PATIENT PERCEPTION TOWARDS ARTIFICIAL INTELLIGENCE IN HEALTHCARE: TRUST, ACCEPTANCE, AND ETHICAL

Garima Pathania[1]* &Dr. Ashok Bansal[2]

*[1]*Assistant Professor, Chitkara College of Sales & Marketing, Chitkara University, Punjab - 140417* ***Email:*** *garima.pathania16@outlook.com*
[2] Assistant Professor, Department of Management, CDOE, Himachal Pradesh University, Shimla – 171005 ***Email:*** *ashokkumarbansal2009@gmail.com*

ABSTRACT
Artificial intelligence (AI) is becoming a regularly used component in healthcare models around the globe and transforming the processes of clinical decision-making, diagnostics, patient monitoring, and administration. Regardless of the rapid development of technology, the introduction of AI into the healthcare industry is highly dependent on patients, their views and trust. The current paper is a review of the literature and a synthesis of the existing empirical data related to the subject of patient perception of AI in the medical facility. Based on the researches made with various demographic groups and in different health settings, the paper will explore determinants that promote patient acceptance or resistance such as transparency, explainability, cultural sensitivity, existing knowledge of technology, and privacy and data security concerns. The analysis also discusses how patient autonomy and algorithmic decision-making interact, how the collaboration between physicians and AI affects patient trust, and how healthcare organizations can act ethically concerning the responsible implementation of AI. The results show that although patients recognize the opportunities of AI to enhance the process of healthcare delivery, the lack of trust, insufficient health literacy, and dehumanization fears do not allow the widespread acceptance of AI. Implications on healthcare policy, the design of human-centered AI systems, and recommendations on improving patient engagement in AI transition are presented at the end of the paper.

Keywords: *artificial intelligence, patient perception, healthcare AI, trust, acceptance, ethics, digital health, physician-AI collaboration, explainability, health equity*

Received: 14/03/2026 *Revised: 24/03/2026* *Acceptance: 02/05/2026* *Publication: 09/05/2026*

1. INTRODUCTION

The artificial intelligence and healthcare sectors have come together, which has created breakthrough opportunities previously found in science fiction. Machine learning algorithms are currently helping radiologists locate early tumors, AI-based natural language processors find meaningful patterns in electronic health records, and patients are now being triaged with AI-powered chatbots which offer twenty-four-hour patient support. Such advances can minimise diagnostic errors, streamline treatment pathways, and democratise care provision at the level of expertise of specialists, in geographies heretofore underserved.

Nevertheless, the technical potential of AI systems and its embracement among the patients are two different realities. Even the clinically proven AI tool may not serve its purpose if the patients are not ready to interact with it, have confidence in its results, or allow their personal health information to be used in its training. The human aspect of AI implementation in healthcare is, thus, not peripheral to the success of any implementation plan, but central.

The perception of patients is a wide range of cognitive, emotional, and social reactions to AI-based healthcare services. It consists of patient perceptions of AI, the extent to which they believe in algorithmic suggestions, their perception of comfort of data sharing, their expectations of physician interaction, and their perceptions of fairness, accountability, and compassion in care delivered by AI. These perceptions are also important to understand by clinicians, technologists, and policymakers who may want to implement AI solutions that are effective as well as accepted and valued by the populations that they will serve.

The paper will be a synthesis of existing literature on AI perception in healthcare by patients. It is structured into five key themes: (1) the conceptual model of patience to medical technology adoption; (2) influences on trust and distrust in AI system; (3) demographic and cultural differences in AI perception; (4) the aspects of transparency, explainability, and the role of physicians; and (5) ethical and privacy issues. The implications to practice, policy, and future research are also discussed in the paper.

2. THEORETICAL FRAMEWORK

2.1 Technology Acceptance Models in Healthcare

A variety of theoretical frameworks has been derived to describe the way people make attitudes to and eventually accept or decline the new technologies. Technology Acceptance Model (TAM) was initially postulated by Davis (1989), who serves on the view that technology adoption relies on the perceived usefulness and the perceived ease of use. Models of TAM such as TAM2 and Unified Theory of Acceptance and Use of Technology (UTAUT) add new constructs like social influence, facilitating conditions and hedonic motivation.

TAM, in the context of healthcare, has been modified to incorporate domain specific variables, including patient health literacy, clinical trust and patient-provider relations. It has been established that the conventional TAM constructs need a major adjustment when used to study populations that have differing degrees of digital fluency or when dealing with sensitive personal health data as the technology. In particular, perceived risk, which is not a construct of original TAM, is a strong predictive of resistance to healthcare AI.

An additional complementary approach is the Health Belief Model (HBM) which is concerned with the subjective judgments elicited by patients concerning the threat to the health and the perceived gains as well as hindrances of a health-promoting behavior or technology adoption. Applied in AI, the HBM implies that patients will be better responsive to AI-based tools when they recognize a personal health advantage, such barriers as complexity and data safety are reduced to a minimum, and their medical professional gives them positive incentives to act are present.

2.2 Trust as a Central Construct

The concept of trust plays a central position in patient-provider relationships, and consequently, patient relationships with AI systems. Trust in AI is a multidimensional construct that includes cognitive and affective trust forms that are anchored in the rational evaluation of the reliability and competence of the system and the feeling of care and benevolence, respectively. In healthcare AI, the fact that AI technologies lack human affect poses a special problem: patients can acknowledge the technical proficiency of AI systems and at the same time question the ability of AI to exercise the type of moral judgment that defines compassionate care.

Trust is always one of the most effective predictors of patient readiness to use AI-based healthcare services, as it is stated in literature. When patients indicate that they have high institutional trust in hospitals and healthcare systems, it is more likely that they are open to AI. On the other hand, historically marginalized people, who have faced systematic discrimination in healthcare, are likely to be more distrustful of AI and believe that algorithm systems can coded existing prejudices and continue the same patterns.

3. PATIENT PERCEPTION: EMPIRICAL EVIDENCE

3.1 General Attitudes Toward Healthcare AI

The results of cross-sectional surveys offered in various nations indicate that there are generally diverse attitudes towards AI in healthcare. A survey study conducted in large scale with people living in the United States, the United Kingdom, and Australia has shown that although the majority of the respondents admitted that AI could significantly enhance the

quality of the diagnosis, only a small portion of them would prefer to see a doctor led by AI rather than a human physician. The patients always said they wanted AI as an addition rather than independent equipment and that they consider it as an addition to clinical practice and not a substitute of physician judgment.

This quantitative image has been supplemented by qualitative research. According to the in-depth interviews with patients in a diverse range of clinical environments, the perception of AI is typically ambivalent and context-dependent. Patients can embrace AI support in the areas that seem to be technical or data-heavy, including imaging analysis, drug handling, or appointment scheduling, and also reject the involvement of AI in the spheres that involve interpersonal sensitivity, including mental health visits, end-of-life care conversations, and the provision of severe diagnoses.

The context of care matters considerably. Patients managing chronic conditions such as diabetes or cardiovascular disease tend to express greater openness to AI monitoring tools, particularly when these tools offer real-time feedback and empower self-management. In contrast, patients in acute or emergency settings frequently express concerns about the reliability of AI decision-support systems under time pressure and uncertainty.

3.2 Factors Influencing AI Acceptance

Several empirically supported factors shape the trajectory of patient acceptance toward healthcare AI. Prior experience with digital health technologies is among the most robust predictors; patients who regularly use health apps, wearables, or patient portals are significantly more likely to report positive attitudes toward AI. This finding underscores the importance of digital health literacy as a foundational precondition for AI adoption.

Age and generational cohort also emerge as significant moderators. Younger patients, who have grown up in digitally saturated environments, tend to exhibit higher levels of comfort with AI-mediated interactions. Older adults, by contrast, more frequently express concerns about technological complexity and the erosion of human connection in care. These generational differences are not, however, immutable; research suggests that targeted digital literacy interventions can substantially improve technology confidence among older populations.

Education level and health literacy are closely associated with AI perception. Patients with higher education levels and stronger health literacy generally demonstrate more nuanced understandings of AI's capabilities and limitations, enabling them to form more calibrated expectations. Low health literacy, conversely, is associated with both overestimation of AI's infallibility and underestimation of its potential benefits, two failure modes that can produce quite different behavioural outcomes.

The nature of the healthcare relationship also plays a significant mediating role. Patients who report high satisfaction with their primary care physician and a strong sense of therapeutic alliance are more likely to accept AI tools endorsed by that physician. The physician's communication style, their willingness to explain how AI is being used and why, and their explicit framing of AI as a collaborative partner rather than a replacement are all associated with greater patient comfort.

3.3 Demographic and Cultural Variations

The perception of AI in health care varies not equally among demographic groups of patients. Several studies have found a difference in gender as some studies have found that women were more concerned with privacy and data security whilst men are more inclined to mention performance and accuracy expectations. These disparities are not absolute and are extremely conditioned by other factors, but they point to the significance of disaggregating survey data by demographic subgroup in developing patient communication approaches.

The cultural background of patients is an important factor influencing the cognitive and emotive constructs within which patients perceive AI in the healthcare system. The impersonality of algorithmic decision-making can be viewed as a cultural violation of the relational care norms in a collectivist society, in which healthcare decisions are often made when the family or community steps into the picture. On the contrary, patients who belong to more individualistic cultures might attach more importance to the privacy-conserving nature of AI-mediated appointments.

The racial and ethnic minorities of Western healthcare have proven to exhibit unique trends of AI perception that are closely connected to a history of medical mistrust. Research in the United States has determined that Black and Hispanic patients, who have traditionally felt unfairly treated and discriminated in healthcare facilities, are more skeptical about the objectivity and fairness of AI diagnostic tools. Such mistrust is justified considering the reported cases of algorithmic bias when using AI systems that are trained on datasets that do not reflect the minority groups.

4. TRUST, TRANSPARENCY, AND EXPLAINABILITY

4.1 The Explainability Gap

The lack of transparency in the algorithmic decision-making processes, which is commonly referred to as the black box problem, is one of the most commonly mentioned obstacles to patient trust in AI. Numerous successful AI systems, especially those trained with deep learning systems, produce results that are not easily interpretable even to their creators. In situations where clinicians cannot provide information about how an AI came to a certain recommendation, patients are deprived of the information scaffolding required to make judgments and accept that recommendation with trust.

Explainable AI (XAI) is an idea that is proposed as a solution to this issue. XAI is a type of AI that aims to explain its reasoning processes in ways understandable to humans to empower clinicians and patients to question the foundation behind algorithmic recommendations. Studies indicate that patients who are provided with explanations about AI-generated diagnostic or treatment advice/ recommendations report much higher degrees of trust and satisfaction than those who are provided with non-explained outputs.

Nevertheless, the formulation of useful descriptions of different groups of patients is also problematic. Patients with lower health literacy can find technical explanations that are satisfactory to an expert clinician confusing or frightening. There is an increasing appreciation that explanations should not just be content-specific, but also medium specific, with visual, narrative and interactive explanations promising potential across various patient subgroups.

4.2 Physician-AI Collaboration and Patient Confidence

The physician-AI relationship structure carries important patient experience and perception implications. There is reason to believe that patients will react in the most positive way to models where AI acts as a decision-support system supervised by a competent clinician, and not as an autonomous decision-maker. As soon as physicians actively convey that they engage in and critically evaluate AI suggestions, patients express increased confidence in the general care process.

The idea of appropriate reliance is informative in this context: both patients and clinicians gain advantages of AI systems that can assist and not replace the human judgment. The presentation of AI recommendations as non-negotiable, or a situation where clinicians are seen to blindly follow the output of an algorithm is linked to patient anxiety and dissatisfaction. Even in the most AI-centric care paths, the issue of physician accountability seems to be a non-negotiable factor of patient-centered AI implementation.

5. ETHICAL CONSIDERATIONS, PRIVACY, AND DATA SECURITY

5.1 Data Privacy and Consent

The AI systems used in healthcare need a great amount of patient data to operate. The gathering, storing, analyzing and sharing of such information provokes serious ethical concerns of privacy, autonomy, and informed consent. Data privacy has proven to be one of the topmost concerns patients raise whenever assessing AI in healthcare through surveys. The aspect that patients fear most is the commercial use of their health information, data breach, and the use of their information to train AI systems without their direct permission.

The regulation of healthcare data differs significantly across jurisdictions. In the European Union, the General Data Protection Regulation (GDPR) provides the strong protection of health information and the right of the patient to an explanation in case of automated decision-making. In other areas, the regulatory frameworks are less established, which gives them the setting where the ethical duty of AI developers and healthcare facilities is less clear. When patients understand that they have regulations to follow, they are more likely to demonstrate more confidence in AI systems, which underscores the protective role of transparent governance.

The nature of informed consent processes should change to meet the demands of healthcare AI. The current frameworks of traditional consent, which have been designed for discrete clinical intervention settings, do not coincide well with data-intensive, continuous AI monitoring. Newer frameworks of dynamic, longitudinal consent, where patients are able to adjust their data-sharing preferences as time goes on, are under pilot in a few jurisdictions, and have demonstrated potential to increase patient autonomy without limiting data utility.

5.2 Algorithmic Fairness and Health Equity

Possible encoding and escalation of health disparities by AI systems is one of the gravest ethical issues facing the discipline. The minority groups may receive less accurate AI diagnostic and predictive tools trained on the majority population data, resulting in unequal rates of diagnosis missed, incorrect treatment suggestions, and discriminatory risk assessment. The studies have reported clinically meaningful performance gaps of AI applications in the areas of skin condition diagnosis, pain management, and cardiac risk prediction, among others.

The patients who represent historically underrepresented groups are extremely sensitive to these issues. They are not ungrounded in their skepticism regarding AI fairness and should be addressed by means of strict algorithmic auditing, a wide variety of training data, as well as involving the community in the creation and testing of AI-based products. It is an ethical responsibility of healthcare institutions to ensure that the implementation of AI does not contribute to the current inequality in the delivery of care and to keep patients informed about the measures that are taken to ensure the fairness.

5.3 The Dehumanization Concern

Throughout the qualitative and quantitative literature, the fear of dehumanization is a common topic in patient discussion of AI: the belief that the introduction of AI into the healthcare system will lead to the decline of the empathic, relational attributes that patients deem the key to good medical practice. Patients often report that compassion, intuition, and moral judgment, which they attribute to competent clinicians, cannot be simulated by machines and must not be put in the service of algorithmic efficiency.

This is not just a sentimental issue. The scientific literature on patient-centered care has long determined that the perceptions of patients that they are understood, respected and cared as an individual are strong predictors of adherence to treatment, health outcomes and wellbeing overall. Healthcare AI deployed in a manner that saves clinicians time in personal contact with patients, or that creates the impression of surveillance and data extraction on care relationships, may jeopardize such therapeutic outcomes.

It is both a moral obligation and a functionality to design AI systems that might complement, but not replace, human presence in healthcare. The system of AI tools that help clinicians save time spent on administrative processes and focus efforts on communication with patients and more sophisticated thinking is a prototype of integration that meets patient values and is likely to produce more positive impressions.

6. DISCUSSION

The synthesis of existing research presented in this paper reveals a consistent pattern: patients are neither uniformly receptive to nor uniformly resistant toward AI in healthcare. Their perceptions are shaped by a complex interplay of individual, relational, institutional, and sociocultural factors that cannot be adequately addressed by a one-size-fits-all approach to AI deployment. Understanding these factors in their specificity is essential for healthcare institutions, policymakers, and technology developers who wish to realize the genuine potential of AI to improve population health.

Trust emerges as the pivotal construct mediating patient perception and behavioral intention. Trust in healthcare AI is not simply a function of technical performance; it is built through transparency, through the quality of clinical communication, through the demonstrable commitment of healthcare institutions to patient privacy and fairness, and through the sustained experience of receiving care that is felt to be attentive and personalized. Building this trust requires sustained investment not only in technical systems but in the human and organizational contexts in which those systems are embedded.

The findings also highlight the critical importance of health equity in AI governance. The enthusiastic adoption of AI in healthcare systems serving predominantly affluent, technologically literate populations must not proceed without parallel attention to the needs and perspectives of communities that have historically been underserved by both medicine and technology. Inclusive design, community-based research partnerships, and mandatory bias auditing are among the tools that can help ensure that AI advances health equity rather than undermining it.

Future research should expand the geographic and demographic scope of patient perception studies, many of which have been conducted predominantly in high-income, Western settings. Longitudinal designs that track changes in perception over time, as patients accumulate experience with AI-integrated care, would provide valuable insights into the dynamics of trust formation and erosion. Qualitative and participatory methods can complement quantitative surveys by eliciting the nuanced, contextually situated meanings patients attach to AI experiences.

7. IMPLICATIONS FOR PRACTICE AND POLICY

Based on the evidence reviewed, several implications for healthcare practice and policy can be identified. First, patient education and health literacy programs must be updated to include AI literacy, equipping patients to engage critically and constructively with AI-mediated care. This requires accessible, jargon-free communication about how AI systems work, what they can and cannot do, and how patient data is used and protected.

Second, clinical training curricula should incorporate competencies in communicating about AI to patients, including how to explain AI-generated recommendations, how to present AI as a collaborative tool rather than an autonomous authority, and how to elicit and respond to patient concerns about AI. Clinician communication is a key mediator of patient trust, and investment in this area is likely to yield significant returns in terms of patient engagement and satisfaction.

Third, regulatory frameworks governing healthcare AI must be strengthened and harmonized across jurisdictions to ensure consistent standards of transparency, accountability, and patient protection. Mandatory algorithmic auditing, explainability requirements, and participatory governance mechanisms that include patient representatives are among the policy instruments that can build the institutional trust necessary for responsible AI adoption.

Fourth, AI developers and healthcare institutions should invest in community engagement processes that involve diverse patient populations in the design, testing, and governance of AI tools. Patient advisory boards, participatory design workshops, and community-based research partnerships can ensure that AI systems are developed in ways that reflect and respect the values, needs, and concerns of the populations they are intended to serve.

8. CONCLUSION

Artificial intelligence holds remarkable promise for enhancing the quality, efficiency, and accessibility of healthcare. Yet its transformative potential can only be realized if patients, the ultimate beneficiaries of these technologies, engage with them willingly and confidently. This paper has argued that patient perception toward AI in healthcare is a multidimensional phenomenon, shaped by trust, transparency, cultural context, prior experience, and ethical concern, that demands serious and sustained attention from the healthcare community.

The challenge is not merely technical but deeply human. Building a future in which AI serves health equity and patient empowerment requires that the design and deployment of these technologies be guided by the values, voices, and lived experiences of the patients they are intended to help. A people-centered approach to healthcare AI, one that places patient dignity, autonomy, and wellbeing at its core, is not simply an ethical aspiration; it is a practical prerequisite for the responsible and effective integration of artificial intelligence into the fabric of modern medicine.

REFERENCES

1. Abramoff, M. D., Lavin, P. T., Birch, M., Shah, N., & Folk, J. C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digital Medicine, 1(1), 39. https://doi.org/10.1038/s41746-018-0040-6
2. Ahuja, A. S. (2019). The impact of artificial intelligence in medicine on the future role of the physician. PeerJ, 7, e7702. https://doi.org/10.7717/peerj.7702
3. Bates, D. W., Saria, S., Ohno-Machado, L., Shah, A., & Escobar, G. (2014). Big data in health care: Using analytics to identify and manage high-risk and high-cost patients. Health Affairs, 33(7), 1123–1131.

4. Blease, C., Kaptchuk, T. J., Bernstein, M. H., Mandl, K. D., Halamka, J. D., & DesRoches, C. M. (2019). Artificial intelligence and the future of primary care: Exploratory qualitative study of UK general practitioners' views. Journal of Medical Internet Research, 21(3), e12802.
5. Choudhury, A., & Asan, O. (2020). Role of artificial intelligence in patient safety outcomes: Systematic literature review. JMIR Medical Informatics, 8(7), e18599. https://doi.org/10.2196/18599
6. Davenport, T., & Kalakota, R. (2019). The potential for artificial intelligence in healthcare. Future Healthcare Journal, 6(2), 94–98.
7. Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, 13(3), 319–340.
8. Dillahunt-Aspillaga, C., Gloss, T., Torres, E., & Ruggiero, K. (2020). Patient perceptions of technology-assisted rehabilitation: A qualitative study. Disability and Rehabilitation: Assistive Technology, 15(6), 669–675.
9. European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council: General Data Protection Regulation. Official Journal of the European Union.
10. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.
11. Garg, A. X., Adhikari, N. K., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., & Haynes, R. B. (2005). Effects of computerized clinical decision support systems on practitioner performance and patient outcomes. JAMA, 293(10), 1223–1238.
12. He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. Nature Medicine, 25(1), 30–36.
13. Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., & Wang, Y. (2017). Artificial intelligence in healthcare: Past, present and future. Stroke and Vascular Neurology, 2(4), 230–243.
14. Khullar, D., Casalino, L. P., & Qian, Y. (2020). Perspectives of patients about artificial intelligence in health care. JAMA Network Open, 3(6), e2014857.
15. Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. Journal of Consumer Research, 46(4), 629–650.
16. Maddox, T. M., Rumsfeld, J. S., & Payne, P. R. O. (2019). Questions for artificial intelligence in health care. JAMA, 321(1), 31–32.
17. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.
18. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31–38.
19. Reddy, S., Fox, J., & Purber, M. P. (2019). Artificial intelligence-enabled healthcare delivery. Journal of the Royal Society of Medicine, 112(1), 22–28.
20. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). 'Why should I trust you?': Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
21. Shortliffe, E. H., & Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. JAMA, 320(21), 2199–2200.
22. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.
23. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. MIS Quarterly, 27(3), 425–478.
24. Wachter, R. (2015). The digital doctor: Hope, hype, and harm at the dawn of medicine's computer age. McGraw-Hill Education.
25. Zou, J., & Schiebinger, L. (2018). AI can be sexist and racist — it's time to make it fair. Nature, 559(7714), 324–326.