International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Human-AI Collaboration in Healthcare: A Review Update, Research Gaps, and Future Agenda

Dr. Madhavi Satish Avhankar
School of Information Technology, Indira University, Pune, India

***Abstract*—Human-AI collaboration in healthcare has continued to develop rapidly since Lai, Kankanhalli, and Ong (2021) published their foundational review and research agenda, which examined 28 papers on the topic through mid-2020 and identified a need for more research on the evolving collaboration process, human-AI complementarity, adoption and perception, and the long-term organizational impact of such collaboration. In the years since, generative AI and large language models have entered clinical workflows, new empirical work has directly examined trust formation among clinicians and patients, and further reviews have synthesized adjacent literatures such as explainable AI (XAI) and AI adoption barriers. This paper revisits Lai et al.'s (2021) research agenda in light of this more recent literature. We summarize what has and has not been addressed since 2021, synthesize findings from ten studies published between 2017 and 2025, and propose an updated set of research priorities for human-AI collaboration in healthcare, with particular attention to trust, generative AI, and multi-stakeholder collaboration.**

***Keywords*—Artificial Intelligence (AI), Healthcare, Human-AI Collaboration, Trust, Literature Review.**

I. INTRODUCTION

Artificial intelligence (AI) systems can now learn from data, identify patterns, and support or make decisions across many domains, including healthcare. As Lai, Kankanhalli, and Ong (2021) argue in their review of the field, healthcare is a particularly important context for human-AI collaboration: it faces a persistent global shortage of physicians, and medical professionals are routinely overworked in ways that can compromise care quality. Human-AI collaboration — in which AI systems work jointly with clinicians as partners rather than simply automating routine tasks — has been proposed as one way to mitigate these pressures, for example through clinical decision support systems (CDSS) that collaborate with physicians on diagnosis.

Lai et al. (2021) conducted the first systematic review to synthesize this literature specifically through the lens of human-AI collaboration (as distinct from AI adoption or automation more broadly), reviewing 28 papers identified from an initial pool of 1,019 records across five major databases and the IS Senior Scholars' Basket of Eight journals.

Their review identified several important gaps: limited research on how human-AI collaboration evolves over time, on the specific competencies required for effective collaboration, on how healthcare professionals perceive AI as a teammate versus a tool, and on collaboration involving three or more parties (e.g., AI, physicians, and patients together). They called, above all, for more theory-driven research on human-AI collaboration in healthcare.

Since that review was conducted, the field has moved quickly — most notably with the emergence of large language models (LLMs) and generative AI in clinical settings, which were not yet part of the literature Lai et al. (2021) could review. This paper has two goals. First, we revisit their research agenda and ask which of the gaps they identified have since been addressed, using literature published between 2021 and 2025. Second, we synthesize this more recent literature, alongside foundational work it builds on, to propose an updated set of priorities for future research.

II. RELATED WORK: THE LAI ET AL. (2021) REVIEW AND WHAT HAS FOLLOWED

Lai et al. (2021) is the primary foundation for this paper. Reviewing literature from computer science, information systems, health informatics, and medicine, they found that while a growing number of studies addressed AI in healthcare, comparatively few addressed human-AI collaboration specifically — that is, cases where AI and human clinicians actively and reciprocally work toward a shared goal, rather than AI simply automating a task or serving as a passive tool. They organized their discussion of trust in healthcare AI around six factors: accuracy and reliability, explainability of decisions, integration with clinical workflow, ethical and privacy considerations, collaborative decision-making, and security of patient data (Lai et al., 2021, Figure 1). We adopt this framework in Section IV below, since it remains, to our knowledge, the most complete synthesis of trust factors specific to human-AI collaboration (as opposed to AI adoption generally) in healthcare.

Several literatures have developed in parallel or since. Vedula and Hager (2017) had earlier introduced "surgical data science" as a field explicitly reframing AI-enabled surgical tools as active collaborators rather than passive

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

instruments — a framing Lai et al. (2021) drew on. Pacis, Subido, and Bugtai (2018) reviewed AI applications in telemedicine, identifying collaborative information analysis as one of four major trends. Miner, Shah, Bullock, Arnow, Bailenson, and Hancock (2019) proposed a related but distinct framework specific to mental health, outlining four models of AI-human integration in psychotherapy delivery, from fully human-delivered care to fully AI-delivered care, and argued that trust, safety, and oversight were prerequisite considerations — a theme Lai et al. (2021) later generalized across healthcare.

More recent reviews have picked up threads Lai et al. (2021) left open. Gomez, Cho, Ke, Huang, and Unberath (2024) conducted a systematic review of interaction patterns in AI-assisted decision-making, including healthcare applications, and found that most existing systems remain largely one-directional — the AI provides output and the human accepts or rejects it — rather than genuinely collaborative, echoing Lai et al.'s (2021) observation that truly interactive collaboration was still rare. Prentzas, Kakas, and Pattichis (2023) systematically reviewed 198 explainable AI (XAI) papers in medicine and found that physician-in-the-loop evaluation remains uncommon, despite explainability being central to Lai et al.'s (2021) trust framework. Nguyen, Churpek, and Wiegmann (2025) reviewed 75 empirical studies on clinician trust in AI and explicitly called for longitudinal research on how trust changes over time — directly addressing one of the specific gaps Lai et al. (2021) identified. Hassan, Kushniruk, and Borycki (2024) conducted a scoping review of AI adoption barriers and facilitators in healthcare more broadly, finding trust to be a central catalyst of adoption. Nong and Platt (2025) surveyed a nationally representative sample of US adults and found generally low patient trust in health systems' use of AI — extending Lai et al.'s (2021) clinician-focused trust discussion to the patient perspective. Finally, Wang and Liu (2025) reviewed human-machine collaboration specifically within Chinese healthcare settings, highlighting regional and regulatory factors not addressed in Lai et al.'s (2021) largely US/Singapore-oriented review.

III. What Has Changed Since 2021

The most significant development since Lai et al.'s (2021) review is the emergence of large language models (LLMs) and generative AI as tools directly used in clinical workflows — for documentation, triage support, and diagnostic reasoning assistance.

This did not exist in a mature form in the literature Lai et al. (2021) could review, and it introduces new questions about collaboration that their original six-factor trust framework did not anticipate in detail: for example, how clinicians should calibrate trust in a system whose outputs can vary across otherwise identical queries, and how to maintain appropriate skepticism toward fluent but sometimes incorrect AI-generated text.

Beyond generative AI specifically, the studies reviewed above show incremental but real progress on several of Lai et al.'s (2021) original open questions. Nguyen et al. (2025) and Nong and Platt (2025) together represent meaningful movement on the trust-related gaps Lai et al. (2021) identified, extending the discussion from clinicians alone to patients as well. Gomez et al. (2024) and Prentzas et al. (2023) provide more systematic evidence — across hundreds of studies combined — for a claim Lai et al. (2021) made from a smaller sample: that most human-AI systems in healthcare remain more automated than collaborative in practice. At the same time, several of Lai et al.'s (2021) other open questions — particularly longitudinal study of collaboration over time, the specific competencies required of healthcare professionals, and collaboration involving three or more parties simultaneously — remain comparatively underexplored in the literature identified for this review, and are carried forward into Section VII below.

IV. Trust In Human-AI Coordination

Figure 1 reproduces, with light adaptation, the six-factor trust framework developed by Lai et al. (2021), which we treat as the starting point for this section rather than an original contribution of this paper. Recent empirical work provides additional support for several of these factors specifically.

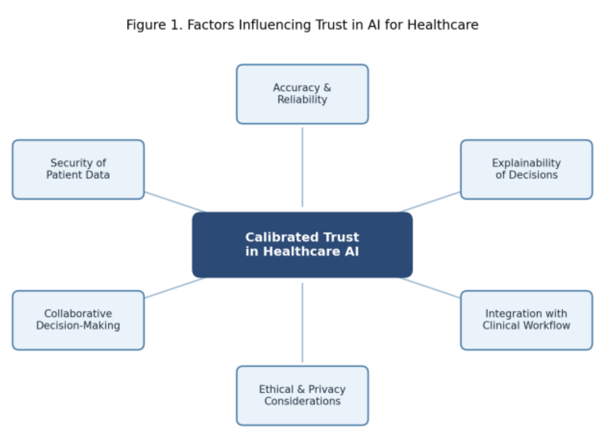

Figure 1. Factors influencing trust in healthcare AI (adapted from Lai, Kankanhalli, & Ong, 2021).

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

1) Accuracy and Reliability: Consistent with this factor, Gomez et al. (2024) found that AI systems perceived as unreliable across similar cases undermine clinician willingness to engage collaboratively rather than simply deferring to or overriding the system.

2) Explainability of Decisions: Prentzas et al. (2023) found that while explainability techniques are increasingly common in medical AI, few studies evaluate whether these explanations actually improve physician understanding or trust in practice — suggesting this factor remains more aspirational than demonstrated.

3) Integration with Clinical Workflow: Hassan et al. (2024) identified workflow disruption as one of the most frequently cited barriers to AI adoption across the 50 studies in their scoping review, reinforcing the centrality of this factor.

4) Ethical and Privacy Considerations: This factor remains consistently cited across the literature reviewed here, though it was not a primary focus of any single study synthesized in this update.

5) Collaborative Decision-Making: Gomez et al.'s (2024) taxonomy of interaction patterns speaks directly to this factor, distinguishing systems that merely present recommendations from those that support genuine back-and-forth exchange — finding the latter still rare.

6) Security of Patient Data: Nong and Platt (2025) found that concerns about data misuse were among the strongest predictors of low patient trust in health systems' AI use, extending this factor's relevance from providers to patients specifically.

V. METHODOLOGY

This paper is a narrative update review rather than a new formal systematic review. We began from Lai et al.'s (2021) systematic review, which searched INFORMS PubsOnline, the AIS Electronic Library, PubMed, Scopus, ACM Digital Library, and the IS Senior Scholars' Basket of Eight journals for literature through June 2020. To identify literature published since, we conducted targeted searches for recent (2021–2025) empirical studies, reviews, and reference works addressing human-AI collaboration, trust, and adoption in healthcare, prioritizing peer-reviewed journal articles and systematic/scoping reviews where available. Search terms combined AI-related keywords (e.g., "artificial intelligence," "machine learning," "large language model") with collaboration- and trust-related terms (e.g., "human-AI collaboration," "clinician trust," "adoption") and healthcare-related terms.

This process identified ten studies (summarized in Table 1) that either directly extend Lai et al.'s (2021) research agenda or provide empirical evidence relevant to their six-factor trust framework. This is a much smaller and less exhaustive set than a full systematic review would produce, and should be read as an illustrative update rather than a comprehensive survey of all literature published since 2021.

VI. FINDINGS FROM THE LITERATURE UPDATE

Table 1 summarizes the ten studies reviewed for this update, spanning 2017 to 2025.

TABLE I
STUDIES REVIEWED FOR THIS UPDATE (IN ADDITION TO LAI ET AL., 2021)

Study	Domain	Method	Key Contribution
Lai, Kankanhalli, & Ong (2021)	Cross-domain (foundational review)	Systematic literature review, n=28 (through June 2020)	Found limited research on the evolving collaboration process, human-AI complementarity, adoption/perception, and long-term organizational impact; called for more theory-driven research.
Miner et al. (2019)	Mental health / psychotherapy	Conceptual / viewpoint	Proposed four models of AI-human integration in psychotherapy delivery (Human Only → AI Only) and argued that safety, trust, and oversight are prerequisite design considerations.
Vedula & Hager (2017)	Surgery	Conceptual / field-defining article	Introduced "surgical data science" as a discipline, framing AI-enabled surgical tools as interactive collaborators rather than passive instruments.
Pacis, Subido, & Bugtai (2018)	Telemedicine	Narrative review	Identified four trends in AI-enabled telemedicine (patient monitoring, health IT, intelligent diagnosis assistance, and collaborative information analysis).
Gomez, Cho, Ke, Huang, & Unberath (2024)	Cross-domain incl. clinical decision-making	Systematic review of interaction patterns	Found that most existing human-AI decision-support tools, including in healthcare, remain largely one-directional; genuinely bidirectional collaboration is still rare.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

Study	Domain	Method	Key Contribution
Prentzas, Kakas, & Pattichis (2023)	Medical imaging / CDSS	Systematic review, n=198	Found model-agnostic explainability techniques dominate current XAI solutions, but that physician-in-the-loop evaluation and real collaboration between clinical and AI experts remain uncommon.
Nguyen, Churpek, & Wiegmann (2025)	Cross-specialty (trust)	Review of empirical research, n=75	Found most existing studies address process- and performance-related trust rather than purpose-related trust, and called for longitudinal research on how trust changes over time — directly echoing Lai et al.'s (2021) call.
Hassan, Kushniruk, & Borycki (2024)	Cross-domain (adoption)	Scoping review, n=50	Identified about 18 categories of barriers and facilitators to AI adoption in healthcare, organized around development, implementation, and governance factors; found trust to be a central catalyst of adoption.
Nong & Platt (2025)	Cross-domain (patient trust)	National survey (US adults)	Found that a majority of surveyed adults reported low trust in their health system's ability to use AI responsibly, and that overall institutional trust predicted AI-specific trust.
Wang & Liu (2025)	Cross-domain (China)	Narrative review	Surveyed recent applications of human-machine collaboration across Chinese healthcare settings, highlighting regional/regulatory factors shaping adoption not addressed in Lai et al. (2021).

Two patterns stand out. First, several studies (Gomez et al., 2024; Prentzas et al., 2023) provide larger-scale, more recent evidence for a claim Lai et al. (2021) made from a smaller literature: that human-AI systems in healthcare remain more automated than genuinely collaborative in practice, five years on. Second, the trust literature has diversified since 2021 — from a clinician-centered discussion in Lai et al. (2021) to work explicitly addressing patient trust (Nong & Platt, 2025) and cross-national/regulatory variation (Wang & Liu, 2025), both of which extend rather than contradict the original framework.

VII. Research Agenda: Building On Lai Et Al. (2021)

Lai et al. (2021) proposed several directions for future research. Below, we revisit each in light of the literature identified for this update.

A. Longitudinal Study of Collaboration Over Time

Largely still open. Nguyen et al. (2025) explicitly call for longitudinal research on how clinician trust evolves with experience, echoing Lai et al. (2021), but note this remains rare in the empirical literature they reviewed.

B. Human-AI Complementarity and Collaborative Outcomes

Partially addressed. Gomez et al.'s (2024) taxonomy provides a more systematic account of how (and how rarely) current systems achieve genuine complementarity, but direct outcome studies comparing collaborative versus non-collaborative designs remain limited in the literature reviewed here.

C. Competencies Required for Collaboration

Still largely open. None of the studies identified for this update directly examine what training or skills healthcare professionals need to collaborate effectively with AI systems — this remains a clear gap consistent with Lai et al. (2021).

D. Perceptions of AI as Teammate vs. Tool

Emerging evidence. Miner et al.'s (2019) framework of AI-human integration models and Gomez et al.'s (2024) interaction taxonomy both speak to this question, though neither directly measures clinician perceptions of AI as a teammate versus a tool in the way Lai et al. (2021) proposed.

E. Multi-Stakeholder Collaboration

Still open. None of the studies identified for this update examine collaboration involving three or more parties (e.g., AI, physician, and patient) simultaneously; this remains one of the clearest open gaps from Lai et al. (2021).

F. Theory-Driven Research and Generative AI

We add one direction not present in Lai et al. (2021): as generative AI and LLMs enter clinical workflows, future research should examine whether existing human-AI collaboration theory — developed largely around deterministic decision-support tools — still applies to systems whose outputs vary across similar inputs, and whether new theoretical constructs are needed.

VIII. Conclusion

This paper revisited Lai, Kankanhalli, and Ong's (2021) foundational review of human-AI collaboration in healthcare in light of literature published since.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435 (Online) Volume 15, Issue 08, August 2026)

We find real but partial progress: trust research has extended from clinicians to patients, larger-scale reviews have reinforced (rather than overturned) the claim that most healthcare AI systems remain more automated than collaborative, and new work has emerged on adoption barriers and cross-national variation. At the same time, several of Lai et al.'s (2021) original open questions — particularly longitudinal study of collaboration, the competencies required of healthcare professionals, and multi-stakeholder collaboration — remain largely unaddressed, and the emergence of generative AI introduces new questions not anticipated in their original framework. We hope this update, and the research agenda in Section VII, offer a useful reference point for researchers building on Lai et al.'s (2021) work.

REFERENCES

[1] Gomez, C., Cho, S. M., Ke, S., Huang, C.-M., & Unberath, M. (2024). Human-AI collaboration is not very collaborative yet: A taxonomy of interaction patterns in AI-assisted decision making from a systematic review. Frontiers in Computer Science, 6, Article 1521066. https://doi.org/10.3389/fcomp.2024.1521066

[2] Hassan, M., Kushniruk, A., & Borycki, E. (2024). Barriers to and facilitators of artificial intelligence adoption in health care: Scoping review. JMIR Human Factors, 11, e48633. https://doi.org/10.2196/48633

[3] Lai, Y., Kankanhalli, A., & Ong, D. (2021). Human-AI collaboration in healthcare: A review and research agenda. In Proceedings of the 54th Hawaii International Conference on System Sciences (pp. 1–10). https://doi.org/10.24251/HICSS.2021.046

[4] Miner, A. S., Shah, N., Bullock, K. D., Arnow, B. A., Bailenson, J., & Hancock, J. (2019). Key considerations for incorporating conversational AI in psychotherapy. Frontiers in Psychiatry, 10, Article 746. https://doi.org/10.3389/fpsyt.2019.00746

[5] Nguyen, O. T., Churpek, M. M., & Wiegmann, D. A. (2025). Factors associated with purpose-, process-, and performance-related trust for artificial intelligence technologies for clinical practice among healthcare professionals: A review of empirical research. Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care, 14(1), 33. https://doi.org/10.1177/2327857925141009

[6] Nong, P., & Platt, J. (2025). Patients' trust in health systems to use artificial intelligence. JAMA Network Open, 8(2), Article e2460628. https://doi.org/10.1001/jamanetworkopen.2024.60628

[7] Pacis, D. M. M., Subido, E. D. C., Jr., & Bugtai, N. T. (2018). Trends in telemedicine utilizing artificial intelligence. AIP Conference Proceedings, 1933(1), Article 040009. https://doi.org/10.1063/1.5023979

[8] Prentzas, N., Kakas, A., & Pattichis, C. S. (2023). Explainable AI applications in the medical domain: A systematic review (arXiv:2308.05411). arXiv. https://arxiv.org/abs/2308.05411

[9] Vedula, S. S., & Hager, G. D. (2017). Surgical data science: The new knowledge domain. Innovative Surgical Sciences, 2(3), 109–121. https://doi.org/10.1515/iss-2017-0004

[10] Wang, W., & Liu, L. (2025). Advances in the application of human-machine collaboration in healthcare: Insights from China. Frontiers in Public Health, 13, Article 1507142. https://doi.org/10.3389/fpubh.2025.1507142

[11] Avhankar, M. S., Pawar, D. J. A., Majalekar, S., & Kedari, S. (2022). Mobile ad hoc network routing protocols using opnet simulator. International Journal on Recent and Innovation Trends in Computing and Communication, 10(1), 1-7.

[12] Pawar, J. A., Avhankar, M. S., Gupta, A., Barve, A., Patil, H., & Maranan, R. (2024, May). Enhancing network security: Leveraging isolation forest for malware detection. In 2024 2nd International conference on advancement in computation & computer technologies (InCACCT) (pp. 230-234). IEEE.

[13] Dey, R., Kassim, S., Mahajan, R. A., Avhankar, M., Dharwadkar, R., Kumbhar, V., & Shendage, J. (2025). Applications of Machine Learning in Islamic Finance. Journal of Information Systems Engineering and Management, 10(26s), 794-804.

[14] Avhankar, M., & Pawar, J. (2023). An experimental comparison of AODV and DSDV routing protocols in MANET. International Journal of Scientific and Research Publications, 13(3).

[15] Kumbhar, V., Keskar, V., Avhankar, M. S., Yarde, P., Walhhekar, R., & Kumbhar, S. A. (2025). Reconnoitering blockchain consensus mechanisms. In Fostering Machine Learning and IoT for Blockchain Technology: Smart Cities Applications, Volume 1 (pp. 65-114). Singapore: Springer Nature Singapore.