

Artificial Intelligence Scribe and Large Language Model Technology in Healthcare Documentation: Advantages, Limitations, and Recommendations

Sarah A. Mess, MD*†‡
 Alison J. Mackey, PhD§
 David E. Yarowsky, PhD¶

Summary: Artificial intelligence (AI) scribe applications in the healthcare community are in the early adoption phase and offer unprecedented efficiency for medical documentation. They typically use an application programming interface with a large language model (LLM), for example, generative pretrained transformer 4. They use automatic speech recognition on the physician–patient interaction, generating a full medical note for the encounter, together with a draft follow-up e-mail for the patient and, often, recommendations, all within seconds or minutes. This provides physicians with increased cognitive freedom during medical encounters due to less time needed interfacing with electronic medical records. However, careful proofreading of the AI-generated language by the physician signing the note is essential. Insidious and potentially significant errors of omission, fabrication, or substitution may occur. The neural network algorithms of LLMs have unpredictable sensitivity to user input and inherent variability in their output. LLMs are unconstrained by established medical knowledge or rules. As they gain increasing levels of access to large corpora of medical records, the explosion of discovered knowledge comes with large potential risks, including to patient privacy, and potential bias in algorithms. Medical AI developers should use robust regulatory oversights, adhere to ethical guidelines, correct bias in algorithms, and improve detection and correction of deviations from the intended output. (*Plast Reconstr Surg Glob Open* 2025; 13:e6450; doi: [10.1097/GOX.00000000000006450](https://doi.org/10.1097/GOX.00000000000006450); Published online 16 January 2025.)

WHAT IS ARTIFICIAL INTELLIGENCE SCRIBE TECHNOLOGY?

As has been extensively discussed in a wide range of forums, artificial intelligence (AI) represents a quantum leap for human efficiency. AI is currently being adapted for generating medical documentation, and early reports suggest these efforts are being adopted

rapidly. The Permanente Medical Group (TPMG), for example, enabled ambient AI scribe technology in 2023, and reported 3442 TPMG physicians using the tool across 303,266 patient encounters in its first 10 weeks.¹ There is clearly a need for evaluation of AI in a medical context. The authors of the current article are a physician early AI adopter, a linguist focusing on human interaction, and a computer scientist with extensive experience with AI and large language models (LLMs). The goal of our collaboration in this article was to share the early adopter's experiences with an AI scribe application, focusing on its advantages and its limitations, along with recommendations for medical AI developers.

HOW DOES AI SCRIBE WORK?

AI scribes are phone-, tablet- or desktop-based applications (eg, Freed, <https://www.getfreed.ai/>; Heidi, <https://www.heidihealth.com/>) that listen to

From *Sarah A. Mess, M. D., LLC, Columbia, MD; †Department of Plastic Surgery, Georgetown University Clinical Faculty, Washington, DC; ‡Department of Plastic Surgery, Johns Hopkins Clinical Instructor, Baltimore, MD; §Department of Linguistics, Georgetown University, Washington, DC; and ¶Department of Computer Science, Johns Hopkins University, Baltimore, MD.

Received for publication August 25, 2024; accepted November 11, 2024.

Presented on AI Scribe at PSTM24, September, 2024, San Diego.

Copyright © 2025 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of The American Society of Plastic Surgeons. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 \(CCBY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal. DOI: [10.1097/GOX.00000000000006450](https://doi.org/10.1097/GOX.00000000000006450)

Disclosure statements are at the end of this article, following the correspondence information.

Related Digital Media are available in the full-text version of the article on www.PRSGlobalOpen.com.

physician-patient interaction and generate an impressively detailed subjective, objective, assessment, and plan (SOAP) note, together with an email summary intended for the patient, which is all typically generated within a minute of the end of an appointment. The AI scribe costs approximately \$100 per month per user versus the average medical scribe salary costs of \$2800 per month.² How is it done (Fig. 1)? First, automatic speech recognition technology converts the audible interaction to text. Next, the LLM being used, for example, Chat Generative Pretrained Transformer 4 (GPT-4) by Open AI, completes the complex task of writing by using artificial neural network algorithms based on potentially trillions of words of both general and medical text and applying them to the automatic speech recognition output of both the physician and patient statements during the encounter.³⁻⁵ AI scribe companies provide an application programming interface with GPT-4 or other LLMs for producing documentation specific to healthcare.⁶ In this article, we report on the early adopter physician author's use of Freed AI, cloud-based software that uploads securely to Microsoft's Azure cloud, which is Health Insurance Portability and Accountability Act-compliant. Freed claims it does not learn from protected health information (PHI), and deletes the notes in 30 days.⁷ Microsoft offers its own AI medical note-taking tool called Nuance DAX Copilot.⁸ Other AI scribe vendors use cloud service providers such as Google Cloud Platform or Amazon Web Services to remotely host the LLM infrastructure.⁹ Google's LLM system used for some medical scribe applications is called

Takeaways

Question: How do artificial intelligence (AI) scribes work and what are the advantages and limitations of using them?

Findings: AI scribes harness large language models, such as Generative Pretrained Transformer 4, trained on large medical records corpora to create medical documentation. They provide benefits in terms of efficiency and more patient-focused encounters. They also pose risks including inaccuracies, biases, and potential breaches of privacy.

Meaning: The article provides an overview of advantages, limitations, and practical recommendations for physicians and medical AI developers to enhance the reliability and safety of AI-generated medical documentation.

Med-Gemini and uses Med-PaLM technology.¹⁰ Amazon Web Service Healthscribe is powered by Bedrock, an umbrella of Foundation Models including the LLMs Claude and Claude Instant.¹¹ The commonly used publicly available version of GPT-4 overcomes being out of date by learning from (and often memorizing) information from millions of people who upload their text when they use ChatGPT.¹² Thus, physicians need to be clear that they should not send PHI through these services, because ChatGPT (and others) essentially retain their input to train, and it may potentially be possible to reidentify and expose the incorporated PHI.⁹ The Food

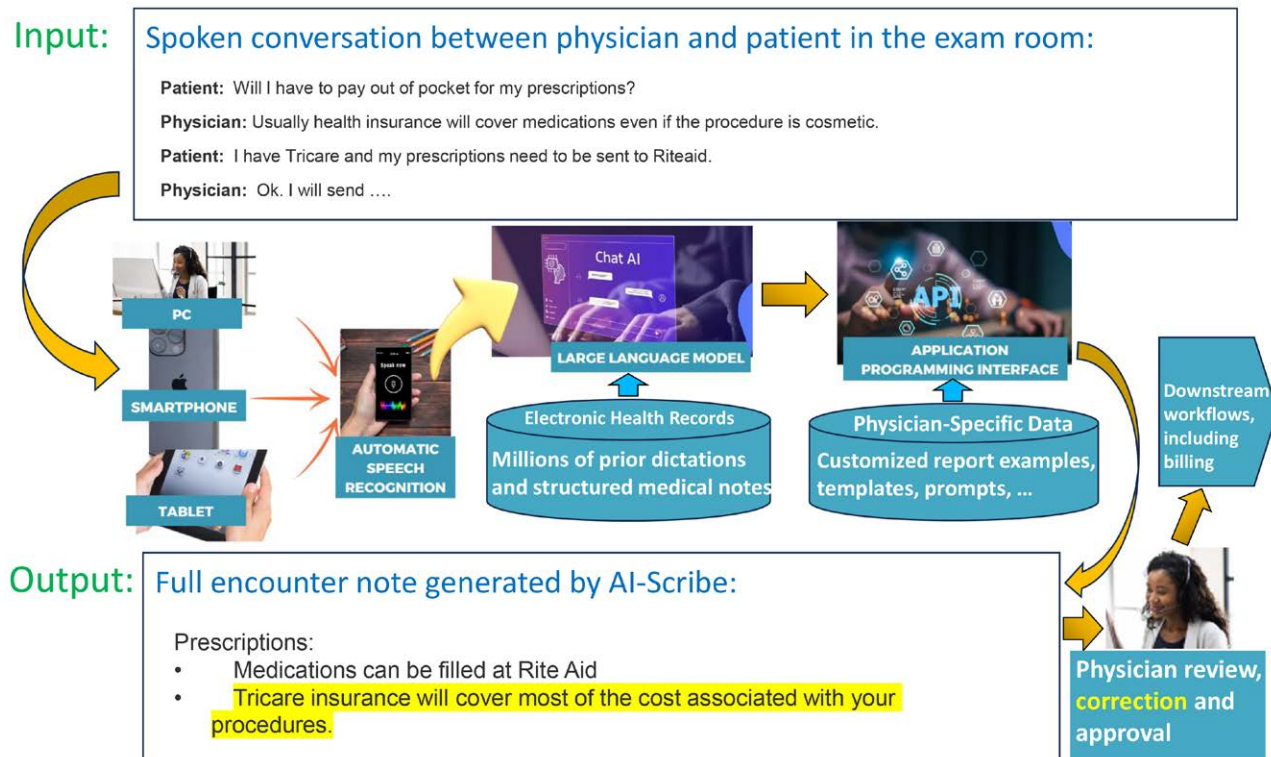


Fig. 1. Diagram of how AI scribe works.

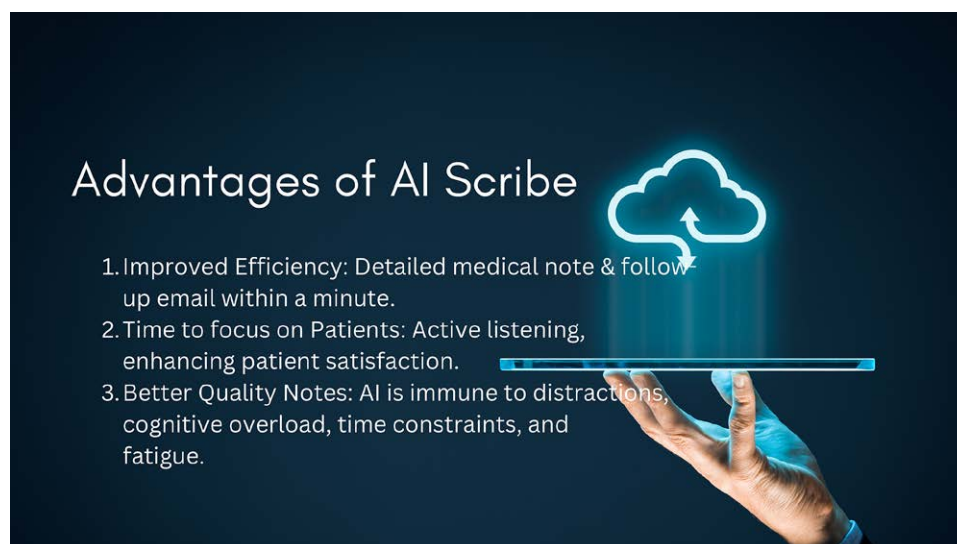


Fig. 2. Advantages of AI scribe.

and Drug Administration (FDA) has purview over the use of AI in healthcare, regulating software as a medical device.^{3,13} However, AI scribe may not fit the definition of software as a medical device, unless it is considered clinical decision support.¹⁴ As with many developments of AI, with the speed at which technology changes, it can be hard for government oversight and regulatory bodies to keep up.

AI SCRIBE IMPROVES PHYSICIAN–PATIENT INTERACTIONS

Harnessing AI to scribe removes the computer screen as a focus point in a patient-physician appointment (Fig. 2). The scribe allows the physician to be a more involved and active listener and communicator through increased observation of patient gestures, sustained eye contact, body language, and nonverbal communications—all done more easily when facing patients and not screens during the interaction. The AI takes on the role of extracting and collating details, making recommendations of diagnostic tests, office visits, referrals to other doctors, prescriptions, activity levels, and instructions for care. An important limitation, though, is that the findings of physical examinations need to be verbalized to be captured by the AI.^{15,16} AI may often also capture a fuller set of clinically relevant information from dialogue with patients than a physician can alone, given that AI is immune to cognitive overload, distraction, fatigue,⁹ and time constraint. AI scribes can also document critical junctures between patients and physicians, such as informed consent or change of surgical plan. AI also helps the physician close the communication loop by drafting an email to be sent to the patient summarizing the appointment. (See appendix, Supplemental Digital Content 1, which displays an email summary created for the patient, <http://links.lww.com/PRSGO/D776>.)

AI SCRIBES AND RECORDINGS

It seems likely that (saved) AI transcripts will become a regular feature of the courtroom and may influence insurance payouts in the near future. There are obvious issues of legality where recordings are concerned, with different countries, and different states of the United States having different regulations about audio recording with and without prior knowledge and consent. Because voices and sounds are not audiorecorded but are transcribed in real time, technically, there is no recording. Currently, automatic transcripts seem to be generally considered a middle ground because the software creating the consultation summary is akin to the already established and accepted practices of having a human medical scribe or another type of virtual assistant in the room. AI scribe companies should obviously consider requirements for disclosure in different jurisdictions. As good practice, prior written informed consent should be obtained before using any type of AI scribe. Freed AI, for example, includes a recommended script for explaining its purpose and obtaining consent from patients. (See appendix, Supplemental Digital Content 2, which displays how to explain AI scribe to patients by Freed AI,⁷ <http://links.lww.com/PRSGO/D777>.)

AI AND PRIMARY CARE

Current commercial AI scribe software is designed for physicians of all types, from primary care to specialized surgeons. Medical AI scribe developers aggregate the knowledge of many fields of medicine and use the best notes of the medical field as target learning material for note content and format. If patients' medical conditions are mentioned in the current inventory of the medical history, AI may provide clinical decision support⁹ and recommend follow-up steps on conditions asynchronous to the consultation, such as “consultation with your gastroenterologist to effectively manage your acid reflux and IBS.” Although GPT-4 has trained on open source medical

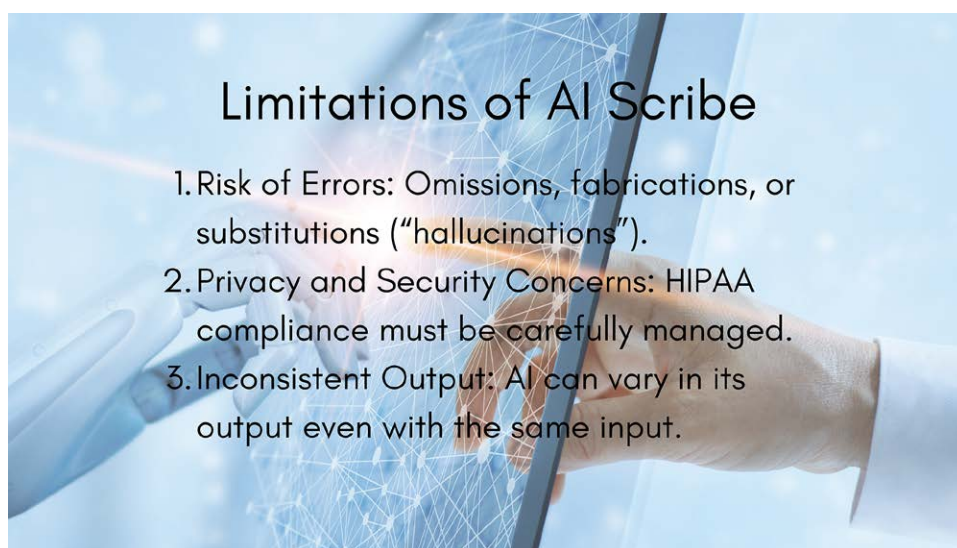


Fig. 3. Limitations of AI scribe.

research literature, websites, podcasts, and videos,¹² it has historically not had access to EHR on private networks for training its general foundation model, necessitating additional clinical-document training by AI scribe vendors.⁶

AI DISTINGUISHES VOICES WITH LIMITATIONS

AI scribe software such as Freed AI usually distinguishes voices (eg, patient and physician) despite currently using only 1 microphone to create the SOAP note. There are some performance limitations though. (Fig. 3) For example, because current standard LLMs are limited to text-based interactions, they cannot currently assess gestures, facial expressions, or other body language. Likewise, conversational pauses, eye contact, back-channeling and a range of other interactional strategies are not assessed.⁴ Often, physician’s comments and explanations may be incorrectly attributed to the patient. Personal, unrelated conversations, part of the chit-chat of human interaction, do not always appear as part of the note, but sometimes they do, and when this happens, they need to be identified and deleted by the proofreader. When multiple interactions over an extended time are captured as a single session, a multiple-patient summary can be generated in error. (See appendix, **Supplemental Digital Content 3**, which displays the multiple-interaction summary which was added to the assessment/plan section. For context, the physician exited patient 0’s room, then entered another room and consulted with patient 1, then entered the corridor to talk to staff. Patients 2, 3, and 4 were related to answering staff questions about other patients. Information is deidentified and the names have been changed, <http://links.lww.com/PRSGO/D778>.)

For these reasons, we recommend that medical AI developers incorporate an introductory phase, where the AI familiarizes itself with the distinct voices of the physician and office staff, along with an alert for when a new voice enters the encounter so that the physician can confirm the encounter is continuing, or tap their end visit button.

AI AND PROCEDURE NOTES

Freed AI offers a SOAP note, but currently no procedure note. This could be intentional due to compliance issues, and/or such a capability may be forthcoming. When attempting to provide the AI with content that can be used for a procedure note, step-by-step details need to be verbalized. Unlike with dictation software, there are no predefined hot words that trigger a phrasal expansion for standard consent or procedure detail. Because procedural notes are important for documenting what was done and major findings, it might be more productive to dictate procedural notes separately or use templates with metaphrases.

AI AND NOTE VARIABILITY

LLMs often vary in their output despite attempts to make uniform input.^{9,17} For example, verbally reviewing the FDA breast implant safety checklist should logically result in a standard list of risks in the AI-generated e-mail to the patient. However, currently, the notes vary from including all the bullet points of the consent to a select few. Medical AI developers should consider ways to achieve more uniform outcomes, perhaps by focusing on emphasis, such as repeating or verbally emphasizing a statement. In the AI field, LLMs are currently understood to be unpredictably sensitive to user prompts and input.¹⁸ It is useful to note that LLM applications can have their “temperature” configured higher for more creativity (and aggressive nonliteral summarization) and lower for more reproducible, literal and deterministic output.⁹

AI AND LLM DEVELOPMENT

Companies developing LLMs are known to keep a tight lid on their methodology,¹⁹ with many aspects of their programming considered close trade secrets. The general “black box” behavior of neural network algorithms needs to be considered in relation to variability. Output is based on complex multilayered learned connections that are opaque²⁰ and generally inscrutable

even to the engineers.²¹ It is difficult to constrain neural networks with established medical knowledge or rules, and incorrect conclusions are often inferred in their unsupervised learning. Currently, the FDA has limited ability to regulate these adaptive algorithms and unsupervised learning for many applications due to both technological and financial constraints.³ “White box” auditing, where testers have full access to internal code and design, has been proposed to give government regulators unrestricted access to the LLMs,²² although this would do relatively little to reduce their inherently opaque and unpredictable behavior. To better understand a physician’s desired style and to point out errors for improved adaptive learning, manual edits could also be utilized as “learn this style” feedback, along with queries to the company. Again, we recommend medical AI developers incorporate an introductory phase to learn about the user, their dictation style, and compliance prerequisites.

AI HALLUCINATIONS: ERRORS OF FABRICATION AND SUBSTITUTION

AI’s hallucinations are usually not in the form of linguistic errors, and thus are not mistakes in the same way as we are used to seeing from human or automatic transcriptions. LLMs are very powerful at being fluent, grammatical, and coherent. However, they sometimes override what they *hear* and put in place what they *think* was intended (or rather, what would be most likely in a given context).²⁰ Examples of problematic hallucinations in the physician author’s AI notes follow. In an email to a patient for implant removal without capsulectomy, the medical AI wrote, “Capsules removed during surgery will be sent to pathology for examination.” However, capsulectomy was explicitly not part of the surgical plan, and was neither discussed nor implied. AI replaced “Aveli for cellulite,” with “Qwo for cellulite” although Qwo was not mentioned. Qwo is also no longer on the market. LLM output often replaces something that is rare or unseen in their training data with something more common and likely in the given context or scenario.²³

Filler brands are not recognized by the AI scribe either, and are almost always replaced with “Botox,” which seems to be treated as a sort of generic word for filler by the AI. In another example, “Sciton BBL,” Broad Band Light for facial skin, was replaced with “Brazilian Butt Lift.” In other words, abbreviations and brand names currently require more guidance to the AI. Hallucinated LLM replacements were again probably based on training data of a few years old.

There are various reasons why LLMs do not update, including freezing the training epoch, paywalls to information, intellectual property issues, or piggybacking on older models. We recommend that retraining or overriding the vernacular of brand names would trigger fewer corrections, and because third-party plug-ins have that ability, it would be helpful for medical AI developers to incorporate this capability.

HOW TO FACILITATE PROOFREADING

With rare or unknown terminology, we recommend that the AI should draw attention to uncertainty and low confidence. This does not seem to be a frequent feature yet. In 9 months of use of Freed AI by the first author, it has only admitted to misunderstanding a word twice, writing “braecerus” in quotes as an explicitly uncertain rendition of the intended “procerus” and placing “bursty” in quotes as an explicitly uncertain rendition of the intended “tuberous” [breasts]. Although there are likely better ways to indicate uncertainty in either the speech recognition or context-driven generation,²⁰ inclusion of quotes as a marker of uncertainty is far preferable to including no indication at all. We recommend medical AI developers incorporate better ways (eg, highlighted text, confidence scores expressed in percentages) to draw physicians’ attention to where the AI has low confidence and prompt the user and/or proofreader to pay attention and issue confirmation or corrections.

AI AND BIAS: RACE, SEX, AND SPONSORSHIP

As has been documented in the AI literature, LLMs may develop sexist and racist biases which are often attributable to the fact that they are learning in part from World Wide Web sources, such as Reddit and social media. ChatGPT, for example, has been observed to attribute occupations requiring higher education to men, high-sentiment terms to “Asian” individuals, and low-sentiment terms to “Black” individuals.²⁴ Biases arise when there is underrepresentation of race and gender in the training data, such as the fact that Black women are less likely to undergo genetic testing of their breast tumor.²⁵ Underrepresentation could be a result of socioeconomic status, incomplete health records, or disparity of care. Biases could also occur with overemphasis on outdated medical practices.³ As an example of potential bias, during note generation for a labiaplasty consultation by the physician author, Freed AI changed “clitoral hood” to “prepuce,” which is either a male anatomical analogue or a gender neutral word but which did not capture the intention of the physician to precisely convey a plan relating to the clitoral hood anatomy.

Sponsorship biases can also arise in healthcare LLMs given the control of companies producing the technology and/or hosting these cloud applications, such as Open AI, Microsoft, Google, and Amazon. The costs associated with developing and maintaining healthcare APIs could make smaller, third-party companies susceptible to sponsorship bias.⁹ Single-source bias occurs when the training data is from one source¹³ and could be an issue with centralization of control over AI.

AI AND THE IMPORTANCE OF CHECKING

As AI technologies make fewer errors, it is natural for users to trust them more and spend less time editing and proofreading. Physicians should always look out for errors where the AI might omit or fabricate information in a subtle or hard-to-detect manner but still with significant falsehood. Therefore, the first author spends more time



Fig. 4. Recommendations for improvement of AI scribe.

proofreading AI's detailed notes and emails than prior charting with templates and drop-down menus. The algorithms depend on pattern matching and frequency rather than abstract human thought and logic, and it is important to understand that their accuracy declines with rare, novel, or intricate concepts.²⁶ AI trains on large datasets: GPT-4 has 13 trillion tokens, where its predecessor GPT-3 had 175 billion parameters to learn from.²⁴ AI searches for emerging patterns and normalizes the average.⁴ Physicians should not become complacent as the technology improves.

AN AI-AUGMENTED FUTURE?

We believe that the age-old medical pedagogy of learning how to write a SOAP will change with increasing use of AI, as will all genres of writing in all fields. Some express concern that overreliance on automation will change and potentially reduce the skills of physicians.²¹ Multimodal LLMs based on image, video, and text are rapidly being developed,¹¹ with, as noted, nonverbal cues likely becoming interpretable by future AI scribes. Teaching communication skills for capture by multimodal LLM might become a component of continuing medical education. Coaching such skills might be profitably built into the AIs themselves by medical AI developers. The landscape of AI is changing so rapidly that next year or even next month will reveal new standards, norms, and ideas, and render some thoughts in the current article obsolete.

CONCLUSIONS

The medical field, like many others, has entered an unheralded era of AI-generated health documentation, data collection, and analysis. The repository of electronic medical record is growing 50% annually, and AI is simultaneously processing such repositories for a growing range of applications, unlocking the huge potential for knowledge discovery from them.^{5,27} Before 2023, LLMs had not had access to large corpora of medical data,³ but now the floodgates have opened. Epic, America's largest

EHR company with 29% of market share and 305 million patient charts, has partnered with Microsoft's Azure Open AI.⁶ Clinical information can be extracted and mined from handwritten notes as well.²⁸ Natural language processing technologies are being harnessed to write clinic notes, respond to patients, provide point-of-care decision support,²⁹ simulate patient encounters for training,⁴ improve workflow, automate the coding of claims, develop treatment protocols, predict complications, diagnose,⁵ make discoveries, and decrease expenditures.⁶ Concomitant with this potential for explosion of knowledge by medical AIs, there is also the potential for exploitation of patient privacy, development of biased algorithms,¹⁶ and fabrication of unreliable content.⁶ Medical AI developers need to use robust regulatory oversights to protect data privacy and security, adhere to ethical guidelines,³ de-bias their algorithms,³⁰ test for accuracy,²² and detect and correct deviations from the intended output (Fig. 4). We recommend physicians adopt a signature disclaimer, for example, "This note was drafted using AI and may potentially contain errors of artificial neural network omission, substitution, or fabrication." Will the healthcare systems installing the software, LLM developers, EHR companies, and makers of healthcare plug-ins to LLM also share responsibility for errors? It seems unlikely. We believe physicians will likely be on the forefront of accountability for errors.¹²

Sarah A. Mess, MD

9821 Broken Land Pkwy #102

Columbia, MD 21046

E-mail: drmess@sarahmessmd.com

Instagram: @sarahmessmd

DISCLOSURES

Dr. Mess is on the speaker's bureau for Allergan and Sciton. She has consulted for Becton, Dickinson, and Company once in the past 2 years and Medical Mutual Liability Insurance Society of Maryland twice in the past year. The other authors have no financial interest to declare in relation to the content of this article.

REFERENCES

1. Tierney AA, Gayre G, Hoberman B, et al. Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. *NEJM Catal Innov Care Deliv*. 2024;5:1–15.
2. Indeed. Available at <https://www.indeed.com/career/medical-scribe/salaries>. Accessed October 10, 2015.
3. Klann JG, Szolovits P. An intelligent listening framework for capturing encounter notes from a doctor-patient dialog. *BMC Med Inform Decis Mak*. 2009;9:1–10.
4. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digit Med*. 2023;6:120.
5. Benítez TM, Xu Y, Boudreau JD, et al. Harnessing the potential of large language models in medical education: Promise and pitfalls. *J Am Med Inform Assoc*. 2024;31:776–783.
6. Edwards B. GPT-4 will hunt for trends in medical records thanks to Microsoft and Epic. Available at <https://arstechnica.com/information-technology/2023/04/gpt-4-will-hunt-for-trends-in-medical-records-thanks-to-microsoft-and-epic>. Accessed July 5, 2024.
7. Summary of Freed Inc.'s Security Structure. *Freed AI*. Available at <https://help.getfreed.ai/en/articles/9346638-summary-of-freed-inc-s-security-structure>. Accessed July 5, 2024.
8. Nuance. Dragon AI clinical solutions: DAX Copilot. 2023. Available at <https://www.nuance.com/healthcare/dragon-ai-clinical-solutions/dax-copilot.html>. Accessed July 5, 2024.
9. Moura L, Jones DT, Sheikh IS, et al. Implications of large language models for quality and efficiency of neurologic care: emerging issues in neurology. *Neurology*. 2024;102:e209497.
10. Corrado G, Barral J. Advancing medical AI with Med-Gemini. 2024. Available at <https://research.google/blog/advancing-medical-ai-with-med-gemini>. Accessed July 7, 2024.
11. Amazon. Amazon Bedrock offers access to multiple generative AI models. 2023. Available at <https://www.amazon.science/news-and-features/amazon-bedrock-offers-access-to-multiple-generative-ai-models>. Accessed July 7, 2024.
12. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI Chabot for medicine. *N Engl J Med*. 2023;388:1233–1239.
13. Rajpurkar P, Chen E, Banerjee O, et al. AI in health and medicine. *Nat Med*. 2022;28:31–38.
14. FDA. Your clinical decision support software: is it a medical device? 2022. Available at <https://www.fda.gov/medical-devices/software-medical-device-samd/your-clinical-decision-support-software-it-medical-device>. Accessed July 7, 2024.
15. Overhage JM, McCallie D, Jr. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. *Ann Intern Med*. 2020;172:169–174.
16. Choi E, Leonard KW, Jassal JS, et al. Artificial intelligence in facial plastic surgery: a review of current applications, future applications, and ethical considerations. *Facial Plast Surg*. 2023;39:454–459.
17. Bafna N, Murray K, Yarowsky D. Evaluating large language models along dimensions of language variation: a systematic investigation of cross-lingual generalization. *arXiv*. 2024.
18. Safranek CW, Sidamon-Eristoff AE, Gilson A, et al. The role of large language models in medical education: applications and implications. *JMIR Med Educ*. 2023;9:e50945.
19. Mitchell M. How do we know how smart AI systems are? *Science*. 2023;381:eadj5957.
20. Salvagno M, Taccone FS, Gerli AG. Artificial intelligence hallucinations. *Crit Care*. 2023;27:180.
21. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA*. 2017;318:517–518.
22. Mökander J, Schuett J, Kirk HR, et al. Auditing large language models: a three-layered approach. *AI Ethics*. 2023;4:1085–1115.
23. González-Corbelle J, Diz AB, Alonso-Moral J, et al. Dealing with hallucination and omission in neural Natural Language Generation: a use case on meteorology. In Proceedings of the 15th International Conference on Natural Language Generation, Santiago de Compostela, Spain; July, 2022.
24. Brown T, Mann B, Ryder N, et al. Language models are few-shot learners. *Adv Neural Inf Process Syst*. 2020;33:1877–1901.
25. Parikh RB, Teeple S, Navathe AS. Addressing bias in artificial intelligence in health care. *JAMA*. 2019;322:2377–2378.
26. Razeghi Y, Logan RL, IV, Gardner M, et al. Impact of pretraining term frequencies on few-shot reasoning. *arXiv*. 2022.
27. Loftus TJ, Tighe PJ, Filiberto AC, et al. Artificial intelligence and surgical decision-making. *JAMA Surg*. 2020;155:148–158.
28. Yehia E, Boshnak H, AbdelGaber S, et al. Ontology-based clinical information extraction from physician's free-text notes. *J Biomed Inform*. 2019;98:103276.
29. Sardar P, Abbott JD, Kundu A, et al. Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. *JACC Cardiovasc Interv*. 2019;12:1293–1303.
30. Hacker P, Engel A, Mauer M. Regulating ChatGPT and other large generative AI models. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL; June 12–15, 2023.