

PERSPECTIVE

The Moment AI Arrived in the Clinic: Insights from the SAIL 2025 Year in Review

Pierre Elias , M.D.,¹ Kameron Collin Black , D.O., M.P.H.,² Payal Chandak , B.A.,³ Elizabeth Healey , Ph.D.,³ Rachit Kumar , Ph.D.,⁴ Michelle M. Li , Ph.D.,³ Hugo Morales , M.D.,⁵ Brett Beaulieu Jones , Ph.D.,⁶ and Emily Alsentzer , Ph.D.²

Received: July 15, 2025; Revised: August 20, 2025; Accepted: August 27, 2025; Published: October 16, 2025

Abstract

The past year saw the unprecedented adoption of AI in clinical practice, most notably through the rapid integration of language models into administrative workflows and increased clinician reliance on chatbots for diagnostic and management recommendations. Recent clinical trials and economic analyses have demonstrated that some AI applications meaningfully improve patient outcomes and offer cost-effective approaches to care. However, not all AI tools have demonstrated a clear benefit. Companies have encountered legal challenges related to AI-driven insurance decisions, and uncertainty remains regarding optimal clinician–AI interaction and effective regulatory oversight. Drawing from the Symposium on AI for Learning Health Systems (SAIL) 2025 Year in Review, this perspective examines six key areas that defined clinical AI from 2024 to 2025. We highlight areas where AI tools demonstrated meaningful clinical and operational impact and where challenges persist. Finally, we highlight critical uncertainties the medical community must confront to ensure the enthusiasm surrounding clinical AI translates into lasting clinical impact.

Introduction

The past year marked a turning point for AI in clinical care, as ambient documentation tools and decision support algorithms moved decisively from pilot projects into widespread implementation across health care systems. Industry surveys indicate that approximately 70% of payers and providers in the United States are pursuing or have already implemented generative AI projects.^{1,2} This rapid adoption has begun reshaping clinical workflows, patient–clinician interactions, and reimbursement policies. Yet, as AI shifts from controlled research settings into complex real-world environments, both its benefits and its limitations have become more evident. This perspective summarizes insights from the 2025 Year in Review at the Symposium on AI for Learning Health Systems (SAIL) 2025, an annual symposium on the opportunities for and barriers to deploying AI in clinical settings. The 2025 meeting convened over 150 participants, including clinicians, informaticists, computer scientists, health system leaders, patients, payers, regulators, and industry stakeholders. The Year in Review was curated by a committee of researchers and clinicians through an extensive review of literature, industry reports, and news articles to identify

Kameron Collin Black, Payal Chandak, Elizabeth Healey, Rachit Kumar, Michelle M. Li, and Hugo Morales contributed equally to this article.

The author affiliations are listed at the end of the article.

Emily Alsentzer can be contacted at ealsentzer@stanford.edu.

themes from the past year. We examine six key areas that defined clinical AI from 2024 to 2025, highlighting where AI tools demonstrated meaningful clinical and operational impact, where they fell short, and the necessary steps to ensure their safe, equitable, and effective integration into clinical practice.

AI for Clinical Documentation

Ambient AI documentation tools emerged as the most widely adopted generative AI application in health care this past year. Among 43 responding health systems, clinical documentation was the only use case reported to have universal adoption activities.³ Studies reported benefits such as reduced burnout and less after-hours work. In one multicenter cohort, burnout dropped from 77 to 51% after implementation of AI scribes.⁴ However, efficiency gains were mixed. A large longitudinal study showed no significant reduction in documentation time or increase in clinical productivity or revenue per visit.⁵ In addition, some AI transcription systems have been prone to hallucinating text, with fabricated clinical content raising safety and privacy concerns.⁶ Clinician satisfaction also remains uncertain. Although ambient systems were generally rated easy to use in subjective surveys, overall attitudes were mixed, with one large academic group reporting roughly equal numbers of promoters and detractors (Net Promoter Score≈0).⁷

Although current systems remain imperfect, the pace of technical improvement is striking, far outpacing the typical adoption trajectories of other clinical technologies. Yet key challenges remain, raising difficult but necessary questions: How should value be measured beyond user satisfaction? What safeguards are needed to prevent hallucinated content from entering the medical record — or clinicians from accepting it uncritically — due to automation bias? And what are the downstream consequences of relying on AI-generated notes, including their impact on care decisions, liability, and future models trained on these records?

AI for Clinical Decision Support

Since 2024, the evidence base for AI decision support has advanced significantly, with some trials now demonstrating clear patient benefits from AI. One study reported a 17% reduction in 90-day mortality among hospitalized patients when an AI electrocardiogram (ECG) alert system flagged high-risk findings and physicians acted on them (3.6% vs. 4.3% death; hazard ratio, 0.83).⁸ A clinical trial of another

model, EchoNext, had twice the positive predictive value in detecting structural heart disease as routine clinical care and found that half of all high-risk patients do not go on to get an echocardiogram, highlighting that many patients may remain undiagnosed.⁹ In heart failure management, one group found that providing physicians with real-time data from a novel wearable sensor led to 38% fewer rehospitalizations in 90 days.¹⁰ These trials exemplify how AI can translate into clinical action: the ECG alerts prompted confirmatory testing and treatment for at-risk patients, and the wearable data guided earlier interventions to prevent decompensation.

However, not all AI-supported clinical decision support tools have demonstrated clear benefit. A multicenter trial of AI-based intracranial hemorrhage triage using head computed tomography scans showed no improvement in sensitivity, specificity, or reporting turnaround times when radiologists were assisted by AI, compared with standard workflows.¹¹ This suggests that some AI tools, even if accurate in retrospective studies, may not help in real workflows. Moving forward, maximizing the effectiveness of AI-based clinical decision support tools will require identifying both the clinical contexts and implementation factors that enable accurate AI predictions to consistently translate into improved patient outcomes.

AI in Insurance Denials

Recent reports indicate that health plans are increasingly using AI to autodenial claims, sometimes with minimal human review,^{12,13} prompting lawsuits against insurers, as well as the rise of new AI-powered services designed to fight insurance denials.¹⁴ Moreover, some patients flagged for early disease by AI have been denied reimbursement due to lack of traditional clinical indicators, resulting in financial distress — or “financial toxicity” — for patients and contributing further to insurers’ legal challenges.¹⁵ Resolving these emerging tensions between AI-driven coverage practices and patient-centered care will require novel regulatory and ethical frameworks to avoid exacerbating existing disparities in health care access.

Clinician–AI Interaction

As clinical AI tools become commonplace, the critical challenge now is learning how clinicians and AI can optimally collaborate to enhance patient care. Does doctor plus AI always beat either alone? Recent trials suggest

the answer depends on the task. In a prospective German screening study (PRAIM study [Prospective Multicenter Observational Study of an Integrated AI System with Live Monitoring], 2025), radiologists using an AI-assisted double-reading workflow found 17.6% more breast cancers than standard double reads.¹⁶ The AI-guided group detected 6.7 cancers per 1000 screened versus 5.7 cancers per 1000 screened in controls, without raising false recalls. This indicates a synergistic effect: the computer complements, rather than replaces, human radiologists.

In contrast, the value of clinician-AI collaboration in diagnostic reasoning remains less certain. Google DeepMind's conversational multiagent system (Articulate Medical Intelligence Explorer, or AMIE) outperformed primary care physicians across diagnosis, management, communication, and empathy in standardized patient scenarios.¹⁷ Similarly, two recent randomized controlled trials showed Generative Pretrained Transformer 4 alone matched physicians using conventional or large language model (LLM)-supported resources in clinical management and surpassed them in diagnostic accuracy.^{18,19} Such results demonstrate AI's potential to exceed clinician performance in knowledge-driven tasks, but there is a real need to ensure clinicians are calibrated to the value of new technologies and use them in real-world settings to determine how they truly fare.

However, these findings do not imply that AI will replace doctors any time soon. Rather, they highlight the complexity of integration, emphasizing that effective clinician-AI workflows require thoughtful protocols and targeted training. Institutions must clearly define tasks AI can safely manage (e.g., initial triage, routine follow-up) and those necessitating human judgment. Naively pairing clinicians with AI may not always yield additive benefits, especially if time constraints, automation bias, or poor interface design undermine meaningful human review. Future research into team dynamics, interface optimization, the presentation of AI uncertainty, and safeguards against overreliance will be crucial. For example, Penda Health in Nairobi, Kenya, is conducting the largest randomized controlled trial of LLM-based clinical decision support to date, testing a "traffic light" safety-net model designed to reduce alert fatigue while preserving clinician autonomy.²⁰ The trial, still ongoing, underscores the importance of real-world clinician-AI evaluation, particularly in resource-constrained settings. Ultimately, this prompts reflection on the core focus of clinical practice. Rather than defining human roles simply as what AI performs poorly, we must actively decide where human expertise is genuinely desired. For example, if AI

demonstrates greater empathy than physicians in addressing patient concerns, should we entrust difficult conversations to AI, or redouble efforts to strengthen clinicians' bedside manner?

Economics of AI in Health Care

Investor enthusiasm for health care AI remains high, with approximately US\$11 billion invested in 2024 alone.^{1,2} By contrast, the broader digital health market has underperformed in recent years, and many hospitals face increasing financial strain as inflation and care costs outstrip reimbursement growth.²¹ Thus far, enthusiasm for AI appears to exceed the available evidence for sustained return on investment (ROI). For health care AI to be economically sustainable, stakeholders must clearly demonstrate value through cost savings or meaningful clinical improvements.

Encouragingly, recent studies offer positive examples of economically impactful AI applications. For instance, an AI-driven ECG system used to screen for low left ventricular ejection fraction was found to be highly cost-effective.²² In the United Kingdom, personalized breast cancer screening driven by an AI risk model was both cost-saving and more effective,²³ while in China, a hierarchical AI screening strategy for cataracts demonstrated superior cost-effectiveness compared with all other strategies.²⁴

However, these successes remain exceptions rather than the rule. Looking ahead, health systems will require better frameworks for evaluating the economic impact of AI across varied contexts. Such frameworks must account for indirect cost savings, such as reduced clinician burnout or prevented diagnostic delays, and enable meaningful ROI assessments even as multiple AI tools are implemented simultaneously across clinical and operational domains.

Regulation and Postdeployment Safety

Many experts now argue for a "life cycle" approach to AI oversight, including supervised deployment, local monitoring, and periodic retraining.²⁵ Recent evidence underscores this need: a 2024 study comparing proprietary and open-source early warning scores found that a lack of transparency regarding model thresholds and performance characteristics made fair evaluation nearly impossible.²⁶ Similarly, researchers in Canada demonstrated that AI models analyzing chest x-rays during the coronavirus

disease 2019 pandemic failed to detect significant population-level data drift unless specialized monitoring tools were applied.²⁷ These examples highlight the urgent need for robust monitoring frameworks to ensure postdeployment safety.

Although regulatory frameworks remain under debate, unregulated AI medical scribes continue to be deployed widely across health care systems. The U.S. Food and Drug Administration has yet to authorize an LLM, as it is unclear whether to evaluate them as new devices or “novel forms of intelligence” that undergo training and evaluation similar to physicians and nurses.²⁸ Should LLMs be held to the same standards as expert human clinicians? Are humans the ideal examiners of LLMs’ behavior? Until a coherent regulatory pathway emerges, clinicians and health systems will shoulder the responsibility and liability for ensuring these tools benefit rather than harm patients.

Outlook

Clinical AI has passed an inflection point. AI has appeared in the back office and at the bedside, and it is here to stay. Every health system and clinician will likely have at least one AI technology they must reckon with in the coming year. Given this rapid proliferation, it is essential to evaluate the consequences of integrating AI into documentation and decision-making, lest we embed new risks rather than deliver lasting benefits. The field now faces three urgent tasks: first, to establish rigorous frameworks for evaluating AI safety, efficacy, and postdeployment performance; second, to conduct meaningful ROI assessments to guide adoption; and third, to redefine the roles of clinicians and institutions so that AI complements rather than undermines human judgment. Achieving these goals will require aligning market incentives with meaningful patient outcomes and developing transparent pathways for oversight. Whether we meet these requirements will determine if AI becomes a durable driver of better health or a short-lived experiment.

Disclosures

Author disclosures are available at ai.nejm.org.

Author Affiliations

¹ Columbia University, New York

² Stanford University, Stanford, CA

³ Harvard Medical School, Boston

⁴ University of Pennsylvania, Philadelphia

⁵ Munai Health, Curitiba, Parana, Brazil

⁶ University of Chicago, Chicago

References

1. Deakers C. Roadmap: healthcare AI. Bessemer Venture Partners. September 25, 2024 (<https://www.bvp.com/atlas/roadmap-healthcare-ai>).
2. Bousleiman R, Sletteland P, Barkatally N, Spencer J, Lamarre T. The AI-powered healthcare experience: mapping the patient journey. 2025 (<https://www.svb.com/trends-insights/reports/artificial-intelligence-ai-in-healthcare/>).
3. Poon EG, Lemak CH, 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;32:1093-1100. DOI: [10.1093/jamia/ocaf065](https://doi.org/10.1093/jamia/ocaf065).
4. Misurac J, Knake LA, Blum JM. The effect of ambient artificial intelligence notes on provider burnout. *Appl Clin Inform* 2025;16:252-258. DOI: [10.1055/a-2461-4576](https://doi.org/10.1055/a-2461-4576).
5. Liu T-L, Hetherington TC, Dharod A, et al. Does AI-powered clinical documentation enhance clinician efficiency? A longitudinal study. *NEJM AI* 2024;1(12). DOI: [10.1056/AIoa2400659](https://doi.org/10.1056/AIoa2400659).
6. Koenecke A, Choi ASG, Mei KX, Schellmann H, Sloane M. Careless whisper: speech-to-text hallucination harms. In: *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY: Association for Computing Machinery, 2024:1672-1681. DOI: [10.1145/3630106.3658996](https://doi.org/10.1145/3630106.3658996).
7. 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:e2460637. DOI: [10.1001/jamanetworkopen.2024.60637](https://doi.org/10.1001/jamanetworkopen.2024.60637).
8. Lin C-S, Liu W-T, Tsai D-J, et al. AI-enabled electrocardiography alert intervention and all-cause mortality: a pragmatic randomized clinical trial. *Nat Med* 2024;30:1461-1470. DOI: [10.1038/s41591-024-02961-4](https://doi.org/10.1038/s41591-024-02961-4).
9. Poterucha TJ, Jing L, Ricart RP, et al. Detecting structural heart disease from electrocardiograms using AI. *Nature* 2025;644:221-230. DOI: [10.1038/s41586-025-09227-0](https://doi.org/10.1038/s41586-025-09227-0).
10. Boehmer JP, Cremer S, Abo-Auda WS, et al. Impact of a novel wearable sensor on heart failure rehospitalization: an open-label concurrent-control clinical trial. *JACC Heart Fail* 2024;12:2011-2022. DOI: [10.1016/j.jchf.2024.07.02](https://doi.org/10.1016/j.jchf.2024.07.02).
11. Savage CH, Tanwar M, Elkassem AA, et al. Prospective evaluation of artificial intelligence triage of intracranial hemorrhage on noncontrast head CT examinations. *AJR Am J Roentgenol* 2024;223:e2431639. DOI: [10.2214/AJR.24.31639](https://doi.org/10.2214/AJR.24.31639).
12. Mello MM, Rose S. Denial-artificial intelligence tools and health insurance coverage decisions. *JAMA Health Forum* 2024;5:e240622. DOI: [10.1001/jamahealthforum.2024.0622](https://doi.org/10.1001/jamahealthforum.2024.0622).
13. Mello MM, Cohen IG. Regulation of health and health care artificial intelligence. *JAMA* 2025;333:1769-1770. DOI: [10.1001/jama.2025.3308](https://doi.org/10.1001/jama.2025.3308).

14. Baxi S, Parikh S, Peterson, M, Ray A. Setting the revenue cycle up for success in automation and AI. McKinsey and Company. July 2023 (<https://www.mckinsey.com/~media/mckinsey/industries/healthcare%20systems%20and%20services/our%20insights/setting%20the%20revenue%20cycle%20up%20for%20success%20in%20automation%20and%20ai/setting-the-revenue-cycle-up-for-success-in-automation-and-ai.pdf?shouldIndex=false&sa=D&source=docs&ust=1752601961196768&usg=AOvVaw3quluyZreULkjW-OcejO10>).
15. Jain SS, Mello MM, Shah NH. Avoiding financial toxicity for patients from clinicians' use of AI. *N Engl J Med* 2024;391:1171-1173. DOI: [10.1056/NEJMp2406135](https://doi.org/10.1056/NEJMp2406135).
16. Eisemann N, Bunk S, Mukama T, et al. Nationwide real-world implementation of AI for cancer detection in population-based mammography screening. *Nat Med* 2025;31:917-924. DOI: [10.1038/s41591-024-03408-6](https://doi.org/10.1038/s41591-024-03408-6).
17. Tu T, Schaeckermann M, Palepu A, et al. Towards conversational diagnostic artificial intelligence. *Nature* 2025;642:442-450. DOI: [10.1038/s41586-025-08866-7](https://doi.org/10.1038/s41586-025-08866-7).
18. Goh E, Gallo R, Hom J, et al. Large language model influence on diagnostic reasoning: a randomized clinical trial. *JAMA Netw Open* 2024;7:e2440969. DOI: [10.1001/jamanetworkopen.2024.40969](https://doi.org/10.1001/jamanetworkopen.2024.40969).
19. Goh E, Gallo RJ, Strong E, et al. GPT-4 assistance for improvement of physician performance on patient care tasks: a randomized controlled trial. *Nat Med* 2025;31:1233-1238. DOI: [10.1038/s41591-024-03456-y](https://doi.org/10.1038/s41591-024-03456-y).
20. Mateen BA, Menon V, Agweyu A, et al. Trials for LLM-supported clinical decisions in African primary healthcare. *Nat Med* 2025;31:2833-2835. DOI: [10.1038/s41591-025-03815-3](https://doi.org/10.1038/s41591-025-03815-3).
21. American Hospital Association. Costs of caring: challenges facing America's hospital in 2025. April 2025 (<https://www.aha.org/costsofcaring>).
22. Thao V, Zhu Y, Tseng AS, et al. Cost-effectiveness of artificial intelligence-enabled electrocardiograms for early detection of low ejection fraction: a secondary analysis of the electrocardiogram artificial intelligence-guided screening for low ejection fraction trial. *Mayo Clin Proc Digit Health* 2024;2:620-631. DOI: [10.1016/j.mcpdig.2024.10.001](https://doi.org/10.1016/j.mcpdig.2024.10.001).
23. Hill H, Roadevin C, Duffy S, Mandrik O, Brentnall A. Cost-effectiveness of AI for risk-stratified breast cancer screening. *JAMA Netw Open* 2024;7:e2431715. DOI: [10.1001/jamanetworkopen.2024.31715](https://doi.org/10.1001/jamanetworkopen.2024.31715).
24. Wu X, Wu Y, Tu Z, et al. Cost-effectiveness and cost-utility of a digital technology-driven hierarchical healthcare screening pattern in China. *Nat Commun* 2024;15:3650. DOI: [10.1038/s41467-024-47211-w](https://doi.org/10.1038/s41467-024-47211-w).
25. Warraich HJ, Tazbaz T, Califf RM. FDA perspective on the regulation of artificial intelligence in health care and biomedicine. *JAMA* 2025;333:241-247. DOI: [10.1001/jama.2024.21451](https://doi.org/10.1001/jama.2024.21451).
26. Edelson DP, Churpek MM, Carey KA, et al. Early warning scores with and without artificial intelligence [published correction appears in *JAMA Netw Open* 2024;4;7(11):e2448969]. *JAMA Netw Open*; 2024;7(10):e2438986. DOI: [10.1001/jamanetworkopen.2024.38986](https://doi.org/10.1001/jamanetworkopen.2024.38986).
27. Kore A, Abbasi Babil E, Subasri V, et al. Empirical data drift detection experiments on real-world medical imaging data. *Nat Commun* 2024;15:1887. DOI: [10.1038/s41467-024-46142-w](https://doi.org/10.1038/s41467-024-46142-w).
28. Blumenthal D, Patel B. How to regulate generative AI in health care. *Harvard Business Review*. September 4, 2024 (<https://hbr.org/2024/09/how-to-regulate-generative-ai-in-healthcare>).