorm at any time. My e-mail is arodman@bidmc.harvard.edu, or you can always stop by my office!

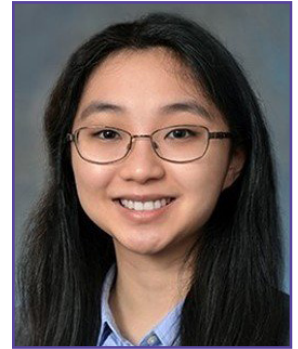


Dr. Adam Rodman

Residency Experiences at BIDMC

Margaret Lie, MD, is a fellow in the MGH-BIDMC Harvard Pulmonary and Critical Care Medicine Program. She completed her medical training at Baylor College of Medicine, followed by a residency in internal medicine at Beth Israel Deaconess Medical Center.

During internal medicine residency at BIDMC, artificial intelligence (AI) was gaining traction in the medical field for research, clinical, and educational purposes. Learning about AI was incredibly important to stay informed about its integration into the practice of medicine. The faculty at BIDMC was incredibly supportive in our endeavor as residents to educate ourselves about this topic. As part of the AI in Medicine Interest Group, we had the privilege to learn about the basics of AI and understand how computers use algorithms and other statistical models to find patterns applicable to certain cases. We held specific discussions on topics such as bias and fairness, clinical decision support, and education and training for our fellow residents. In addition, we presented to participants in the 2023 BIDMC Academy Fall Retreat on Artificial Intelligence in Medical Education and held an Academic Half Day for the Internal Medicine residents in the spring of 2024 to hone skills in prompt design, introduce OpenEvidence, discuss AI-related policy, and learn how to implement AI in database research. We even integrated ChatGPT-4 as a discussant in a Clinicopathological Case presentation.



Dr. Margaret Lie

To disseminate the usefulness of AI and machine learning on a broader scale, Dr. Adam Rodman, Dr. Byron Crowe, and I collaborated on an Academic Medicine Last Page article and provided specific cases where AI can be helpful for medical education. We described the numerous applications of AI to training/learning, curriculum design, and assessment/feedback. It all starts with selecting a high-performing model, such as GPT-4, and then building prompts to define a role and state a command for the model. Through an iterative process, prompts can be further specified to target goals such as creating an outline for a presentation, sifting through feedback, and generating example cases. Practice is key to getting a satisfactory output, which is highly dependent on the quality of the inputs set by the user. The inclusion of more detail and specifiers in the prompt helps the model formulate a response into the desired format/structure and focus.

Several of my co-residents have conducted research on large language models associated with AI and their performance on standardized testing and clinical scenarios; they are also studying how physicians make clinical decisions with the availability of these tools. AI allows for increased collaboration broadly - not only with physicians across the world to study and implement AI judiciously but also with researchers and computer scientists outside of medicine.

What is a 'prompt'?

- Prompts are instructions telling what the model to do - just like we 'prompt' people to do things with language.
- For new AI, prompts can be delivered in 'natural language'. This is different from prior tools that require software code.
- This means anyone can simply write instructions to ask the AI for an output.
- But not all prompts are created equal...

Prompt Design

Some prompts get **much** better outputs than others.

Learning how to build a prompt allows you to unlock the full potential of AI. Good prompt design can be done with a few simple principles and some testing

(you've probably heard of this as 'prompt engineering' which is technically more advanced and requires coding, but same general idea)

How do we design prompts?

How AI Can Be Useful for Medical Educators Right Now

By Dr. Adam Rodman

Given how new AI in health professions education is—and how many conversations gravitate to what AI means for the future—the fact that generative AI can be useful now for educators often gets lost. Despite ChatGPT being the fastest-growing app in history, we're still firmly in "early adopter" territory in medical education, with an intimidating road ahead. We will host faculty development courses across the medical center to build skills, but here is a framework you can use to experiment with AI right now.

- ▶ **Choose the right model.** For most medical education tasks, frontier models—the highest-performing models capable of human-like text output—are the best choice. Fortunately, many options are free. ChatGPT, powered by GPT-4o, is available in free and paid versions. Claude 3.5 from Anthropic is a personal favorite for educational purposes, with reliable outputs and remarkable creativity. Google's Gemini Advanced is also powerful, with a large context window (to upload larger documents). Finally, Meta's Llama 3 400b will soon be released, adding open-source to these frontier models.
- ▶ **Learn the basics of prompting.** Unlike traditional programming languages, you communicate with models through prompts in natural language, making it easy to start. There are basic principles of prompt design, which we've written about in academic medicine here:

https://journals.lww.com/academicmedicine/citation/9900/harnessing_generative_artificial_intelligence_for.855.aspx.

- ▶ **Create cases and multiple-choice questions.** LLMs can create cases for educational sessions and even multiple-choice questions for low-stakes assessment. While these need expert editing, they can save time preparing for educational settings.
- ▶ **An extra participant in journal club.** All major commercial models now include documents in their context window, enabling retrieval-augmented generation (RAG). While base models may hallucinate, RAG generates from text and can cite sources. I use this when running journal clubs, uploading PDFs of studies and supplementary material to "chat" with the text, asking questions about inclusion criteria or the analysis used.
- ▶ **Sort through course feedback.** Educators must sift through large amounts of textual information. While I wouldn't rely on LLMs for high-stakes feedback, for lower-stakes tasks, like teaching evaluations, basic prompting can provide useful sentiment analysis and insights.
- ▶ **Teaching clinical reasoning.** Language models excel at demonstrating clinical reasoning due to the similarities between human cognition and their token prediction architecture. This makes them helpful for case conferences, expanding differentials, and explaining decisions. They can also serve as tutors for case-based learning with medical students.
- ▶ **Don't put private data into public LLMs.** Beyond regulations like HIPAA or FERPA, anything input into a model could be used by the technology company. Exercise caution when entering intellectual property like manuscripts or curricula. BIDMC is working on secure models, but these are not yet available.
- ▶ **Experiment!** Generative AI has exciting properties that make it useful across domains, but with limitations. While there are big questions about the impact on health professions education, for now, the best approach is to experiment, think critically, and have fun!

Spotlight on Scholarship - Leveraging AI Tools

In this issue's Spotlight on Scholarship, we discuss some tips for using generative AI tools to enhance the efficiency of your work in education research and scholarship. In this issue we will cover how to use ChatGPT to address an important question you need to ask while designing a quantitative study: how many people do I need in my sample?

When you are designing your study, you want to be sure that your sample size will provide enough statistical power to allow you to make valid inferences from your statistical tests. Or, if you have a fixed sample size, you may want to find out how much statistical power you will have based on this sample size. For this you need to carry out sample size and power calculations. There are many online programs and software packages to use to address this question, but ChatGPT can make this easy! Here is what you need to tell ChatGPT to get the answers you need:

- Sample size (if you want to estimate power)
- Expected mean difference (if you know this from prior research)



Amy Sullivan, EdD

- Effect size you expect to detect (typically we want to be able to detect medium effect size differences, which for surveys is often a half point (0.5) difference on a 1-5 rating scale)
- Alpha level – typically $p < .05$ (two tailed)
- Power – (if you want to identify needed sample size) typically .80

The examples below show how to determine (1) the best sample size to achieve adequate statistical power, or (2) the statistical power you will have, given an existing sample size.

Want to learn more about statistical power? Check out this link: <https://statisticsbyjim.com/hypothesis-testing/sample-size-power-analysis/>.

Give it a try! And if you need help with this or any part of your scholarship efforts, sign up for a meeting with an education research expert at: <https://www.shapiroinstitute.org/education-research-consult>

1. Example of how to determine sample size to achieve adequate statistical power

I want to be able to detect a difference of .5 between 2 groups on a 1-5 response scale. The standard deviation is 1, alpha <.05 (two tailed), power .80. What should the sample size be to detect a medium and small effect size difference?

To detect a difference of 0.5 between two groups on a 1-5 response scale with a standard deviation of 1, an alpha level of less than 0.05 (two-tailed), and a power of 0.80, the required sample sizes per group are as follows:

- For a medium effect size ($d = 0.5$): Approximately 64 participants per group.
- For a small effect size ($d = 0.2$): Approximately 394 participants per group.

These sample sizes ensure that the study has sufficient power (80%) to detect the expected differences with the specified alpha level.

2. Example of how to determine power based on an existing sample size

I have a study comparing two groups, 64 students in each group. I expect to find a .5 difference between groups on a 1-5 response scale, $P < .05$ (two tailed). Standard deviation is 1. What is the power of my study?

For a study comparing two groups, each with 64 students, expecting to find a 0.5 difference between groups on a 1-5 response scale, with a standard deviation of 1 and an alpha level of 0.05 (two-tailed), the power of the study is approximately 0.801 (80.1%).

This indicates that your study has sufficient power to detect the expected difference with the specified parameters.

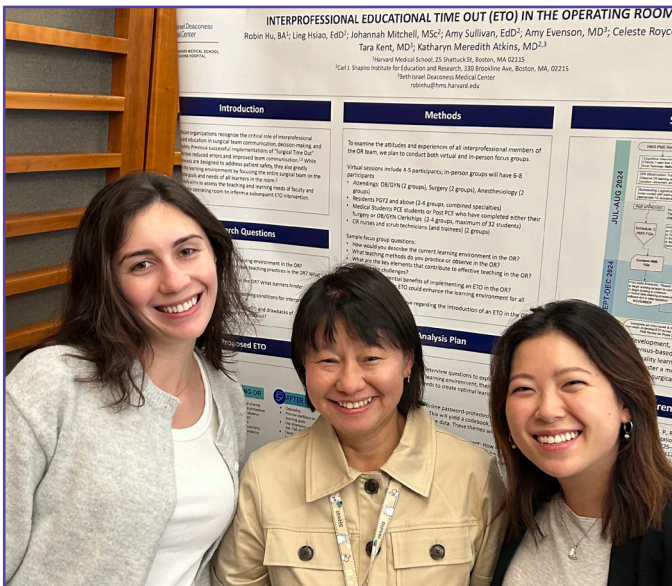
Shapiro Institute awarded Macy Foundation Grant

The Shapiro Institute for Education and Research is proud to announce that it has been awarded a prestigious President's Grant from The Josiah Macy Jr. Foundation. This grant will support the Shapiro Institute's upcoming Millennium Conference 2025, titled "Artificial Intelligence: Prompts, Hallucinations, and the Future of Medical Education." The conference will be held in April 2025 and will include teams of educators from eight medical schools across North America. Request for applications to attend Millennium Conference 2025 expected in November.

The Millennium Conference will explore the transformative impact of AI on health professional education. The Josiah Macy Jr. Foundation, the only national foundation dedicated solely to improving health professional education, is aligned with the Shapiro Institute's mission to enhance public health through innovative educational initiatives and has supported activities of the Institute in the past.

BIDMC Research Retreat Poster Presentation

Harvard Medical student Robin Hu recently presented her poster on 'Interprofessional Educational Time Out in the OR' at the 2024 BIDMC Research Retreat. Robin's work was assisted by Ling Hsiao, EdD, EdM, and Johannah Mitchell, MSc, from the Shapiro Institute's Office of Education Research.



Dr. Meredith Atkins Receives Stone Award



Dr. Meredith Atkins

The Carl J. Shapiro Institute is thrilled to announce that our very own Dr. Meredith Atkins as the 2024 recipient of the S. Robert Stone Award for Excellence in Teaching. The Stone Award was the first Harvard teaching award established at one of its academic affiliates and is a manifestation of BIDMC's strong commitment to its educational mission. It was established in honor of the late past board chair of Beth Israel Hospital by his children. Dr. Atkins is an Obstetrician-Gynecologist and educator at BIDMC and an Assistant Professor at Harvard Medical School. She is the Associate Dean for Curriculum at HMS and an Associate Director of the Shapiro Institute.

As one of her students recently wrote, "Dr. Atkins not only goes above and beyond as a teacher but is a mentor to all students regardless of specialty of interest. She inspires and encourages her students. She is fully invested, and I mean truly invested, in our learning. She is incredibly deserving for this award!"

Dr. Richard Schwartzstein Receives HMS Outstanding Career Achievement in Medical Education Award



Congratulations to Dr. Schwartzstein who will receive the 2024 Outstanding Career Achievement in Medical Education Award at Harvard Medical School (HMS) on October 22nd at the annual Awards Celebration. This award recognizes a member of the HMS faculty who has demonstrated an outstanding level of career achievement as an educator of Harvard medical students through their efforts in the Program in Medical Education. The award is among HMS's most visible commendations of teaching and educational leadership accomplishments and reflects in a tangible way how much Dr. Schwartzstein is valued by his students and by HMS.

Shapiro Institute Welcomes China's Fudan University

On September 24, the Shapiro Institute had the pleasure of welcoming a delegation from Fudan University in China. Dr. Molly Hayes provided an insightful presentation, highlighting the rich history of Beth Israel Deaconess Medical Center (BIDMC), its longstanding affiliation with Harvard Medical School, and the unique aspects that set BIDMC apart. Dr. Hayes emphasized BIDMC's commitment to excellence in quality care, and its leadership in medical education.



Shapiro Institute Hosts Israel's Weizmann Institute of Science

The Shapiro Institute had the pleasure of hosting the Weizmann Institute of Science to help design an innovative MD, PhD program in Israel using the latest advances in medical education to accomplish a seamless integration of basic and medical sciences. The group met with Dr. Schwartzstein to talk about critical thinking, toured the Shapiro Simulation Center, learned about interprofessional teaching and artificial intelligence in medical education during their visit.



HARVARD
MEDICAL SCHOOL

Beth Israel Lahey Health 
Beth Israel Deaconess Medical Center

Carl J. Shapiro Institute for
Education and Research

**PRINCIPLES OF MEDICAL EDUCATION
MAXIMIZING YOUR TEACHING SKILLS**



**Join us for
#HMSMedEd24**

Interactive Learning - Bedside Teaching - Effective Mentoring
Impactful Feedback - Developing Curriculum - Assessing Learners
AI in Teaching - Learner Engagement

**Oct. 23-25
2024**

REGISTER: <https://medicaleducators.hmscme.com>

DEPARTMENT NEWS

Congratulations PCE Class of 2024

A reception was held in the Leventhal Conference Room on Sept. 20 to celebrate the Principal Clinical Experience (PCE) Class of 2024. PCE is a 12-month comprehensive curriculum during which HMS students develop the foundational basis for their medical knowledge while seeing patients in clinical settings. A big thank you to our PCE staff for your exceptional work supporting our AY 24 class, and congratulations to all our graduates. Please keep in touch and best of luck in all your future endeavors!



As the PCE celebration wrapped up, one student remarked, "We may leave the BI, but the BI will never leave us".

GME CORNER

GME Wellness Committee Seminar Series

The GME Wellness Committee invites you to the first of our financial fitness seminar series this year on Contract Negotiations! The seminar will be held in a Hybrid format, zoom and in person, on Wed Oct 9th.

Please RSVP using the link below to receive more details:

<https://biturl.top/QVvlfu> or reach out to Ritika Parris at rparris@bidmc.harvard.edu with any questions.

Additional upcoming sessions will include:

Dec 2024: Student Loans and Cash Flow Part 1

Feb 2025: Disability Insurance


March 2025 (virtual): Buying vs Renting

April 2025: Cash Flow and Savings Part 2

*From all of us at the Shapiro Institute for Education and Research,
we hope you enjoy the beautiful New England fall season!*



**Carl J. Shapiro Institute for
Education and Research**

Follow Us!
X  @BIDMC_Education
www.shapiroinstitute.org