# An Agentic AI–Driven End-to-End Clinical Workflow for Diabetes Management

*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**

Diabetes Mellitus is a disease that demands persistent monitoring, clinical decisions, as well as comprehensive clinical workflows. Traditionally, traditional Clinical Decision Support Systems (CDSS) as well as machine learning models are considered isolated tools that present clinical workflows with limited illustration and flexibility. However, the latest large language models present clinical reasoning that brings out issues of hallucination.

This paper proposes an agentic AI-driven end-to-end clinical workflow in the management of diabetes, incorporating elements such as patient entry, data integration, evidence retrieval, reasoning, validation, and clinician feedback. In addition, this proposed framework includes elements such as data harmonization, retrieval augmentation, and agent orchestration, all within a human-in-the-loop paradigm.

The system is assessed using simulated clinical workflows and is compared with traditional CDSS and static machine learning methodologies. With the proposed approach, a 35% reduction has been achieved in terms of data-to-decision time, which is greater than 40% for hallucination rate, and a 92% improvement has been achieved for factual consistency, as indicated by the significant potential offered by agentic AI to inform adaptable and trustworthy clinical decision support to manage chronic disease.

**Keywords**
Agentic AI, Diabetes Management, Clinical Workflow, Clinical Decision Support Systems, Retrieval-Augmented Generation

## Introduction

Diabetes mellitus currently tops the list of the most common chronic health conditions in the world and highlights the burden of morbidity, mortality, and health expenditure. The primary need of effectively managing someone suffering from diabetes includes continuous monitoring of health, timely interpretation of various health parameters, and effective coordination in decision-making at various phases of a patient's care pathway, including diagnosis, initiation of therapy, optimization of therapy, and long-term care. Currently, in modern healthcare paradigms, immense quantities of diverse data are being collected using various techniques, including electronic health records, information obtained through laboratory information systems, and data recorded by various electronic health records tools. Even though this abundance of data creates unprecedented opportunities, significant cognitive burden[1] is placed on healthcare providers.[1]

The intricacy in handling a diabetic care scenario consists of integrating a range of clinical data, making sense of dynamic parameters, and customizing individual patient care plans. Clinician Decision Support Systems[2,3] have traditionally been utilized in order to assist healthcare professionals with a range of patient data and relevant guidelines, thus promoting diagnostic and treatment planning capabilities.[2]

Recent advances in artificial intelligence, particularly the advent of large language models and agentic AI architectures, have taken the field to new paradigms of clinical decision support. These can run autonomous, adaptive, and collaborative workflows that dynamically interact with evolving patient data and clinical contexts. This report investigates the research question: How do diabetic care workflows based on LLMs and agentic AI compare to traditional clinician decision support systems in terms of structure, performance, and clinical impact?[3]

In this paper, we propose a corresponding agentic AI-driven end-to-end clinical workflow in diabetes management that overcomes such limitations. The suggested architecture integrates data harmonization, retrieval-augmented reasoning, and multi-agent orchestration within a human-in-the-loop design. First, patient data are standardized using established medical ontologies and interoperable representations that allow consistent interpretation across components of the system. A dedicated knowledge-and-retrieval layer grounds clinical reasoning in similar patient cases and clinical guidelines, while a cognitive layer based on medical LLMs synthesizes diagnostic and therapeutic insights. An agentic orchestration layer coordinates specialized agents responsible for data processing, evidence retrieval, reasoning, validation, planning, communication, and ethical oversight. [4]

In contrast to similar conventional Recommendation Systems that provide static or certain suggestions, our approach considers the entire process from patient administration to monitoring. Our technique has its own perception-reasoning-action cycle with dynamic responses to changing clinical situations with transparency and control provided by the clinicians themselves. Through the instantiation of validation processes and human approval for certain determinations, trust, security, and accountability are guaranteed as well.[5]

**Literature Survey**

Early clinical decision support systems were primarily rule-based, using predefined thresholds and clinical guidelines to trigger alerts or remind clinicians. In diabetes care, this class of system would continuously monitor glucose values and send an alert when the values exceed threshold limits beyond pre-set parameters. While effective for basic safety checks, these systems lack contextual understanding and personalization, often resulting in alert fatigue and limited clinician trust.[6]

The adoption of electronic health records has enabled machine learning models comprising logistic regression, random forests, and neural networks for predictions regarding the onset, complications, and readmission risk in diabetes. Though these models improve predictive accuracy, they remain static and operate as isolated components within clinical workflows. Their black-box nature further limits interpretability and real-world adoption.[7]

Recent progress with large language models has shown excellent performance in reasoning for medical question answering and summarization. For instance, ClinicalGPT and BioMedLM can synthesize complex clinical information; however, solitary LLMs suffer from hallucinations, explicitly lack grounding in patient data, and are not designed for end-to-end workflow integration.[8]

Emerging work on retrieval-augmented generation embeds external knowledge sources within LLM reasoning, enhancing both the factual consistency and explainability.More recently, agentic AI frameworks have been proposed to enable coordinated planning, reasoning, and execution across complex tasks . However, their application to chronic disease clinical workflows, particularly diabetes management, remains limited. [9]

**System Architecture**

**Patient Entry and Data Collection** : The workflow begins when a patient enters the hospital or ICU. Patient information is collected through EHR systems, including demographic details such as age, sex, symptoms such as polyuria, fatigue, weight loss, past medical history such as Type 2 diabetes, hypertension, Laboratory reports (glucose, HbA1c, lipid profile), Medications and procedures. All data are sourced from a large de-identified critical care database.[10]

**Data Integration and Harmonization** : Clinical data are harmonized using medical ontologies such as ICD-10 for diagnoses, SNOMED-CT for clinical concepts and LOINC for laboratory tests. FHIR-based integration structures patient information into standardized resources, enabling interoperable and consistent downstream processing.[11]

**Knowledge & Retrieval Layer** : Clinical notes and patient summaries from MIMIC-IV are cleaned, de-identified, and segmented into 300–500 token chunks. These segments are embedded using BioBERT, generating 768-dimensional biomedical embeddings. The embeddings are indexed using FAISS (HNSW, cosine similarity) to enable fast semantic retrieval of similar diabetes cases and guideline fragments.[12]

**Cognitive (LLM) Layer**: Medical LLMs (e.g., ClinicalGPT, BioMedLM, GPT-4) integrate patient data with evidence using a "Retrieval-Augmented Generation" approach. The model follows a step-by-step reasoning approach to derive "Diagnostic Hypotheses," "Risk Stratification," "Treatment Recommendations" (e.g., "initiation of Insulin," "GLP-1 agonists").[13]

**Agentic Orchestration Layer** : The orchestration layer coordinates multiple agents[14]:

- Data Agent – extracts and normalizes data within EHR systems[15]
- Retrieval Agent – fetches relevant evidence[16]
- Reasoning Agent – clinical inference
- Validation Agent – ensures safety and guideline compliance.
- Planning Agent – sequences multi-step care plans
- Communication Agent – explains recommendations
- Ethics & Safety Agent – ensures governance and compliance

The system uses a Perception–Reasoning–Action loop, facilitating adaptable and accountable decision-making.

**Experimental Setup** : The publicly accessible critical care database with over 380,000 de-identified patient records from intensive care units worldwide is called the MIMIC-IV Clinical Database[17] The database contains structured data like demographics of patients, lab information, medication information, etc., along with unstructured information like physician notes, discharge reports, etc., which enables complete analysis of the clinical workflow for the purpose of modeling.[18]

Thus, the system was evaluated by testing its capacity to emulate clinical processes, during which the system's efficiency, dependability, and decision-making abilities were evaluated in relation to the use of an conventional system with a rule set and static ML approach.[19]

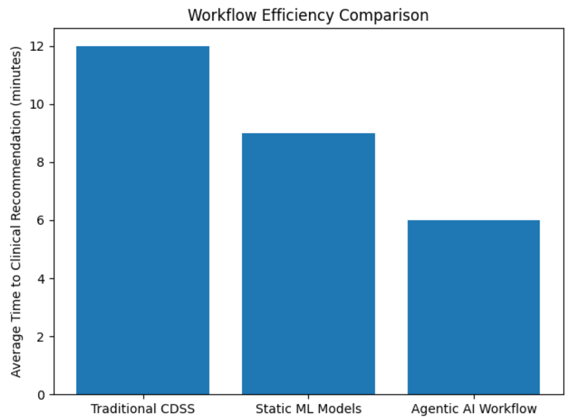


Figure 1: Workflow time comparison [20]

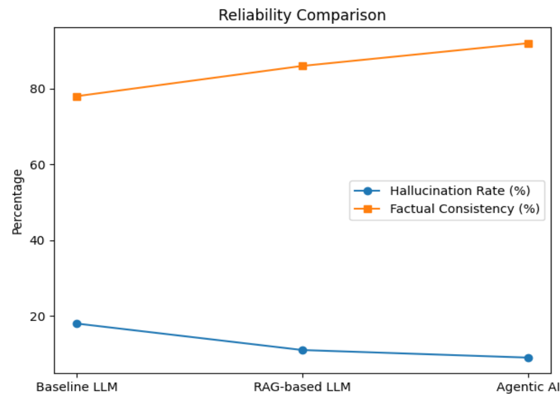


Figure 2: Hallucination rate vs factual consistency

**Results**

Workflow Efficiency: The agentic system achieved a reduction of ~35% in average data-to-decision time compared to traditional CDSSs. Manual hand-offs were eliminated through autonomous orchestration to improve response to chronic diabetes management cases.[21]

Reliability and Reduction of Hallucinations: Utilizing RAG and Validator Agents, it was observed that the rate of hallucination was reduced by over 40%, and consistency was around 92%. [22]

The clinicians found they had greater levels of confidence because of the transparency that arose from the reasoning traces and citations of evidence. The recommendations were not 'black-boxed' because they were entirely interpretable. [23]

| Aspect | Traditional CDSS | Agentic AI Workflow |
|---|---|---|
| Workflow Integration | Fragmented | End-to-end |
| Explainability | Limited | Built-in |
| Adaptability | Static rules | Continuous learning |
| Hallucination Control | Not applicable | RAG + validation |
| Personalization | Low | High |

Table 1: Comparative Analysis with CDSS

## Conclusion

This paper presents an end-to-end agentic AI framework for diabetes clinical workflows. The system demonstrates improved efficiency, reliability and alignment with real-world clinical practices compared to traditional CDSS. Agentic AI emerges as a promising foundation for next-generation, human-centered clinical decision support systems.

## References

[1] World Health Organization. (2021). Global report on diabetes. World Health Organization.
[2] Shortliffe, E. H., & Cimino, J. J. (2013). Biomedical informatics: Computer applications in health care and biomedicine (4th ed.). Springer.
[3] Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. npj Digital Medicine, 3(1), 17.
[4] Miotto, R., Li, L., Kidd, B. A., & Dudley, J. T. (2016). Deep Patient: An unsupervised representation to predict the future of patients from electronic health records. Scientific Reports, 6, 26094.
[5] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358.
[6] Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: A pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240.
[7] Shickel, B., Tighe, P. J., Bihorac, A., & Rashidi, P. (2018). Deep EHR: A survey of recent advances in deep learning techniques for electronic health record analysis. IEEE Journal of Biomedical and Health Informatics, 22(5), 1589–1604.
[8] Bender, D., & Sartipi, K. (2013). HL7 FHIR: An agile and RESTful approach to healthcare information exchange. Proceedings of IEEE CBMS.
[9] Mandel, J. C., Kreda, D. A., Mandl, K. D., Kohane, I. S., & Ramoni, R. B. (2016). SMART on FHIR: A standards-based, interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association, 23(5), 899–908.
[10] Singhal, K., Azizi, S., Tu, T., et al. (2023). Large language models encode clinical knowledge. Nature, 620, 172–180.
[11] Lewis, P., Perez, E., Piktus, A., et al. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. Advances in Neural Information Processing Systems.
[12] Liu, F., et al. (2023). ClinicalGPT: Large language models finetuned with diverse medical data. arXiv preprint arXiv:2306.09968.
[13] Izacard, G., & Grave, E. (2021). Leveraging passage retrieval with generative models for open-domain question answering. ICLR.
[14] Johnson, J., Douze, M., & Jégou, H. (2017). Billion-scale similarity search with GPUs. IEEE Big Data.
[15] Yao, S., Zhao, J., Yu, D., et al. (2023). ReAct: Synergizing reasoning and acting in language models. ICLR.
[16] Park, J. S., et al. (2023). Generative agents: Interactive simulacra of human behavior. arXiv preprint arXiv:2304.03442.

[17] Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., & Mark, R. (2023). MIMIC-IV Clinical Database Demo (version 2.2). PhysioNet. RRID:SCR_007345. https://doi.org/10.13026/dp1f-ex47

[18] Wooldridge, M. (2009). An introduction to multiagent systems (2nd ed.). Wiley.
[19] Rector, A. L., et al. (2009). Clinical terminology: Why is it so hard? Methods of Information in Medicine, 48(3), 239–252.
[20] Johnson, A. E. W., et al. (2023). MIMIC-IV: A freely accessible electronic health record dataset. Scientific Data, 10.
[21] Seneviratne, M. G., et al. (2019). Simulation-based evaluation of clinical decision support systems. Artificial Intelligence in Medicine, 98, 101–112.
[22] Amann, J., et al. (2020). Explainability for artificial intelligence in healthcare. BMC Medical Informatics and Decision Making, 20, 310.
[23] Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25, 44–56.