International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

Harnessing AI Governance and Mechanism to Bridge Trust Among the Healthcare Stakeholders

Manmeet Singh*, Vice President-IT & CIO, Lambda Therapeutic Research Ltd. Ahmedabad.

Abstract - The current age of artificial intelligence (AI) has shown its vital potential in diverse field like science and communication etc. Nowadays, healthcare industry has not remained untouched with the wider application of AI, where it supports Clinicians with decision making related to disease diagnostics and prediction of future health risks along with customized and watchful treatments to each patient. The inclusion of technology comes with hidden complexity of lack of confidence in its application to patients. Previously, studies have been carried out independently to understand Artificial Intelligence governance, explainable Artificial Intelligence and challenge for its implication to the human as subject of treatment, none of these have combined all three perspectives into one. This article suggests to create and investigate a mechanism that combines both, Artificial Intelligence governance (AIG) and Explainable Artificial Intelligence (XAI) to enhance the trust among all stakeholders of healthcare industries.

Keywords: Artificial Intelligence governance (AIG), Explainable Artificial Intelligence (XAI), Health risks, Clinicians, Patients.

1. INTRODUCTION:

In modern healthcare industry, Artificial Intelligence is fetching much attention due to development of new skills such as machine learning (ML), deep learning, natural language processing (NLP) and generative artificial intelligence in early detection and diagnosis of diseases, fine-tuning treatments for the patients making it more inclusive [1]. Today, many hospitals incorporates Artificial Intelligence in critical decision assistance, image (MRI, CT, X-Ray) interpretation to envisaging problems in medical applications of electronic health records and in tele-medicines as well as telehealth [2], yet the final decision making relies heavily on the human intelligence and interventions known popularly as Human-in-the-loop, where AI predicts while most of the decisions gets relied on the human, overruling the machine learning decision [3]. The concept of supervising the AI is of utmost importance as it can make multiple mistakes despite being an excellent coding assistant, still every command and code needs to be supervised properly. Hence, the lack of confidence in medial practitioner for these AI- based models is due to questions like accountability, if the system fails or privacy concerns causing the leakage of patient data as well as their law-abiding mechanism due to patents laws. Hence, in spite of having tools and technologies, it is not welcomed automatically either by the clinicians or other stakeholders such as patients and their assistants trust advice that comes from a system whose inner workings are hidden. These worries become bigger when a wrong decision could harm a patient. Another important factor for this hesitation is that the functioning of Artificial Intelligence models is like a "black box." with a predictive nature where the recommendations are not backed by detailed reasoning to the patients and their assistants [4, 5]. As a result, clinical practitioners may not rely only on the model while making choices. Other factors such as regulators, professional groups and patients push healthcare providers for the illustration of clarity, accountability and authenticity of the Artificial Intelligence as fair and ethical. In order to respond these concerns, researchers across the globe are turning to two approaches: Artificial Intelligence governance (AIG) and Explainable Artificial Intelligence (XAI). Artificial intelligence governance sets up rules, oversight, risk plans and clear lines of responsibility to set Artificial Intelligence in a responsible way [6, 7], whereas explainable artificial intelligence makes systems more readable and are elaborated to make people understand the decision making [8, 9]. The convergence of these two advanced approaches can play an important role in enhancing trust among the different stakeholders to use artificial intelligence. However, this congression can be way more effective than conventional independent approaches. Earlier, few studies have examined the synergistic efficacy of AI governance mechanisms and explainability features of AI in building conviction among many healthcare stakeholders. Still there is real gap that checks healthcare-specific integrated frameworks. This lack creates a significant issue for healthcare organizations willing to have transparent evidence-based assistance to keep

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

trustworthy AI systems at place. Hence, it becomes inevitable to explore the convergence of AI Governance and Explainable AI framework for healthcare. Here, we are inspecting how governance mechanisms and explainability dimensions of artificial intelligence can enhance trust among healthcare professionals, caregivers, and patients, pushing them towards the adoption of AI. By combining both of these, governance and explainability in one model, can expand the area of responsible AI and its adoption in healthcare technology.

2 AI GOVERNANCE IN HEALTHCARE

The transition of traditional healthcare services into wider positioning of artificial intelligence into various domains like patients' data management, clinical diagnostics administrative and financial have been instrumental in modern day healthcare systems. It includes the setting of principles for the structures, rules, standards and checks to keep AI development and integration into day-to-day usage in a fair manner with certain properties such as openness, responsible usage maintaining privacy along with humanized control with better compliances and proper risk handling associated with clinical industries [10]. The role of AI governance becomes even more while considering the involvement of generative AI and large multi modal model (LMMs) in prediction of image-based diagnostics that utilizes the predefined algorithms. This includes a combination of various committees, policy frameworks, technical oversighting control and a continuous updating workflow based on the distinct requirements to protect people's health and safety in real time. In an regulated environment, healthcare organizations have to follow the standards decided by Word Health Organization (WHO) for the Ethics and Governance of AI for Health to protect the human autonomy, accountability and transparency as per their six basic principles [11]. In India, Indian Council of Medical Research (ICMR) has also outlined certain guidelines for the usage of the AI in a responsible and transparent manner as its ten principle for the healthcare industry [12]. With these guidelines in place, it becomes exciting to that the healthcare industries are observing these guidelines, as there is no report that can give a clear idea about this aspect in reference to Indian healthcare industry. Although many private and public institution are deploying such compliances. Also, new guidelines like Digital Personal Data Protection (DPDP) Act, 2023 are enforcing the laws to the healthcare industry in terms of the patient data privacy and protection, while in Europe already General Data Protection Regulation (GDPR) law which is active since May 2018 demands confidentiality and securing personnel information and breach to GDPR will results into substantial fines. Hence, this is intriguing to understand and elaborate on the role of AI governance in the era of digital technology.

3. EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI):

Being in Healthcare Life Sciences industry (HLS) since last 21 years, the major lacuna that I have observed in the usage of AI in such commercial places is the lack of confidence in the working mechanisms of the AI- based system and their interpretations. This is due to lack of information a.) how an AI based system behaves, b.) how it predicts? c.) Are they prediction made by AI are personalized or a general observation based on the prior information rather than ignoring the human behavioral complexity, d.) How and what makes an AI-based system to come to a conclusion? Thus, it can be interpolated that the biggest challenge in AI -based system integration in healthcare Common dimensions of explainability includes the transparency and interpretability along with the quality user-centered explanations leading to a better understanding and greater confidence among the professional as well as naïve stakeholders regarding the humanized AI based interpretations. This remains still elusive and needs a detailed elaboration to understand which kinds of explanations work best or how much explanation is needed to build trust and drive adoption.

4. TRUST IN HEALTHCARE AI

Building confidence is a requirement for using AI in healthcare. This can be incorporated in multiple ways like, cognitive trust that defines the capability and dependability of the system, and dependable, whereas affective trust grows when people feel the system is kind and designed for them in a personalized manner. On the other end, believing in the organization's structure, governance and proper defined guidelines build Institutional trust. Many studies treat trust as a fixed result of a moving process that is shaped by governance, explainability, openness and rules [13]. Moreover, there is a recent report where the integration of six attributes, like ability, compassion, veracity, consistency, functionality, and privacy, and alleged risks are considered to the important to create a core confidence or trust among the stakeholders of the health care industry [14]. Based on these studies, it becomes pertinent to construct such models to show how trust is built and changes among different healthcare groups to enhance confidence into the AI governance in healthcare industry. Trust is anticipated to serve as a mediator linking governance and explainability to adoption outcomes. Specifically, governance and explainability may indirectly influence both intention and

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

actual system usage through trust-enhancing pathways among patients and their caregivers, a major stakeholder of this Healthcare industry.

5. ADOPTION OF AI IN HEALTHCARE INDUSTRY

In recent times, Global healthcare industry has quietly adopted the Artificial intelligence. This has been intriguing to understand that using AI in healthcare depends not on usual tech acceptance factors but also, on trust how risky it feels, ethics and rules. As healthcare AI gets more independent and powerful the choice to adopt relies more on governance and explainability. There are several models like, Technology Adoption Model (TAM) and Unified Theory of Acceptance and Use of Technology (UTAUT) that can be adopted or integrated to improvise the healthcare system. [15, 16]. These theories signify the concept of adoption as how and why any AI based technology will be adopted and used of a system as a futuristics approach. The major area where these AI-based systems are being incorporated are online health tools, electronic health records (EHRs), telemedicine platforms, and diagnostic systems due to perceived ease and usefulness of the technology-based systems [17]. One of the important aspects is AI literacy, which make the patients, caregivers, medical practitioners, pharmacist as well as administrators to understand the real sense, critically assess, ethical governance as well as the two-way interaction with it. It even becomes more important in the developing countries lie, India where the availability of low-cost internet and social media has influenced the choices of patients and their next-to-kin in undergoing the medical as well as pharmaceutical preferences.

6. DISCUSSION

Nowadays, there is an increasing demand for human-centered healthcare AI that can bridge the communication among the different stakeholders regarding its usage at different levels of the healthcare industry. Unlike the previous studies that investigate governance, explainability and trust separately, there is a dire need to integrate these sole components within a unified explanatory model, where these can be advanced to the responsible AI by demonstrating how governance structures and explainability mechanisms jointly contribute to confidence formation. This integration will also extend and take support from the technology by incorporating the `AI governance and explainability as backgrounds to trust and adoption. The expected association will also support the growing recognition that technical performance solely is insufficient to ensure successful healthcare AI implementation. Hence, institutions must simultaneously address governance, ethical and human-centered considerations. Furthermore, this will also help to understand the healthcare informatics literature by adopting a multi-participant perspective in addition to research that has primarily focused on clinicians whereas this integration will explicitly incorporate the patients and caregivers along with the other stakeholder such as medical practitioners, radiologists and pharmaceutical partners etc. Thus, this is another approach which will offer a comprehensive understanding of healthcare AI adoption dynamics within increasingly complex healthcare ecosystems.

7. LIMITATIONS

Despite being one of the most sought integrations, there are many limitations, that should be acknowledged. The study carried out can be more on self-reported perceptions leading to response bias which can give a false interpretation with reference to global scenario due to its cross-sectional nature of sampled data which can restrict the interpretation. Another limitation focuses on the changing nature of the healthcare industries due to continuous research in the mentioned sector backed by life science, medical and pharmaceutical domain, hence making a universal AI tool to be integrated in a pharmaceutical industry may not do well in the other industry despite being in the same domain. Also, differences in culture and management of an individual hospital can impact the trust of people of that particular arena to use the AI-embedded patient care system across healthcare systems. Moreover, it can be dependent on the variable views of people based on the different country's technological advancement, that can give them a variable confidence to rely on the integration of the AI into healthcare system. Hence, it is of utmost required to look for the long-term changes at a comparative level in the different countries and accommodate the confident AI usage in the future studies in real life settings.

10. CONCLUSION

In current time, with the advent of artificial intelligence has changed the traditional working models of modern healthcare industries. Still, there are major concerns of their working methodology these AI- enabled or integrated system are using especially in the patients and their caregivers due to the AI literacy in the developing countries. Still, it is a mammoth task to put in the AI-enabled system to place in the hospitals which may need a better trust build-up to all the stakeholders by ensuring the supervised reliability, clarity and safety.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 11 , November - 2026

11. DECLARATION STATEMENT

a. Ethical Approval and Consent to Participate: Not Applicable.

b. Consent for Publication: Author declares no competing interests (financial or others) in relation to this manuscript. There is no conflict of interest.

c. Funding: This research did not receive any specific grant from any funding agencies.

12. DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were generated during the current study.

13. REFERENCES:

1. Bokolo AJnr, Denecke K, Gabarron E (2026) A patient-centered approach to inclusive AI in healthcare: potential benefits, challenges, and recommendations. Health Technol 16:681–703. https://doi.org/10.1007/s12553-026-01080-z
2. Wang L, Zhang S, Xu N, He Q, Zhu Y, Chang Z, Wu Y, Wang H, Qi S, Zhang L, Shi Y, Qu X, Zhou X, Song J (2025) Role of artificial intelligence in medical image analysis. Chinese Medical Journal 138:2879–2894. https://doi.org/10.1097/CM9.0000000000003824
3. Olawade DB, Plabon SB, Ojo A, Ogunbona MA, Makanjuola BD, Olasilola OR (2026) Human in the loop artificial intelligence in healthcare: applications, outcomes, and implementation challenges. International Journal of Medical Informatics 213:106362. https://doi.org/10.1016/j.ijmedinf.2026.106362
4. Wadden JJ (2022) Defining the undefinable: the black box problem in healthcare artificial intelligence. J Med Ethics 48:764–768. https://doi.org/10.1136/medethics-2021-107529
5. Hassan R, Nguyen N, Finserås SR, Adde L, Strümke I, Støen R (2025) Unlocking the black box: Enhancing human-AI collaboration in high-stakes healthcare scenarios through explainable AI. Technological Forecasting and Social Change 219:124265. https://doi.org/10.1016/j.techfore.2025.124265
6. Hussein R, Zink A, Ramadan B, Howard FM, Hightower M, Shah S, Beaulieu-Jones BK (2026) Advancing healthcare AI governance through a comprehensive maturity model based on systematic review. npj Digit Med 9:236. https://doi.org/10.1038/s41746-026-02418-7
7. Alami H, Pozelli Sabio R, Pérez EJ, Gagnon M-P, Langlois L, Denis J-L, Malas K, Rivard L, Savoldelli M, Ag Ahmed MA, Fortin J-P (2026) Artificial Intelligence Governance in Health Systems: Systematic Review of Frameworks and Integrative Model Proposal. J Med Internet Res 28:e87448–e87448. https://doi.org/10.2196/87448
8. Mondal R, Ghosh SK, Ponnalagu RN (2026) Explainable Artificial Intelligence in Smart Healthcare. In: Banerjee S, Saha S, Biswas S, Debnath NC (eds) AI in Smart and Secure Healthcare. Springer Nature Switzerland, Cham, pp 45–71
9. Shiddik MdAB (2026) Explainable Artificial Intelligence in Healthcare: Current Landscape, Challenges, and Future Directions. Health Science Reports 9:e72172. https://doi.org/10.1002/hsr2.72172
10. Reddy S, Allan S, Coghlan S, Cooper P (2020) A governance model for the application of AI in health care. Journal of the American Medical Informatics Association 27:491–497. https://doi.org/10.1093/jamia/ocz192
11. Goodman KW, Litewka S, Malpani R, Pujari S, Reis AA (2023) Global health and big data: The WHO's artificial intelligence guidance. S Afr J Sci 119:. https://doi.org/10.17159/sajs.2023/14725
12. Kundu T, Bardhan M (2025) Artificial intelligence in neurology, ethics, recent guideline, and law-an Indian perspective. Front Neurol 16:1515041. https://doi.org/10.3389/fneur.2025.1515041
13. Mertz M, Toskovich K, Shields G, Attema G, Dumond J, Cameron E (2025) Exploring trust factors in AI-healthcare integration: a rapid review. Front Artif Intell 8:1658510. https://doi.org/10.3389/frai.2025.1658510
14. Dickens J, Uvet H, Bauman A, Kwon HY (2026) The big six model of AI trust: a unifying framework and 20-item scale for knowledge creation. Journal of Knowledge Management 30:2827–2862. https://doi.org/10.1108/JKM-02-2025-0222
15. Rouidi M, Elouadi AE, Hamdoune A, Choujtani K, Chati A (2022) TAM-UTAUT and the acceptance of remote healthcare technologies by healthcare professionals: A systematic review. Informatics in Medicine Unlocked 32:101008. https://doi.org/10.1016/j.imu.2022.101008
16. Lee AT, Ramasamy RK, Subbarao A (2025) Understanding Psychosocial Barriers to Healthcare Technology Adoption: A Review of TAM Technology Acceptance Model and Unified Theory of Acceptance and Use of Technology and UTAUT Frameworks. Healthcare 13:250. https://doi.org/10.3390/healthcare13030250
17. Martins MDS, Filho BADC, Sousa MFD, Mendonça AVM, Andrade NFD, Filha SRVDF, Costa LDD, Silva JPFD, Soares JNF, Xavier IBC, Farias JS (2026) The use of the unified theory of acceptance of technology model for telemedicine: A systematic review for developing a research agenda for future studies. DIGITAL HEALTH 12:20552076261435081. https://doi.org/10.1177/20552076261435081

(This work is licensed under a Creative Commons Attribution 4.0 International License.)