TELEHEALTH AND MEDICINE TODAY

EDITORIALS AND DISCUSSIONS

AI Agents in Healthcare: The Need for Governance

Tomer Jordi Chaffer, MSc[1], Joe Littlejohn, MD[2], Muthu Ramachandran, PhD[3,4] and Claudia Lamschtein, MD[5]

[1]Faculty of Law, McGill University, Quebec, Canada; [2]Zucker School of Medicine, New York, NY, USA; [3]Research Consultant at Forti5 Tech and at Self-Evolving Software (SES) Systems Group, London, UK; [4]University of South Africa (UniSA), Pretoria, South Africa; [5]Department of Psychiatry, Faculty of Medicine, University of Manitoba, Manitoba, Canada

Corresponding Author: Tomer Jordi Chaffer, Email: tomer.chaffer@mail.mcgill.ca

DOI: https://doi.org/10.30953/thmt.v10.622

Keywords: AI agents, AI ethics, governance by design, multi-agent systems

Received: August 25, 2025; Accepted: November 21, 2025; Published: December 4, 2025

Artificial intelligence (AI) is reshaping healthcare, from improving diagnostic accuracy in radiology and pathology to streamlining clinical workflows and enabling more personalized patient care. Pressures such as workforce shortages, aging populations, and rising costs have accelerated the need for technological solutions that can extend human capacity.[1,2] Yet much of today's healthcare AI remains narrow in scope: static decision aids, image classifiers, or chat-based triage systems optimized for discrete tasks.

Evolution in AI Agents

The next phase of this evolution is emerging in the form of AI agents.[3] Built on the perceive–reason–act–learn loop, these systems move beyond single-task automation to autonomous, goal-directed activity that can adapt across contexts.[4,5] Rather than functioning solely as clinical tools, large language model (LLM)-powered agents are poised to collaborate with clinicians, coordinate care, and engage directly with patients—signaling a profound shift in how healthcare is delivered.

This profound shift is most visible in telehealth, which provides a vivid lens for understanding the impact of AI agents because it collapses the traditional boundaries of time, place, and specialty in care delivery.[6] By moving beyond episodic encounters, agent-enabled telehealth evolves from a conduit of information into a dynamic platform for continuous, coordinated care.

Voice-Enabled Assistants

One of the most immediate applications is the rise of voice-enabled assistants or conversational diagnostic agents,[7] which combine advanced speech recognition, natural language processing, and secure interoperability with electronic medical records (EMRs). A patient's history of present illness (HPI) forms the backbone of any clinical encounter, and AI-enabled platforms now support asynchronous data collection, allowing patients to provide their history through conversational interfaces or structured questionnaires before the telehealth session. This ensures that clinicians enter the visit with a clear understanding of the patient's concerns, enabling more efficient and focused virtual encounters.

During intake, voice-enabled systems also capture responses via ambient listening and transcribe them into relevant clinical fields, generating a narrative HPI that can be reviewed for accuracy before being seamlessly incorporated into the EMR. In parallel, chart preparation—traditionally performed by medical assistants or support staff—can be augmented by AI.[8] By reviewing prior visit notes, lab results, imaging, and medication lists, AI agents can generate organized chart summaries for providers to review ahead of the appointment. This kind of virtual prep work ensures a smooth workflow, reduces redundancy, and supports more informed decision-making.

Beyond intake and chart review, voice-enabled assistants further streamline care by supporting verbal queries

for lab, pathology, and imaging results, as well as enabling spoken physician order entry.[9,10] Collectively, these capabilities illustrate how AI agents can optimize telehealth encounters end-to-end, freeing clinicians to focus on complex diagnostic and therapeutic decisions.

Real-Time Applications

In parallel, and given the increasing importance of continuous monitoring and remote patient care, smart health devices are being reshaped by AI agents that can process and act upon health data in real time.[11] For instance, with the expansion of wearable devices and home monitoring tools, telehealth visits increasingly rely on data beyond patient-reported symptoms. AI agents can synthesize streams from blood pressure cuffs, glucose monitors, and fitness trackers into concise summaries that highlight trends, outliers, and alerts, presenting this information in an actionable format that supports more personalized recommendations while preventing clinicians from being overwhelmed by raw data.[12]

Building on this capacity for real-time synthesis, a particularly promising innovation is the agentic health wallet—a convergence of digital health management and AI. Anchored in the self-sovereign patient paradigm of Healthcare 4.0,[13] these wallets serve as secure interfaces for data exchange between patients and providers. Enhanced with agentic capabilities, prototypes such as Pluxee and Inrupt health wallets extend beyond passive record-keeping to actively manage treatment, deliver personalized recommendations, and facilitate consent and data sharing across care settings.[14] Together, these developments are redefining how health data are collected, synthesized, and translated into actionable care across the patient journey.

Changing Perceptions

As these examples suggest, AI agents are increasingly envisioned not merely as clinical tools but as "**colleagues"** within healthcare teams.[15] Their integration raises important considerations around collaboration, delegation, and trust, particularly as they assume roles in workflow coordination and patient interaction. This evolution invites a re-examination of the ethical foundations of medicine. While the Hippocratic Oath emphasized patient welfare, confidentiality, and non-maleficence, today's context requires us to uphold these commitments while also considering the role played by non-human actors who cannot themselves assume moral responsibility.[16]

This shift demands a bioethical framework attentive to core medical principles—autonomy, beneficence, non-maleficence, and justice—to ensure these agents support patient rights, clinical benefit, safety, and fairness. For patients, trust will likely be a decisive factor in whether agentic AI becomes an accepted partner in care.[17] Unlike traditional clinical tools, AI agents will increasingly interact directly with patients, shaping their understanding of illness, treatment, and consent. Transparency about AI agents' capabilities, limitations, and decision-making processes must therefore be embedded in patient interactions to ensure informed consent and uphold these ethical imperatives.

The Next Frontier

Looking ahead, the emergence of multi-agent systems marks the next frontier in the evolution of AI agents in healthcare.[18] This will involve multiple AI agents interacting with one another and with clinicians to coordinate across specialties, synchronize data streams, and negotiate priorities in real time. For instance, a multiagent conversation framework was developed by Chen and colleagues to be used in clinical practice to perform diagnostic tasks, highlighting the potential of multiagent systems to simulate clinical teamwork in healthcare settings.[19] Moritz and colleagues describe decentralized multi-agent systems for healthcare (MASH) as distributed networks of expertise, with agents specializing in areas such as diagnostics, treatment planning, or patient monitoring. Each agent would be trained on domain-specific datasets—for example, a radiology agent limited to imaging or a medication management agent restricted to prescribing data.[20] When additional information is required, agents could request controlled access through authentication within the trusted MASH network, thereby enabling personalized, high-quality care.

This decentralized approach to multi-agent systems addresses a critical challenge in AI healthcare deployment: the risk of perpetuating bias through centralized data and decision-making. Singh and colleagues at the MIT Media Lab at the Massachusetts Institute of Technology argue that distributed AI architectures can mitigate bias by drawing on heterogeneous datasets and enabling evaluation across diverse contexts.[21] Rather than consolidating data, compute resources, and governance in single entities, decentralized AI enables collaboration across organizations while preserving data ownership and privacy. This reinforces the MASH framework's potential, suggesting that specialized agents operating across different institutions could not only maintain data privacy but also reduce algorithmic bias through their distributed, heterogeneous nature. Together, these developments suggest a future in which AI agents not only collaborate with clinicians but also with one another, forming adaptive networks that expand human capacity for care.

Governance

As the scope of healthcare AI agents expands from discrete use cases toward more interconnected and adaptive roles, new governance models capable of ensuring accountability, transparency, and alignment are essential.[22,23]

Here, technical capability and clinical integration must be matched by adaptive oversight and flexible regulatory mechanisms.[24] Different jurisdictions have already signaled diverging philosophies in this regard. In the United States, as outlined in the recent AI Action Plan, a "try-first" orientation emphasizes rapid deployment and iterative learning, privileging innovation and speed of adoption.[25] Yet this orientation does not exempt healthcare applications from stringent compliance requirements. Agentic AI systems that process patient data must adhere to existing regulatory frameworks, most notably the Health Insurance Portability and Accountability Act (HIPAA), which governs the handling of Protected Health Information (PHI). Ensuring compliance with HIPAA entails rigorous safeguards for data privacy, security, and access control.[26]

By contrast, the European Union has formalized a governance-first approach through the AI Act, which requires comprehensive risk management prior to deployment. Its risk-based classification framework assigns AI applications to minimal, limited, high, or unacceptable categories, with healthcare systems generally designated as "high risk."[27] This designation triggers stringent safeguards, including transparency documentation, interpretable decision pathways, reliance on representative and bias-mitigated datasets, and mandatory human oversight. These requirements respond directly to the challenges posed by advanced general-purpose models, particularly LLMs, which often function as opaque black boxes.

A forward-looking approach to governance emphasizes the principle of governance by design,[28] whereby directives are embedded throughout the lifecycle of agent development, validation, deployment, and monitoring. Concretely, governance by design should address six interrelated domains: (1) rigorous testing protocols with clearly defined scope of practice, (2) transparency in decision-making pathways and clear delineation of liability, (3) systematic mitigation of bias together with measures to ensure accessibility, (4) comprehensive safeguards for privacy and cybersecurity, (5) clinical oversight to ensure AI augments rather than replaces clinical judgment, and (6) continuous monitoring to detect errors, drift, or unintended consequences. This framework ensures that patient health information is protected in compliance with HIPAA and other regulations, AI algorithms are trained on diverse datasets to avoid disparities in care delivery, and AI outputs remain explainable and actionable for both providers and patients, thereby reinforcing trust and accountability in clinical practice.

Yet governance cannot remain limited to the oversight of individual agents. As agents are increasingly deployed in coordinated or competitive multi-agent settings, the focus must extend to the dynamics of interaction.[29,30] Emergent behavior at the system level may amplify risks even when each individual agent behaves within its intended parameters.[31] For example, two agents may independently act safely but, when interacting, reinforce diagnostic bias or generate conflicting recommendations that undermine clinician trust. These dynamics demand systematic study within a nascent field that is being termed the sociology of humans and machines, which examines how autonomous systems and human actors co-construct patterns of coordination, conflict, and meaning.[32] Governance therefore requires a dual orientation: ensuring reliability of components while also monitoring and constraining the collective patterns that arise when agents operate together and in human-agent teams.

For hospitals, startups, and regulators, the governance challenge is immediate. Hospitals and telehealth providers must embed governance checkpoints into procurement and deployment, asking vendors to demonstrate scope of practice, liability pathways, and evidence of bias mitigation. Startups should adopt governance by design as a market differentiator, building transparency, patient-facing consent features, and monitoring tools into their products from the outset. Regulators, meanwhile, need to go beyond static compliance frameworks and invest in adaptive oversight that can keep pace with the learning and interaction patterns of multi-agent systems. Practical steps include developing standardized auditing protocols,[33] establishing registries for certified healthcare agents,[34] and requiring reporting of adverse events linked to AI-mediated care. The high-stakes, tightly regulated nature of medicine ensures that governance innovations developed here can inform broader frameworks for the agentic web, a distributed, interactive internet ecosystem in which AI agents persistently plan, coordinate, collaborate, and execute goal-directed tasks.[35,36] Governance thus emerges as a foundational mechanism for building trust in autonomous systems that will increasingly shape the future of care.

Funding

None.

Conflicts of Interest

The authors have no conflicts of interest to report.

Contributors

All authors contributed to the conceptualization of the manuscript at various stages of its development. Tomer Jordi Chaffer drafted the initial manuscript, provided governance strategies, and integrated feedback from co-authors. Dr. Joe Littlejohn contributed key clinical insights into the challenges of integrating AI agents into clinical workflows. Dr. Muthu Ramachandran provided technical and strategic advice on the integration of AI agents into healthcare systems. Dr. Claudia Lamschtein provided a human-centered commentary on the ethical

dilemmas of integrating AI agents into clinical practice. All authors reviewed, revised, and approved the final version of the manuscript for publication.

Data Availability Statement (DAS), Data Sharing, Reproducibility, and Data Repositories

This editorial does not contain any primary data analysis or new datasets.

Application of AI-Generated Text or Related Technology

ChatGPT (GPT-4o model) was used to assist with grammatical correction, spelling, and editorial refinement throughout the preparation of this manuscript. Claude was used to support adherence to the Vancouver citation style.

Acknowledgments

The authors thank the BHTY Corporate Advisory Council members for their support and encouragement.

References

1. Jing AB, Garg N, Zhang J, Brown JJ. AI solutions to the radiology workforce shortage. npj Health Syst. 2025;2(1):20. https://doi.org/10.1038/s44401-025-00023-6
2. Jiang Y, Black KC, Geng G, Park D, Zou J, Ng AY, et al. MedAgentBench: a virtual EHR environment to benchmark medical LLM agents. NEJM AI. 2025;2:AIdbp2500144. https://doi.org/10.1056/AIdbp2500144
3. Karunanayake N. Next-generation agentic AI for transforming healthcare. Inform Health. 2025;2(2):73–83. https://doi.org/10.1016/j.infoh.2025.03.001
4. Lee YC. Rethinking artificial intelligence in medicine: from tools to agents. Clin Exp Emerg Med. 2025;12(2):101–3. https://doi.org/10.15441/ceem.25.125
5. Qiu J, Lam K, Li G, Acharya A, Wong TY, Darzi A, et al. LLM-based agentic systems in medicine and healthcare. Nat Mach Intellig. 2024;6(12):1418–20. https://doi.org/10.1038/s42256-024-00944-1
6. Ahmed Kamal M, Ismail Z, Shehata IM, Djirar S, Talbot NC, Ahmadzadeh S, et al. Telemedicine, e-health, and multi-agent systems for chronic pain management. Clin Pract. 2023;13(2):470–82. https://doi.org/10.3390/clinpract13020042
7. Tu T, Schaekermann M, Palepu A, Saab K, Freyberg J, Tanno R, et al. Towards conversational diagnostic artificial intelligence. Nature. 2025;642:442–50. https://doi.org/10.1038/s41586-025-08866-7
8. Lee C, Vogt KA, Kumar S. Prospects for AI clinical summarization to reduce the burden of patient chart review. Front Digit Health. 2024;6:1475092. https://doi.org/10.3389/fdgth.2024.1475092
9. Adams SJ, Acosta JN, Rajpurkar P. How generative AI voice agents will transform medicine. npj Digit Med. 2025;8(1):353. https://doi.org/10.1038/s41746-025-01776-y
10. Leng Y, He Y, Amini S, Magdamo C, Paschalidis I, Mukerji SS, et al. A GPT-4o-powered framework for identifying cognitive impairment stages in electronic health records. npj Digit Med. 2025;8(1):401. https://doi.org/10.1038/s41746-025-01834-5
11. Borkowski AA, Ben-Ari A. Multiagent AI systems in health care: envisioning next-generation intelligence. Fed Pract. 2025;42(5):188–94. https://doi.org/10.12788/fp.0589
12. González-Rivas JP, Seyedi SA, Mechanick JI. Artificial intelligence enabled lifestyle medicine in diabetes care: a narrative review. Am J Lifestyle Med. 2025:15598276251359185. https://doi.org/10.1177/15598276251359185
13. Chaffer TJ, Littlejohn J, Nadarasa A, Lamschtein C. The self-Sovereign patient as a cornerstone of healthcare 4.0. Blockchain Healthc Today. 2025;8(2):414. https://doi.org/10.30953/bhty.v8.414
14. Inrupt. Pluxee & Inrupt partner to transform employee healthcare benefits [Internet]. Inrupt.com. [cited 2025 Aug 22]. Available from: https://www.inrupt.com/case-study/pluxee-redefines-healthcare-benefits-with-healthy-lifestyle-data-wallet
15. Zou J, Topoi EJ. The rise of agentic AI teammates in medicine. Lancet. 2025;405(10477):457. https://doi.org/10.1016/S0140-6736(25)00202-8
16. Felländer-Tsai L. AI ethics, accountability, and sustainability: revisiting the Hippocratic oath. Acta Orthopaed. 2020;91(1):1–2. https://doi.org/10.1080/17453674.2019.1682850
17. Juravle G, Boudouraki A, Terziyska M, Rezlescu C. Trust in artificial intelligence for medical diagnoses. Prog Brain Res. 2020;253:263–82. https://doi.org/10.1016/bs.pbr.2020.06.006
18. Chen L, Zhang Y, Feng J, Chai H, Zhang H, Fan B, et al. AI agent behavioral science. arXiv preprint arXiv:2506.06366. 2025.
19. Chen X, Yi H, You M, Liu W, Wang L, Li H, et al. Enhancing diagnostic capability with multi-agents conversational large language models. NPJ Digit Med. 2025;8(1):159. https://doi.org/10.1038/s41746-025-01550-0
20. Moritz M, Topol E, Rajpurkar P. Coordinated AI agents for advancing healthcare. Nat Biomed Eng. 2025;9:432–8. https://doi.org/10.1038/s41551-025-01363-2
21. Singh A, Lu C, Gupta G, Behari N, Chopra A, Blanc J, et al. A perspective on decentralizing AI. MIT Media Lab; 2025.
22. Chaffer TJ, Goins CV, II, Okusanya B, Cotlage D, Goldston J. Decentralized governance of autonomous AI agents. arXiv preprint arXiv:2412.17114. 2024.
23. Taylor RA. AI agents, automaticity, and value alignment in health care. NEJM AI. 2025;2(8):AIp2401165. https://doi.org/10.1056/AIp2401165
24. Freyer O, Jayabalan S, Kather JN, Gilbert S. Overcoming regulatory barriers to the implementation of AI agents in healthcare. Nat Med. 2025;31:3239–43. https://doi.org/10.1038/s41591-025-03841-1
25. United States. AI action plan [Internet]. 2025. Available from: https://www.whitehouse.gov/wp-content/uploads/2025/07/Americas-AI-Action-Plan.pdf
26. Neupane S, Mittal S, Rahimi S. Towards a HIPPA compliant agentic AI system in healthcare. arXiv preprint arXiv:2504.17669. 2025.
27. Borrelli M, Musch S, Kohn B, Mishra A, Paul S, Chaffer TJ, et al. EU AI act, harnessing digital-twin technology: how the innovative AI legislation provides regulatory safeguards to support the synergy between digital twin technology and general purpose AI systems. SSRN Elect J. 2025. https://doi.org/10.2139/ssrn.5357031
28. Joshi H. AI governance by design for agentic systems: a framework for responsible development and deployment. 2025. Preprints.org.

29. de Witt CS. Open challenges in multi-agent security: towards secure systems of interacting AI agents. arXiv preprint arXiv:2505.02077. 2025.
30. Ashery AF, Aiello LM, Baronchelli A. Emergent social conventions and collective bias in LLM populations. Sci Adv. 2025;11(20):eadu9368. https://doi.org/10.1126/sciadv.adu9368
31. Chen YJ, Albarqawi A, Chen CS. Reinforcing clinical decision support through multi-agent systems and ethical AI governance. arXiv preprint arXiv:2504.03699. 2025.
32. Tsvetkova M, Yasseri T, Pescetelli N, Werner T. A new sociology of humans and machines. Nat Hum Behav. 2024;8(10):1864–76. https://doi.org/10.1038/s41562-024-02001-8
33. Hinostroza Fuentes VG, Karim HA, Tan MJ, AlDahoul N. AI with agency: a vision for adaptive, efficient, and ethical healthcare. Front Dig Health. 2025;7:1600216. https://doi.org/10.3389/fdgth.2025.1600216
34. Singh A, Ehtesham A, Raskar R, Lambe M, Chari P, Grogan JJ, et al. A survey of AI agent registry solutions. arXiv preprint arXiv:2508.03095. 2025.
35. Yang Y, Ma M, Huang Y, Chai H, Gong C, Geng H, et al. Agentic web: weaving the next web with AI agents. arXiv preprint arXiv:2507.21206. 2025.
36. Chaffer TJ. Know your agent: governing AI identity on the agentic web. SSRN Elect J. 2025. https://doi.org/10.2139/ssrn.5162127