Real-World Challenges of Intelligent Clinical Workflows

Rashmi S[1], Updendra Kumar[2]

[1] Lincoln Global Postdoctoral Research (LGPR) , Lincoln University College, Malaysia ; [2] Lincoln Global Postdoctoral Research (LGPR) , Lincoln University College, Malaysia & Assistant Professor, Department of Computer Science & Engineering, Institute of Engineering & Technology, Lucknow ;

[1] pdf.rashmi@lincoln.edu.my, [2]upendra.ietlko@gmail.com

Abstract:

The rise of Agentic AI and LLMs has led to new possibilities for intelligent, adaptive and context-aware healthcare systems. The framework has been proposed, which involves integration of agentic reasoning, inference enhanced by the use of retrievals, interoperable FHIR representation, and clinician guided planning for the purpose of supporting dynamic, patient-centered workflows. Despite obvious advantages in terms of automated reasoning, patient monitoring and personalized decision-making, implementation of the system in the clinical environment is not free from specific challenges. This paper provides critical discussion of limitations and barriers involved in deploying the framework in clinical practice. In addition, issues regarding regulatory compliance, ethical responsibility, clinicians' trust, explainability and automation bias will be discussed. The main conclusion of this study is that the development of next generation intelligent healthcare systems represented by an intelligent framework has a good potential but still needs resolving deployment-related challenges in regard to safety, interoperability, accountability and scalability to become truly clinically usable.

Keywords: Agentic AI; LLMs; challenges; Clinical workflows; FHIR; Healthcare

---

Introduction

The rapid development of Artificial Intelligence (AI), including Large Language Models (LLMs) and Agentic AI, has revolutionized healthcare applications, including clinical decision-support and research. Although the current implementations of AI in medicine have shown promise when applied to prediction analysis, diagnosis support, and workflow process automation, most of them still function within narrow scopes without providing contextual awareness, adaptivity, or real-time connectivity with EHRs, IoMT devices, and clinicians. Some new developments in agentic systems seek to solve these problems by offering autonomous reasoning, dynamic planning, and constant interaction with health-related records, Internet-connected equipment, and healthcare specialists. One possible example of such a solution may be the proposed architecture using LLMs and Agentic AI[1]

On the one hand, it seems the framework has some promising features. On the other hand, however, there exist many limitations and potential barriers to its actual implementation in practice. First of all, a healthcare setting is a very sensitive environment in which any mistake, hallucination, bias, or malfunction in a decision-making procedure can have serious consequences for patients' health and well-being. Secondly, healthcare institutions are characterized by high heterogeneity of infrastructure, data standardization practices, and compliance needs. Finally, autonomous agentic AI systems require solving a series of problems associated with explainability, accountability, transparency, and safety. Thus, discussing the implementation challenges and barriers is important before moving on to large-scale deployment.

Related work

Recent studies have explored the increasing prominence of LLMs within healthcare processes. For instance, Artsi et al. identified that GPT models led to enhanced productivity, patient engagement, and record-keeping in healthcare practices. However, issues such as generalization, interoperability, and regulatory compliance still prevailed [2]. Similarly, Longhurst et al. stressed that generative AI models in the healthcare domain necessitate rigorous testing, governance, and human supervision prior to clinical deployment [3]. Research based on Retrieval-Augmented Generation (RAG) have revealed an improvement in the factual accuracy of medical question-answer applications using verified medical information sources [4].

The current literature on research on Agentic AI for healthcare emphasizes its autonomy and adaptability. According to Wang et al., the use of AI-based clinical extraction could result in substantial reductions in transcription time without sacrificing the quality of structured data [5]. Nevertheless, performance drops in cases of unstructured material have been noted. In turn, Habicht et al. analyzed GPT-4-powered therapy support software and noticed an increase in patient engagement. However, they recorded decreasing user retention rates [6]. Hence, these studies highlight the potential of AI technologies to improve clinical practices but also point out various obstacles connected with dependability, trustworthiness, compatibility, and governance of AI systems in real-life situations.

Moreover, FHIR as a standard for interoperability between information systems becomes crucial for integration. Nonetheless, implementation studies show that it is difficult to preserve semantic consistency in different healthcare facilities with heterogeneous EHR structures and specific documentation methods [7]. Literature also shows that monitoring and clinician-in-the-loop designs along with proper reasoning are needed to avoid hallucinations and automation bias in healthcare AI systems [8].

Challenges and Results

Conventional CDSS solutions were inflexible and context unaware [9]. However, transformer models have been employed to develop conversational and semantically aware AI in healthcare, still functioning primarily as passive assistants.

The proposed design[1], unlike typical AI systems, involves agentic AI, retrieval-augmented inference, interoperable interfaces using FHIR, and context-aware planning. The proposed clinically guided planner would constantly learn from patient-related data and provide recommendations on how to manage patients' chronic diseases (e.g., diabetes). As opposed to independent LLMs, the proposed system will try to limit the hallucinations associated with language models via retrieval grounding and reasoning pipelines.

Nevertheless, even systems based on similar designs like Med-PaLM, GatorTron, and EHR assistants powered by LLMs prove the challenges associated with scaling up such technologies in terms of resource-intensive training, privacy risks, and model management[10]. Current AI solutions in healthcare demonstrate that clinicians are unwilling to rely on AI-generated insights due to their lack of transparency and explainability. Thus, although this system represents an innovative architecture, its implementation inherits numerous existing problems.

There are various challenges associated with implementing the proposed frmaework in clinical settings. The first significant limitation that should be discussed is data heterogeneity. Each hospital uses its own EHR systems, which have different formats, terminology and approaches to documentation. Although the proposed framework is based on the application of FHIR-based pipelines to achieve interoperability, it is quite hard to properly map legacy healthcare data into the standardized format. Incomplete or low quality data could negatively impact reasoning performed by the proposed system[1].

Scalability and computational requirements pose a barrier to the widespread adoption . It should be noted that agentic architectures rely on constant context interpretation, planning, memory management and reasoning. These processes impose considerable computing overheads, which might be increased when deploying the system in real time. What is more, healthcare institutions of low digital maturity might experience certain difficulties while implementing complex agentic architectures.

Clinical AI applications have been recently recognized as Software-as-a-Medical-Device (SaMD). This means that clinical systems have to undergo validation, become explainable and comply with healthcare-related regulations like HIPPA and GDPR [11]. Moreover, being constantly adapted to changing contexts, the system becomes increasingly hard to manage and validate. Accountability issues also arise when AI-based systems influence patients' health. Thus, it is unclear whether clinician or AI algorithm is responsible for the mistake.

Clinician adoption of new technologies poses an additional challenge to be considered. If the system relies heavily on black box algorithms and makes it hard for physicians to understand what is happening inside, clinicians would be unwilling to use it in practice. Furthermore, over-reliance on automation might cause automation bias, which is defined as over-reliance on AI suggestions. In order to prevent these issues from happening, the system needs to incorporate explainable reasoning capabilities, overrides, and confidence scoring.

Notwithstanding all its limitations, the framework has much promise in terms of personalized medicine, saving money and improving workflow. Its ability to combine retrieval-based reasoning and physician guidance is particularly impressive. Nevertheless, future implementation has to take into consideration various regulatory, ethical, computational and social aspects of agentic architectures.

Conclusions

The framework provides an important milestone towards the development of next-generation intelligent healthcare systems that make use of both Agentic AI and LLMs. The framework is poised to address many challenges that characterize conventional methods of incorporating artificial intelligence into the healthcare system due to its contextualized reasoning, retrieval-enhanced inference, and interoperability provided by the use of the FHIR standard. Additionally, due to the ability of the framework to adapt dynamically depending on specific patients and facilitate decision making, the framework offers an attractive architecture for healthcare.

However, there are a lot of obstacles that may arise when it comes to the application and use of the above-mentioned framework in actual practice settings. Such obstacles may include issues relating to the precision, validity, transparency, explainability, and scalability of the technology, as well as gaining the trust of clinicians in the use of AI for decision-making in healthcare and regulating the use of this technology for ethical governance. Future research is needed to tackle the issue of validation and monitoring, as well as the problem of clinician trust in AI-based healthcare solutions.

References

1. R. S and U. Kumar, “An Agentic AI–Driven End-to-End Clinical Workflow for Diabetes Management,” *Sustainable Global Societies Initiative*, vol. 1, no. 3, 2026. https://vectmag.com/sgsi/paper/view/211
2. Y. Artsi, V. Sorin, B. S. Glicksberg, P. Korfiatis, G. N. Nadkarni and E. Klang, “Large Language Models in Real-World Clinical Workflows: A Systematic Review of Applications and Implementation,” *medRxiv*, 2025. https://doi.org/10.1101/2025.06.10.25329323

3. C. A. Longhurst, N. H. Shah, V. Patel, A. Y. Ng and A. Rajkomar, “Generative AI in Health Care: Decision Making, Implementation, and Policy Considerations,” *NEJM AI*, vol. 1, no. 5, 2023. https://doi.org/10.1056/AI-S2300233
4. A. Thomo, “PubMed Retrieval with RAG Techniques,” *Studies in Health Technology and Informatics*, vol. 316, pp. 652–653, 2024. https://doi.org/10.3233/SHTI240498
5. B. Wang, J. Lai, H. Cao, F. Jin, Q. Li, M. Tang, C. Yao and P. Zhang, “Enhancing the Interoperability and Transparency of Real-World Data Extraction in Clinical Research,” *European Heart Journal Digital Health*, vol. 5, no. 6, pp. 712–724, 2024. https://doi.org/10.1093/ehjdh/ztae066
6. J. Habicht, L. M. Dina, J. McFadyen, M. Stylianou, R. Harper, T. U. Hauser and M. Rollwage, “Generative AI-Enabled Therapy Support Tool for Improved Clinical Outcomes and Patient Engagement,” *Journal of Medical Internet Research*, vol. 27, 2025. https://doi.org/10.2196/60435
7. H. Mandel et al., “FHIR-Based Interoperability for AI-Driven Clinical Systems,” *Healthcare Informatics Research*, vol. 30, no. 1, pp. 15–27, 2024.
8. R. Ratwani, D. Classen and C. Longhurst, “The Compelling Need for Shared Responsibility of AI Oversight,” *JAMA*, vol. 332, no. 10, pp. 787–788, 2024. https://doi.org/10.1001/jama.2024.12630
9. E. Shortliffe and J. Cimino, *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*, Springer, 2021. https://doi.org/10.1007/978-3-030-58721-5
10. K. Singhal et al., “Large Language Models Encode Clinical Knowledge,” *Nature*, vol. 620, pp. 172–180, 2023. https://doi.org/10.1038/s41586-023-06291-2
11. U.S. Food and Drug Administration, “Software as a Medical Device (SaMD): Clinical Evaluation Guidance,” *FDA Guidelines*, 2024. https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd