Integrating AI Scribes into Medical Education: Guardrails for Preserving Clinical Reasoning

Jane Abernethy, MD, MBE[1], Anna Shah, MD[2], Belinda Chen, MD[1], Stasia Reynolds, MD[1], Scott M Wright, MD[1], and Paul O'Rourke, MD, MPH[1]

[1]Johns Hopkins Bayview Medical Center, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; [2]Division of General Internal Medicine, University of South Florida, Tampa, FL, USA

ABSTRACT

Clinical documentation is a cornerstone of physician training, not only as a record of care but as a catalyst for clinical reasoning. For medical trainees, writing notes compels them to prioritize information, justify decisions, and synthesize fragmented data into coherent narratives. With the emergence of artificial intelligence (AI) scribes that auto-generate clinical notes from ambient audio, the task of composing notes is increasingly outsourced, raising questions about its impact on education. At our internal medicine residency program, we piloted a 6-month implementation of an AI scribe tool with 48 residents, generating nearly 1000 notes. We propose seven best practices—mapped to the Accreditation Council for Graduate Medical Education (ACGME) Core Competencies—with the goal that AI scribes support, rather than erode, the development of reflective practice and diagnostic thinking. These include establishing baseline documentation skills, structured AI training, critical review of AI-generated notes, and new opportunities for feedback. In this formative moment, educators must guide learners to use AI as a scaffold for reasoning, not a substitute for it.

KEY WORDS: artificial intelligence; AI scribe; medical education; ambulatory care; clinical reasoning

WRITING REFLECTS THINKING

For trainees, writing a clinical note is more than just producing documentation for the record—it is a catalyst for clinical reasoning.[1] It forces learners to choose what matters, justify decisions, and shape a narrative out of fragments of information. But what happens when that cognitive work is outsourced to an artificial intelligence (AI) scribe during earliest stages of professional development?

AI programs capable of generating clinical notes are entering exam rooms with increasing frequency.[2] These tools, which convert ambient audio into structured documentation, promise to reduce clerical burden and return clinicians' attention to the patient.[3] Early studies suggest that AI scribes may improve physician satisfaction, lower burnout rates, and enhance the patient experience by freeing clinicians to listen more deeply and connect more personally.[4–6] Recent evidence also indicates that large language models now perform at a level comparable to medical students on standardized clinical reasoning and knowledge assessments.[7]

Yet as these technologies become more prevalent, their use in medical education raises pressing questions. AI-generated notes may not precisely reflect a trainee's overall reasoning, memory, or decision-making framework.[8] In clinical encounters shaped by nuance, uncertainty, and emotional complexity, AI may miss essential cues or emphasize the wrong details.[9] Errors can creep in subtly, and algorithmic biases encoded in training data specific to race, gender, and historically marginalized groups may influence documentation in ways that escape immediate detection.[10] Beyond accuracy, a deeper concern arises about the fate of writing or presentation skills, clinical judgment, and reflective practice when the work of composing the note is no longer done by the learner. While the educational value of note writing had already been diluted before AI's arrival—by template-driven structures, note bloat, cut-and-paste practices, and the boilerplate text required by regulatory agencies—AI further threatens writing as a space for genuine reasoning and reflection, and its role as an effective tool for learning.[11]

In clinical education, writing is not just a method of recording—it is a method of thinking. Writing allows learners to synthesize information, wrestle with ambiguity, and begin to see their clinical thinking take shape.[12] The clinical note is where diagnostic acumen is sharpened. When this process is delegated to AI without careful oversight, we may risk short-circuiting the development of core competencies.[13] Trainees may become passive editors rather than active authors of care.[8] The clinical note—once a place where the learner made sense of illness and context—risks becoming a stylized output, detached from the inner work of reasoning.

Professional societies have begun to articulate principles to guide the responsible use of AI in clinical practice, but most offer limited direction for educational settings.[14] The 2025 Society of General Internal Medicine (SGIM) position

statement on AI urges clinicians to "insist on high standards of transparency and evidence" and cautions against using generative AI tools for clinical decisions unless those choices can be justified to patients and peers.[15] While these recommendations establish important ethical guardrails, they do not address how ambient AI tools—such as AI scribes—should be supervised or taught within training programs. Recent frameworks, including Gin and colleagues' *Educational Strategies for Clinical Supervision of Artificial Intelligence Use* (*N Engl J Med*, 2025), begin to fill this gap by outlining models for faculty oversight, learner reflection, and prevention of "deskilling" in AI-assisted environments.[8] Together, these guidelines highlight the growing imperative to pair implementation of ambient AI technologies with intentional educational design and faculty development, ensuring that efficiency gains do not come at the expense of clinical reasoning or professional growth.

A "JUST-IN-TIME" PILOT IN PRACTICE

At our internal medicine residency program, we piloted the use of an AI scribe tool—*Abridge*—for 6 months in a longitudinal outpatient clinic.[16] Forty-eight residents were offered the opportunity to use the tool after completing at least 6 months of clinical training and documentation experience. Thirty-one chose to do so, generating nearly 1000 clinical notes during the pilot period. Trainees received a brief, just-in-time orientation on responsible use, data privacy, and best practices for editing AI-generated notes led by a chief resident and Abridge representatives. They were provided a standardized template for their notes, which incorporated a problem-based assessment and plan section. All notes were reviewed and finalized by the resident before submission. Experienced faculty who had previously been responsible for routinely reviewing documentation quality provided oversight. Our precepting model was not impacted by AI scribe implementation.

Overall, residents and faculty regarded the pilot as a valuable and instructive experience, even as it surfaced new challenges in documentation and reflection. Many of our residents who used the technology reported that AI scribes reduced after-hours documentation and allowed them to be more present during patient visits. Yet some trainees expressed a growing detachment from the notes they signed. "It captured what I said," one resident reflected, "but not necessarily what I meant." Reviewing AI-generated notes days after the visit, some found the documentation more challenging to review and less representative of the patient's story. Subtle threads of reasoning—why a specific test was chosen, how a diagnosis was prioritized—were often missing, even when the AI-generated text appeared coherent.

Faculty noticed these patterns too. The diagnostic "voice" of the trainee—the structure of a differential, the articulation of medical uncertainty, the attention to psychosocial detail—was quieter and more generic. In some cases, the resident's reasoning seemed obscured by the fluency of the AI's language. As one preceptor put it, "The note sounded professional, but I couldn't see the full evolution of what the resident was actually thinking." Faculty also voiced concerns regarding the lack of published curricula or guidelines to direct residents in developing proficiency with the AI scribe tools; many remarked that even attending clinicians have not yet figured out how to use the technology effectively. In this evolving landscape, learners may receive inconsistent modeling and feedback, compounding the risk that essential cognitive skills are bypassed rather than established within a solid foundation.

To inform subsequent curricular design, we collected informal feedback from residents, coaching faculty, and 6 other educational scholars through coaching conversations, monthly faculty meetings, a 3-hr faculty mini-retreat, presentation at a monthly academic conference, and specific questioning and email reflections with the lead author. After the pilot, we also developed a standardized note-evaluation rubric, adapted from existing outpatient and inpatient documentation assessment tools, to facilitate the evaluation of self-generated and AI-generated note quality.

GUARDRAILS FOR EDUCATIONAL USE

We do not suggest abandoning AI scribes in education. When used thoughtfully, these tools can enhance learning by freeing trainees to focus on patient connection and by offering new avenues for feedback. In our pilot, several preceptors highlighted that reviewing AI-generated transcripts facilitated targeted coaching on communication, empathy, and diagnostic strategy—an unanticipated educational benefit. But these tools must be integrated with pedagogical intent, not just passive acceptance.

Seventeen clinician-educator faculty, including longitudinal preceptors and chief residents, observed use of the tool and helped identify best practices for implementation. These guidelines, mapped to the ACGME Core Competencies, are designed to ensure that AI scribes support—rather than subvert—clinical education (Table 1).[17]

Ensure Foundational Note-Writing Competence First

ACGME Core Competency: Patient Care, Medical Knowledge, Practice-based Learning and Improvement. Before introducing AI scribes, learners must demonstrate their ability to independently write clear, accurate, and clinically meaningful notes. This skill reflects their capacity for clinical reasoning, synthesis, and communication. Without this foundation, AI risks becoming a crutch that bypasses the hard work of diagnosis and decision-

Table 1 Dos and Don'ts of AI Scribe Use in Medical Education

Encouraged AI practice	Discouraged AI practice
Do ensure foundational note-writing competence first	Don't introduce AI scribes too early
Do provide structured training on AI scribe technology	Don't let trainees skip the editing step
Do mandate active editing and review	Don't use AI as a replacement for feedback
Do use AI notes for reflection and learning	Don't assume AI notes are always accurate or complete
Do foster critical thinking about humanizing narratives	Don't neglect legal, ethical, and privacy considerations
Do teach mastery of customization and adaptation	
Do leverage AI capabilities for more authentic and effective clinical feedback	

making. Based on our collective experience in working with learners for decades, we speculated that a 6-month mark, along with a minimum of 50 patient encounters, served as an appropriate initial benchmark for readiness, acknowledging that some highly skilled trainees may be ready to incorporate AI support earlier. By contrast, others will require additional time and experience and may benefit from a phased adoption with delayed access to the assessment portion of the note that requires significant diagnostic reasoning. As a result of this pilot, we have instituted a program to review the quality of self-generated notes prior to allowing AI scribe use. Careful, individualized assessment by preceptors is essential to determine true readiness for this transition.

Provide Structured Training on AI Scribe Technology

ACGME Core Competency: Practice-based Learning and Improvement, Systems-based Practice. Trainees should not be expected to intuitively understand how to use AI scribes to enhance one's growth as a clinician or improve patient care. Structured training is necessary to ensure they grasp how the technology functions, what its limitations are, and how to use it responsibly. Structured training should include explicit instruction on how large language models process language, the common types of bias they may introduce, and how to identify and correct factual or contextual inaccuracies. Several national and scholarly frameworks provide adaptable guidance for developing such curricula in graduate medical education, including the Association of American Medical Colleges (AAMC)'s Principles for Responsible use of AI in and for Medical Education, the American Medical Association (AMA)'s guidance on augmented intelligence in medical education, and the competency-based framework proposed by Russell et al.[18–20] Recent scholarship also offers practical strategies for teaching AI literacy, including exercises that guide learners to compare AI- and clinician-generated notes to detect inaccuracies, bias, or lapses in patient-centeredness.[21] By incorporating these structured, reflective activities, residency programs can move beyond simple editing tasks to cultivate deliberate, informed oversight of AI documentation tools. In short, a mini-curriculum or course should be offered by training programs.

Mandate Active Editing and Review

ACGME Core Competency: Patient Care, Practice-based Learning and Improvement, Professionalism. As emphasized by Gin et al (2025) in the *New England Journal of Medicine*, educators must help trainees "supervise" AI tools rather than defer to them—approaching AI outputs as *drafts to be interrogated* rather than conclusions to be accepted.[8] AI systems can sometimes hallucinate details, misattribute speakers, or reproduce stigmatizing language, particularly around social or behavioral domains, underscoring the need for vigilance to ensure accuracy and integrity.[22] At the same time, AI scribes can also prompt greater attentiveness to patient identity and consent—for example, by asking clinicians to confirm pronouns for recording. Such prompts reinforce the ethical principles of respect, accountability, and transparency, even as they highlight the necessity of human oversight and clinical ownership. Ultimately, by actively editing and evaluating AI-generated notes, trainees not only have the opportunity to improve the quality of documentation but also deepen their understanding of patient care.

Use AI Notes for Reflection and Learning

ACGME Core Competency: Practice-based Learning and Improvement, Medical Knowledge. AI-generated notes can serve as powerful tools for reflection. By comparing their own notes and recollections from the encounter with the document produced by AI, trainees can identify differences in structure, emphasis, and reasoning.[8] While we did not have a formal reflection curriculum, individual faculty coaches routinely incorporated these comparisons into existing coaching or feedback sessions. An example reflective prompt used was: "As you review the note generated by AI, what is one insight captured that you might have missed, and one aspect of your reasoning or communication that AI did not convey?" By prompting trainees to compare an AI-generated note with their own and reflect on how well each captures the patient encounter, educators embed a cycle of experience, reflection, concept-formation, and experimentation; this may foster deeper listening, integration of knowledge, and mindful documentation.[23]

Foster Critical Thinking About Humanizing Narratives

ACGME Core Competency: Patient Care, Medical Knowledge, Interpersonal and Communication Skills. AI tools may inadvertently omit or de-emphasize important elements of the patient's story, particularly those related to social determinants of health, emotional context, or non-clinical, personal details.[9] When these dimensions are lost, documentation risks reducing complex lives to lists of symptoms and plans. Trainees should be encouraged to examine how well AI-generated notes capture the patient's voice and to consider what is missing that might deepen understanding or empathy. This process invites reflection-on-action, transforming documentation review into a practice of critical awareness.[24] Empathy and attentiveness to the patient's story are core components of clinical skill integral to accurate diagnosis, effective communication, and ethical care.[25] By engaging in reflective comparison grounded in a humanizing lens, trainees ensure that the patient's narrative remains central to their practice of medicine.

Teach Mastery of Customization and Adaptation

ACGME Core Competency: Practice-based Learning and Improvement, Systems-based Practice. Trainees should learn to engage with AI scribes intentionally. By thoughtfully customizing templates and prompts, trainees can offload clerical tasks which enables them to sharpen their focus on listening, identifying key questions, integrating knowledge, and clearly communicating clinical insights to patients and the care team.[26,27]

Leverage AI Capabilities for More Authentic and Effective Clinical Feedback

ACGME Core Competency: Communication skills, Practice-based Learning. AI scribes enable indirect observation in a new form. With trainee permission, faculty can review transcripts of patient interactions—offering a window into clinical reasoning and communication that might otherwise be inaccessible. Though previous literature on ethical, legal, and psychological safety guidelines related to recording trainee interactions should be thoughtfully considered for this application, this may allow for rich, targeted feedback without the pressures of in-room supervision.[28]

LOOKING AHEAD

Decisions made now about the introduction of AI scribes to trainees will shape not only how future physicians document care, but how they learn to think, reason, and relate to patients. As these technologies evolve, they will undoubtedly gain new capabilities that can further enhance clinical accuracy, empathy, and efficiency in patient care. At the same time, learners themselves are becoming increasingly fluent in using AI tools to augment their performance, streamline workflow, and sharpen clinical insight. The challenge for educators, then, is not whether AI will be part of training—but how to guide its use in ways that deepen, rather than dilute, professional formation.

The risk of AI scribes is not merely that there can be inaccuracies—it is that they might make us forget what the act of writing teaches us. Writing a clinical note involves both the recording of what was done and also the discovery of what we truly believe. It is during this process where thinking becomes tangible. In the age of AI, we must ensure that this formative space is not lost, but reimagined. Particularly for learners, it is imperative that they are pushed to discover for themselves how to partner with this valuable technology to synthesize the complexity of clinical encounters into orderly notes that still retain their meaning, nuance, and humanity. Faculty will be required to rethink not just what they teach about documentation, but when and how they engage with learners, recognizing that some rich teaching moments may now arise after the clinical encounter, once the note is complete. If we are intentional, AI scribes can serve as important scaffolds—tools that support deeper reasoning, richer storytelling, and more reflective clinical practice. But if we are not, we risk erasing one of the most powerful learning tools in medicine: the act of writing itself.

Corresponding Author: *Jane Abernethy, MD, MBE; Johns Hopkins Bayview Medical Center, Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA (e-mail: jaberne1@jhmi.edu).*

REFERENCES

1. **Richardson KM, Cristiano JA, Schafer KR, Shen E, Burns CA.** Writing Is Thinking: Implementation and Evaluation of an Internal Medicine Residency Clinical Reasoning and Documentation Curriculum. Med Sci Educ. 2022;32(4):773-777. https://doi.org/10.1007/s40670-022-01570-5.

2. **Poon EG, Harris Lemak C, Rojas JC, Guptill J, Classen D.** Adoption of artificial intelligence in healthcare: survey of health system priorities, successes, and challenges. J Am Med Inform Assoc. 2025;32(7):1093–1100. https://doi.org/10.1093/jamia/ocaf065.

3. **Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Wilson Hannay SB, Castilla K, Lau CS, Kipnis P, Liu V, Lee K.** Ambient artificial intelligence scribes: Learnings after 1 year and over 2.5 million uses. NEJM Catalyst. 2025. https://doi.org/10.1056/CAT.25.0040.

4. **Shah SJ, Crowell T, Jeong Y, et al.** Physician Perspectives on Ambient AI Scribes. JAMA Netw Open. 2025;8(3):e251904. https://doi.org/10.1001/jamanetworkopen.2025.1904

5. **Duggan MJ, Gervase J, Schoenbaum A, Hanson W, Howell JT III, Sheinberg M, Johnson KB.** Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Netw Open. 2024;7(12):e2451234. https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2830383.

6. **Tierney AA, Gayre G, Hoberman B, Mattern B, Ballesca M, Kipnis P, Liu V, Lee K.** Ambient artificial intelligence scribes to alleviate the burden of clinical documentation. NEJM Catal. 2024;5(3). https://catalyst.nejm.org/doi/pdf/10.1056/CAT.23.0404?download=true.

7. **Strong E, DiGiammarino A, Weng Y, et al.** Chatbot vs Medical Student Performance on Free-Response Clinical Reasoning Examinations. JAMA Intern Med. 2023;183(9):1028-1030. https://doi.org/10.1001/jamainternmed.2023.2909.

8. **Gin B, Boscardin C, Abdulnour R-E.** Educational Strategies for Clinical Supervision of Artificial Intelligence Use. N Engl J Med. 2025;393(8):786-797. https://doi.org/10.1056/NEJMra2503232.

9. **Marron JM, Rosenbaum L.** AI-driven clinical documentation — Driving out the chitchat? N Engl J Med. 2023;389(22):2053–2055. https://doi.org/10.1056/NEJMp2416064.

10. **Cross, J. L., Choma, M. A., & Onofrey, J. A.** Bias in medical AI: Implications for clinical decision-making. PLOS Digital Health. 2024;3(11):e0000651. https://doi.org/10.1371/journal.pdig.0000651.

11. **Rule A, Chiang J, Newton Dame R, et al.** Length and Redundancy of Outpatient Progress Notes. JAMA Netw Open. 2021. https://doi.org/10.1001/jamanetworkopen.2021.15334.

12. **Charon R.** "Narrative and Medicine." N Engl J Med. 2004;350(9):862–864. https://doi.org/10.1056/NEJMp038249.

13. **Zhai C, Wibowo S, Li LD.** The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. Smart Learn Environ. 2024;11:28. https://doi.org/10.1186/s40561-024-00316-7.

14. **Furfaro D, Celi LA, Schwartzstein RM.** Artificial intelligence in medical education: a long way to go. Chest 2024;165:771-774. https://doi.org/10.1016/j.chest.2023.11.028.

15. **Crowe B, Shah S, Teng D et al.** Recommendations for Clinicians, Technologists, and Healthcare Organizations on the Use of Generative Artificial Intelligence in Medicine: A Position Statement from the Society of General Internal Medicine. J Gen Intern Med. 2025;40:694–702. https://doi.org/10.1007/s11606-024-09102-0.

16. Abridge AI, Inc. Ambient AI for clinical documentation. 2024. https://www.abridge.com/.

17. Accreditation Council for Graduate Medical Education. (2025). The Milestones guidebook: Competency-based medical education and milestones development. ACGME. https://www.acgme.org/globalassets/MilestonesGuidebook.pdf.

18. Association of American Medical Colleges. Principles for the Responsible Use of Artificial Intelligence in and for Medical Education. AAMC. https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use. Accessed January 8, 2026.

19. Report of the Council on Medical Education, CME Report 4-A-19: Augmented Intelligence in Medical Education (Resolution 317-A-18). American Medical Association; 2019. https://www.ama-assn.org/system/files/cme-report-4-a19-annotated.pdf. Accessed January 8, 2026.

20. **Russell RG, Novak LL, Patel M, Garvey KV, Craig KJ, Jackson GP, Moore D, Miller BM.** Competencies for the Use of Artificial Intelligence–Based Tools by Health Care Professionals. Acad Med. 2023;98(3):348-356. https://doi.org/10.1097/ACM.0000000000004963.

21. **Hui Rincón E, Jiménez D, Chavarro L-A, Pérez Flórez JM, Romero Tapia ÁE, Jaimes Peñuela CL.** Mapping the use of artificial intelligence in medical education: a scoping review. BMC Med Educ. 2025;25:526. https://doi.org/10.1186/s12909-025-07089-8.

22. **Topaz M, Peltonen LM, Zhang Z.** Beyond human ears: navigating the uncharted risks of AI scribes in clinical practice. NPJ Digit Med. 2025;8(1):569. Published 2025 Sep 24. https://doi.org/10.1038/s41746-025-01895-6.

23. **Kolb DA.** Experiential Learning: Experience as the Source of Learning and Development. 2nd ed. Englewood Cliffs, NJ: Prentice Hall; 1984.

24. **Schön DA.** The Reflective Practitioner: How Professionals Think in Action. New York, NY: Basic Books; 1983.

25. **Riess H.** The Science of Empathy. J Patient Exp. 2017;4(2):74-77. https://doi.org/10.1177/2374373517699267.

26. **van Buchem MM, Kant IMJ, King L, Kazmaier J, Steyerberg EW, Bauer MP.** Impact of a Digital Scribe System on Clinical Documentation Time and Quality: Usability Study. JMIR AI. 2024;3:e60020. Published 2024 Sep 23. https://doi.org/10.2196/60020.

27. **Genes N, Sills J, Heaton HA, Shy BD, Scofi J.** Addressing Note Bloat: Solutions for Effective Clinical Documentation. J Am Coll Emerg Physicians Open. 2025;6(1):100031. Published 2025 Jan 13. https://doi.org/10.1016/j.acepjo.2024.100031.

28. **Boscardin CK, Gin B, Golde PB, Hauer KE.** ChatGPT and Generative Artificial Intelligence for Medical Education: Potential Impact and Opportunity. Acad Med. 2024;99(1):22-27. https://doi.org/10.1097/ACM.0000000000005439.