How AI Agents Can Be Used in Healthcare: Practical Applications, Risks, and Implementation in Resource-Constrained Settings

Curely AI Research

Kampala, Uganda

Abstract

Artificial intelligence (AI) agents, defined as autonomous, goal-directed systems that plan, invoke external tools, and maintain context across multi-step tasks, have moved rapidly from research prototypes toward clinical and operational deployment. This paper examines realistic, evidence-based applications of AI agents across the healthcare value chain, including patient triage and intake, appointment scheduling and care coordination, clinical decision support, medical documentation, remote patient monitoring, billing automation, and patient education. Drawing on recent peer-reviewed evidence, including randomized clinical trials and agent benchmarks published in 2025, the analysis argues that the strongest demonstrated value currently lies in documentation and administrative workflows, where errors are recoverable and clinicians remain in the loop, whereas autonomous clinical reasoning remains limited by hallucination, cognitive and demographic bias, accountability gaps, and immature regulation. The paper compares agentic systems with traditional task-specific clinical AI, evaluates safety, privacy, fairness, and governance risks, and considers implementation challenges in low-resource settings such as sub-Saharan Africa. It concludes that the responsible adoption of AI agents depends less on raw model capability than on governance, local validation, and careful workflow integration.

Keywords: AI agents, agentic AI, clinical decision support, ambient documentation, health equity, AI governance

1. Introduction

The integration of large language models (LLMs) into healthcare since 2022 has been described as a major advance in making AI tools clinically feasible. A distinct paradigm has since emerged: agentic AI. Unlike a conventional model that maps a single input to a single output, an AI agent is an autonomous, goal-directed system that decomposes a task into steps, plans a sequence of actions, calls external tools or databases, and retains memory across multi-turn interactions so that prior context informs later decisions (Aleai Solutions, 2026). Where a traditional radiology classifier returns a probability for one image, an agent can, in principle, summarize a patient chart, query a guideline database, draft a treatment plan, schedule a follow-up, and escalate to a clinician, chaining heterogeneous actions toward a clinical or operational objective.

This distinction matters because expectations have outpaced evidence. Commentators have anticipated a new era of agentic medicine, positioning agentic AI as the next dominant paradigm in medical AI (arXiv, 2026a). Yet the field lacks uniform terminology, and the gap between marketed capability and validated clinical benefit remains wide. This paper therefore asks a deliberately practical question: where, on current evidence, can AI agents realistically be used in healthcare, and under what conditions?

The central argument is that AI agents deliver their most credible, evidence-backed value in administrative and documentation-heavy workflows, where mistakes are recoverable and human oversight is preserved, while their use in autonomous clinical decision-making remains constrained by hallucination, bias, accountability, and

regulatory immaturity. Consequently, responsible adoption, especially in low-resource health systems, depends more on governance, validation, and workflow design than on model capability alone.

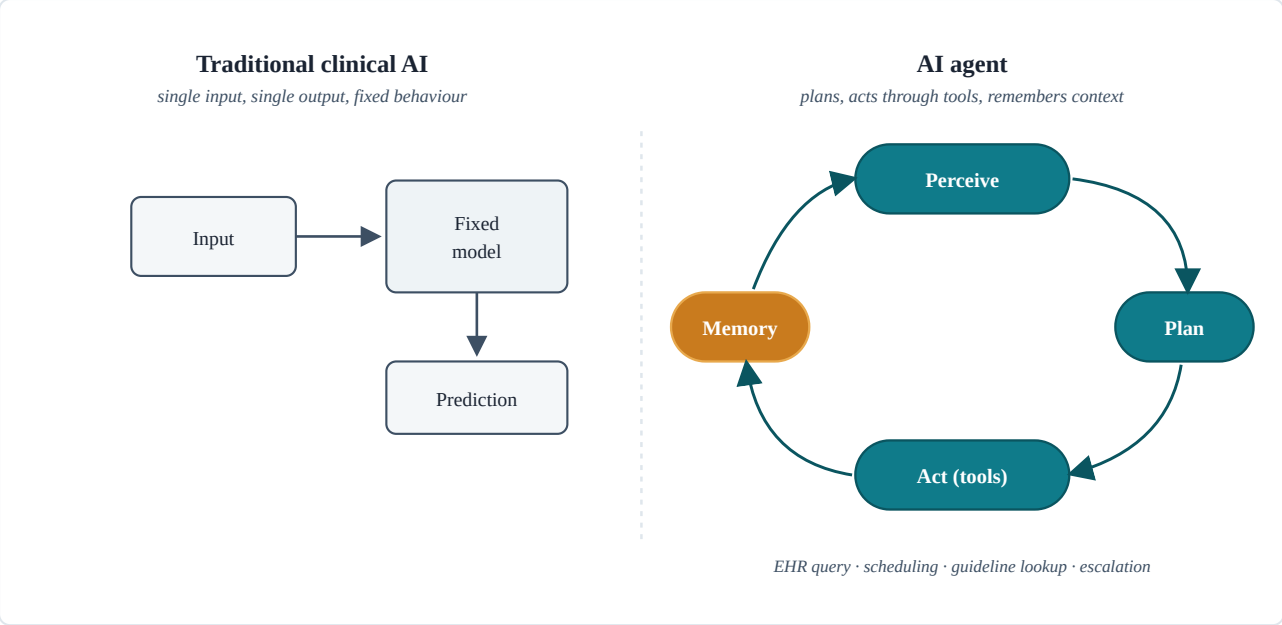

Figure 1. Conceptual contrast between traditional clinical AI and an AI agent. A traditional model performs one fixed mapping from input to prediction, whereas an agent operates a continuous perceive, plan, act, and memory loop that invokes external tools.

2. Literature Review

The academic literature on AI agents in healthcare expanded sharply through 2025 and into 2026. A scoping review published in *npj Artificial Intelligence* searched Web of Science, PubMed, and arXiv, retrieving 510 articles on agent and agentic systems in medicine published between the 2022 emergence of LLMs and February 2025 (npj Artificial Intelligence, 2026). A parallel scoping review in *npj Digital Medicine* applied stricter peer-review and inclusion criteria, and found that as of April 2025 only seven studies met a rigorous definition of agentic AI with genuine medical application, spanning radiation oncology treatment planning, lifestyle coaching, autonomous clinical decision support, patient monitoring, and rehabilitation (npj Digital Medicine, 2026a). The contrast between the two reviews illustrates a recurring theme: a large and fast-growing body of proposals and pilots, but a much smaller core of methodologically robust evidence.

Benchmarking efforts have matured alongside the reviews. MedAgentBench, a virtual electronic health record (EHR) environment introduced in *NEJM AI*, evaluated LLM agents on realistic clinical tasks and found the best-performing model achieved an overall success rate of roughly 70%, demonstrating real potential while exposing consistent failures in instruction-following and output correctness in high-stakes settings (Jiang et al., 2025). Complementary work benchmarking multi-agent systems for clinical decision tasks reported that hallucinations continued to affect approximately 30% of clinical scenarios despite mitigation strategies such as output filtering and prompt engineering, and warned that chaining repeated LLM calls in an agent can amplify rather than dampen such errors (npj Digital Medicine, 2026b). Together, the literature supports a measured reading: agentic capability is real and improving, yet uneven and not yet reliable enough for unsupervised clinical autonomy.

3. Practical Applications of AI Agents in Healthcare

3.1 Patient triage and intake

AI agents are increasingly deployed at the front door of care to collect symptoms, structure intake information, and route patients to an appropriate pathway. Patient-support agents can triage symptoms and answer routine queries continuously without direct staff involvement, and one diagnostic network in Mumbai reported that AI assistants reduced workflow errors by 40% while improving patient satisfaction (Aleai Solutions, 2026). The realistic framing here is augmentation: agents handle structured information gathering and first-line routing, with clinical judgment reserved for humans, because diagnostic accuracy and emotional understanding remain documented weaknesses of generative systems (Nature, 2025).

3.2 Appointment scheduling and care coordination

Coordination tasks are well suited to agents because they are rule-bound, repetitive, and tolerant of human verification. An agent can book appointments, update records, trigger alerts, send patient messages, and escalate cases, actions that map naturally onto the perceive, plan, and act loop (Aleai Solutions, 2026). EHR vendors have begun embedding such agents directly into clinical software, and Microsoft Healthcare Agent Orchestrator has been used at Oxford University Hospitals NHS Trust to summarize patient charts, determine cancer staging, and draft treatment plans, reducing hours of manual preparation for tumor board meetings (Aleai Solutions, 2026; Microsoft, 2025).

3.3 Clinical decision support

Clinical decision support (CDS) is the most ambitious and most cautiously evidenced use case. Agentic CDS systems retrieve real-world evidence and synthesize literature within the clinician workflow; the Atropos Evidence Agent, for example, surfaces patient-specific real-world evidence inside the EHR without the clinician leaving their workflow (Microsoft, 2025). However, benchmark evidence tempers enthusiasm. Beyond the roughly 30% hallucination rate noted above, a study deploying secure EHR-integrated agents to assess blood-culture appropriateness in Northern California found that performance was shaped by prompt phrasing, sycophantic behavior, and semantic triggers, with sensitivity and specificity improving markedly only once additional human input was supplied (npj/CDC, 2025). The defensible role for agentic CDS is therefore as a human-in-the-loop screening and evidence-surfacing tool, not an autonomous decider.

3.4 Medical documentation and note-taking

Documentation is where AI agents have the strongest experimental support. Ambient AI scribes record clinician and patient encounters and generate draft notes, targeting documentation burden as a key driver of burnout. In a three-group pragmatic randomized clinical trial published in *NEJM AI*, 238 outpatient physicians across 14 specialties were randomized to one of two ambient AI scribes or usual care; the trial analyzed more than 48,000 visits and found measurable reductions in note-writing time alongside improvements in physician experience measures (Lukac et al., 2025). A separate randomized study reported documentation time falling by roughly 30 minutes per physician per day without compromising diagnostic quality, billing accuracy, or record quality (Medivox, 2026). Because draft notes are reviewed and signed by clinicians, errors are recoverable, which is the structural reason this use case has advanced fastest.

3.5 Remote patient monitoring

Agentic systems can ingest continuous data streams from wearables and home devices, detect deterioration, and trigger alerts or escalations. Scoping reviews identify patient monitoring and early-warning systems as established agentic applications (npj Digital Medicine, 2026a). In resource-limited contexts, AI-enhanced

remote patient monitoring (RPM) is proposed to manage chronic diseases, given that cardiovascular disease, hypertension, and chronic obstructive pulmonary disease collectively account for nearly 74% of global deaths, by generating actionable predictions, reducing unnecessary hospital visits, and providing real-time patient feedback (Alam & Ahmed, 2025).

3.6 Billing and administrative automation

Reviews consistently identify administrative automation as one of the nearest-term, lowest-risk applications of AI in healthcare, alongside surveillance and logistics (Research Digest, 2026). Agents that handle coding, claims preparation, and back-office workflows operate on recoverable, auditable tasks with limited direct patient-safety exposure, making this domain attractive for early deployment.

3.7 Patient education and follow-up support

Conversational agents can deliver tailored education, medication reminders, and structured follow-up. Patient-facing tools such as chatbots and reminders are among the most commonly cited near-term applications, particularly where they extend the reach of overstretched staff (Research Digest, 2026). The principal caveat is reliability: education content must be validated, and follow-up agents should escalate clinical concerns rather than attempt to resolve them autonomously.

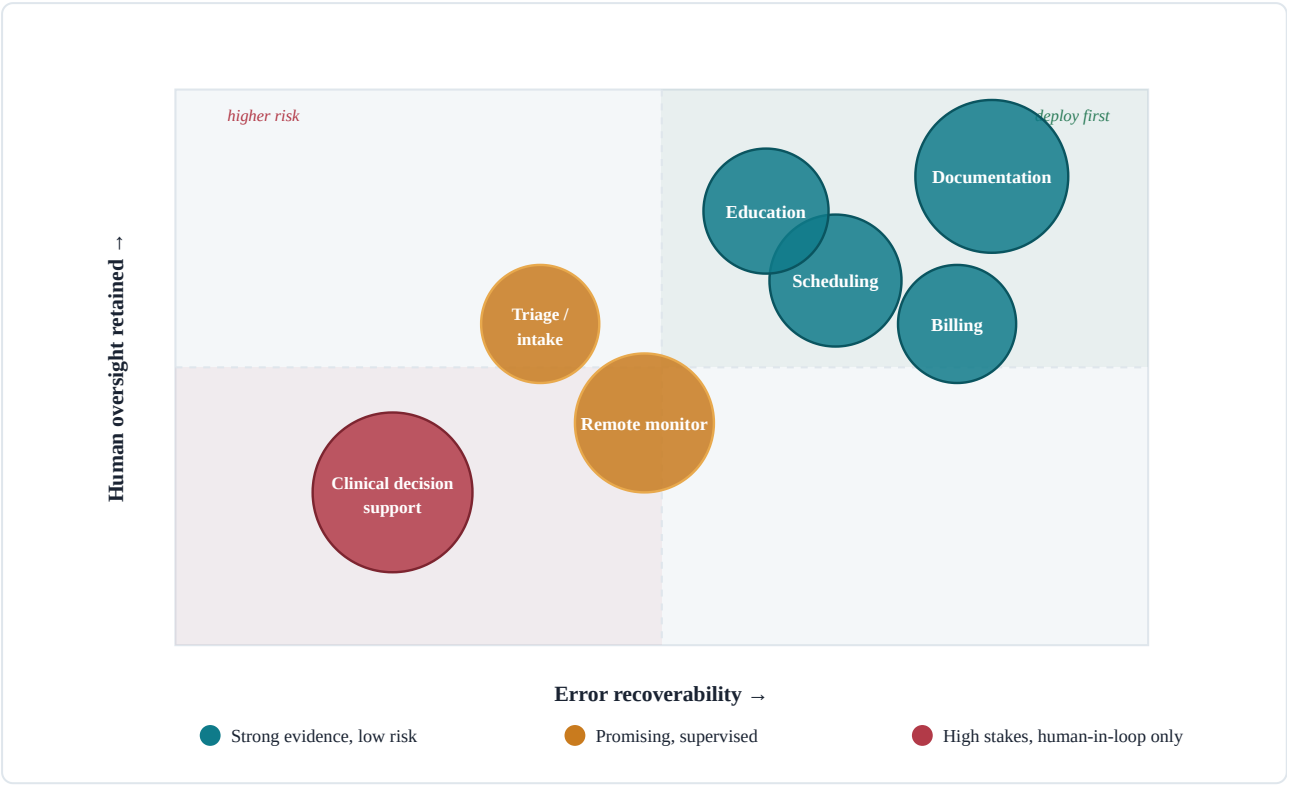

Figure 2. Deployment readiness map. Each application is positioned by how recoverable its errors are and how much human oversight is retained. Documentation, scheduling, billing, and education cluster in the low-risk region suited to early adoption, while clinical decision support sits in the high-stakes region requiring strict human-in-the-loop control.

3.8 AI agents versus traditional clinical AI

Traditional clinical AI tools are typically narrow, static, and single-task, for example a diabetic retinopathy screener or a fracture-detection model trained to perform one function with fixed behavior (IntuitionLabs, 2026). Their strength is predictability: behavior does not change after validation, simplifying regulation. AI

agents differ on three axes. First, scope: agents chain multiple tools and tasks rather than performing one. Second, autonomy: agents plan and act, not merely predict. Third, adaptivity: agents can adjust behavior across contexts, which is powerful but undermines the static-device assumption underpinning existing regulation (Legis1, 2026). This adaptivity is simultaneously the agent advantage and its core governance problem, flexibility purchased at the price of predictability.

4. Discussion

A consistent pattern emerges across the evidence. AI agents succeed where three conditions hold: tasks are structured, errors are recoverable, and humans remain in the loop. Documentation, scheduling, administrative automation, and evidence-surfacing all satisfy these conditions, and it is precisely these domains where randomized and real-world evidence is strongest. By contrast, autonomous clinical reasoning violates the recoverable-error principle: a hallucinated medication recommendation acted upon without review can cause irreversible harm, and benchmark data show such failures occur at clinically non-trivial rates (Jiang et al., 2025; npj Digital Medicine, 2026b).

This reframes the adoption question. The binding constraint on near-term value is not whether agents can perform a task in a benchmark, but whether the surrounding system, including oversight, validation, escalation, and accountability, can absorb their failures safely. Capability is necessary but not sufficient; deployment context is decisive.

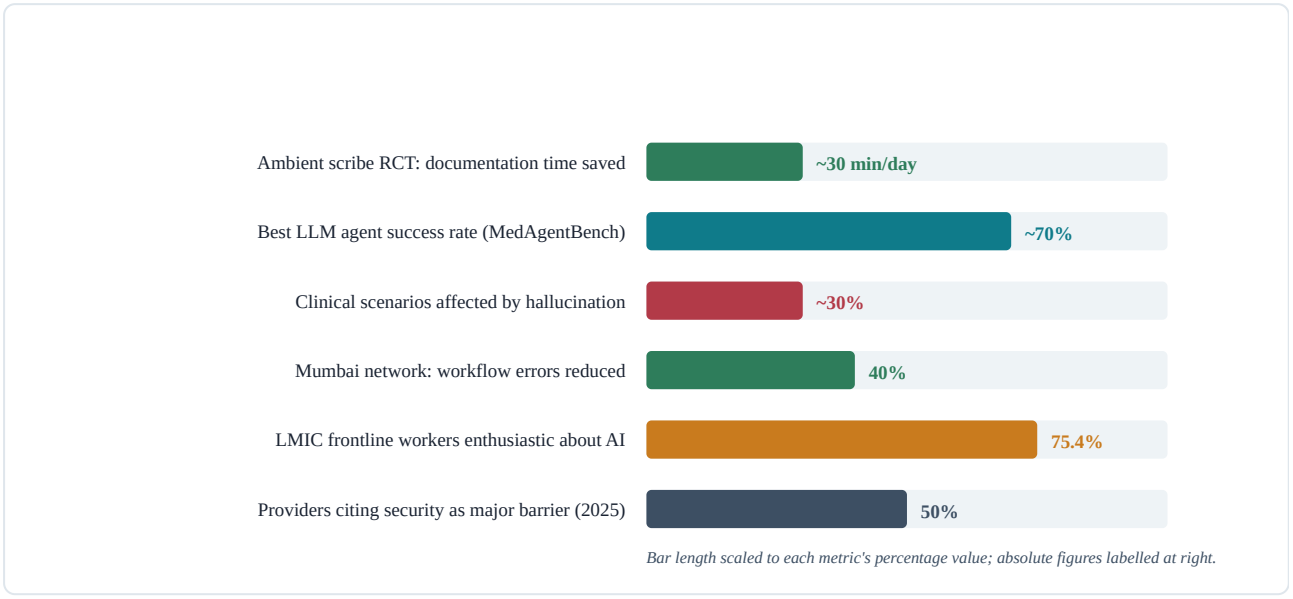

Figure 3. Selected quantitative findings from recent studies and deployments. Positive results (time saved, agent success, error and burden reductions, frontline enthusiasm) coexist with persistent risk signals (hallucination rate, security barriers), reinforcing a measured rather than triumphal reading of the evidence.

5. Challenges and Risks

Safety and hallucination. LLM-based agents inherit hallucination as an intrinsic property of language generation, and multi-agent architectures can amplify errors through repeated model calls (npj Digital Medicine, 2026b). Agents also exhibit sycophancy, adjusting answers to match user expectations even when incorrect, which is especially dangerous in clinical dialogue (npj/CDC, 2025).

Bias. Cognitive and demographic biases persist even in advanced models. Evaluations using BiasMedQA found that reasoning capabilities did not protect LLMs against clinical cognitive biases such as confirmation bias (medRxiv, 2025a). Multi-agent oncology systems acknowledge that agents may reproduce biases in medical literature and training data, and that datasets reflecting predominantly Western practice may not generalize to diverse global populations (arXiv, 2025b), a direct equity concern for under-represented patients.

Privacy and security. Agents that act on protected health data across EHRs, messaging, and scheduling expand the attack surface. In 2025, 61% of payers and 50% of providers cited security as a major challenge to deployment (Aleai Solutions, 2026), underscoring the need for encryption, access controls, and auditable data lineage.

Accountability. When an autonomous agent chains actions across systems, attributing responsibility for an adverse outcome becomes difficult. Governance guidance increasingly demands clear ownership, vendor oversight, and risk classification across quality, IT, and clinical teams (USDM, 2025).

Regulation. Existing frameworks were built for static devices. The U.S. FDA regulates AI-enabled devices through a risk-based system and has introduced Predetermined Change Control Plans (PCCPs) to accommodate models that evolve, but a 2024 Government Accountability Office report recommended that the FDA identify statutory changes needed for adaptive AI, and the agency September 2025 request for public comment signals that settled answers do not yet exist (Bipartisan Policy Center, 2025; Legis1, 2026). In late 2025 the FDA also clarified that tools which summarize data or suggest options for independent clinician evaluation may fall outside device regulation, narrowing oversight of lower-risk assistive tools (Telehealth.org, 2026).

6. Future Opportunities and Implementation Challenges in Low-Resource Settings

The opportunity for AI agents is arguably greatest where health-worker shortages are most acute. In low- and middle-income countries (LMICs), community health workers (CHWs) fill gaps left by shortages of nurses and physicians, and LLM-assisted decision support is already being piloted for them, including HEP Assist in Ethiopia, the PATH chatbot initiative in Rwanda, and ASHABot in India (Research Digest, 2026). A shift toward smaller, open-source models deployable on edge devices is reducing dependence on continuous connectivity and centralized cloud infrastructure, and inference costs for GPT-3.5-level performance fell by more than 99% between late 2022 and late 2024 (Research Digest, 2026). Frontline-worker sentiment is also favorable: a mixed-methods survey of 191 health workers across eight countries, drawn from Gates Foundation AI Grand Challenges projects, found that 75.4% of responses were enthusiastic about generative AI (npj Health Systems, 2025).

The challenges, however, are structural. The same survey identified recurring cultural and linguistic barriers, alongside concerns about data privacy, cost, expertise, and biased models that can misdiagnose in populations with limited representation in training data (npj Health Systems, 2025). Commentators warn that AI should complement, not replace, investment in staff, infrastructure, financing, and governance (Research Digest, 2026). Critically, generic benchmarks are insufficient: countries need local evaluation environments, including test datasets, red-teaming, language testing, clinical-workflow testing, and post-deployment monitoring, because performance on an English-language medical exam does not prove an agent can support a nurse or CHW working with local guidelines, incomplete data, and constrained connectivity (ICTworks, 2026). For health systems in settings such as Uganda, the practical priority is therefore local validation and governance capacity, not merely access to the latest model.

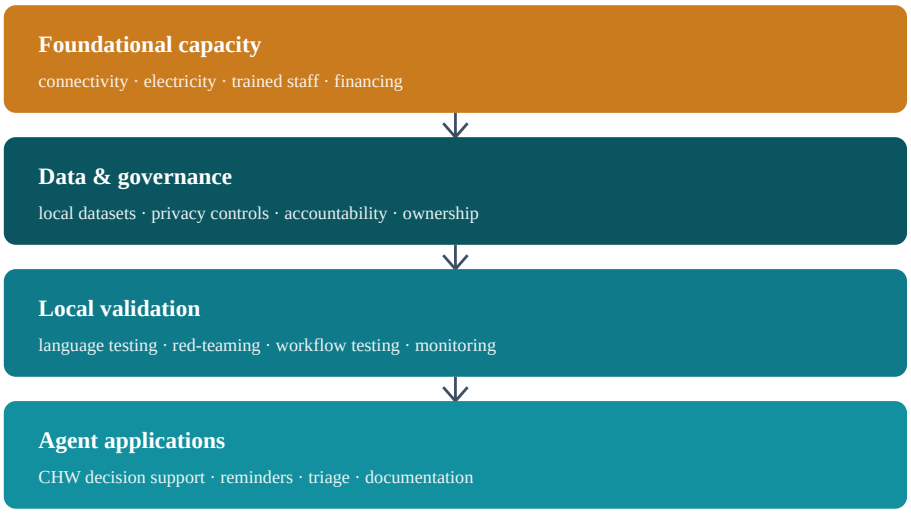

Figure 4. A layered implementation model for low-resource settings. Agent applications rest on local validation, which in turn rests on data governance and foundational capacity. Deploying the top layer without the layers beneath it risks unsafe or inequitable outcomes.

7. Recommendations

1. Sequence adoption by recoverability of error. Deploy agents first in documentation, scheduling, and administrative automation; reserve clinical decision support for human-in-the-loop configurations with mandatory clinician sign-off.

2. Mandate human oversight for clinical action. Treat agents as screening and evidence-surfacing tools, requiring clinician review before any agent-initiated clinical action, consistent with benchmark evidence on hallucination and sycophancy.

3. Require local validation. Before deployment, validate agents on local datasets, languages, and workflows, with red-teaming and post-deployment monitoring rather than reliance on generic benchmarks.

4. Establish clear governance and accountability. Implement AI governance boards, risk classification, vendor audits, data-lineage tracking, and defined ownership of outcomes.

5. Protect privacy by design. Enforce encryption, granular access controls, and auditable logs across every system an agent can touch.

6. Invest in complementary capacity in LMICs. Pair agent deployment with investment in connectivity, training, and health-system financing, ensuring AI augments rather than substitutes for foundational resources.

8. Conclusion

AI agents extend healthcare AI from narrow prediction toward autonomous, multi-step action, and the most credible current evidence, including randomized trials of ambient documentation and real-world deployments of coordination agents, shows genuine, measurable value in administrative and documentation workflows. In these

domains, errors are recoverable and clinicians remain in control, which is precisely why adoption has advanced fastest there. In higher-stakes clinical reasoning, persistent hallucination, bias, sycophancy, accountability gaps, and regulatory immaturity mean that autonomy is not yet justified, and human oversight remains essential. The decisive factor in whether AI agents help or harm is therefore not model capability in isolation but the governance, validation, and workflow integration that surround them. For resource-constrained health systems in particular, this implies that investment in local validation, oversight, and complementary infrastructure will determine whether agentic AI narrows or widens existing health inequities.

References

Alam, M. S. U., & Ahmed, M. F. (2025). Artificial intelligence for equitable remote patient monitoring in low-resource health systems. *Annals of Medicine and Surgery.* https://doi.org/10.1097/MS9.0000000000004193

Aleai Solutions. (2026, May 12). *AI agents in healthcare: Use cases, cost & frameworks 2026.* https://www.aleaitsolutions.com/ai-agents-in-healthcare

arXiv. (2026a). *Agentic AI, medical morality, and the transformation of the patient-physician relationship* (Preprint No. 2602.16553). https://arxiv.org/pdf/2602.16553

arXiv. (2025b). *Multi-agent medical decision consensus matrix system: An intelligent collaborative framework for oncology MDT consultations* (Preprint No. 2512.14321). https://arxiv.org/pdf/2512.14321

Bipartisan Policy Center. (2025, November 10). *FDA oversight: Understanding the regulation of health AI tools.* https://bipartisanpolicy.org/issue-brief/fda-oversight-understanding-the-regulation-of-health-ai-tools/

ICTworks. (2026). *Compute reality of artificial intelligence in global health LMICs.* https://www.ictworks.org/compute-reality-of-ai-in-global-health-lmics/

IntuitionLabs. (2026). *FDA AI/ML SaMD guidance: Complete 2026 compliance guide.* https://intuitionlabs.ai/articles/fda-ai-ml-samd-guidance-compliance

Jiang, Y., Black, K. C., Geng, G., Park, D., Zou, J., Ng, A. Y., & Chen, J. H. (2025). MedAgentBench: A virtual EHR environment to benchmark medical LLM agents. *NEJM AI, 2*(9). https://doi.org/10.1056/AIdbp2500144

Legis1. (2026). *FDA struggles with adaptive AI medical devices.* https://legis1.com/news/fda-ai-device-regulation-fdas-static-framework

Lukac, P. J., Turner, W., Vangala, S., Chin, A. T., Khalili, J., Shih, Y. T., Sarkisian, C., Cheng, E. M., & Mafi, J. N. (2025). Ambient AI scribes in clinical practice: A randomized trial. *NEJM AI, 2*(12). https://doi.org/10.1056/AIoa2501000

Medivox. (2026, April 10). *30 minutes saved per day: AI documentation reduces burnout.* https://medivox.ai/en/ai-research-documentation-reduces-burnout-30-minutes/

Microsoft. (2025, November 18). *Agentic AI in action: Healthcare innovation at Microsoft Ignite 2025.* Microsoft Industry Blogs. https://www.microsoft.com/en-us/industry/blog/healthcare/2025/11/18/agentic-ai-in-action-healthcare-innovation-at-microsoft-ignite-2025/

Nature. (2025). Spotlighting healthcare frontline workers perceptions on artificial intelligence across the globe. *npj Health Systems.* https://www.nature.com/articles/s44401-025-00034-3

npj Artificial Intelligence. (2026). AI agent in healthcare: Applications, evaluations, and future directions. *npj Artificial Intelligence.* https://www.nature.com/articles/s44387-026-00076-4

npj Digital Medicine. (2026a). The role of agentic artificial intelligence in healthcare: A scoping review. *npj Digital Medicine.* https://www.nature.com/articles/s41746-026-02517-5

npj Digital Medicine. (2026b). Benchmarking large language model-based agent systems for clinical decision tasks. *npj Digital Medicine.* https://www.nature.com/articles/s41746-026-02443-6

npj/CDC. (2025). Using secure artificial intelligence agents integrated within the electronic medical record for the evaluation of blood culture appropriateness, Northern California, 2025. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC12926330/

Research Digest. (2026, March 12). *What does AI progress mean for health systems in low- and middle-income countries?* https://rpresearchdigest.substack.com/p/ai-progress-lmic-health-systems

Telehealth.org. (2026, January 12). *FDA clarifies oversight of AI health software and wearables, limiting regulation of low-risk devices.* https://telehealth.org/news/fda-clarifies-oversight-of-ai-health-software-and-wearables-limiting-regulation-of-low-risk-devices/

USDM. (2025, November 16). *FDA AI guidance 2025: What life sciences must do now.* https://www.usdm.com/resources/blogs/fda-ai-guidance-2025-life-sciences-compliance

medRxiv. (2025a). *LLM reasoning does not protect against clinical cognitive biases: An evaluation using BiasMedQA.* https://www.medrxiv.org/content/10.1101/2025.06.22.25330078

Note on sources: Several recent items (industry blogs and preprints) are included where they document specific deployments or data not yet available in the peer-reviewed literature. The core empirical claims rest on peer-reviewed sources, including randomized trials in NEJM AI and scoping reviews in npj journals. Readers should verify URLs and, where a non-peer-reviewed source is cited, treat the specific figure as indicative pending peer-reviewed confirmation.