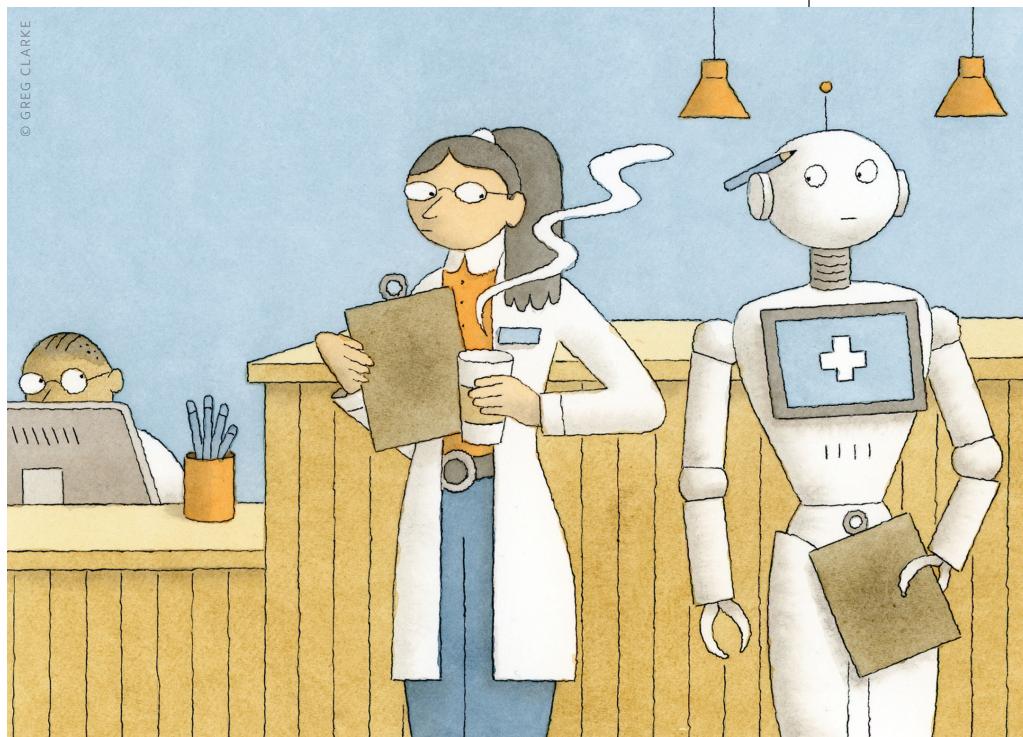


STEVEN E. WALDREN, MD, MS

The Promise and Pitfalls of AI in Primary Care

Programs like ChatGPT have the potential to greatly diminish your administrative burden. But how do you get started — and can you trust them?

Near the end of 2022, a group called OpenAI launched ChatGPT, a large language model (LLM) artificial intelligence (AI) chatbot. It may have seemed like a novelty at first (e.g., “Write a poem about Medicare in the style of *Hamilton*”), but it soon became apparent that ChatGPT and AI models like it could have huge implications for education, business, and even medicine.



By January 2023, ChatGPT had become the fastest-growing consumer software application in history, reaching 100 million users in just two months (TikTok held the previous record at nine months).¹ Before the end of the year, other companies launched similar products, such as Google’s Bard and Microsoft’s Bing Chat (now Copilot).

AI is here to stay and will likely become more embedded in our daily lives in the coming years. If used properly, it could be a

ABOUT THE AUTHOR

Dr. Waldren is chief medical informatics officer for the American Academy of Family Physicians. Author disclosure: no relevant financial relationships.

tremendous boon to primary care physicians, potentially ridding us of administrative tasks that are a leading source of burnout.² But, as with any new technology, there are downsides. This article seeks to illuminate some of the ways AI can help primary care practices now and in the near

AI models are always learning, but it's not advisable to use the current models to guide clinical decision-making.

future—and some of the ways AI could be downright dangerous.

WHAT IS AI?

At its most basic, AI is when computers try to mimic how the human brain works, learning from the information (data) they take in and becoming progressively more capable.

AI has existed in various forms for decades, but what's different about ChatGPT and other LLMs is the sheer amount of data they are able to process and their ability to be "generative." Generative AI can take a prompt from a user (an input as text, image, etc.) and can output almost instantly a novel response based on what it has learned from a massive corpus of existing data. Using Google or another traditional search engine is like looking through books in a library yourself and copying down what one author wrote. But using a generative AI program is like having an assistant who can look through all the books in the library and synthesize all of

that information into a brand new answer.

LLMs are generative AI models trained on enormous volumes of text. The training process allows the model to learn statistical relationships between words and phrases. It then uses these relationships to predict the most likely next word given the user's prompt (and the prior words it just generated). In its most simplistic sense, it is a fancy autocomplete model like you see in smartphone texting applications, where the phone predicts what you may want to type next based on phrases you've used in the past.

One of the reasons AI researchers are so interested in LLMs is the potential for "emergence," which is when an AI model can accomplish tasks that it was not explicitly trained to perform. There is some debate among academics about whether the current models have achieved true emergence, but there is no denying that LLMs can generate responses far beyond what people assumed they could accomplish.

What does that mean for health care? It's not entirely clear yet, but the technology is moving fast. Early LLMs could barely pass the U.S. Medical Licensing Examination, but more recent models, such as Google's medically focused Med-PaLM 2, have achieved relatively high scores.^{3,4} Some of the leading EHR companies are also testing ways to integrate generative AI within their programs.⁵

USER BEWARE

Before we get to how the new generative AI models can help, we should understand how they could harm. First, it is important to remember that these models were trained to generate the best next word (probabilistically speaking)—not to understand logic, the scientific method, or medical questions. Second, their learning is only as good as the data used to train the model (a common maxim in computer science is "garbage in, garbage out," which means that any shortcomings in the data used to create a program will manifest themselves in the program's execution). This leads to two of the biggest problems with current generative AI products: bias and hallucinations.

Any significant biases in the data can be learned by the model. Then the model's

KEY POINTS

- New artificial intelligence (AI) systems such as ChatGPT can reduce administrative burden, but their current shortcomings make it inadvisable to use them to aid clinical decision making.
- Tasks AI can help with now include drafting prior authorization requests, rewriting medicolegal forms in more patient-friendly language, and explaining normal test results.
- Proprietary or HIPAA-protected information should only be submitted to closed, private AI systems, not open systems such as ChatGPT.

responses will be informed by these same biases, which is why you may have read reports of chatbots producing conversations that are racist, sexist, homophobic, or otherwise awful.⁶ Bias in medicine is well-documented, even in clinical guidelines.⁷ Therefore it would not be surprising for generative AI models trained on existing scientific literature to perpetuate these biases. AI developers are designing and implementing tactics to confront this challenge, but AI users should be conscious of the potential for bias in the responses.

The second shortcoming is when LLMs make up something that is not true. AI literature calls this a “hallucination,” which conveys the concept that the AI does not seem to “know” it is being untruthful (i.e., lying). If confronted by the user (with a subsequent prompt), the model is likely to continue to respond as if the hallucination were true, or respond like a toddler and deny it did anything wrong. This type of behavior makes sense. The model was trained to predict the next best word and learned from the vast amount of human text, not all of which adds up. But hallucinations are a very serious obstacle for being able to use LLMs in medicine. For example, in one high-profile instance, ChatGPT created an entire fake data set to support a hypothesis about ophthalmologic care.⁸

Generative AI models are always learning, and each iteration is generally more capable than the last, but it's not advisable to use the current models to guide clinical decision making. You must be able to carefully double-check the AI's answers, and after doing that you've likely wasted more time than you saved. Plus, surveys show most patients are uncomfortable with the idea of doctors using AI to inform treatment decisions.⁹ Fortunately, surveys also show most doctors are similarly wary of it.¹⁰

COMMON USES IN PRIMARY CARE

Now that we've provided the necessary caveats about AI in medicine, it's time to get to the fun part: how generative AI can help family physicians with some of the tasks they most despise. (If you want to experiment with generative AI as you read this article, you can create a free account

at <https://chat.openai.com> or <https://bard.google.com>, but make sure to follow the safeguards described in the box below.)

With a quick browse through the web, we can find news stories, journal articles, blog posts, and forums that discuss the possible uses of LLMs in health care.¹¹⁻¹⁵ These range from performing administrative tasks to generating communications for patients to translating medical jargon. Here are some of the use cases.

- **Rewriting medical or legal forms** in patient-friendly language. For example, you might ask the AI program to “Rewrite this informed consent form for those who read at an eighth-grade level: [insert text]” or “Create a new informed consent form for those with low health literacy.”

- **Summarizing information** such as a

THREE SAFEGUARDS FOR USING AI IN MEDICAL PRACTICE

1. Use artificial intelligence (AI) large language models (LLMs) when the physician or other user is able to easily verify the accuracy of the AI output. For example, it is easy for a physician to look at an AI-generated office visit note and quickly verify whether it is accurate and complete. But when using LLMs to generate initial drafts of messages to patients about lab results or post-diagnosis/post-procedure instructions, first ask, “Can I independently verify the accuracy of the AI response?” and “Does verifying it take less effort and time than generating the output myself?”

2. Do not enter any protected health information or private organizational information into open online LLMs, such as ChatGPT and Google's Bard. For those cases, instead use an LLM embedded in a company focused on health care solutions, such as an EHR vendor, that will operate under a HIPAA business associate agreement. Do your due diligence on the company by asking them questions about the safety of their solution, including their processes to ensure accuracy. You should also plan to verify the output because you are still liable for the safety of your patients. It is essential to protect patient privacy and organizational security. The information entered into an AI model is not safeguarded from public view unless specifically noted, as in a proprietary model.

3. Use the LLM only in low-risk situations. Clinical uses are not recommended in primary care at this point. But independent physicians or physicians in leadership positions could consider leveraging LLMs for administrative functions, for example, creating employee policy documents or generating newsletters for teams. Verification of the information is still needed in these cases. Consider the LLM response a first draft that you must edit, which is still usually much faster than creating a document from scratch.

patient's medical record, a report, insurer policies and regulations, and journal articles. An example would be asking an AI embedded in your EHR to "Give me all the information on [patient X] pertaining to diabetes" or asking ChatGPT to "Summarize this journal article: [insert text]."

- **Generating initial drafts of patient communications** such as responding to portal questions, explaining test results, providing general education on chronic disease care, or explaining new diagnoses. Researchers have found that ChatGPT often responds to patient questions with more empathy than physicians (the machines don't have the same time constraints as us).¹⁶ Still, when it comes to test results, you might want to explain abnormal results yourself and reserve AI for explaining normal results ("Explain normal results for an electrocardiogram").

- **Searching for information within a trusted source** such as the medical record ("Has the patient had a colonoscopy in the last 10 years?") or an evidence-based guideline ("Using the following guideline, what is best course of treatment for a patient with [condition]? The guideline is [text of the guideline]"). While this might seem like using AI to aid clinical decision making, it's

insurance appeals, etc. For example, "Write a letter to [insurance company] requesting authorization for a patient to get an MRI of the left knee." To strengthen your prior authorization request, ask the AI program to reference scientific literature that supports it (but remember to double-check for AI hallucinations), or paste in the insurance company's template or copies of similar requests that were successful in the past and tell the program to use them as models.

- **Generating documentation from an audio recording of an office visit.** There are already AI products on the market that act as virtual scribes, recording the appointment, transcribing it in its entirety, creating a summary, and placing it in the patient's record.¹⁷

Even the uses described above require safeguards (see page 29), such as considering the AI-generated text to be a draft that you must review for accuracy. I would not recommend just firing up ChatGPT, for instance, and using it immediately in practice. Although its makers have added options to keep your chat history private, conversations with ChatGPT are still recorded temporarily and the program has suffered privacy breaches in the past.¹⁸ So, while it might be fine for rewriting generic informed consent forms, any information that is proprietary to your organization or HIPAA-protected should go through an AI platform covered under a HIPAA business associate agreement. And, as noted, current AI models can produce "hallucinations." The consequences may not be as dire for administrative tasks versus clinical ones, but it's still something to be alert to.

AI models can produce "hallucinations." The consequences may not be as dire for administrative tasks versus clinical ones, but it's something to be alert to.

actually using AI to search and curate the trusted guideline that is aiding your decision making.

- **Populating clinical registries.** AI programs within EHRs can increasingly take on this data entry task, using medical records to find and place the appropriate patients on the registry ("Find all patients who have billed for services involving [insert ICD-10 codes] in the past two years and put them in a spreadsheet").

- **Generating initial drafts of referral letters, prior authorization requests,**

LOOKING FORWARD

In my mind, there is no question LLMs will have a prominent position in medicine over the next several years. We are already at a place where there is too much information for humans to manage in health care. Having AI that can summarize and review every piece of information and never forget a single data point can significantly improve health outcomes and decrease the cognitive burden on physicians. Having AI that can handle administrative tasks will free physicians from the EHR and paperwork and allow them to focus on the patient and care delivery. At least one

university is already offering a combined doctor of medicine/master of science in artificial intelligence degree to help prepare physicians for this future.¹⁹

Yet, I also think AI presents significant peril. As long as the financial incentives of medicine are misaligned, there are market pressures to leverage innovations such as LLMs to do things that are not in the best interest of patients and primary care (such as insurers allegedly denying claims based on AI algorithms).²⁰ Because of AI's promise and peril, I believe primary care physicians must become educated about it and its application in medicine. Family physicians should weigh in on the design, development, and deployment of AI in medicine to ensure it is more helpful than harmful to patients, primary care physicians, and practices. **FPM**

1. Hu K. ChatGPT sets record for fastest-growing user base — analyst note. Reuters. Feb. 2, 2023. Accessed Dec. 18, 2023. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>
2. Abid R, Salzman G. Evaluating physician burnout and the need for organizational support. *Mo Med*. 2021;118(3):185-190.
3. Singhal K, Azizi S, Tu T, et al. Large language models encode medical knowledge. *Nature*. 2023;620(7972):172-180.
4. DePeau-Wilson M. Google AI performs at "expert" level on U.S. medical licensing exam. *MedPage Today*. March 14, 2023. Accessed Feb. 8, 2024. <https://www.medpagetoday.com/special-reports/exclusives/103522>
5. Nelson H. How four EHR vendors are leveraging generative AI in clinical workflows. *Tech Target*. Dec. 1, 2023. Accessed Jan. 31, 2024. <https://www.techtarget.com/searchhealthit/feature/How-4-EHR-Vendors-Are-Leveraging-Generative-AI-in-Clinical-Workflows>
6. Wolf ZB. AI can be racist, sexist, and creepy. What should we do about it? *CNN.com*. March 18, 2023. Accessed Dec. 18, 2023. <https://www.cnn.com/2023/03/18/politics/ai-chatgpt-racist-what-matters/index.html>
7. Brown J, Beesley R. It's time: six steps to creating an anti-racist clinic. *Fam Pract Manag*. 2023;30(4):23-26.
8. Naddaf M. ChatGPT generates fake data set to support scientific hypothesis. *Nature*. 2023;623(7989):895-896.
9. Reed T. Patients worry about how doctors may be using AI: survey. *Axios*. Dec. 13, 2023. Accessed Dec. 18, 2023. <https://wwwaxios.com/2023/12/13/survey-patients-wary-of-generative-ai>
10. Bettelheim A. Majority of doctors worry about AI driving clinical decisions, survey shows. *Axios*. Oct. 31, 2023. Accessed Dec. 18, 2023. <https://wwwaxios.com/2023/10/31/doctors-ai-clinical-decisions-concern>
11. Frist B. How generative AI — a technology catalyst — is revolutionizing healthcare. *Forbes*. Aug. 1, 2023. Accessed Dec. 20, 2023. <https://www.forbes.com/sites/billfrist/2023/08/01/how-generative-ai-a-technology-catalyst-is-revolutionizing-healthcare/>
12. Liu S, McCoy AB, Wright AP, et al. Leveraging large language models for generating responses to patient messages. *medRxiv*. 2023;2023.07.14.23292669.
13. Berger E, Dries M. Beyond hype: getting the most out of generative AI in healthcare today. *Bain & Company*. Aug. 7, 2023. Accessed Dec. 20, 2023. <https://www.bain.com/insights/getting-the-most-out-of-generative-ai-in-healthcare/>
14. Flavius NC. Using ChatGPT to speed up patient communications. *Reddit r/medicine*. August 2023. Accessed Dec. 20, 2023. https://www.reddit.com/r/medicine/comments/15v3y11/using_chatgpt_to_speed_up_patient_communications/
15. Reed T. A year after launching, ChatGPT is already changing medicine. *Axios*. Nov. 29, 2023. Accessed Dec. 20, 2023. <https://wwwaxios.com/2023/11/29/chat-gpt-health-care-medicine-clinical-diagnosis>
16. Ayers JW, Poliak A, Dredze M, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA Intern Med*. 2023;183(6):589-596.
17. Reed T. AI might be listening during your next health appointment. *Axios*. Sept. 22, 2023. Accessed Dec. 21, 2023. <https://wwwaxios.com/2023/09/22/ai-listening-doctor-mental-health>
18. Mauran C. ChatGPT rolls out important privacy options. *Mashable*. April 25, 2023. Accessed Dec. 21, 2023. <https://mashable.com/article/openai-chatgpt-chat-history-privacy-setting>
19. Stringer M. Nation's first degree combining medicine, AI comes to San Antonio. *Axios*. Sept. 19, 2023. Accessed Dec. 21, 2023. <https://wwwaxios.com/local/san-antonio/2023/09/19/first-artificial-intelligence-medicine-degree-texas>
20. Emerson J. UnitedHealth, Cigna face lawsuits over alleged automated claims denials. *Becker's Healthcare*. Nov. 27, 2023. Accessed Dec. 21, 2023. <https://www.beckerspayer.com/payer/home-page/unitedhealth-cigna-face-lawsuits-over-alleged-automated-claims-denials.html>

Family physicians should weigh in on the design, development, and deployment of AI in medicine to ensure it is more helpful than harmful.

Send comments to fpmedit@aafp.org, or add your comments to the article online.