Help Us!

We're offering this module free of charge to help clinicians get familiar with AI in clinical care. We'd love your feedback—please consider completing this short, optional survey. It's part of an educational research project, and no personal information is collected. Thanks for considering! (https://yalesurvey.ca1.qualtrics.com/jfe/form/SV_a9JH1v642zjeL42)

Artificial Intelligence in Clinical Care

Jaideep S. Talwalkar, MD, Wies Rafi, MSc, Andrew Loza, MD, PhD

"The real problem is not whether machines think but whether men do."
—B.F. Skinner

Learning Objectives

1. Develop a working vocabulary for discussing artificial intelligence (AI) in clinical care
2. Critically evaluate major clinical use cases for AI, including ambient documentation, inbox support, clinical decision support, and summarization
3. Consider uses of AI that preserve diagnostic reasoning, communication skills, and professional accountability rather than promoting cognitive offloading and deskilling
4. Apply practical safeguards for clinical and educational AI use, including verification, calibration, cognitive forcing, privacy review, and local governance
5. Discuss the climate implications of AI in health care

Primary Reference

1. Maddox TM, Embí P, Gerhart J, Goldsack J, Parikh RB, Sarich TC. Generative AI in medicine - evaluating progress and challenges. N Engl J Med. 2025;392(24): 2479-2483. (https://www.nejm.org/doi/abs/10.1056/NEJMsb2503956)

Authors' note: Artificial intelligence was used in the creation of this module for ideation (GPT-4o, Claude Sonnet 4.5), literature search (OpenEvidence), and revising language. GPT-4o and Sonnet 4.5 were accessed using a generative AI tool housed within Yale University's secure infrastructure in which inputs are not added to the external training data set.

Case One

Your health system is using ambient AI scribes to help clinicians with their chart notes. You're reviewing a note that the tool just created for your last patient. "This history is outlined really nicely," you think, but when you mentioned "hydroxyzine," the system registered "hydralazine," and your discussion about housing instability is omitted from the assessment and plan.

1. What types of artificial intelligence are commonly used in medical tools? Why is basic knowledge of AI and related terminology important for all clinicians?

Artificial intelligence (AI) is a broad term describing computer systems that perform tasks typically requiring human intelligence. Most modern medical AI tools are built using machine learning, a subset of AI in which algorithms learn relationships from large datasets rather than through explicit programming. Deep learning is a form of machine learning that uses layered computational structures called neural networks, loosely modeled after biological neurons, to detect complex patterns in data such as medical images, clinical notes, or other signals (e.g., EKGs). These systems learn by analyzing large quantities of training data (e.g., clinical notes, radiology images, laboratory results, medical literature) which allow the algorithm to identify statistical patterns and improve performance on specific tasks. Many healthcare applications rely on natural language processing (NLP), a subtype of AI in which computers understand and generate human language, enabling tasks such as summarizing clinical encounters, extracting diagnoses from notes, or answering questions about medical knowledge.

Newer systems that utilize generative AI are increasingly appearing in healthcare workflows. Generative AI can create new content such as text and images based on patterns reviewed during training. Many generative systems in healthcare are built on large language models (LLMs), which are deep learning models trained on enormous collections of written text and then fine-tuned to produce coherent responses to prompts. Tools such as chatbots are interfaces built on top of these models that allow users to interact conversationally with AI systems. Together, these technologies have led to advances in AI-powered ambient documentation tools, patient messaging systems, and decision-support applications.

While the distinctions among these technical concepts were previously in the realm of developers, increasingly, clinicians need a working knowledge of the technology to be effective and responsible consumers. Familiarity with these concepts is important because these systems influence clinical documentation, patient communication, and diagnostic and management reasoning. A basic understanding about how these tools are built can help clinicians recognize risks such as algorithmic bias, hallucinated outputs, and overreliance on automation, while enabling thoughtful and responsible integration of AI into patient care.

Advancements in technology hold great promise but many published studies focus on narrow knowledge tasks rather than real clinical workflows, real patient data, calibration, or deployment performance. The clinician must ask and understand, “What task is this system actually optimized for?” before concluding that a tool is clinically trustworthy.

2. What are ambient scribes, what do they seem to do well, and what are their current limits?

In clinical settings, ambient scribes are workflow tools that capture spoken clinical encounters and using speech recognition models called Speech-to-Text models plus generative language models to draft structured clinical documentation. Numerous studies have demonstrated benefits to practicing clinicians: decreased documentation burden, reduced time writing notes, increased productivity, increased focus on the patient, decreased cognitive load during patient visits, and decreased burnout.

However, the early literature on ambient scribes also emphasizes a need for proofreading and editing. Ambient notes may be inaccurate, over- or underinclusive, stylistically rigid, subtly misleading, or built for generic situations mismatched from a specific clinical circumstance. Commercial tools are built for mass use but may not fit the needs or style of a specific user. For example, a system trained to create outputs in a succinct style following a traditional “SOAP” format may omit information about

development, family structure, structural determinants of health, or anticipatory guidance. Exclusion of these details may be appropriate in some contexts, but could change the meaning of the encounter in others. At worst, the omissions and altered context and tone of a note created by an ambient scribe could exaggerate certainty or suggest erroneous diagnoses with real implications on billing, handoffs, medico-legal interpretation, patient perceptions, and future clinicians' understanding of the patient. Therefore, these tools are best framed as drafting assistants whose outputs require careful clinician review.

Current best practice dictates disclosure of ambient scribes to patients at the point of care. More broadly, disclosure of use of AI tools in professional settings is a broadly acceptable practice. However, these practices may shift as the use of AI tools becomes assumed by patients. For example, clinicians do not typically make overt disclosures about other technology that is now standard in clinical settings, such as a computer with the electronic health record.

3. What is "cognitive deskilling" and how might the use of ambient scribes contribute to it? What strategies might help prevent it?

Documentation is a cognitive act in addition to being clerical work. Many clinicians, especially during training and early in their careers, refine their synthesis while writing their assessment, noticing gaps in the history, inconsistencies in the timeline, questions they should have asked, or future contingency plans. The removal of all friction from note generation risks a reduction in clinical reflection, the impact of which is unclear. Additionally, the creation of a well-crafted output by an automated system can lead to automation bias, the tendency to rely on technology even at the expense of one's own judgment (e.g., following directions of mapping software instead of the shortcut near your house). Over time, automation can lead to loss of an acquired skill, or deskilling (e.g., forgetting there was a shortcut). In medicine, many authors have described "cognitive deskilling" as technology has advanced, with additional warnings that generative AI may also promote "mis-skilling" (learning the wrong thing), or "never-skilling" (never learning a critical skill in the first place). A trainee who sees the note as a clerical task, and only ever learns to edit the output of an ambient scribe, may miss opportunities to build reasoning and reflection skills. Similar concerns have been raised about other uses of generative AI in medicine, such as platforms that create differential diagnoses before the clinician has gone through the cognitive process of diagnostic reasoning on their own.

In educational settings, there are safeguards that hold promise to protect against deskilling. Banning use of technology altogether is problematic because it denies trainees the opportunity to learn how to use these tools. Delayed or graded introduction of AI tools within training programs may be a more practical strategy. For example, a trainee would need to demonstrate foundational competence in note-writing prior to being granted access to an ambient scribe. Cognitive forcing strategies have also been shown to protect against deskilling. This technique requires trainees to first think through the problem or perform the task without the assistance of technology. The technology assistance is then added as a layer, often in real time. For example, the trainee would write a note themselves and then review the note created by the ambient scribe, reflecting not only on the case, but also how the AI system might have impacted their decision making (positively, negatively, or not at all). Alternatively, the trainee could be "forced" to create a prioritized differential diagnosis before consulting AI suggestions, with deliberate reflection on how the AI outputs aligned with the situation. Finally, after coming up with their own ideas, the trainee could use the model like an expert collaborator, asking it to produce competing arguments, alternative explanations, or Socratic requests for deeper reasoning.

Deskilling is not limited to trainees though they are at higher risk than seasoned clinicians because their skills are less ingrained. The same risks and safeguards apply to clinicians at any career phase, especially when it comes to learning a new skill.

The success of these protective strategies requires discipline and self-awareness among clinicians who have an abundance of tools at their disposal. Ultimately, the goal is to embed AI tools in workflows as a cognitive partner rather than a cognitive replacement.

Case Two

You are preparing to see your next patient, a healthy 20-year-old college athlete coming in for a sports physical. Your electronic health record has a built-in clinical decision support tool to help review and analyze chart data. When you open the chart, an alert pops up: "The patient had an EKG performed during an emergency room visit one year ago. Automated analysis has detected a prolonged QT interval. Consider a follow-up EKG for further evaluation."

---

4. What is clinical decision support, and how should clinicians interpret the outputs of these systems?

Clinical decision support (CDS) includes digital tools that assist clinicians with diagnosis, risk estimation, medication safety, preventive care, and management decisions. AI-enhanced CDS refers to systems whose recommendations arise from machine-learning models, probabilistic prediction, or generative synthesis rather than from fixed rule sets alone. Conceptually, this is important because many clinicians think of CDS as either "the alert that pops up in the EHR" or "the chatbot that suggests diagnoses," when in fact it spans a much wider spectrum, from narrowly scoped dosing support to complex prediction models.

AI-CDS is compelling because clinicians routinely face information overload, fragmented records, and time pressure. A well-designed model may detect weak but important signals across many variables and may outperform unaided clinicians on narrowly defined tasks (e.g., detecting hidden signal on an EKG). Human-machine partnership can therefore be genuinely valuable, especially when the tool is tuned to a specific use case and embedded in a logical and supportive workflow.

Clinicians should conceptualize AI-powered CDS as augmentation rather than delegation. The tool may improve signal detection, memory support, or pattern recognition, but it does not assume the clinician's responsibility. The clinician must still decide whether the CDS recommendation is accurate and applies to the patient in front of them. For example, an output may be reasonable at the population level yet inappropriate for the individual patient.

These concepts are similar to how clinicians already apply recommendations from expert guidelines or population-based studies to a specific patient. However, given the polished presentation or numerical precision from a CDS tool, automation bias can lead to ascribing more validity than a model deserves. To counteract this tendency, clinicians will need to learn to challenge CDS suggestions with questions to convert passive acceptance into active appraisal. What data is this output based on? What does the model not know about this patient? What is the cost of being wrong in this specific direction? In this case, a model trained on an adult data set may not consider that longer than "normal" QTc may be normal in young athletes. Blind acceptance of the alert without consideration of the clinical situation could lead to unnecessary anxiety or exclusion from participation. On the other hand, the cost of missing a true

conduction defect could be catastrophic and the intervention (repeat EKG) is relatively inexpensive and non-invasive.

5. What are other emerging uses of AI within electronic health records?

AI-powered tools have been developed to help with inbox management. Using natural language processing, these tools can triage messages and generate draft responses. To date, implementation studies primarily focus on feasibility rather than large-scale deployment, though early results suggest that AI-assisted messages reduce cognitive load, improve efficiency, and offer more empathy than responses drafted by physicians alone. Chart summarization is another emerging application that can be used for pre-charting, rounding, or discharge summary creation. Studies have revealed reductions in documentation time and cognitive load. However accuracy varies and it remains critical for the clinician to remain in-the-loop to verify all information (e.g., ensure that no hallucinations or omissions of critical information are present). This presents a workflow challenge, because unlike clinician review of ambient scribe outputs, verification will require careful chart review, which may offset workload benefits of the tool.

6. As unemotional tools, why can AI systems magnify bias and disparities?

Algorithmic bias in health care can arise from inherent bias or selection bias that is present in training data, labeling inconsistencies, differences in access to care, documentation artifacts, and design decisions that encode historical inequities. In clinical care, an apparently neutral system can reproduce or magnify existing disparities embedded in the data.

To help address this problem, many institutions have developed AI governance systems which provide a structured review process of each new algorithm or digital health tool. Part of this assessment focuses on existing data and planned monitoring for model fairness and bias. Proposals for new AI and statistical models are often required to submit a Model Card, which requires documentation of how the model was developed, what data were used for training and validation, known limitations, and performance across various subgroups. Governance frameworks may also require prospective monitoring to detect difference in reported and real-world performance or emergent biases and patient populations and clinical practices evolve.

Systems built with “explainability” can also help address this problem. Explainability refers to the degree to which systems are designed with transparency of their inner workings such that outputs of AI algorithms are transparent and interpretable. Explainability allows the user to understand what the model is doing, where it may fail, and how to contest its output. However, as systems become increasingly complicated, it is impractical for clinicians to understand all the underlying details. For the front-line clinician, a broad awareness of the potential for bias should inform interpretation of outputs. As noted above, at the point of care, AI tools are best used to assist with rather than replace clinician decision-making. For the subgroup of clinicians with deeper technical knowledge, awareness of the issues that contribute to algorithmic bias can help them implement design solutions to reduce these problems or guide procurement decisions towards products with sufficient explainability.

Case Continued

A student in your office is impressed by the decision support offered by your electronic health record. She shares that she is collecting EKGs to eventually build a chatbot to help her and her classmates get better at EKG interpretation.

7. What privacy, professionalism, and governance considerations are at play here?

Students are increasingly using personalized chatbots to guide their learning. These tools can offer customized feedback that is available anytime. However, there are important data security considerations with the use of any AI tool. Users should be aware of what the platform does with information entered into it. Most consumer-facing large language models absorb user inputs into their training data. Depending on the type of data being entered into the system, this can be problematic (e.g., loss of intellectual property when entering a manuscript draft), unethical (e.g., entering someone else's private information), or illegal (e.g., entering identifiable student, employment, or patient data). Many health systems and universities have enterprise-level access to AI tools like ambient scribes and large language models. In these instances, users must be aware of the data-security clearance for the specific system and whether the system functions within a "walled garden" (i.e., data does not leave the secure computing environment of the organization). Even direct-to-consumer commercial platforms that advertise being HIPAA or FERPA-secure may not be cleared for use within an institution, so it is critical to clarify governance at the local level.

Governance considerations operate within a broader institutional framework that frontline clinicians rarely see. Before an AI tool reaches a clinical workflow, the administration in a health system typically evaluates it through procurement and compliance processes designed to protect patients, staff, and the organization. In practice, this presents a major challenge since vendors of AI products may be unable to answer traditional questions (conceived for non-AI technology) about how their systems were built, validated, or tested for failure.

Most frontline clinicians have no involvement in procurement decisions, but awareness of this process is important. At the point of care, institutional guidance can help clinicians select among countless available tools. A product that has been formally evaluated and cleared through institutional governance is more likely to meet privacy and credibility expectations than one that has not. Additionally, clinicians are often the first to notice when a tool behaves in unexpected ways, and institutional governance channels are the appropriate mechanism for raising those concerns. Finally, as new tools emerge, formal requests from clinicians for use at the institutional level can lead to broader access and better integration with other resources.

When evaluating or inquiring about an AI tool, the following questions reflect the kind of institutional due diligence that protects patients and supports responsible adoption:

- What patient population was this model trained on, and how similar is it to ours?
- Has this tool been tested for performance differences across race, sex, age, or language?
- What happens when the model is uncertain? Does it communicate that uncertainty to the user?
- Is this tool cleared for use within our institution's data security environment?
- Who is responsible for monitoring this tool's performance after deployment, and how are errors reported?

- What is the process for turning this tool off if a safety concern is identified?

These questions will not always have complete answers. If a vendor is unable or unwilling to answer them, this may be a signal that a tool is not ready for clinical deployment, regardless of how compelling its performance benchmarks appear.

In this case, unless the student's chatbot is being built with a HIPAA-secure platform within the university or health system, it is illegal to load patient information into an AI tool. Without permission from the patients themselves, doing so may also be unethical. Even if all visible identifiable information were removed from the EKGs and patients were to give permission, it is possible that meta-data (organizational data in the file itself) embedded in the files could be identifiable. Finally, within a health system, it is likely that governance procedures would dictate that the student would need permission from the organization to build (health system) and study (institutional review board) such a tool.

Case Continued

Hearing this, another student in the office declares "please just use the EKG textbook. We're killing the planet with all this AI use!"

---

8. When it comes to environmental stewardship, what is the "AI paradox" in health care?

Healthcare contributes a substantial share of greenhouse gas emissions in the United States and AI systems are further increasing energy utilization. AI systems require energy-intensive data centers, large-scale computation, and complex supply chains. Additionally, AI systems increase water utilization due to the cooling needs of the data centers that power them. Modern computing also involves hardware requiring mineral extraction, manufacturing, and e-waste disposal. The environmental burden falls disproportionately on low- and middle-income countries and communities. AI may accelerate climate change, which is already shaping patterns of disease, displacement, and health inequity.

The AI paradox is the tension between these planetary stressors and AI's potential to make health care more efficient. Well-designed systems may reduce waste and/or improve health by improving scheduling, avoiding redundant work and testing, prioritizing high-risk patients, and supporting preventive care. Absent longer term analyses, it is premature to draw conclusions about overall risk and benefit related to AI in healthcare and environmental stewardship.

Institutional decisions are likely to have larger impact on energy efficiency than those at the individual level. Pooling computational resources across institutions, sharing development efforts for model-training, signposting relative energy utilization of different models, and having deliberate conversations with vendors about environmental impact are some strategies that might help. At the individual level, before deploying an AI solution, there are variety of questions a user can ask. What problem is this solving? What is the non-AI alternative? Does a solution to this problem exist before I build a new one? Could a smaller or more task-specific model achieve the same goal? Ultimately, at the level of the institution or individual, resource intensity should be matched by demonstrable educational or clinical value.

Finally, in this early phase of widespread AI adoption, clinicians who avoid using modern tools risk "never-skilling" for competencies that will be essential for practice in the near future. While an argument can be made for avoidance of frivolous AI use, it may be these low-stakes, low-stress uses that help

novices build confidence and skills that will allow them to apply AI tools with confidence in higher stakes (clinical) settings.

9. **Optional conversation prompts for further discussion.**

   - Name one current AI use in your current professional workflow that feels high value. Name one that feels high risk.

   - How do you feel about charting after seeing patients? What has been your experience with ambient scribes?

   - What should we disclose to patients about our use of AI in their care? Do we disclose use of other technologies (PubMed searches, Electronic Health Record, etc.)? Is there a difference?

Additional References

1. Abdulnour RE, Gin B, Boscardin CK. Educational strategies for clinical supervision of artificial intelligence use. N Engl J Med. 2025;393: 786-797.
2. Abernethy J, Shah A, Chen B, Reynolds S, Wright SM, O'Rourke P. Integrating AI scribes into medical education: guardrails for preserving clinical reasoning. J Gen Intern Med. Published online February 02, 2026.
3. Albrecht M, Shanks D, Shah T, et al. Enhancing clinical documentation with ambient artificial intelligence: a quality improvement survey assessing clinician perspectives on work burden, burnout, and job satisfaction. JAMIA Open. 2025;8(1): ooaf013.
4. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20: 310.
5. Bedi S, Liu Y, Orr-Ewing L, et al. Testing and evaluation of health care applications of large language models: a systematic review. JAMA. 2025;333(4): 319-328.
6. Bowker D, Torti J, Goldszmidt M. Documentation as composing: how medical students and residents use writing to think and learn. Adv Health Sci Educ Theory Pract. 2022;28(2):453-475.
7. Bratan T, Heyen NB, Hüsing B, Marscheider-Weidemann F, Thomann J. Hypotheses on environmental impacts of AI use in healthcare. J Clim Chang Health. 2024;16: 100299.
8. Buçinca Z, Malaya MB, Gajos KZ. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. Proc ACM Hum Comput Interact. 2021;5(CSCW1):188.
9. Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28: 231-237.
10. Duggan MJ, Gervase J, Schoenbaum A, et al. Clinician experiences with ambient scribe technology to assist with documentation burden and efficiency. JAMA Netw Open. 2025;8(2): e2460637.
11. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19(1):121-127.
12. Lucas HC, Upperman JS, Robinson JR. A systematic review of large language models and their implications in medical education. Med Educ. 2024;58(11): 1276-1285.

13. McCoy LG, et al. Understanding and training for the impact of large language models and artificial intelligence in healthcare practice: a narrative review. BMC Med Educ. 2024;24: 1096.

14. Meskó B, Topol EJ. The imperative for regulatory oversight of large language models (or generative AI) in healthcare. NPJ Digit Med. 2023;6:120.

15. Olson KD, Meeker D, Troup M, et al. Use of ambient AI scribes to reduce administrative burden and professional burnout. JAMA Netw Open. 2025;8(10): e2534976.

16. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med. 2019;380(14): 1347-1358.

17. Richie C, et al. Environmental impacts of artificial intelligence in health care: considerations and recommendations. Health and Technology. 2025;15(6).

18. Risko EF, Gilbert SJ. Cognitive offloading. Trends Cogn Sci. 2016;20:676-688.

19. Shah NH, Entwistle D, Pfeffer MA. Creation and adoption of large language models in medicine. JAMA. 2023;330(9): 866-869.

20. Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. J Am Med Inform Assoc. 2025;32(2): 375-380.

21. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA. 2018;320(21): 2199-2200.

22. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1): 44-56.

23. Tun HM, Rahman HA, Naing L, Malik OA. Trust in artificial intelligence-based clinical decision support systems among health care workers: systematic review. J Med Internet Res. 2025;27: e69678.

24. Wornow M, Xu Y, Thapa R, et al. The shaky foundations of large language models and foundation models for electronic health records. NPJ Digit Med. 2023;6(1): 135.

25. Zhui L, et al. Impact of large language models on medical education and teaching adaptations. JMIR Med Inform. 2024;12: e55933.

Resources

1. Educational resources on AI for clinicians from AAFP. (https://www.aafp.org/family-physician/practice-and-career/managing-your-practice/artificial-intelligence.html)
2. Information on AI in pediatrics care from AAP. (https://www.aap.org/en/practice-management/health-information-technology/artificial-intelligence-in-pediatric-health-care/?srsltid=AfmBOop153mGzvZ-IrI2qu9NOWPz_oz70f6Ntk7nFc178zyQG85b0CTN)
3. ACP artificial intelligence resource hub for clinicians. (https://www.acponline.org/clinical-information/clinical-resources-products/artificial-intelligence-ai-resource-hub)
4. Digital health resources from AMA. (https://www.ama-assn.org/practice-management/digital-health)