REGULATING ARTIFICIAL INTELLIGENCE IN HEALTHCARE: BALANCING INNOVATION, PATIENT SAFETY, AND ETHICAL ACCOUNTABILITY

Soham Sandeep Joshi, LL.M., Symbiosis Law School, Pune[1]

ABSTRACT

Artificial Intelligence (AI) is rapidly reshaping healthcare by enhancing diagnostic accuracy, enabling predictive treatment, and optimising patient monitoring. While these advancements promise efficiency and accessibility, they simultaneously expose healthcare systems to unprecedented legal and regulatory challenges. In India, the absence of binding legislation—contrasted with structured regimes such as the U.S. FDA's SaMD Action Plan, EHDS Regulation, 2025 and the EU's AI Act—has left governance fragmented and largely dependent on non-binding ethical guidelines, most notably the ICMR 2023 framework. This regulatory shortfall raises pressing concerns regarding liability for algorithmic errors, patient safety, bias mitigation, and data privacy. Through a doctrinal and comparative analysis, this research identifies the central regulatory gap in India: the over-reliance on voluntary codes without enforceable safeguards. The study argues for a risk-based, legally binding framework incorporating mandatory audits, post-market monitoring, and clear accountability rules. Ultimately, the findings underscore the imperative for India to move beyond aspirational ethics toward robust regulation that balances innovation with patient-centric protections, thereby aligning healthcare AI with both constitutional principles and global best practices.

Keywords: Artificial Intelligence in Healthcare; ICMR Guidelines; Data Protection; Algorithmic Accountability; Regulatory Framework; India

[1] Soham Sandeep Joshi, LL.M., Symbiosis Law School, Pune.

INTRODUCTION

Artificial Intelligence (AI) has rapidly become a transformative facilitator of change in the healthcare sector, fundamentally transforming clinical decision-making processes, diagnostic practices, treatment policies, and patient monitoring processes. From AI-powered radiological devices able to detect imperfections never seen before to forecasting models enabling the early diagnosis of ailments, this technology can potentially bring about enhanced efficiency, accessibility, and cost-effectiveness in medical services. However, this revolution is met by a multi-faceted suite of legal, ethical, and regulatory issues, particularly in countries like India, where healthcare delivery is decentralised and regulatory infrastructure remains underdeveloped. Unlike the United States and the European Union, which have adopted systematic approaches in the form of the FDA's Software as a Medical Device (SaMD) Action Plan, European Health Data Space Regulation, 2025 and the EU Artificial Intelligence Act, respectively, India currently relies almost solely on ethical guidelines and fractured data protection laws. Such a lack of regulation creates questions about patient safety, accountability for algorithm errors, and the protection of confidential health data. Against this background, the planned research seeks to methodically review deficiencies around AI governance in India's healthcare landscape and suggest framework principles balancing innovation and comprehensive safeguards around trust, transparency, and patient rights.

LITERATURE REVIEW AND METHODOLOGY: ARTIFICIAL INTELLIGENCE IN HEALTHCARE

The present study is conducted by applying the doctrinal research methodology, supplemented by a comparative and case study analysis method. Artificial intelligence (AI) progress in the field of healthcare is representative of the path that is characterised by early experimentation, up-to-date advances, and thorny regulatory and ethical challenges. Initial expert programs, including Stanford's MYCIN of the 1970s, showed diagnostic judgment superior to that of non-specialist practitioners, yet never became integrated into clinical practice due to liability and accountability issues.[2] Subsequent programs, including INTERNIST-1, and later iterations such as CADUCEUS and DXplain, had better diagnostic decision-making abilities, yet were

[2] MYCIN: the beginning of artificial intelligence in medicine, Telefonica Tech (2024), https://telefonicatech.com/en/blog/mycin-the-beginning-of-artificial-intelligence-in-medicine.

limited by their inflexible, rule-based architecture.[3] Initial attempts included demonstration of concept feasibility, yet indicated the complexities of practical AI application.

Our modern era began with natural language processing and machine learning breakthroughs, as in IBM's Watson. It was first celebrated for winning at Jeopardy! Yet Watson for Oncology demonstrated the danger of too much exuberance: too little training data, unsuitable clinical workflow integration, and "unsafe" advice ended in much medical scepticism and eventual demise of a many-billion-dollar program.[4] It is in such cases that technological advances can falter in clinical settings, absent thorough verification and practical rollout plans.

Approval landmarks such as the 2018 FDA clearance of IDx-DR for diabetic retinopathy screening marked the transition from experimental prototype to regulated medical device. [5]Regulation lags behind the reality of technology. Nicolas Terry identifies the incompleteness of the US frameworks that divide FDA regulation of devices from state medical licensing.[6] Adaptive AI that can learn from deployment data rejects frozen "locked" models that must be repeatedly re-approved after updating. By comparison, the EU has adopted a risk-based approach, combining GDPR's strict data protection rules with the proposed AI Act, under which healthcare AI is classified as "high risk" and subject to lifecycle regulation.[7]

Theory proposals, like W. Nicholson Price II's "Four Roles" framework, posit artificial intelligence as extending the boundaries of medicine, making expert knowledge democratic, automating quotidian operations, and making scarce resources more optimal.[8] That classification transcends the usual substitute/replace commentary, proposing that AI may usher in a change in medical practice rather than replicate it. Empirical research verifies that possibility; AI systems matched the precision of dermatologists in skin cancer detection and facilitated scalable screenings for diabetic retinopathy, particularly in resource-scarce

[3] AI's Ascendance in Medicine: A Timeline, Cedars-Sinai (2023), https://www.cedars-sinai.org/discoveries/ai-ascendance-in-medicine.html.
[4] Henrico Dolfing, *IBM Watson: From Healthcare Canary to a Failed Prodigy* (2023), https://healthark.ai/wp-content/uploads/2023/11/IBM-Watson-From-healthcare-canary-to-a-failed-prodigy_1.pdf.
[5] FDA permits marketing of artificial intelligence-based device to detect certain diabetes-related eye problems, FDA (Apr. 11, 2018), https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye.
[6] Nicolas Terry, *Of Regulating Healthcare AI and Robots*, 21 Yale J.L. & Tech. 133 (2019).
[7] Janos Meszaros, Jusaku Minari & Isabelle Huys, The Future Regulation of Artificial Intelligence Systems in Healthcare Services and Medical Research in the European Union, 13 Front. Genetics 927721 (2022).
[8] W. Nicholson Price II, Artificial Intelligence in the Medical System: Four Roles for Potential Transformation, 21 Yale J.L. & Tech. 122 (2019).

locations. But that success is qualified by concerns of algorithmic bias and issues of generalisability.

Algorithmic bias is among the most significant unsolved problems. Obermeyer et al. showed racial bias in cost-based healthcare prediction algorithms, with reduced spending among Black patients incorrectly being seen as reduced need for healthcare.[9] Comparable biases have been shown for cardiovascular risk scoring, chest X-ray interpretation, and dermatological diagnosis, with model precision typically reduced for minority populations.[10] Though post-hoc bias detection methodologies are available, active bias prevention during system design is still underdeveloped.

Another gap concerns validation and generalizability. Most AI systems are trained on retrospective datasets from well-resourced institutions, limiting their performance in diverse healthcare environments. Price emphasises contextual bias—algorithms optimised for one clinical setting may underperform when transferred elsewhere.[11] Prospective and adaptive validation frameworks are essential but largely absent in current regulatory schemes.

Liability and accountability complicate integration even further. Courts and academics argue whether responsibility for damage from AI ought to be that of manufacturers, hospitals, or practising individuals. Liability in Taylor v. Intuitive Surgical depended not only upon product flaws but also upon literally requiring institutional oversight and adequacy of educational preparation, hinting at similar controversy for healthcare AI.[12] The doctrine of the corporate practice of medicine and the learned intermediary rule complicate even more whether AI systems themselves are "practising medicine."

Three significant disputes enliven scholarship today. First, does AI substitute for human expertise with implications for regulation and professional practice? Substitution models raise questions of licensing, and augmentation stresses human supervision. Second, frozen versus adaptive regulation reveals the contradiction between safety and innovation, in that learning systems that are constantly learning cannot be frozen at the moment of authorisation. Third,

[9] Ziad Obermeyer et al., Dissecting racial bias in an algorithm used to manage the health of populations, 366 Science 447 (2019).
[10] Understanding, Identifying and Mitigating Algorithmic Bias in Healthcare, Accuray (2023), https://www.accuray.com/blog/overcoming-ai-bias-understanding-identifying-and-mitigating-algorithmic-bias-in-healthcare.
[11] *Ibid, 7.*
[12] Supra, 5.

privacy versus innovation mirrors the conflict between the data minimisation spirit of GDPR and the need for AI to operate with large, varied datasets.[13] The "right to explanation" exacerbates the conflict, in that black-box systems find it challenging to meet transparency obligations.

Lastly, international harmonisation is absent. The EU AI Act, future U.S. reforms, and future Asian frameworks diverge in their philosophies, threatening market segmentation and inconsistent patient protections.[14] The patchwork obstructs the worldwide release of useful technologies while creating holes for accountability.

In the Indian context, the Indian Council of Medical Research (ICMR) published the Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare in 2021, a landmark development for the Global South, as it was the first national framework.[15] The frameworks articulate the importance of patient safety, algorithmic transparency, and accountability while embracing the principle of "human oversight" of clinical decision-making. The guidelines recommended inclusivity in dataset curation to address bias, required informed consent when engaged in AI-based interventions, and indicated that IRB approval would likely be appropriate for intervening in research utilising AI. The framework importantly emphasised the privacy and security of data in accordance with the newly developing data protection legislation in India. Though not binding, the ICMR framework foreshadows a shift in policy towards ethical stewardship of AI practices, placing India as a frontrunner of developing countries, addressing the normative quandaries of AI in healthcare.

Overall, literature identifies promise as well as pitfalls of AI in healthcare. Traditional systems had defined feasibility but failed to test for liability; contemporary devices demonstrate empirical efficacy yet reveal underlying problems of bias, validation, and regulation. As much as the research of Price, Terry, and Meszaros et al. demonstrates important theoretical and regulatory critique, much is yet to be done to design integrated governance frameworks capable

[13] The Intersection of GDPR and AI and 6 Compliance Best Practices, Exabeam (2022), https://www.exabeam.com/explainers/gdpr-compliance/the-intersection-of-gdpr-and-ai-and-6-compliance-best-practices/.

[14] Arya.ai, *Policies and Regulations Around AI Usage: Interpretation and Impact* (2022), https://arya.ai/research/ai-regulation.

[15] Indian Council of Medical Research, *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research & Healthcare* (2021), available at https://www.icmr.gov.in/icmrobject/uploads/Guidelines/1724842648_ethical_guidelines_application_artificial_intelligence_biomed_rsrch_2023.pdf.

of tipping the balance between innovation and safety in varied healthcare settings. Mitigating bias, prevention, adaptive validation, and cross-border coordination will be the keys to guaranteeing that AI's revolutionary promise is actualised without derogating equity or trust.

ISSUES ADDRESSED AND OBJECTIVES

Medical artificial intelligence represents a peculiar combination of legal, ethical, and regulatory challenges, which are under-addressed in the Indian environment. Primarily, there is an issue of liability and responsibility: if an AI-aided diagnosis system provides a detrimental suggestion, no one knows who is liable. Should it be the creator, the hospital, or the doctor who trusted the system? Existing negligence and product liability laws in India do not give any clear answer. Yet another urgent topic is algorithmic bias and fairness. Research after research has demonstrated AI systems performing worse among minority groups if trained on biased data sets. And in India, where healthcare accessibility is already segmented between urban–rural and socio-economic gradients, rampant use of AI might widen the gap further.

Another issue regards post-market surveillance and validation. Unlike traditional medical devices, AI modules continue to mature through learning, bringing the possibility of "algorithmic drift." It implies India lacks adaptive regulatory frameworks, meaning there are no proper mechanisms available to ensure ongoing safety and reliability after an AI module is deployed in real-world hospital contexts.

Finally, issues of data governance and privacy become significant challenges. The 2023 Digital Personal Data Protection Act enshrines basic privacy protections; nevertheless, no account is taken of the specific vulnerabilities of medical data under artificial intelligence regimes, of secondary use for training purposes and of transnational transfers. These issues, in aggregate, highlight a central gap in India's AI healthcare statute. While the ICMR directives mark a beginning, the lack of a statutory basis, enforcement, and harmonisation with general healthcare law leaves patients vulnerable and innovation at large.

The main objectives of this study can be summarised into a few points.

i. To analyse existing Indian legal and ethical instruments.

ii. To conduct a comparative study of international regulatory approaches.

iii. To propose a **risk-based, patient-centric regulatory model.**

iv. To recommend **institutional and legislative reforms.**

FINDINGS

The Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare define Artificial Intelligence as, *"a system's ability to correctly interpret external data and to use those learnings to achieve specific goal and tasks through flexible adaption"*[16] The literature suggests that India is at a crucial juncture in regulating Artificial Intelligence (AI) in healthcare. While international regimes like the U.S. FDA's[17] *Software as a Medical Device (SaMD) Action Plan* and the European Union's[18] The proposed AI Act is moving toward risk-based and lifecycle regulation, and India continues to rely primarily on non-binding ethical frameworks. The Ethical Guidelines issued by the ICMR are a classic example of India still being in the process of accepting the broader implications of AI, and the stakeholders are still studying this phenomenon in detail. However, when evaluated against the operational and legal gaps highlighted in global scholarship, it becomes evident that these progressive guidelines remain inadequate as substitutes for enforceable statutory regulation.

The guidelines stress the inevitable AI revolution in the healthcare field and list various sectors that would be majorly affected by the advent of AI. This revolution is expected to improve the healthcare delivery systems by making it affordable to the general public. Some instances where AI can enhance patient care are Computed Tomography (CT) scans, which radiologists can efficiently diagnose using an AI mechanism. Mammography scans can now predict the onset of breast cancer before any visual symptoms appear by using AI. When it is so evident that AI cannot be averted, it becomes necessary to regulate its use. It is a settled position that the case is not such that an AI would never fail; there have been recorded instances wherein the AI provided a wrong diagnosis. The question arises of legal liability, as the law on this factor is still unclear, since AI cannot be held liable for its diagnoses and judgments. It is evidenced in the said guidelines that around 10% of deaths are accounted for by misdiagnoses

[16] Indian Council of Medical Research, *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare* (2023), available at https://www.icmr.gov.in/icmrobject/custom_data/pdf/Ethical-guidelines/Ethical_Guidelines_AI_Healthcare_2023.pdf.
[17] Supra, 4.
[18] Supra, 6.

of a disease in a patient.[19] Use of AI can significantly reduce the misdiagnoses and may facilitate an easy and early detection of an ailment. However, it becomes imperative to note that, as mandated by the EU, these AI diagnoses must always be done under human supervision to avoid any unwanted incident.

The scope and purpose of these guidelines are very well enunciated, and it is not to "limit innovation" but to encourage "effective yet safe development." Not only this, but also these guidelines aim to ensure safe deployment and adoption of AI-based technologies in the field of biomedical research and healthcare. Furthermore, these guidelines apply to all biomedical research involving human participants and are to be religiously followed by health professionals, researchers, technicians, hospitals, and all stakeholders in the medical industry. The guidelines further advance the concept of Responsible AI, as a secure and responsible AI would be necessary; however, it is a matter of concern as to how it will be incorporated and employed in the current technology. The broader implications of these guidelines and the gaps in implementing these can be understood through the following points.:

Patient Safety and Algorithmic Transparency: The guidelines stress that AI systems shall not replace human participation in diagnosis but should function under the principle of "human oversight" as also implemented by the EU. The rationale behind this principle is to maintain transparency so that the diagnosis will be explainable to both the physicians and the patients. However, the guidelines fail to mandate the explainability feature in practice. Thus, this creates a gap between normative aspirations and the practicality of enforceable standards. Since these are just guidelines, they lack the statutory force of law.

The case of Bias: A salient aspect of the ICMR Guidelines is the recognition that inclusiveness is needed in the curation of datasets. Recognising the under-representation of specific populations in AI training data can lead to disproportionate impact owing to inequity in Indian healthcare. India's socio-demographic diversity and the rural-urban divide are particularly interesting. That said, the guidelines only "recommend" methods of inclusive dataset curation without articulating how the curation would be enforced or incorporating periodic audits of AI systems deployed.

Data Privacy: The guidelines align with the freshly enacted Digital Data Protection Act, 2023.

[19] Supra, 15.

The government plans to bring the Digital Information Security in Healthcare Act bill. It is envisioned that these guidelines, along with the enactments, would fortify the privacy of the healthcare data of individuals. The Guidelines assert that individual privacy and protection of personal health data must be protected at all stages of AI development and deployment. As medical data are extremely sensitive, the guidelines require anonymisation of patient identifiers (including metadata and on-image data) before release, with identifiable formats only sanctioned for clinical utility. Patients must own their data, including the right to access, amend, or withdraw consent, and must be informed about the nature and purpose of their data usage, as well as the safeguards in place concerning the data itself. Special care is emphasised with respect to predictive algorithms and biometric data, requiring explicit consent, with additional security and ethical approval of the use of such data. Manufacturers take on a duty to avert privacy harms by ensuring there is no risk of reidentification of subjects, and to correct the risk of data loss, leakage, or re-use of the data undisturbed. Data cannot be repurposed without new consent, and any mass-scale deployment operationalisation must be pre-approved and assessed for impact on human rights, ethics, and privacy. Collectively, these stipulations outline both the moral and legal responsibilities of maintaining the integrity of health data, particularly given the evolving context of data protection laws in India in light of the IT Act, 2000 and the forthcoming DISHA and PDP.

Liability: As discussed earlier, one of the significant concerns of applying the AI-based technology in healthcare is legal liability if an undesired outcome occurs during a patient's healthcare. The law on this point is still unclear; however, the guidelines acknowledge that accountability shall be shared by all the stakeholders and institutions involved in healthcare and medical research. However, no provisions have been made for shifting responsibility in cases of AI-induced harm. This absence of statutory clarity poses a considerable risk of leaving the patients without any remedy in case of an injury.

CONCLUSION AND RECOMMENDATIONS

Between the optimism of technological progress and the dangers of legal uncertainty lies the actual battleground for development, deployment and encouragement of use of AI in healthcare in India. AI in healthcare poses an excellent opportunity for the talent in our country, but it is also probably the most urgent issue that needs to be regulated by the legislature. While the ICMR, 2023 guidelines represent an admirable effort in self-regulation by the medical

community, it is to be noted that these are only guidelines and they lack enforcement, as without any legal or statutory backing, these would remain aspirational. We need a strong legal framework to overcome this legislative gap. With the advent of AI, which is ever-changing and dynamic, we need a comprehensive legal framework along with these guidelines and the DPDP Act, which would ensure patient safety by mandating accountability and sustaining the public trust. A risk-based classification adopted by the EU in its AI Act would provide a safe utilisation of these AI technologies under human supervision. Even though the jurisprudential development of the liability concept concerning AI-induced harm is yet to be fortified by the judicial systems worldwide, India can still incorporate specific provisions protecting patients from bias, unwanted surveillance, and informed consent. By doing so, innovation in this field will be channelised in the correct direction. We can set an example to the global south by granting remedies to the patients who have faced AI-based harm, which would be based on the basic jurisprudential concepts of justice, equity and a conscience that it is the patient who remains the central consideration for any innovation, which would fortify the objective of preserving the patient rights.