Concept Paper

Muli-Agent AI Systems in Healthcare: Technical and Clinical Analysis

Andrew Borkowski [*] and Alon Ben-Ari

Concept Paper

Multi-Agent AI Systems in Healthcare: Technical and Clinical Analysis

Andrew A. Borkowski [1],* and Alon Ben-Ari [2]

[1] VA Sunshine Healthcare Network, Tampa, FL, USA

[2] VA Northern California Healthcare System, Sacramento, CA, USA

* Correspondence: andrew.borkowski@va.gov

Abstract: Scaling the operation is the single greatest challenge in every enterprise. The more complex the enterprise, the greater the challenge. Scaling in healthcare has been a consistent challenge, with the need to meet clinical endpoints with less staff, increased costs of resources, regulatory oversight, and the goal of improving access to care. The article explores the emerging paradigm of multi-agent AI systems in healthcare, representing a significant leap beyond traditional Large Language Models. We comprehensively analyze the potential of these systems to revolutionize patient care, streamline administrative processes, and support complex clinical decision-making. The paper describes a hypothetical sepsis management system comprising seven specialized AI agents, each handling specific aspects of patient care, from data collection and diagnosis to treatment recommendations and resource management. We also examine applications in chronic disease management and hospital patient flow optimization. The technical implementation of these systems is discussed, including the use of advanced Large Language Models, inter-agent quality control measures, implementing guardrails, self-reflection, integration with Electronic Health Records, and the importance of explainable AI for decision transparency. In addition to describing the promising potential benefits, such as enhanced diagnostic accuracy and personalized treatment plans, we address significant challenges, including data quality assurance, workflow integration, and ethical considerations. We conclude the article by highlighting future directions, such as integrating IoT devices and developing more sophisticated natural language interfaces. Our work underscores the transformative potential of multi-agent AI systems in healthcare while emphasizing the need for rigorous validation, ethical oversight, and a patient-centered approach in their development and implementation.

Keywords: AI-agent; multi AI agents; large language models; healthcare; clinical operation; agentic systems

Introduction

The landscape of artificial intelligence is rapidly evolving, with Large Language Models (LLMs) marking a significant milestone in our ability to process and generate human-like responses to natural language prompts. This breakthrough, however, is merely the beginning of a more profound transformation in AI capabilities. We now find ourselves at the frontier of a new paradigm: AI Agents.

AI Agents represent a leap forward from traditional LLM applications. While definitions may vary slightly among tech giants, the core concept remains consistent: these are autonomous software entities designed to interact with their environment, make rational choices, and execute tasks based on predefined goals [1–3]. What sets AI Agents apart is their combination of sophisticated components and their assembly in different architectures. At their heart lies a Large Language Model for generating responses, which is augmented by a suite of tools to optimize workflow and complete tasks, memory capabilities for personalized interactions, and autonomous reasoning abilities. This powerful combination allows AI Agents to plan, create subtasks, gather information, and learn iteratively from their own experiences or other AI Agents.

The true potential of this technology becomes apparent when we consider the orchestration of multiple AI Agents working in concert. This concept, known as multi-AI Agent systems, introduces a new level of flexibility and capability in tackling complex tasks. Frameworks such as Autogen, CrewAI, and LangChain offer various ways to configure these agent networks, allowing for hierarchical, sequential, conditional, or even parallel task execution [4–6]. This adaptability opens up a world of possibilities across various industries, but perhaps nowhere is the potential impact more exciting and profound than in healthcare.

The application of AI Agents in healthcare presents an opportunity to revolutionize patient care, streamline administrative processes, and support complex clinical decision-making. In this paper, we will explore three key scenarios that illustrate the transformative potential of this technology. We will examine a hypothetical sepsis management system, delve into chronic disease management, and explore the application of AI Agents in optimizing hospital patient flow. As we examine these scenarios, we will provide a detailed look at the technical implementation challenges, including the integration with existing healthcare IT systems, data privacy considerations, and the crucial role of explainable AI in maintaining trust and transparency.

Implementing AI agents in healthcare is challenging. We will address significant hurdles, including ensuring data quality and mitigating bias, seamlessly integrating these systems into existing clinical workflows, and navigating the complex ethical considerations that arise when deploying autonomous systems in healthcare settings. Looking to the future, we will explore exciting developments, such as integrating Internet of Things (IoT) devices to provide real-time patient data and developing more sophisticated natural language interfaces to enhance human-AI collaboration.

The journey of AI Agents in healthcare is just beginning, and it promises to be transformative. As we stand on the brink of this new era, we invite you to explore with us the fascinating intersection of artificial intelligence and medical care. This paper aims to provide a comprehensive overview of the current state, potential applications, and future directions of AI Agents in healthcare, offering insights valuable to researchers, clinicians, and policymakers alike.

1. Architecture of Multi-Agent AI Systems in Healthcare: Management of Sepsis

Sepsis, a systemic inflammatory response secondary to an infectious agent, has seen relatively steady mortality rates for several decades despite advancements in broad-spectrum antibiotics, imaging, and life support systems. The complexity of maintaining optimization in clinical settings has hindered progress in managing this critical condition. Previous attempts to develop predictive sepsis models have proven challenging [7]. Here, we propose a plausible system of multi-AI agents working in concert to provide comprehensive patient monitoring and care.

The Data Collection and Integration Agent is powered by a controlled vocabulary to specify all data items. It is tasked with cleaning, transforming, and organizing patient data from various structured and unstructured sources. This agent prepares succinct summaries of consultant notes, formatting data for both human and machine consumption. All numerical data is presented graphically, including relevant historical data. The agent digitally captures all orders in a structured manner using a specified controlled vocabulary, which feeds into the output of other agents, i.e., documentation treatment, risk stratification, and, more importantly, supplies the data structures for future training.

The Diagnostic Agent. Critically ill disease manifests in multiple attributes of a wide array of tests ranging from plain chest X-ray films, CT scans, blood cell composition, plasma chemistries, and microscopic evaluation of specimens. Furthermore, parameters of life support could inform on disease severity, or a source of recommendation management. This offers a wide array of visual and numerical data to be used as input for computation, recommendation, and further training.

For example, to evaluate fluid overload in plain chest X-ray films or tissue histopathology slides, an agent can leverage deep learning models such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) to analyze images like radiographs and histopathology slides [8,9]. Recurrent Neural Networks (RNNs) or Transformer models process sequential data like time-series

vital signs. The agent also implements ensemble methods that combine multiple machine learning algorithms for robust diagnosis.

The Risk Stratification Agent assesses severity and predicts potential outcomes.

Morbidity and mortality risks are calculated using an established scoring system and individualized based on the history of other agents' conditional similar patients. These are presented graphically, with major risk factors highlighted for explainability.

The Treatment Recommendation Agent utilizes a reinforcement learning toolset supplemented by up-to-date clinical guidelines. Reinforcement learning builds on historical data, served in a data structure that implements a conserved vocabulary from patients with similar clinical features and their management over time. Training is also conducted on the patient's physiological data up to that point in time. All recommendations are presented on a dedicated user interface in a human-readable format, along with recommendations for editable, orderable items, references, and full-text snippets from the literature. Stop rules are implemented such that agents stop computing if confidence in recommendations is too broad or no clear pathway can be computed with certainty, prompting human mitigation.

The Resource Management Agent coordinates hospital resources using constraint programming techniques for optimal resource allocation. It utilizes queueing theory models to predict and manage patient flow [10] and implements genetic algorithms for complex scheduling problems [11].

The Monitoring and Alert Agent tracks patient progress and alerts staff to changes. It uses anomaly detection algorithms to identify unusual patterns in patient data and implements time series forecasting models, such as ARIMA and Prophet, to predict future patient states. The agent also utilizes stream processing techniques for real-time data analysis [12,13].

The Documentation and Reporting Agent maintains comprehensive medical records and generates reports. It employs advanced NLP techniques for automated report generation, uses advanced Large Language Models fine-tuned on medical corpora for narrative creation, and implements information retrieval techniques to efficiently query patient records.

2. Clinical Case Studies

2.1. Sepsis Management

To illustrate the functionality of a multi-agent system, let us examine its application in managing sepsis. The Data Collection and Integration Agent continuously aggregates patient data from various sources, normalizing and time-stamping it for consistent processing. The Diagnostic Agent analyzes this integrated data in real time, applying sepsis criteria and utilizing a deep learning model trained on a large dataset of sepsis cases to detect subtle patterns.

The Risk Stratification Agent calculates severity scores like qSOFA, SOFA, APACHE II, and NEWS-2 upon detecting a possible sepsis case [14]. It predicts the likelihood of specific outcomes and estimates the potential trajectory of the patient's condition over the next 24-48 hours. Based on this assessment, the Treatment Recommendation Agent suggests an initial treatment plan, including appropriate antibiotics, fluid resuscitation protocols, and vasopressor recommendations if indicated.

Concurrently, the Resource Management Agent checks the availability of necessary resources and prioritizes allocation based on the severity of all current cases. The Monitoring Agent tracks the patient's response to interventions in real-time, alerting the care team to any concerning changes or lack of expected improvement. Throughout this process, the Documentation Agent ensures that all actions, responses, and outcomes are meticulously recorded in a structured format, generating real-time updates for the patient's electronic health record and preparing summary reports for handoffs between care teams.

2.2. Administrative Workflow Support for Clinical Management

Modern healthcare operations are resource-intensive, requiring coordination of advanced imaging, procedures, lab testing, and professional consultations [15]. The administrative effort to support these activities is often overlooked, yet it is crucial for timely and efficient care delivery.

Multi-agent systems with scheduling assistants, radiology report summarizers, and recommenders for future steps based on imaging findings can serve as scalers to support this critical aspect of daily operations.

Consider the scenario of computed tomography for a patient with a chronic cough. A multi-agent system could schedule the CT appointment, author a summary of the results based on the radiology interpretation, and suggest therapies or consults to the CT-ordering clinician. Another AI agent could then schedule future appointments, including patient feedback and confirmation.

2.3. Operation Optimization Multi-Agent AI System

Hospitals are complex entities that must function at different scales and respond in an agile, timely manner at all hours, deploying staff at various positions [16]. A system of AI Agents can receive signals from sensors monitoring foot traffic in the emergency department and trauma unit, as well as the availability of operating room staff, equipment, and Intensive Care Unit (ICU) beds. Smart sensors make such monitoring possible through well-designed Internet of Things (IoT) networks. These networks benefit from advances in adaptive and consensus networking algorithms and recent advances in bioengineering and biocomputing [17]. Utilization of these resources can be modeled conditionally on the load they are expected to serve.

For example, in the scenario of abdominal imaging for suspected abdominal obstruction, an AI agent tasked with scheduling CT scans could time the patient's arrival based on acuity. Another AI agent could alert staff transporting the patient to the CT scan, with the next location contingent on a clinical decision to proceed to the operating room. Yet another AI agent could summarize radiology interpretations and alert the surgeon and anesthesia team to a potential case, while others could notify operating room staff of equipment needs or reserve a bed on the floor or in the ICU.

In this paradigm, AI Agents facilitate more precise and timely communication between multiple staff members who must work together as a team.

3. Technical Implementation

3.1. Large Language Models (LLMs)

Each agent in the system utilizes a different LLM optimized for its specific task. For example, the Diagnostic Agent uses an optimized Large Language Model, pre-trained on a large corpus of biomedical literature and fine-tuned on a dataset of confirmed sepsis cases and their presentations [18]. It implements few-shot learning techniques to adapt to rare or atypical presentations. The Treatment Recommendation Agent also uses LLM, employing a retrieval-augmented generation approach to access the latest clinical guidelines during inference. The Documentation Agent uses another advanced language model, fine-tuned on a large corpus of high-quality medical documentation, implementing controlled text generation techniques and utilizing a separate smaller model for real-time error checking and correction.

3.2. Inter-Agent Quality Control

Agents are implemented to learn from experience and the experience of other agents. Agents are equipped with quality assurance user-defined rule-based/model-based systems for quality assurance with clear stopping rules for human involvement and mitigation.

Sophisticated quality control measures bolster the system's reliability, including ensemble techniques in result comparison, redundancy for critical tasks, and automatic human review for disagreements above a certain threshold. Each agent provides a calibrated confidence score with its output, which is used to weight inputs in downstream tasks and trigger additional checks for low-confidence outputs.

A dedicated Quality Control Agent monitors outputs from all agents, employing both supervised and unsupervised anomaly detection techniques. Feedback loops allow agents to provide feedback on the quality and utility of information received from other agents. The system implements a multi-armed bandit approach to dynamically adjust the influence of different agents based on their

performance over time [19] and periodically retrains agent models using federated learning techniques.

3.3. Integration with Electronic Health Record (EHR) Systems

Seamless EHR integration is crucial for practical implementation. The system has secure API access to various EHR platforms, implements OAuth 2.0 for authentication, and uses HTTPS with perfect forward secrecy for all communications [20]. It works with HL7 FHIR to ensure interoperability and uses SNOMED CT for clinical terminology to ensure semantic interoperability across different EHR systems [21,22].

The system implements a multi-level approval system for write-backs to EHRs, with different thresholds based on the information's criticality. It utilizes digital signatures to ensure the integrity and non-repudiation of AI-generated entries and implements blockchain technology to create an immutable, distributed ledger of all AI system actions [23].

3.4. Explainable AI and Decision Transparency

To ensure transparency in decision-making processes, the system applies techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations) to provide insights into agent decision-making processes [24–26]. It provides customized visualizations for different stakeholders and allows users to explore alternative decision paths through "what-if" scenario modeling [27].

The system provides clear, calibrated confidence indicators for each recommendation or decision, implementing a novel "confidence calibration" agent that continuously monitors and adjusts confidence scores based on observed outcomes.

3.5. Continuous Learning and Adaptation

The system employs several techniques to remain current with evolving medical knowledge. Federated Learning allows learning from diverse datasets across multiple institutions without compromising patient privacy [28]. A/B Testing is used to safely deploy and compare new agent versions in controlled settings, implementing multi-armed bandit algorithms to efficiently explore new models while minimizing potential negative impacts. Human-in-the-loop learning and Active Learning techniques are used to incorporate feedback from healthcare professionals and efficiently solicit expert input on the most informative data points [29].

4. Clinical Implications

The implementation of multi-agent AI systems in healthcare has several potential benefits, including enhanced diagnostic accuracy, personalized treatment, improved efficiency, continuous monitoring, and resource optimization. For instance, a recent review of AI sepsis predictive models exhibited superior results to standard clinical scoring methods like qSOFA [30]. In oncology, such systems can result in more tailored treatments, enhancing outcomes [31]. Implementing an ambient dictation system can improve workflow and prevent provider burnout [32].

5. Challenges and Ethical Considerations

Despite the potential benefits, several challenges must be addressed in implementing multi-agent AI systems in healthcare. These include ensuring data quality and mitigating bias, integrating the systems with clinical workflows, improving the interpretability of AI systems, addressing ethical and legal considerations, and conducting rigorous clinical validation.

Strategies to address these challenges include implementing rigorous data validation processes, prioritizing extensive user experience research, developing hybrid models that combine deep learning with more interpretable techniques, establishing ethics boards to oversee system development and deployment, and conducting multi-center randomized controlled trials [33–35].

6. Future Directions

As multi-agent AI systems in healthcare evolve, several exciting directions emerge. These include integrating the Internet of Things (IoT) and wearable devices, developing more sophisticated natural language interfaces, and applying these systems to predictive maintenance of medical equipment.

7. Conclusions

The advent of multi-agent AI systems in healthcare represents a paradigm shift in how we approach patient care, clinical decision-making, and healthcare management. While these systems offer immense potential to transform healthcare delivery, their development and implementation must be guided by rigorous scientific validation, ethical considerations, and a patient-centered approach. As we navigate this complex landscape, the ultimate goal remains clear: harnessing the power of artificial intelligence to improve patient outcomes, enhance the efficiency of healthcare delivery, and ultimately advance the health and well-being of individuals and populations worldwide.

References

1. What are AI Agents?- Agents in Artificial Intelligence Explained - AWS. Amazon Web Services, Inc. Accessed September 21, 2024. https://aws.amazon.com/what-is/ai-agents/
2. What Are AI Agents? | IBM. July 3, 2024. Accessed September 21, 2024. https://www.ibm.com/think/topics/ai-agents
3. Agent AI. Microsoft Research. Accessed September 21, 2024. https://www.microsoft.com/en-us/research/project/agent-ai/
4. AutoGen | AutoGen. Accessed September 21, 2024. https://microsoft.github.io/autogen/
5. Crew AI. Accessed September 21, 2024. https://www.crewai.com/
6. LangChain. Accessed September 21, 2024. https://www.langchain.com/
7. Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Intern Med*. 2021;181(8):1065-1070. doi:10.1001/jamainternmed.2021.2626
8. Willemink MJ, Roth HR, Sandfort V. Toward Foundational Deep Learning Models for Medical Imaging in the New Era of Transformer Networks. *Radiol Artif Intell*. 2022;4(6):e210284. doi:10.1148/ryai.210284
9. Waqas A, Bui MM, Glassy EF, et al. Revolutionizing Digital Pathology With the Power of Generative Artificial Intelligence and Foundation Models. *Lab Investig J Tech Methods Pathol*. 2023;103(11):100255. doi:10.1016/j.labinv.2023.100255
10. Moreno-Carrillo A, Arenas LMÁ, Fonseca JA, Caicedo CA, Tovar SV, Muñoz-Velandia OM. Application of Queuing Theory to Optimize the Triage Process in a Tertiary Emergency Care ("ER") Department. *J Emerg Trauma Shock*. 2019;12(4):268. doi:10.4103/JETS.JETS_42_19
11. Pongcharoen P, Hicks C, Braiden PM, Stewardson DJ. Determining optimum Genetic Algorithm parameters for scheduling the manufacturing and assembly of complex products. *Int J Prod Econ*. 2002;78(3):311-322. doi:10.1016/S0925-5273(02)00104-4
12. Sardar I, Akbar MA, Leiva V, Alsanad A, Mishra P. Machine learning and automatic ARIMA/Prophet models-based forecasting of COVID-19: methodology, evaluation, and case study in SAARC countries. *Stoch Environ Res Risk Assess*. 2023;37(1):345-359. doi:10.1007/s00477-022-02307-x
13. Samosir J, Indrawan-Santiago M, Haghighi PD. An Evaluation of Data Stream Processing Systems for Data Driven Applications. *Procedia Comput Sci*. 2016;80:439-449. doi:10.1016/j.procs.2016.05.322
14. Asmarawati TP, Suryantoro SD, Rosyid AN, et al. Predictive Value of Sequential Organ Failure Assessment, Quick Sequential Organ Failure Assessment, Acute Physiology and Chronic Health Evaluation II, and New Early Warning Signs Scores Estimate Mortality of COVID-19 Patients Requiring Intensive Care Unit. *Indian J Crit Care Med*. 2022;26(4):466-473. doi:10.5005/jp-journals-10071-24170
15. Khan S, Vandermorris A, Shepherd J, et al. Embracing uncertainty, managing complexity: applying complexity thinking principles to transformation efforts in healthcare systems. *BMC Health Serv Res*. 2018;18(1):192. doi:10.1186/s12913-018-2994-0
16. Plsek PE, Greenhalgh T. The challenge of complexity in health care. *BMJ*. 2001;323(7313):625-628. doi:10.1136/bmj.323.7313.625
17. Kouchaki S, Ding X, Sanei S. AI- and IoT-Enabled Solutions for Healthcare. *Sensors*. 2024;24(8):2607. doi:10.3390/s24082607

18. Saab K, Tu T, Weng WH, et al. Capabilities of Gemini Models in Medicine. Published online May 1, 2024. doi:10.48550/arXiv.2404.18416
19. Villar SS, Bowden J, Wason J. Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges. *Stat Sci*. 2015;30(2):199-215. doi:10.1214/14-STS504
20. What is OAuth 2.0 and what does it do for you? Auth0. Accessed September 21, 2024. https://auth0.com/intro-to-iam/what-is-oauth-2
21. Index - FHIR v5.0.0. Accessed September 21, 2024. https://www.hl7.org/fhir/
22. Home. SNOMED International. Accessed September 21, 2024. https://www.snomed.org
23. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences—A scoping review. *Int J Med Inf*. 2020;134:104040. doi:10.1016/j.ijmedinf.2019.104040
24. Ribeiro MT, Singh S, Guestrin C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. Association for Computing Machinery; 2016:1135-1144. doi:10.1145/2939672.2939778
25. Ekanayake IU, Meddage DPP, Rathnayake U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Stud Constr Mater*. 2022;16:e01059. doi:10.1016/j.cscm.2022.e01059
26. Alabi RO, Elmusrati M, Leivo I, Almangush A, Mäkitie AA. Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP. *Sci Rep*. 2023;13(1):8984. doi:10.1038/s41598-023-35795-0
27. Otto E, Culakova E, Meng S, et al. Overview of Sankey flow diagrams: Focusing on symptom trajectories in older adults with advanced cancer. *J Geriatr Oncol*. 2022;13(5):742-746. doi:10.1016/j.jgo.2021.12.017
28. Fereidooni H, Marchal S, Miettinen M, et al. SAFELearn: Secure Aggregation for private FEderated Learning. In: *2021 IEEE Security and Privacy Workshops (SPW)*. ; 2021:56-62. doi:10.1109/SPW53761.2021.00017
29. Linton DL, Pangle WM, Wyatt KH, Powell KN, Sherwood RE. Identifying Key Features of Effective Active Learning: The Effects of Writing and Peer Discussion. *CBE—Life Sci Educ*. 2014;13(3):469-477. doi:10.1187/cbe.13-12-0242
30. Yang HS. Machine Learning for Sepsis Prediction: Prospects and Challenges. *Clin Chem*. 2024;70(3):465-467. doi:10.1093/clinchem/hvae006
31. Liao J, Li X, Gan Y, et al. Artificial intelligence assists precision medicine in cancer treatment. *Front Oncol*. 2023;12. doi:10.3389/fonc.2022.998222
32. Tierney AA, Gayre G, Hoberman B, et al. Ambient Artificial Intelligence Scribes to Alleviate the Burden of Clinical Documentation. *NEJM Catal*. 2024;5(3):CAT.23.0404. doi:10.1056/CAT.23.0404
33. Borkowski AA, Jakey CE, Thomas LB, Viswanadhan N, Mastorides SM. Establishing a Hospital Artificial Intelligence Committee to Improve Patient Care. *Fed Pract Health Care Prof VA DoD PHS*. 2022;39(8):334-336. doi:10.12788/fp.0299
34. Isaacks DB, Borkowski AA. Implementing Trustworthy AI in VA High Reliability Health Care Organizations. *Fed Pract Health Care Prof VA DoD PHS*. 2024;41(2):40-43. doi:10.12788/fp.0454
35. Han R, Acosta JN, Shakeri Z, Ioannidis JPA, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. *Lancet Digit Health*. 2024;6(5):e367-e373. doi:10.1016/S2589-7500(24)00047-5