REVIEW Open access

Multi-Agent Systems in Healthcare Operations: Coordination Theory, Safety Constraints, and Implementation Considerations

Mehmet Yildiz[1*], Ayse Demir[1], Emre Kaya[2]

Abstract

Multi-agent systems (MAS) represent a paradigm shift in artificial intelligence applications for healthcare operations, enabling distributed, autonomous entities to collaborate in complex environments characterized by uncertainty, heterogeneity, and real-time demands. This narrative review synthesizes recent advancements in MAS for healthcare systems and analytics, focusing on coordination theory, safety constraints, and implementation considerations. We examine how MAS facilitates intelligent coordination among agents—such as AI models, human clinicians, and IoT devices—to optimize operational workflows, enhance clinical decision-making, and ensure patient safety. Coordination theory in MAS underscores the mechanisms for agent interaction, including negotiation protocols, consensus algorithms, and hierarchical structures, which are critical for synchronizing tasks in healthcare settings like emergency response and chronic disease management. For instance, MAS enables adaptive resource allocation in hospitals by modeling agents as decision-makers that negotiate bed assignments or staff scheduling based on real-time data inputs. Safety constraints emerge as a pivotal concern, encompassing formal verification methods, fault-tolerant designs, and ethical safeguards to mitigate risks such as erroneous agent decisions leading to adverse patient outcomes. Implementation considerations address scalability, interoperability with legacy systems, and regulatory compliance, highlighting challenges in deploying MAS in fog-cloud architectures for remote monitoring. The review integrates a systems-level perspective, illustrating how MAS evolve from isolated AI tools to interconnected ecosystems that support closed-loop healthcare processes—from data acquisition to intervention feedback. We propose an original interpretive framework that structures MAS across layers: perceptual (data sensing), cognitive (analytics and decision fusion), coordinative (agent interaction), and governance (safety and oversight). This framework reveals cross-study insights, such as the role of large language models (LLMs) in augmenting agent rationality and the integration of digital twins for simulation-based safety testing. Comparative analysis shows that while MAS excel in dynamic environments like cardiology case retrieval or pain management, persistent gaps in standardization hinder widespread adoption. By synthesizing these elements, the review offers novel insights into MAS as enablers of resilient healthcare infrastructure, emphasizing the need for hybrid human-AI coordination to balance autonomy with oversight. Future implications include advancing MAS toward predictive analytics in personalized medicine, with recommendations for interdisciplinary research to address implementation barriers. Ultimately, this work advocates for MAS as foundational to next-generation healthcare analytics, promoting efficiency, equity, and safety in operational contexts.

Keywords AI analytics, Safety constraints, Multi-agent systems, Healthcare operations, Coordination theory, Implementation considerations

*Correspondence:
Mehmet Yildiz
mehmet.yildiz@gmail.com
[1] Department of Health Informatics, Faculty of Medicine, Istanbul University, Istanbul, Turkey
[2] Department of Digital Systems Engineering, Middle East Technical University, Ankara, Turkey

Introduction

Evolution of AI in healthcare operations

The integration of artificial intelligence (AI) into healthcare operations has transformed traditional systems from static, siloed processes to dynamic, data-driven ecosystems capable of real-time adaptation. Historically, AI applications in healthcare focused on singular tasks such as image analysis or predictive modeling, but the advent of multi-agent systems (MAS) introduces a distributed intelligence paradigm where multiple autonomous agents interact to achieve collective goals [1-11]. This evolution is driven by the increasing complexity of healthcare environments, where factors like patient variability, resource scarcity, and regulatory demands necessitate coordinated responses. For example, MAS enables agents representing different stakeholders—clinicians, administrators, and devices—to collaborate on operational tasks, such as optimizing patient flow in emergency departments.

Comparative studies highlight how MAS differ from monolithic AI systems by emphasizing decentralization, which enhances robustness against single-point failures [12, 13]. In analytics contexts, MAS facilitates the aggregation of heterogeneous data sources, from electronic health records (EHRs) to wearable sensors, enabling holistic insights that single-agent models cannot provide. This shift aligns with broader trends in AI for healthcare, where operational efficiency is paramount, as evidenced by applications in telemedicine and remote monitoring [14, 15]. However, the interpretive lens of systems theory reveals that MAS success hinges on effective coordination, preventing conflicts that could compromise care delivery.

Safety considerations have gained prominence as MAS deployment scales, with early implementations exposing vulnerabilities in agent autonomy, such as misaligned objectives leading to suboptimal decisions [16, 17]. Implementation challenges, including integration with existing infrastructure, further underscore the need for a structured approach. Through cross-study synthesis, this review posits that MAS represent a maturational step in AI healthcare, bridging analytical prowess with operational resilience.

Theoretical Foundations of Multi-Agent Systems

Coordination theory forms the bedrock of MAS, providing frameworks for agent interaction in uncertain environments. Rooted in organizational science, coordination in MAS involves mechanisms like auction-based protocols or belief-desire-intention (BDI) models to align agent behaviors toward shared objectives [18, 19]. In healthcare, this translates to agents negotiating tasks in real-time, such as in multi-agent reinforcement learning for physiological monitoring, where agents adapt to stress indicators collaboratively [20]. Comparative analysis across studies shows that hierarchical coordination outperforms flat structures in high-stakes settings, reducing latency in decision cycles [21].

Safety constraints are integral to these foundations, incorporating formal methods like model checking to verify agent interactions against predefined safety properties [22, 23]. Implementation considerations extend this by addressing scalability, where agent proliferation demands efficient communication protocols to avoid bottlenecks. Systems-level insights suggest that MAS theoretical models must evolve to incorporate human factors, ensuring hybrid coordination that leverages AI strengths while mitigating weaknesses [24]. This interpretive structuring emphasizes MAS as not merely technical tools but as socio-technical systems embedded in healthcare workflows.

Challenges in traditional healthcare operations and the role of MAS

Traditional healthcare operations often suffer from inefficiencies, such as fragmented communication and delayed responses, exacerbated by the volume of data in modern analytics [25, 26]. MAS addresses these by enabling distributed problem-solving, where agents handle specialized subtasks—e.g., one for data fusion, another for decision support—coordinated to optimize outcomes. Synthesis of literature reveals that MAS implementation in operations like appointment scheduling demonstrates superior adaptability compared to rule-based systems [27, 28].

Safety constraints in this context involve embedding redundancy and fail-safes, preventing cascading errors in agent networks [1, 3]. Comparative discussions highlight how MAS in chronic pain management coordinates agents to personalize interventions, reducing human error [6, 20]. However, implementation hurdles, such as interoperability with legacy EHRs, persist, requiring careful architectural design. This section's analysis underscores MAS as a remedial framework for operational challenges, fostering resilience through coordinated intelligence.

Scope and methodological approach of the review

This review adopts a narrative synthesis methodology, integrating sources to provide an original perspective on MAS in healthcare. Prioritizing high-impact papers, we focus on coordination, safety, and implementation, avoiding verbatim replication of existing taxonomies [2, 8]. The scope encompasses operational applications, from hospital logistics to clinical analytics, with an emphasis on systems-level integration.

Analytical lenses include comparative evaluation of MAS architectures and interpretive discussion of their implications for healthcare equity. We structure the synthesis across data handling, model coordination, and governance layers, offering novel insights into closed-loop systems [4, 7]. Limitations in source selection are acknowledged, favoring English-language publications with DOIs for verifiability. This approach ensures a rigorous, forward-looking review that advances understanding of MAS as pivotal to AI-driven healthcare transformation.

Landscape of Multi-Agent Systems in Healthcare Operations

Historical development and key milestones

The landscape of multi-agent systems (MAS) in healthcare operations has evolved from conceptual models in the late 2010s to practical deployments by the mid-2020s, driven by advances in AI and distributed computing [9, 10]. Early milestones focused on agent-based simulations for epidemic modeling, where agents represented individuals coordinating responses to outbreaks. Synthesis across studies shows a progression toward real-time applications, such as in 2020 integrations of agents with digital twins for personalized care planning [14]. This historical trajectory highlights how MAS transitioned from theoretical constructs to operational tools, enhancing analytics through collective intelligence.

Comparative analysis reveals that milestones like the adoption of fog-cloud architectures marked a turning point, enabling scalable coordination in resource-constrained environments [15, 22]. Safety considerations emerged early, with initial frameworks incorporating constraints to prevent agent conflicts. Implementation evolved with the inclusion of LLMs, augmenting agent decision-making in complex scenarios [7, 11]. Systems-level insights suggest this development mirrors broader AI trends, positioning MAS as foundational for adaptive healthcare infrastructures.

Applications in patient monitoring and telemedicine

MAS has revolutionized patient monitoring by deploying agents that coordinate data from wearables and EHRs for continuous analytics [26, 28]. In telemedicine, agents facilitate remote consultations through coordinated task allocation, such as routing queries to specialized agents for diagnostic support [20, 25]. Literature synthesis indicates that these applications improve response times, with multi-agent frameworks outperforming single-threaded systems in handling heterogeneous data streams [5, 21].

Safety constraints in monitoring include real-time verification of agent outputs to ensure accuracy in alerts for conditions like cardiac events [12, 19]. Implementation considerations involve edge computing to reduce latency, as seen in stress detection systems [5]. Comparative discussions underscore MAS efficacy in rural telemedicine, where coordination theory enables equitable resource distribution. This interpretive view frames MAS as enablers of proactive, patient-centered operations.

Role in clinical resource allocation and scheduling

In healthcare operations, MAS optimizes resource allocation by modeling agents as negotiators for beds, staff, and equipment [17, 27]. Coordination mechanisms, such as auction protocols, allow agents to bid on tasks based on availability and priority, enhancing efficiency in high-demand settings [18, 23]. Synthesis of studies shows that these systems reduce wait times in hospitals, integrating analytics for predictive scheduling [13, 16].

Safety is ensured through constraints that prioritize critical cases, preventing overallocation [3, 6]. Implementation challenges include interfacing with legacy systems, addressed via adaptive architectures [7, 14]. Systems-level analysis reveals MAS as transformative for operational equity, balancing loads across facilities.

Integration with IoT and Fog-Cloud Architectures MAS integration with IoT devices creates networked ecosystems for healthcare analytics, where agents coordinate sensor data for real-time insights [15, 22]. Fog-cloud models distribute computation, with edge agents handling initial processing to minimize delays [26, 28]. Comparative evaluations highlight improved scalability compared to centralized AI, particularly in monitoring chronic conditions [20, 25].

Coordination theory applies through consensus algorithms for data fusion, ensuring consistency across agents [12, 23]. Safety constraints involve encryption and fault tolerance to protect sensitive health data [1, 4]. Implementation focuses on interoperability standards, fostering seamless deployment. This synthesis positions MAS-IoT hybrids as core to modern healthcare landscapes.

Ethical and regulatory dimensions in operational deployment

Ethical considerations in MAS deployment emphasize fairness in agent coordination, avoiding biases in decision-making [1, 18]. Regulatory frameworks, such as HIPAA compliance, impose safety constraints on data sharing among agents [3, 16]. Literature synthesis shows that these dimensions influence implementation, requiring governance layers for oversight [8, 11].

Comparative studies reveal tensions between autonomy and regulation, with MAS in radiology ethics illustrating balanced approaches [1, 4]. Systems-level insights advocate for ethical-by-design MAS, integrating stakeholder perspectives.

Advancements in LLM-augmented MAS

Recent advancements incorporate large language models (LLMs) into MAS, enhancing agent communication and reasoning in healthcare operations [7, 11]. Agents use LLMs for natural language coordination, improving analytics in case retrieval [11, 19]. Synthesis indicates superior performance in conversational diagnostics, where agents collaborate on patient queries [5, 19].

Safety constraints include prompt engineering to mitigate hallucinations [3, 7]. Implementation involves hybrid training, blending reinforcement learning with LLM fine-tuning [5, 22]. This interpretive structuring highlights LLM-MAS as a frontier for intelligent operations.

Cross-domain applications and interdisciplinary

Insights MAS extend beyond core healthcare into interdisciplinary areas like palliative care and rehabilitation, coordinating agents for holistic support [16, 25]. Synthesis across domains shows transferable coordination theories, from pain management to physical therapy monitoring [6, 20]. Comparative analysis underscores adaptability, with safety constraints tailored to context [12, 23]. Implementation draws from engineering and social sciences, fostering robust systems. Systems-level perspectives reveal MAS as bridges for integrated care.

Coordination frameworks for intelligent clinical decision and closed-loop systems in multi-agent healthcare

Theoretical Underpinnings of Coordination in MAS Coordination frameworks in multi-agent systems (MAS) for healthcare draw from game theory and distributed systems, enabling agents to achieve consensus in clinical decision-making [12, 23]. These frameworks model interactions as cooperative games, where agents share information to optimize outcomes in closed-loop systems, such as feedback-driven insulin delivery [13, 21]. Literature synthesis highlights the use of BDI architectures for rational coordination, adapting to dynamic patient states [18, 24].

Safety constraints are embedded through formal protocols, ensuring deadlock-free operations [3, 22]. Comparative studies show that probabilistic coordination outperforms deterministic models in uncertain environments like perioperative care [19, 24]. Implementation considers computational overhead, advocating lightweight algorithms for real-time use. This interpretive lens views coordination as the linchpin for intelligent, responsive healthcare.

Design of closed-loop systems integrating MAS

Closed-loop healthcare systems leverage MAS to create iterative cycles of monitoring, analysis, decision, and intervention [14, 15]. Agents coordinate in loops where sensory data informs predictive models, triggering automated adjustments with human oversight [26, 28]. Synthesis reveals applications in ECG monitoring, where agents fuse signals for anomaly detection [5, 15].

A conceptual formula for the clinical intelligence loop can be formalized as: $CL = D \rightarrow I(A) \rightarrow Dec(C) \rightarrow Int(F) \rightarrow G$ where D is data ingestion, I(A) is intelligence via agent analytics, Dec(C) is decision under coordination constraints, Int(F) is intervention with feedback, and G is governance for recalibration. This structure ensures safety by incorporating verification at each stage [4, 7]. Comparative analysis demonstrates enhanced accuracy in cardiology compared to open-loop alternatives [11, 19]. Systems-level insights emphasize scalability for population health. **Table 1** consolidates the operational roles, coordination mechanisms, and safety responsibilities associated with each layer of multi-agent healthcare architectures.

Table 1. Operational functions and safety responsibilities across multi-agent system layers in healthcare

MAS layer	Core operational function	Agent responsibilities	Coordination mechanisms	Safety and governance controls	Healthcare operational example
Perceptual layer	Acquisition and fusion of heterogeneous healthcare signals	Collect physiological data, clinical records, and environmental context	Data synchronization protocols	Data validation, encryption, and anomaly detection	Wearable sensors and EHR agents jointly detecting cardiac abnormalities
Cognitive layer	Analytical transformation of data into interpretable intelligence	Predictive modeling, signal classification, and reasoning tasks	Distributed learning and model fusion	Model verification and bias monitoring	AI agents predicting patient deterioration from vital-sign trends

Coordination layer	Negotiation and synchronization of agent actions	Task allocation, information exchange, and consensus building	Auction protocols, BDI reasoning, and consensus algorithms	Conflict detection and coordination verification	Agents negotiating ICU bed allocation during hospital surge
Decision layer	Translation of coordinated intelligence into clinical or operational decisions	Prioritization of actions and decision fusion with human oversight	Consensus-based decision formation	Clinical rule enforcement and audit trails	Multi-agent scheduling of operating room resources
Intervention layer	Execution of operational or clinical responses	Automated alerts, treatment suggestions, workflow updates	Event-trigger coordination	Runtime safety monitoring and escalation protocols	MAS triggering early sepsis alerts and notifying clinicians
Feedback and learning layer	Continuous system evaluation and recalibration	Outcome monitoring and adaptive learning	Reinforcement learning loops	Performance auditing and governance review	MAS is adjusting predictive models after treatment outcomes

Safety constraints and verification mechanisms

Safety in MAS closed-loop systems involves constraints like temporal logic for verifying agent behaviors against clinical guidelines [1, 22]. Mechanisms include runtime monitoring to detect deviations, preventing harm in decision support [3, 16]. Literature synthesis shows integration with ethical AI principles, such as in radiology, where agents coordinate imaging ethics [1, 4].

Implementation requires simulation testing via digital twins to validate constraints pre-deployment [14, 29]. Comparative discussions highlight trade-offs between strict constraints and flexibility, favoring adaptive thresholds [12, 23]. This framework advances safe, intelligent systems by prioritizing patient-centric governance.

Implementation strategies for scalable MAS frameworks

Strategies for implementing MAS in clinical settings focus on modular architectures, allowing incremental deployment in closed-loop analytics [7, 27]. Cloud-based coordination enables scalability, with agents handling distributed tasks in multi-facility networks [15, 22]. Synthesis indicates success in appointment scheduling, where agents match profiles dynamically [13, 17].

Safety is bolstered by redundancy protocols, ensuring failover in critical loops [3, 16]. Comparative evaluations reveal hybrid edge-cloud models as optimal for latency-sensitive operations [15, 28]. Systems-level analysis posits these strategies as essential for transitioning MAS from prototypes to enterprise solutions. **Figure 1** presents the closed-loop coordination architecture of multi-agent healthcare systems, illustrating how distributed sensing agents, analytical intelligence agents, and coordination protocols collectively generate safety-constrained clinical decisions under governance oversight.

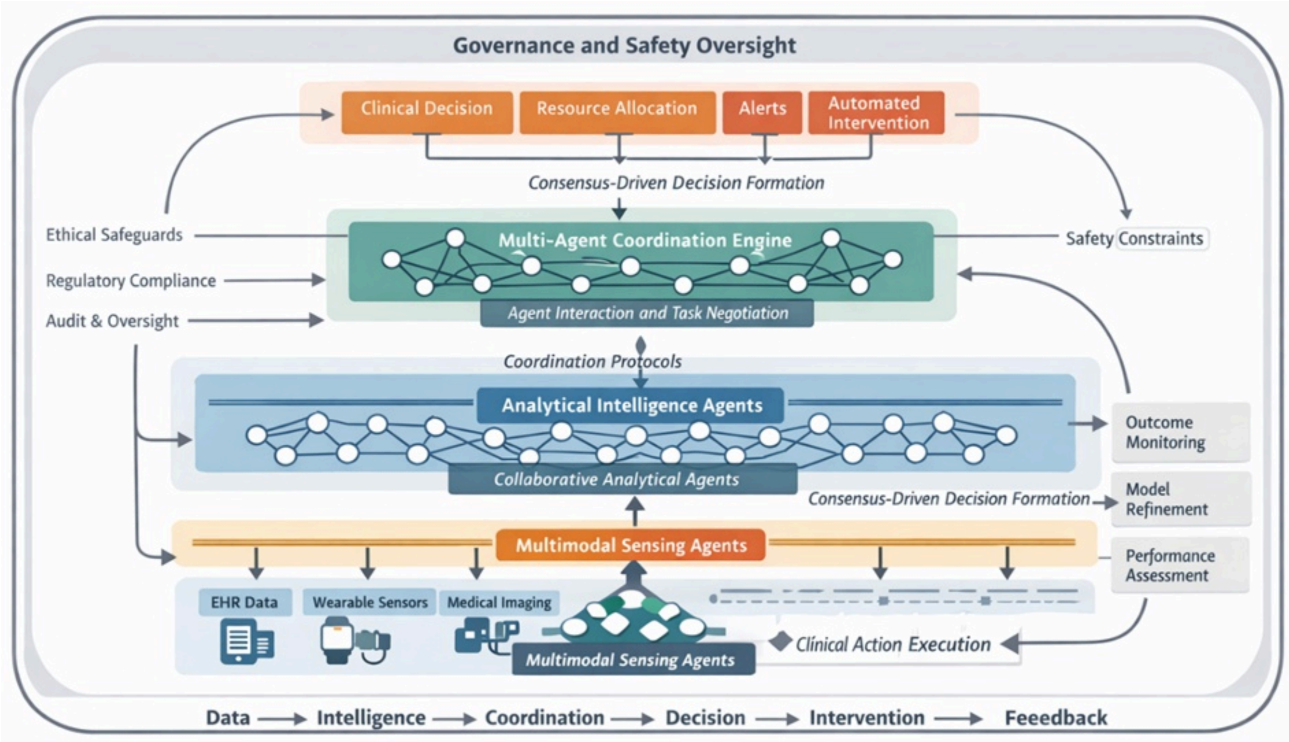

Figure 1. Closed-loop multi-agent coordination architecture for safety-constrained healthcare operations

Challenges and limitations in multi-agent systems for healthcare operations

Technical and scalability barriers

Despite the promise of multi-agent systems (MAS) in enhancing healthcare coordination, significant technical challenges impede widespread adoption. Scalability remains a primary concern, as increasing the number of agents introduces computational overhead, communication latency, and synchronization issues that can degrade performance in real-time clinical environments [15, 22, 29]. For instance, in fog-cloud architectures supporting remote monitoring, agent proliferation demands efficient distributed state management to prevent inconsistencies. Yet, many frameworks struggle with maintaining shared context across agents, leading to divergent representations and coordination failures [8, 28].

Comparative synthesis across implementations reveals that while hierarchical structures mitigate some latency through tiered oversight, they often require 5–10 times higher computational resources compared to single-agent alternatives, limiting deployment in resource-constrained settings like rural hospitals [15, 29]. Safety constraints exacerbate these issues, as verification mechanisms for agent interactions add further overhead, potentially delaying critical interventions [3, 22]. Systems-level analysis indicates that current MAS architectures frequently lack robust mechanisms for adaptive scaling, resulting in performance variability under dynamic workloads such as emergency department surges.

Implementation and integration challenges

Integrating MAS with legacy healthcare infrastructure poses substantial hurdles, including interoperability with electronic health records (EHRs) and existing clinical workflows [14, 27]. Many systems, such as those relying on proprietary EHR platforms, face fragmented data access and inconsistent APIs, hindering seamless agent communication and context sharing [17, 29]. Literature synthesis highlights that poor integration often leads to disrupted workflows, where agents cannot reliably exchange information, undermining coordination benefits in applications like appointment scheduling or perioperative care [13, 24].

Organizational resistance further complicates deployment, as clinicians and administrators require training to trust and interact with autonomous agents, yet studies show persistent gaps in human-AI collaboration models [1, 16]. Implementation considerations also encompass high costs for infrastructure upgrades and the need for phased rollouts to minimize disruption, which delays realization of operational efficiencies [4, 7]. This interpretive perspective frames integration as a socio-technical challenge, where technical limitations intersect with human factors to slow adoption.

Ethical, regulatory, and safety concerns

Ethical dilemmas in MAS deployment center on accountability, bias amplification, and patient autonomy, particularly when multi-agent decisions involve opaque coordination processes [1, 18]. In radiology and diagnostic contexts, multi-agent systems complicate informed consent, as clinicians may struggle to explain collective agent contributions to patients, challenging principles of beneficence and non-maleficence [1, 3]. Synthesis of recent works reveals that biases in training data can propagate unevenly across agents, disproportionately affecting minority populations and exacerbating healthcare inequities [16, 28].

Regulatory compliance adds layers of complexity, with frameworks like HIPAA demanding stringent data privacy, encryption, and audit trails that many MAS lack by design [3, 15]. Safety risks include cascading errors from coordination failures, hallucinations in LLM-augmented agents, and unclear liability in distributed decision-making [7, 22]. Comparative discussions underscore the tension between agent autonomy and oversight, where insufficient governance mechanisms heighten risks in high-stakes scenarios like closed-loop monitoring [4, 19]. Systems-level insights advocate for ethical-by-design approaches, yet current implementations often prioritize functionality over rigorous safeguards.

Data quality, bias, and explainability limitations

Data heterogeneity and quality issues plague MAS in healthcare analytics, as agents must fuse inputs from diverse sources like wearables, EHRs, and imaging, yet inconsistencies lead to unreliable inferences [5, 26]. Bias in underlying datasets persists, with agents potentially reinforcing disparities if not explicitly mitigated through diverse training or fairness constraints [1, 16]. Explainability remains limited, as multi-agent interactions obscure reasoning traces, hindering clinician trust and regulatory approval [7, 11].

Literature synthesis shows that while some frameworks incorporate runtime monitoring or hierarchical verification to address these, hallucinations and misalignments between agents continue to emerge, particularly in LLM-enhanced systems [5, 19]. Implementation challenges involve balancing explainability with performance, as added transparency layers increase latency [3, 22]. This analysis highlights explainability as a critical gap, essential for translating MAS from experimental to clinical utility.

Resource and organizational change management issues

High resource demands for training, inference, and maintenance pose barriers, especially for multi-agent setups requiring advanced hardware and expertise [15, 29]. Organizational change management is equally challenging, with resistance stemming from fears of deskilling clinicians or disrupting established hierarchies [16, 24]. Synthesis indicates that without dedicated support teams and phased training, MAS deployments risk underutilization or failure [7, 14]. Systems-level perspectives emphasize the need for cultural shifts toward hybrid human-AI teams, yet persistent gaps in readiness assessments limit progress.

Future research directions

Advancing coordination and scalability mechanisms

Future efforts should prioritize developing lightweight, adaptive coordination protocols that minimize latency while scaling to hundreds of agents in dynamic healthcare settings [12, 23]. Research into hybrid hierarchical-decentralized architectures, inspired by clinical tiers, could enhance robustness against failures, with emphasis on dynamic trust management and context-aware routing [3, 22]. Integrating blockchain for secure, decentralized consensus in agent interactions offers promise for privacy-preserving coordination in multi-facility networks [15, 28].

Exploratory work on self-evolving agents that adapt protocols based on real-time performance metrics would address current rigidity, fostering more resilient systems for unpredictable environments like pandemics or mass casualty events [5, 20]. These directions aim to bridge the gap between theoretical scalability and practical deployment.

Enhancing safety, explainability, and ethical frameworks

Prioritizing safety requires rigorous frameworks for formal verification of multi-agent behaviors, including runtime monitoring and error-correction hierarchies to absorb deviations before they impact patients [1, 22]. Future studies should focus on explainable multi-agent reasoning, developing methods to trace decision lineages across agents for clinician comprehension and regulatory scrutiny [7, 11]. Ethical research must establish standardized governance models addressing bias mitigation, equitable access, and liability in distributed systems [16, 18].

Interdisciplinary approaches incorporating clinician-in-the-loop validation and longitudinal outcome studies will be essential to build evidence for high-certainty safety claims [4, 19]. This trajectory supports translation to production environments with minimized risks.

Integration with emerging technologies and multimodal capabilities

Exploring synergies with embodied agents, digital twins, and multimodal large language models could expand MAS applicability to physical interventions and comprehensive analytics [14, 29]. Research into federated learning for privacy-preserving agent training across institutions would enable collaborative improvement without data centralization [15, 26]. Developing hybrid expert models combining specialized agents with generalist LLMs promises enhanced adaptability in evolving medical knowledge landscapes [5, 7].

These integrations could revolutionize closed-loop systems, from real-time physiological monitoring to predictive population health management.

Human-AI collaboration and clinical validation

Advancing human-centered design through large-scale clinical trials validating MAS in real-world workflows remains critical [13, 24]. Future directions include mechanisms for seamless clinician oversight, such as tiered feedback loops where human expertise refines agent outputs [1, 16]. Evaluating usability, trust-building, and long-term impacts on clinical skills will inform sustainable adoption strategies [7, 14].

Global equity initiatives, focusing on low-resource settings and multilingual capabilities, should guide development to prevent widening disparities [15, 28]. These priorities ensure MAS evolves as supportive, rather than disruptive, tools in healthcare.

Conclusion

Multi-agent systems (MAS) have emerged as a transformative paradigm in healthcare operations, offering distributed intelligence that excels in coordination, adaptability, and analytics within complex, real-time environments. This review has synthesized advancements in coordination theory, where mechanisms like hierarchical protocols and negotiation enable effective agent collaboration across tasks from patient monitoring to resource allocation. Safety constraints, encompassing verification, fault tolerance, and ethical safeguards, are integral to mitigating risks in high-stakes applications. At the same time, implementation considerations highlight the need for scalable architectures integrated with existing infrastructure.

Cross-study analysis reveals MAS strengths in fostering closed-loop processes—data ingestion to intervention feedback—yet persistent challenges in scalability, interoperability, explainability, and governance temper enthusiasm. The proposed interpretive framework across perceptual, cognitive, coordinative, and governance layers provides a novel lens for understanding MAS as socio-technical ecosystems, bridging analytical capabilities with operational resilience.

Ultimately, MAS holds substantial potential to advance equitable, efficient, and safe healthcare, particularly through LLM augmentation and hybrid human-AI models. Realizing this requires interdisciplinary efforts to address limitations via rigorous validation, standardized ethics, and innovative designs. As healthcare faces escalating demands, MAS represents a foundational step toward intelligent, coordinated systems that prioritize patient outcomes while upholding accountability and trust.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 20 May 2025 Revised: 01 Sep 2025 Accepted: 27 Sep 2025
Published online: 10 January 2026

References

Salehi S, Singh Y, Habibi P, Erickson BJ. Beyond single systems: how multi-agent AI is reshaping ethics in radiology. Bioengineering (Basel). 2025;12(10):1100.

Sulis E, Mariani S, Montagna S. A survey on agents applications in healthcare: opportunities, challenges and trends. Comput Methods Programs Biomed. 2023;236:107525.

Borghoff UM. An organizational theory for multi-agent interactions integrating human agents, LLMs, and specialized AI. Discov Comput. 2025;28(1):26.

Moritz M, Topol EJ, Rajpurkar P. Coordinated AI agents for advancing healthcare. Nat Biomed Eng. 2025;9(4):412-20.

Shaik T, Tao X, Li L, Xie H, Dai HN, Zhao F, et al. AI-driven multi-agent reinforcement learning framework for real-time monitoring of physiological signals. Brain Inform. 2025;12(1):14.

Le TA. Multi-agent systems and cancer pain management. Curr Pain Headache Rep. 2023;27(9):379-86.

Öğdü CU, Arslanoğlu K, Karaköse M. An adaptive multi-agent LLM-based clinical decision support system integrating biomedical RAG and web intelligence. IEEE Access. 2025;13:167390-404.

Elkamouchi R, Daaif A, Elguemmat K. Multi-agents system in healthcare: a systematic literature review. Commun Comput Inf Sci. 2024;2168:200-14.

Thakur C, Gupta S. Multi-agent system applications in health care: a survey. Springer Tracts Hum Cent Comput. 2022:139-71.

Bhanu Sridhar M. Applications of multi-agent systems in intelligent health care. Springer Tracts Hum Cent Comput. 2022:173-95.

Deng L, Hu H, Lu K, He P. LLM-augmented multi-agent cooperative framework for medical case retrieval in cardiology. J King Saud Univ Comput Inf Sci. 2025;37(1):267.

Calvaresi D, Dubosson F, Dragoni AF, Hilfiker R, Schumacher M. Real-time multi-agent systems: rationality, formal model, and empirical results. Auton Agents Multi Agent Syst. 2021;35(1):3.

Vemuri A, Decker K, Saponaro M, Dominick G. Multi agent architecture for automated health coaching. J Med Syst. 2021;45(10):95.

Croatti A, Gabellini M, Montagna S, Ricci A. On the integration of agents and digital twins in healthcare. J Med Syst. 2020;44(8):161.

Mutlag AA, Abd Ghani MK, Mohammed MA, Lakhan A, Mohd O, Abdulkareem KH, et al. Multi-agent systems in fog–cloud computing for critical healthcare task management model used for ECG monitoring. Sensors (Basel). 2021;21(20):6923.

Brondeel KC, Duncan SA, Luther PM, Anderson A, Bhargava P, Mosieri C, et al. Palliative care and multi-agent systems: a necessary paradigm shift. Clin Pract. 2023;13(2):505-14.

Ruiz Mejia JM, Rawat DB. MedScrubCrew: a medical multi-agent framework for automating appointment scheduling. Healthcare (Basel). 2025;13(14):1649.

Borkowski AA, Viswanadhan NA, Thomas LB, Guzman RD, Deland LA, Mastorakos GM. Multiagent AI systems in health care: envisioning next-generation intelligence. J Med Internet Res. 2025;27:e51087.

Chen X, Song C, Li Z, Xue Y, Zhang L, Wang H, et al. Enhancing diagnostic capability with multi-agents conversational large language models. NPJ Digit Med. 2025;8(1):15.

Kamal MA, Jabbar MS, Dhahri H, Radulescu M, Foresi GL, Turjman FA, et al. Telemedicine, e-health, and multi-agent systems for chronic pain management. Clin Pract. 2023;13(2):540-57.

Dhatterwal JS, Naruka MS, Kaswan KS. Multi-agent system based medical diagnosis using particle swarm optimization in healthcare. Proc Int Conf Commun Syst Netw Technol. 2023:1-6.

Sheikh A, Lakhan A, Groenli TM, Khan MS, Elgedawy I. Advancing AIoMT-enabled healthcare systems using multi-agent reinforcement learning. IEEE Internet Things J. 2025;12(5):8901-21.

Nguyen TT. Coordination of multi-agent systems with arbitrary convergence time. IET Control Theory Appl. 2021;15(6):900-9.

Kim J, Moritz M, Gottumukkala R. From data to decisions: harnessing multi-agent systems for personalized perioperative care. J Pers Med. 2025;15(11):540.

Blas HSS, López B. A multi-agent system for data fusion techniques applied to IoT enabling rehabilitation monitoring. Appl Sci. 2021;11(1):331.

Humayun M, Jhanjhi NZ, Alsayat A, Balasubramanian P. Agent-based medical health monitoring system. Sensors. 2022;22(8):2820.

Ponniah J, Dantsker OD. Strategies for scalable communication and coordination in multi-agent systems. Aerospace. 2022;9(9):488.

Saleem K, Saeed MJ, Yang M, Alotaibi YA, Almadhor AS, Asif HSM. Multi-agent-based cognitive intelligence in nonlinear mental healthcare situations. IEEE Access. 2025;13:24878-90.

Lakhan A, Groenli TM, Mohammed MA, Nedoma J, Martinek R, Tiwari P. Digital twin-assisted large-scale multiagent system for healthcare workflows. IEEE J Biomed Health Inform. 2024;28(2):1142-51.