

INTERNATIONAL JOURNAL FOR LEGAL RESEARCH AND ANALYSIS



Open Access, Refereed Journal Multi Disciplinary
Peer Reviewed

www.ijlra.com

DISCLAIMER

No part of this publication may be reproduced, stored, transmitted, or distributed in any form or by any means, whether electronic, mechanical, photocopying, recording, or otherwise, without prior written permission of the Managing Editor of the *International Journal for Legal Research & Analysis (IJLRA)*.

The views, opinions, interpretations, and conclusions expressed in the articles published in this journal are solely those of the respective authors. They do not necessarily reflect the views of the Editorial Board, Editors, Reviewers, Advisors, or the Publisher of IJLRA.

Although every reasonable effort has been made to ensure the accuracy, authenticity, and proper citation of the content published in this journal, neither the Editorial Board nor IJLRA shall be held liable or responsible, in any manner whatsoever, for any loss, damage, or consequence arising from the use, reliance upon, or interpretation of the information contained in this publication.

The content published herein is intended solely for academic and informational purposes and shall not be construed as legal advice or professional opinion.

**Copyright © International Journal for Legal Research & Analysis.
All rights reserved.**

ABOUT US

The *International Journal for Legal Research & Analysis (IJLRA)* (ISSN: 2582-6433) is a peer-reviewed, academic, online journal published on a monthly basis. The journal aims to provide a comprehensive and interactive platform for the publication of original and high-quality legal research.

IJLRA publishes Short Articles, Long Articles, Research Papers, Case Comments, Book Reviews, Essays, and interdisciplinary studies in the field of law and allied disciplines. The journal seeks to promote critical analysis and informed discourse on contemporary legal, social, and policy issues.

The primary objective of IJLRA is to enhance academic engagement and scholarly dialogue among law students, researchers, academicians, legal professionals, and members of the Bar and Bench. The journal endeavours to establish itself as a credible and widely cited academic publication through the publication of original, well-researched, and analytically sound contributions.

IJLRA welcomes submissions from all branches of law, provided the work is original, unpublished, and submitted in accordance with the prescribed submission guidelines. All manuscripts are subject to a rigorous peer-review process to ensure academic quality, originality, and relevance.

Through its publications, the *International Journal for Legal Research & Analysis* aspires to contribute meaningfully to legal scholarship and the development of law as an instrument of justice and social progress.

PUBLICATION ETHICS, COPYRIGHT & AUTHOR RESPONSIBILITY STATEMENT

The *International Journal for Legal Research and Analysis (IJLRA)* is committed to upholding the highest standards of publication ethics and academic integrity. All manuscripts submitted to the journal must be original, unpublished, and free from plagiarism, data fabrication, falsification, or any form of unethical research or publication practice. Authors are solely responsible for the accuracy, originality, legality, and ethical compliance of their work and must ensure that all sources are properly cited and that necessary permissions for any third-party copyrighted material have been duly obtained prior to submission. Copyright in all published articles vests with IJLRA, unless otherwise expressly stated, and authors grant the journal the irrevocable right to publish, reproduce, distribute, and archive their work in print and electronic formats. The views and opinions expressed in the articles are those of the authors alone and do not reflect the views of the Editors, Editorial Board, Reviewers, or Publisher. IJLRA shall not be liable for any loss, damage, claim, or legal consequence arising from the use, reliance upon, or interpretation of the content published. By submitting a manuscript, the author(s) agree to fully indemnify and hold harmless the journal, its Editor-in-Chief, Editors, Editorial Board, Reviewers, Advisors, Publisher, and Management against any claims, liabilities, or legal proceedings arising out of plagiarism, copyright infringement, defamation, breach of confidentiality, or violation of third-party rights. The journal reserves the absolute right to reject, withdraw, retract, or remove any manuscript or published article in case of ethical or legal violations, without incurring any liability.

TOWARDS SAFE AND ACCOUNTABLE AI USE IN INDIAN HEALTHCARE: A LEGAL AND ETHICAL ANALYSIS

AUTHORED BY- SUMIT PATNAIK, RISHI VERMA & SUBHI SHANKAR

Introduction

Artificial Intelligence ('AI') means the '*capability of computer systems or algorithms to imitate intelligent human behavior*'¹. This meaning entails within it a broad definition and gives us a vague idea of the plethora of uses that AI is being used in human-dominated fields. One of the sectors is healthcare, which is the most crucial as it is tasked with protecting and saving people's lives. AI has been increasingly used with credible efficiency in this sector, ranging from diagnostics, treatment, such as surgeries, etc. AI holds immense importance due to the Indian healthcare system being plagued by a high disease burden, shortage of medical professionals, and logistical inefficiencies, and has the ability to revolutionize the healthcare sector.

Case studies from Kerala² have been showing great promise in the integration of AI in the healthcare sector. Early diagnosis has been excellent and has shown great efficiency in Kerala. Ranging from neurorehabilitation, advanced monitoring of kidney-related problems, these are some of the advantages that people in Kerala are getting. Indian startups like SigTuple and large hospital chains³ have been deploying algorithms for analyzing blood samples, X-Rays, etc., with great efficiency. However, there is an even bigger picture which has been continuously ignored by the Indian regulatory authorities.

This blog addresses core issues such as arbitrary import regulations, lack of Indian datasets, inherent problems of AI medical devices, and following that, the authors strive to provide solutions by drawing from comprehensive research and international jurisprudence.

The importance of this discussion becomes clearer when AI is not viewed merely as another software product, but as a technology which can influence decisions concerning a person's

¹ Artificial Intelligence, Merriam-Webster, <https://www.merriam-webster.com/dictionary/artificial%20intelligence> (Last visited Aug. 1, 2025).

² KERALA CALLING INFORMATION & PUBLIC RELATIONS DEPARTMENT, <https://keralacalling.kerala.gov.in/?p=2584> (last visited July 12, 2025).

³ Karan Kashyap, How AI Is Impacting India's Healthcare Industry, Forbes (July 27, 2025, 8:24 PM), <https://www.forbes.com/sites/krnkashyap/2025/02/09/how-ai-is-impacting-indias-healthcare-industry/>.

body, health, and sometimes life itself. A wrong recommendation in an entertainment application may inconvenience a user, but a wrong recommendation in a medical setting can delay diagnosis, alter treatment, or create a false sense of safety. Therefore, the level of caution expected from regulators cannot be the same as that expected for ordinary digital applications. The authors believe that the central question should not only be whether an AI medical device works in a controlled environment, but whether it can work safely, consistently, and fairly when used on the diverse Indian population.

Internationally, the developing approach is to treat AI in healthcare as a lifecycle issue rather than a one-time approval issue. Ethical governance requires attention at the stage of designing the system, collecting data, training the model, validating performance, deploying it in hospitals and monitoring it after deployment. The World Health Organization has similarly emphasized that ethics and human rights should remain at the centre of the design, deployment, and use of AI for health. This approach is important for India because an AI system can change as its data, software, or environment changes, meaning that a device which was safe at the time of approval may not always remain equally reliable in a new setting.⁴

India already has several pieces of a possible governance structure. The Medical Devices Rules, 2017 regulate medical devices, while the ICMR has issued ethical guidelines specifically addressing the application of AI in biomedical research and healthcare. The problem is that these mechanisms operate across different areas and do not by themselves create a complete framework covering the entire AI lifecycle. The ICMR guidelines themselves recognise the involvement of creators, developers, researchers, clinicians, ethics committees, institutions, sponsors, and funding organisations. This makes it evident that AI healthcare governance cannot be placed on one stakeholder alone. A coordinated framework would be necessary so that responsibility is shared but also clearly identifiable.⁵

How Medical AI Devices Are Approved in India

In India, although AI devices are being implemented widely in the healthcare sector, still imports 70% of its medical devices⁶ and the manufacturing power of India is very low in this sector. Large volumes of AI devices are imported into the country, making the regulations

⁴World Health Organization, Ethics and Governance of Artificial Intelligence for Health, (2021), <https://www.who.int/publications/i/item/9789240029200> (last visited Oct. 8, 2026).

⁵Indian Council of Medical Research, Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare, (2023), <https://www.icmr.gov.in/ethical-guidelines-for-application-of-artificial-intelligence-in-biomedical-research-and-healthcare?q=ICS> (last visited Oct. 8, 2026).

⁶Markan, S., Verma, Y., Indian medical device sector: insights from patent filing trends, 3 BMJ Innov 167, 167 (2017), <https://innovations.bmj.com/content/bmjinnov/3/3/167.full.pdf>.

crucial. Delving into dataset concerns and specific challenges, it is imperative to understand the regulatory framework that governs medical AI devices in India.

The Medical Devices Rules, 2017,⁷ regulate the market of the medical device sector in India, which is also being termed as a ‘sunrise’ sector⁸ due to its growing importance. AI MedTech devices were not treated as medical devices and were not subject to a regulatory framework before 2018. ‘Software’ was included in the definition of ‘medical device’ when the amendment of CDSCO notified that 60 medical devices,⁹ connected to software, were now to be regulated under Medical Device Rules, 2017. However, there is a twist in the tale. Even after the introduction of these rules, there is still laxity shown by Indian regulatory authorities, especially in the case of Rule 63, which waives the requirement of undertaking clinical investigations for imported medical devices that do not have a ‘predicate’ device in the market. Various countries have regulatory guidelines for e.g. USA, EU etc., regarding non-device software functions. At this juncture, it is proposed that India should take an active step in the regulation of this ‘sunrise’ sector to avoid any fatal mishap in the future. The question is whether Rule 63 will apply to AI/ML medical devices, and if it is applied, then what are the complications considering the Indian population has varied immunities. In order to address all these concerns, the Indian legislature can mandate pre-marketing clinical investigations and optional post-market studies as well. This can be even made better when more localized Indian data sets would be shared with such devices, which is discussed further.

The first problem lies in drawing a clear boundary between a medical device and ordinary software. AI applications may be used for appointment scheduling, record keeping, billing, or other administrative functions, while another application may analyse a scan and suggest the possibility of a disease. Both may use AI, but their risk to a patient is clearly different. Therefore, the regulation should focus substantially on the intended medical purpose and the consequence of the output. The recent guidance of the Central Drugs Standard Control Organisation on Medical Device Software under the Medical Devices Rules, 2017 provides a useful example of this approach by discussing standalone software, embedded software, AI/ML-based software, intended use, risk-based classification, clinical evidence, and post-

⁷ Medical Devices Rules, 2017, G.S.R. 78(E) (India)

⁸ Press Information Bureau, <https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=1925776>, (last visited July 23, 2025).

⁹ Central Drugs Standard Control Organisation, https://cdsco.gov.in/opencms/resources/UploadCDSCOWeb/2018/UploadPublic_NoticesFiles/Final%20Classification%20of%20MD.pdf (last visited July 27, 2025)

marketing requirements.¹⁰

This is particularly relevant because many AI products are no longer installed only on a physical medical device. They can operate through cloud systems, mobile applications, general-purpose computers, or networks which interact with other datasets. This makes conventional device-centric regulation more difficult. A hospital may effectively be using a medical AI system through a subscription or cloud service, while the underlying model may be updated remotely by the developer. The patient may never see the physical location of the system or even know that the model has changed. A regulatory system must therefore be capable of identifying the relevant software and its intended use, even when the technology is distributed across multiple platforms.

The requirement of clinical evidence becomes even more important in this context. An AI system can show high mathematical accuracy but still perform differently in clinical practice. A model may have been trained using thousands of images and may produce an impressive statistical result, yet this does not automatically establish that the device is safe for every patient population. The World Health Organization has developed a framework focusing on training, validation and evaluation of AI-based medical devices, emphasising evidence generation from development through post-market surveillance. India can draw from this approach by ensuring that the regulatory assessment asks not only whether the algorithm performs well, but also whether the evidence supports its intended use in the population for which it is being marketed.¹¹

The same principle should apply after approval. AI-based devices may be modified, retrained, or updated over time. The United States Food and Drug Administration has recognised that AI/ML-based software raises distinct regulatory questions because of its iterative and data-driven nature and has moved toward a total product lifecycle approach. India can use this idea to ensure that substantial changes in an approved system trigger appropriate review. The aim should not be to require a fresh clinical trial for every minor software correction, as that would be impractical, but there should be a clear distinction between minor maintenance and changes

¹⁰Central Drugs Standard Control Organisation, Guidance Document on Medical Device Software under MDR-2017, CDSCO/MD/GD/MDSW/01/2026 (July 21, 2026), https://www.cdsc.gov.in/opencms/export/sites/CDSCO_WEB/Pdf-documents/Guidance-document-on-Medical-Device-Software-under-MDR-2017.pdf (last visited Oct. 8, 2026).

¹¹World Health Organization, Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation, (2021), <https://www.who.int/publications/i/item/9789240038462> (last visited Oct. 8, 2026).

which can materially alter the output of the device.¹²

Cybersecurity must also be treated as a regulatory concern because the safety of AI healthcare does not depend only on the algorithm's accuracy. Medical AI systems may be connected to hospital networks, cloud services, patient records, or other medical devices. A cyber incident can therefore compromise both confidentiality and patient safety. The evolving regulatory approach to medical software increasingly recognises cybersecurity, quality management, change management, and post-market monitoring as part of the product lifecycle. India should similarly view cybersecurity as a patient safety issue rather than merely as an information technology concern.

Lack of Indian datasets

AI devices in general, work on a set of datasets which was already provided to the algorithm; hence, what was inserted as a dataset into the AI medical device is very crucial. In the concerned case, when AI devices used in the West are introduced to the unregulated Indian markets, the effectiveness of the device can be undermined. To elaborate upon this issue, we look at some real-life examples.

New research was presented at the National Cancer Research Institute¹³, however, researchers found something peculiar with this research. This was the lack of training data of patients with darker skin. Another study published in The Lancet Digital Health¹⁴ highlighted the potential that AI holds in cancer detection and, more importantly, the images and patients that have been used to develop such algorithms that will determine the effectiveness of that machine in that field. A glaring example would be the case of Pulse oximeter¹⁵ which was even more exacerbated during the COVID-19 pandemic. Pulse oximeters are devices used to measure oxygen levels; however, it was found that these devices can misdiagnose oxyhaemoglobin in patients with darker pigmented skin, as melanin plays a huge part in these results.¹⁶ However,

¹²U.S. Food and Drug Administration, FDA Releases Artificial Intelligence/Machine Learning-Based Software as a Medical Device Action Plan, (Jan. 12, 2021), <https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan> (last visited Oct. 8, 2026).

¹³ Molly Andrews, Julia Bakker, Juanita Bawagan and Diana Cano Bordajandi, *The year 2035: What is the future of cancer research and treatment?*, The Institute of Cancer Research, (July 28, 2025, 3:14 PM), <https://www.icr.ac.uk/research-and-discoveries/cancer-blogs/detail/science-talk/the-year-2035-what-is-the-future-of-cancer-research-and-treatment>

¹⁴ David Wen, et al., *Characteristics of publicly available skin cancer image datasets: a systematic review*, 4 The Lancet 64, 64 (2022), <https://www.thelancet.com/action/showPdf?pii=S2589-7500%2821%2900252-1>.

¹⁵ Sylvia E K Sudat, et al., *Racial Disparities in Pulse Oximeter Device Inaccuracy and Estimated Clinical Impact on COVID-19 Treatment Course*, 192 American Journal of Epidemiology 703, 703-713 (2023).

¹⁶ Al-Halawani, et al., *A review of the effect of skin pigmentation on pulse oximeter accuracy*, 44 Physiological measurement, (2023).

device developers don't take into account these types of data, and it is then fed into the algorithms, ultimately being used in real-world scenarios which results in ineffectiveness. This shows a great disparity in the type of data which is being fed into the algorithms. Eventually, these devices are imported into India without proper regulations and are increasingly being used in crucial health sectors.

The dataset problem, therefore, is not simply a technical problem. It is also a problem of equality. When a model is trained mainly on one type of population, its errors may not be distributed equally. A model may appear highly accurate when the overall population is considered while performing considerably worse for a smaller group which is not represented properly in the training material. This is important in India because population diversity is not limited to one category. There can be substantial differences between communities, geographic regions, age groups, occupational exposure, nutritional conditions, and patterns of disease. A model that does well in a hospital in one metropolitan area cannot automatically be presumed to perform equally well in a district hospital serving a different population.

The same issue applies to gender and age. Certain illnesses present differently in men and women, and the availability of images or medical records from particular age groups may vary. If a model has fewer examples from elderly patients, for instance, its performance for that group may be uncertain. The problem can become more serious in paediatric medicine because children are not simply smaller versions of adults and medical conditions may present differently. It is therefore important that dataset design is connected to the intended population of the device. The question should be: on whom is this machine supposed to work, and has it actually been trained and evaluated on people who resemble that population?

There is also a difference between having a dataset and having a good dataset. Large volumes of data do not automatically produce a reliable AI system. Data can contain missing values, errors, inconsistent labels, duplicate entries, or institutional patterns which have nothing to do with the disease itself. If those patterns are absorbed by the algorithm, the system may appear successful during testing while actually relying on irrelevant signals. This means that data governance must include quality checks, clear labelling practices, documentation of how the data was collected, and proper separation between training and evaluation data. The purpose should be to prevent the model from simply learning the characteristics of the hospital or the dataset instead of learning the medical condition.

India also has an advantage in this area because of the scale at which healthcare data is generated. Public hospitals, private hospitals, laboratories, registries and digital health programmes all create different forms of information. Yet the existence of large amounts of

data does not mean that developers should have unrestricted access to it. Health data is highly sensitive, and data sharing must take into account privacy, consent, security, purpose limitation, and the risk of re-identification. The Digital Personal Data Protection Act, 2023 recognises the importance of protecting digital personal data while permitting lawful processing for legitimate purposes. This means that India has to solve two problems simultaneously: creating useful datasets for healthcare AI while preventing such datasets from becoming a source of uncontrolled collection or misuse.¹⁷

The Health Data Management Policy developed under the National Digital Health Mission also provides a useful principle because it focuses on security and privacy by design and recommends interoperable and federated approaches for digital health data. A federated ecosystem can allow different institutions to participate in a larger health data network without necessarily creating one massive central repository containing every patient's information. Such a model can be useful for AI because it can encourage collaboration while reducing some of the risks associated with centralisation. The precise architecture, however, must be supported by clear governance rules so that the use of data for AI does not go beyond the purposes for which patients reasonably expect their information to be processed.¹⁸

With data challenges and import regulations discussed, we now turn to the inherent challenges posed by AI's opaque nature and subsequently look into ways on solving it.

The Problem of Black Box

The Black Box, considered as one of the most contested issues in Artificial Intelligence, arises when AI systems, based on machine learning, makes decisions or predictions without revealing their internal working.¹⁹ In today's time, in the case of AI machines based on machine learning, their algorithm creates patterns and provides results by relying on a vast number of datasets. But in the process, it actually produces its results without revealing its internal processes, which means users are incapable of seeing the process behind the result, which has led to it. This contributes towards creating a situation of unclear liability, where it is difficult to navigate who is holding the liability for unforeseen situations in case an error occurs.

¹⁷Digital Personal Data Protection Act, 2023, No. 22 of 2023 (India), <https://www.meity.gov.in/static/uploads/2024/02/Digital-Personal-Data-Protection-Act-2023.pdf> (last visited Oct. 8, 2026).

¹⁸National Digital Health Mission, Health Data Management Policy, https://abdm.gov.in/static/media/health_management_policy_bac9429a79.80f74bc3e039c00acd4f.pdf (last visited Oct. 8, 2026).

¹⁹ Matthew Kasinski, What is black box Artificial Intelligence, IBM (Aug. 5, 2025, 15:33), <https://www.ibm.com/think/topics/black-box-ai>.

For example, in a study based on a multimodal model trained on Cancer Genome Atlas (TCGA), comprising of cancers submitted from sites across the world. it was observed that despite model's great performance, its internal processes remained opaque. It was also revealed that the model was predicting prognosis not only from disease biology, but also on the basis of other sources such as site-specific patterns and batch effects, which are not uniform everywhere, as these patterns are based on hidden institutional signatures.²⁰

Thus, there are high chances that when such models are used on other sites, there can be many variations in the end result. This shows that AI models may give results based on factors unknown to us, and due to the black box problem, we can't even have a track of them.

The black box problem therefore goes beyond the question of whether a programmer can understand the code. In many AI systems, even if the general architecture of the model is known, the exact reason for a particular output may not be obvious. This creates a major challenge in healthcare because a doctor may need to justify why a certain patient was diagnosed, referred, or treated in a particular way. If the answer is merely that the machine predicted it, the traditional relationship between evidence and medical decision-making becomes weaker. The National Institute of Standards and Technology, through its AI Risk Management Framework, treats risk management as a process covering governance, mapping, measurement, and management of AI risks. This broader approach is useful because it shifts the focus from simply trying to open the black box to creating systems in which risks can be identified and controlled even when complete technical explanation is difficult.²¹

The importance of risk-based governance can also be seen in the European Union's approach. The EU AI Act places certain systems within a high-risk category and imposes stronger requirements relating to risk management, data governance, technical documentation, record keeping, transparency, human oversight, accuracy, robustness, and cybersecurity. India need not copy the European model completely, but the underlying principle can be adopted. The higher the possible harm from an AI system, the greater should be the regulatory requirements. A tool that merely sorts appointments cannot be placed in the same category as a system which influences cancer diagnosis, identifies a stroke, or controls part of a medical device.²²

There must also be an identifiable human responsibility behind every high-risk AI decision.

²⁰ Frederick M. Howard et al., *Multimodal deep learning: An improvement in prognostication or a reflection of batch effect*, 41 *Cancer Cell* 5, 5 (2023), [https://www.cell.com/cancer-cell/pdf/S1535-6108\(22\)00522-0.pdf](https://www.cell.com/cancer-cell/pdf/S1535-6108(22)00522-0.pdf).

²¹National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, (2023), <https://www.nist.gov/itl/ai-risk-management-framework> (last visited Oct. 8, 2026).

²²Regulation (EU) 2024/1689 of the European Parliament and of the Council of June 13, 2024, *Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act)*, <https://eur-lex.europa.eu/eli/reg/2024/1689> (last visited Oct. 8, 2026).

The use of a machine should not create a situation where the hospital blames the doctor, the doctor blames the software, and the manufacturer blames the user for not following the instructions. The existing principles of medical negligence already require a healthcare professional to exercise reasonable care and skill. The Supreme Court's decision in *Jacob Mathew v. State of Punjab* emphasised the duty of care expected from medical practitioners and the distinction between an error of judgment and negligence. In the AI context, this should encourage a model of shared accountability in which the doctor remains responsible for exercising clinical judgment while the developer remains responsible for the safety and accuracy of the product within its intended use.²³

Informed consent

In India, as per a 2020 survey by the Indian Journal of Medical Ethics, many patients in rural India are not aware of the use of AI in their diagnosis and treatment.²⁴ Informed consent in the context of AI refers to information that should be furnished to a person regarding a particular test, treatment that would involve the assistance of AI-related machines and software.²⁵ Patients should be informed of the involvement of AI, especially given to many problems such treatment suffers from, such as widespread black box problem, over-reliance of medical practitioners on AI-based machines, etc.

For ex.- In a study, 27 radiologists after analyzing 50 mammograms, provided the Breast Imaging Reporting and Data System (BI-RADS) with the assistance of AI. It was observed that when AI gave correct BI-RADS suggestions for mammograms, inexperienced doctors had nearly 80% accuracy, but got reduced to less than 20% when AI was wrong. Similarly, in the case of experienced doctors, the accuracy dropped from 82% to 45.5% when AI gave wrong results. This shows how problematic over-reliance on AI is, as it is affecting even the doctors with more than 15 years of experience.²⁶

The mammography example highlights a problem which can be described as automation bias. When a machine appears to be objective and mathematically advanced, a doctor may give its

²³Jacob Mathew v. State of Punjab, (2005) 6 SCC 1, <https://indiankanoon.org/doc/871062/> (last visited Oct. 8, 2026).

²⁴ Shambhavi Naik & Bharath Reddy, *A Framework for the Governance of AI in Healthcare for India: An Exploratory Model*, Indian Journal of Medical Ethics, (May 8, 2025), <https://ijme.in/articles/a-framework-for-the-governance-of-ai-in-healthcare-for-india-an-exploratory-model/?galley=print>.

²⁵ Dariush D Farhud, Shaghayegh Zokaei, *Ethical Issues of Artificial Intelligence in Medicine and Healthcare*, 50 Iranian Journal of Public Health 1, 1-5 (2021), <https://pmc.ncbi.nlm.nih.gov/articles/PMC8826344/pdf/IJPH-50-i.pdf>.

²⁶ Applied Radiology, <https://appliedradiology.com/articles/ai-based-decision-support-can-lead-to-biases-in-mammography-reading> (Last visited Aug. 3, 2025).

recommendation more weight than it deserves. The risk is not limited to inexperienced doctors. Even experienced professionals may be influenced when the technology is integrated into a routine workflow and its recommendation is displayed before independent assessment. The concern is therefore not that doctors will stop thinking altogether, but that the threshold for questioning an AI output may gradually become higher. In high-pressure settings, this can be particularly dangerous because speed and workload may encourage acceptance of the machine's suggestion.

This also changes the meaning of informed consent. Traditionally, a patient is informed about a diagnosis, procedure, material risks, and alternatives. With AI-assisted care, the patient should additionally know when an AI system is involved in a material part of the process. The patient does not necessarily need a technical lesson on neural networks, but should understand that the recommendation is generated with the assistance of an automated system and that the doctor is expected to exercise independent clinical judgment. This becomes especially important where the AI output is uncertain, where multiple treatment choices are possible, or where the use of AI materially changes the nature of the decision.

This example clearly shows the importance of informed consent in today's AI-assisted healthcare where patients need to be aware of its involvement due to the potential risks of misdiagnosis, opaque nature, and diminished human oversight.

The Indian legal understanding of informed consent already provides a foundation on which this principle can be built. In *Samira Kohli v. Dr. Prabha Manchanda*, the Supreme Court treated informed consent as an important part of the doctor-patient relationship and discussed the need for consent to be based on adequate information about the nature and purpose of the treatment. AI-assisted treatment introduces a new element into this relationship because the process may involve an additional decision-making layer which is not directly visible to the patient. The existing doctrine can therefore be adapted so that a patient is informed not only about the medical procedure, but also about material technological assistance involved in it.²⁷ The connection with privacy is also important. Health information can reveal some of the most intimate aspects of a person's life, including illnesses, genetic characteristics, mental health information, reproductive choices, and family history. The Supreme Court in *Justice K.S. Puttaswamy (Retd.) v. Union of India* recognised privacy as constitutionally protected and linked it with dignity, liberty, and individual autonomy. In the context of AI healthcare, autonomy means more than keeping medical records secret. It also concerns the person's ability

²⁷*Samira Kohli v. Dr. Prabha Manchanda*, (2008) 2 SCC 1, <https://indiankanoon.org/doc/438423/> (last visited Oct. 8, 2026).

to understand and influence how important decisions about their body are made. The use of AI should therefore not become a hidden process which removes the patient from the decision-making relationship.²⁸

Consent should also be meaningful for patients who are not technologically literate. A consent form filled with terms such as machine learning, training data, model drift, or algorithmic bias would not satisfy the real purpose of informed consent. Information should be given in a simple form and, where required, in local languages. The patient should be told what the system does, the extent to which it assists the doctor, whether it can make or recommend a decision, and what can be done if the result appears incorrect. The objective must be to make the person capable of asking sensible questions rather than merely obtaining a signature on a form.

The right to human review should also be expressly protected in high-risk healthcare settings. If a patient is told that an AI system has produced a particular result, there should be a clear understanding that the final clinical decision is subject to review by a qualified healthcare professional. This is important not only for patient confidence but also for accountability. Human oversight cannot be reduced to a doctor clicking an acceptance button. The doctor should have the competence and practical opportunity to question the output when it conflicts with symptoms, medical history, or other evidence.

In emergency situations, however, the law should recognise that full consent may not always be possible before immediate care is provided. AI should not become a reason for delaying urgent treatment. The better approach would be to require advance disclosure through hospital policies, general consent processes where appropriate, and clear information given after the emergency when a person is in a position to understand what happened. The law must therefore remain flexible enough to protect autonomy while recognising the realities of emergency medicine.

We now turn to possible and practical solutions by providing an eagle's eye view on preceding discussions while also looking at the way forward for Indian healthcare to assimilate AI without any adversities.

²⁸Justice K.S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1, https://api.sci.gov.in/pdfdate/index1.php?dno=350712012&dt=2017-12-15&filename=supremecourt%2F2012%2F35071%2F35071_2012_Order_15-Dec-2017.pdf (last visited Oct. 8, 2026).

SOLUTIONS

1. Building Stronger Dataset Ecosystems for Indian AI

Considering the issue of lack of localised data sets, the authors are proposing a three-fold solution. First and foremost, there is a need to revamp the data collection method. Although researchers have found multiple studies about the absence of ethnic datasets in the AI devices, developers have turned a blind eye towards this, resulting in devices underperforming in different markets. Hence, localised datasets need to be created by keeping in mind the stratification of ethnicity and gender. Variety of datasets should be collected and actively used by the developers while developing the algorithms of AI/ML devices.

The first step should be to develop organised and trustworthy Indian datasets rather than expecting every individual developer to create a dataset from the beginning. Universities, teaching hospitals, research institutions, public health authorities, and private hospitals can participate in common standards for data collection. The data should be properly labelled and accompanied by information about the population from which it was collected. This would make it easier to identify whether a model has been evaluated on the kind of patients for whom it is intended.²⁹

At the same time, the collection of localised data should not result in excessive data collection. The principle should be quality and relevance rather than quantity for its own sake. A developer should not collect every possible piece of information about a patient merely because it may be useful in the future. The data required should be connected to the intended medical purpose, and the use of health information should remain consistent with the applicable privacy framework. This will help India avoid a situation where the solution to algorithmic bias creates an equally serious problem of uncontrolled health surveillance.

Secondly, India needs to enforce stricter import regulations on devices rather than taking a shortcut in such a critical field. It needs to mandate localized datasets for feeding the algorithms of such devices, ultimately ensuring that the device functions healthily. Therefore, regulation and enforcement need to be stricter.

Thirdly, the method of post-market studies which means that after deployment of such AI devices in the relevant market, monitoring should be continued to understand the performance of the device. The machine responds when there is input of more and more data, hence changing the way it functions, which will help us in understanding and collecting localised data and how

²⁹Ministry of Health and Family Welfare, National Digital Health Blueprint, (2019), <https://www.mohfw.gov.in/newshighlights/final-report-national-digital-health-blueprint-ndhb> (last visited Oct. 8, 2026).

AI reacts in a new market. This needs to be kept in mind by regulators and, therefore, should be practiced more often.

Post-market surveillance should ideally be treated as a continuing regulatory obligation for high-risk AI medical devices. The manufacturer should periodically collect information on performance, complaints, adverse events, subgroup errors, software changes, and other incidents which may indicate that the device is not working as originally intended. The regulator should have the power to require corrective action, additional testing, restrictions, or withdrawal when serious problems are identified.³⁰

Hospitals should also participate in post-market monitoring. A regulator cannot identify every problem from manufacturer reporting alone. Hospitals are the places where the technology is actually used, and doctors may be the first people to notice repeated errors or unusual results. There should therefore be a simple mechanism for healthcare professionals to report suspected AI failures, just as adverse medical events are reported for other regulated products. The reporting process should not be designed to punish a doctor for raising concerns, because fear of liability can discourage honest reporting.

2. Tackling the Black Box problem

India, to deal with the black box problem, can take certain measures, such as the regulation of datasets being made to AI systems, and also contribute towards the development of Explainable AI (XAI).

In terms of the regulation of datasets, India can opt for risk-based classification of AI systems similar to the EU, where they have classified such AI systems into 4 categories - Unacceptable, high risk, limited, and minimal, based on their risk level.³¹ This can help in regulating the quality and risk assessment of information being made available to AI systems, which can help us in mitigating the problem of black box as AI systems gives their results on the basis of their datasets only. For ex.- In the EU, AI machines posing a significant risk of harm to health have been referred to as High-Risk AI systems under Article 6 of the AI Act and would have to go through stricter obligations such as high level of robustness, cybersecurity, and accuracy, logging of activity to ensure traceability of results, etc.³²

³⁰World Health Organization, Regulatory Considerations on Artificial Intelligence for Health, (2023), <https://www.who.int/publications/i/item/9789240078871> (last visited Oct. 8, 2026).

³¹ EU Artificial Intelligence Act, <https://artificialintelligenceact.eu/high-level-summary/> (Last visited Aug. 4, 2025).

³² EU Digital Strategy, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> (Last visited 4 August, 2025).

India can adopt this idea without necessarily reproducing every category in the EU legislation. A simple Indian framework could classify healthcare AI according to the potential harm associated with its use. High-risk systems could include tools that influence diagnosis, treatment selection, emergency triage, or the operation of life-supporting equipment. Such systems should require stronger evidence, documentation, monitoring, and human oversight. Medium-risk systems could be subject to lighter but still meaningful requirements, while low-risk systems may largely rely on existing general compliance requirements. A risk-based system would also prevent regulatory resources from being wasted on technologies which pose little danger to patients.

Documentation should be a major component of this model. A high-risk developer should maintain records of training methodology, validation results, intended population, known limitations, major software changes, cybersecurity controls, and post-market findings. The purpose is not to force developers to disclose their entire source code. Instead, regulators need enough information to understand how the product was built, how it was tested, and where it is expected to fail. This information becomes especially important when an AI system causes a serious adverse event.

Human oversight should also be proportionate to risk. In a low-risk system, a doctor may simply review the output when needed. In a high-risk system, human oversight should be designed into the workflow itself. The user should be able to intervene, override an output, or stop the system where necessary. The AI system should not make it practically impossible to question its decision because the hospital has become dependent on automated processing.

Further, to counter the black box effect and build trust, Explainable AI (XAI) is also emerging as a promising solution, which refers to designing AI systems in a way that their set of processes and results created by machine algorithms can easily be comprehended and trusted by human users.³³ By focusing on the development of AI systems in an explainable manner, we can address the Black box problem effectively, as increased transparency and interpretability of such AI systems would enable users to understand the reasoning behind a decision that has caused an error.³⁴

Explainability should therefore be treated as a tool for accountability and not merely as a technical feature. The most useful explanation may differ depending on who is receiving it. A regulator may need detailed information about how the system was validated, while a doctor

³³ Violet Turri, *What is Explainable AI*, SEI Blog (Aug. 4, 2025, 11:40 PM), <https://www.sei.cmu.edu/blog/what-is-explainable-ai/>.

³⁴ Pantelis Linardatos et al., *Explainable AI: A Review of Machine Learning Interpretability Methods*, 23 Entropy 18, (2021), <https://www.mdpi.com/1099-4300/23/1/18>.

may need to know which clinical factors contributed to a particular prediction. A patient may simply need to understand that an AI system assisted the assessment, what the system was designed to identify, and whether the final decision remained with a doctor. The same AI system can therefore require several levels of explanation.

However, XAI should not be seen as a complete answer to bias or inaccuracy. An understandable explanation does not prove that the underlying model is correct. A system can provide a very persuasive explanation for a wrong prediction. Therefore, explainability must operate together with clinical validation, representative datasets, performance monitoring, and human judgment. The authors propose that India should not allow explainability to become a box-ticking requirement. The central question should always remain whether the explanation helps a qualified user detect, understand, and appropriately respond to the limits of the AI system.

To understand XAI's importance in a better way, we can refer to a skin cancer study that used an optimized Convolutional Neural Network (CNN) on a HAM10000 dataset to identify seven skin lesions and implemented Grad-CAM, Grad-CAM++ as explanation techniques, which are XAI models. The study saw that these CNN models not only provided better accuracy by highlighting disease areas, but also gave the reasoning behind the model's decision, which made AI decisions more understandable for non-technical patients and physicians.³⁵

Further, in case of absence of XAI models, if the CNN model had classified a particular lesion as melanoma, and a dermatologist, on the other hand, had classified it as merely a nevus. Then, it would have created a lot of confusion and uncertainty due to the absence of an explanation for the model's decision. But here, the model's explanation that "*if the color is constant, the classifier is less likely to declare it as a melanoma.*"³⁶ greatly helps the dermatologist in reassessing and arriving at a confirmed decision. Hence, by focusing on the development of XAI, India can make AI more transparent and trustworthy, especially in high-stakes areas such as Healthcare.

The example also shows why the doctor should not be placed in a position where disagreement with the machine is treated as an error by default. A clinical professional may have information which is not included in the model, such as symptoms, medical history, physical examination findings, or recent changes in the patient's condition. A good healthcare AI system should therefore support clinical reasoning rather than suppress it. Institutional policies should make

³⁵ Krishna Mridha et al., *An Interpretable Skin Cancer Classification Using Optimized Convolutional Neural Network for a Smart Healthcare System*, 11 *IEEE Access* 41003, 41003-41018 (2023), <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10107401>.

³⁶ *ibid*

it clear that a doctor is permitted to reject an AI recommendation where professional judgment indicates that it is unreliable.

There should also be an internal process for repeated disagreements between doctors and AI systems. If clinicians repeatedly override a product because its recommendations appear inaccurate in a particular hospital or patient group, that pattern is itself useful evidence. Hospitals should record such incidents and, where necessary, report them to the manufacturer or regulator. This would create an additional feedback mechanism through which AI systems can be evaluated using real-world clinical experience rather than depending only on laboratory metrics.

Trust should therefore be based on evidence, not on the appearance of sophistication. An AI system may be highly advanced technically and still be unsuitable for a particular clinical setting. The combination of explainability, clinical evidence, performance logs, and human oversight can create a more realistic form of trust. The patient should trust the healthcare process because there are mechanisms capable of detecting and correcting errors, not because the machine is presented as infallible.

3. Codifying Consent

One of the key-issue in the integration of AI with healthcare lies in the fact that India lacks defined provisions for informed consent in situations with involvement of AI in the treatment. This problem extends to other countries such as the USA, EU, who not addressed the problem completely, as they also don't have specific legislations and provisions directly governing informed consent in cases of AI-assisted healthcare.³⁷ Although EU under its Article 22 of the GDPR has tried to address the problem as it seeks to protect people from being subject to decisions that are solely based on automated processing, and which can have legal effects for the individual.³⁸ However, the article does not take into consideration the cases where the AI machines are assisting medical practitioners, as it only takes into consideration the scenario where the decisions are taken solely by AI.

India can address this gap by expressly recognising AI involvement as part of the information that may become material to a patient's consent. The legislation need not prescribe one universal consent form for every AI tool. Instead, it can define circumstances in which disclosure is mandatory. These could include situations where AI materially contributes to

³⁷ Julia Amann et al., *Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective*, 20 BMC Medical Informatics and Decision Making 310, (2020), <https://bmcmmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6>.

³⁸ General Data Protection Regulation, <https://gdpr-info.eu/art-22-gdpr/> (Last visited 4 August, 2025).

diagnosis, determines a risk score that affects treatment, recommends a course of treatment, or operates a device which can significantly affect patient safety. For low-risk administrative uses, the same disclosure requirement may not be necessary.

A consent framework should also distinguish between AI assistance and autonomous decision-making. In many current healthcare environments, AI is intended to support a doctor rather than replace the doctor. The law should reflect that distinction while ensuring that the presence of human oversight is genuine. If a hospital describes a system as a decision-support tool but its doctors are effectively required to follow its output because of institutional policy, then the formal description of the technology does not match the practical reality. Regulation should therefore focus on how the system is actually used.

A useful mechanism would be to require hospitals to maintain an AI register identifying the high-risk tools being used, their intended purpose, the manufacturer, the date of the latest validation, and the person or committee responsible for oversight. This would make the use of AI more transparent within an institution. It would also help patients and regulators understand which systems are being used in practice instead of relying only on information supplied by individual doctors.

Thus, the author suggests that India should introduce such provisions and legislation that mandates medical practitioners to inform patients regarding the involvement of AI in their tests and treatments in a complete way for their informed decision-making.

Any new legal provision should also clarify responsibility among the developer, hospital, and healthcare professional. The allocation cannot be identical in every case because the source of the failure may differ. A developer may be responsible where the device contains a design defect or the manufacturer failed to disclose a known limitation. A hospital may bear responsibility where it deployed a system without adequate validation or staff training. A doctor may be responsible where the doctor ignored clear clinical evidence or used the system outside its intended purpose. A clear framework would allow liability to follow the actual source of the harm instead of creating automatic blame for the person closest to the patient.

At the same time, the law should avoid creating an incentive for doctors to completely avoid AI. Excessive personal liability for using a regulated device may discourage clinicians from adopting technologies which could genuinely improve healthcare. The better approach is to require reasonable verification and oversight. Where a doctor has acted within the accepted medical practice, used the device for its intended purpose, considered the patient's condition independently, and followed appropriate safety procedures, the mere fact that an AI prediction turned out to be wrong should not automatically make the doctor liable.

Hospitals should therefore develop internal governance frameworks for AI use. An AI system should be approved for a specific use, assigned an institutional owner, evaluated for staff training requirements, and periodically reviewed for performance. Ethics committees and clinical governance bodies can have an important role in this process. The ICMR guidelines already recognise governance, ethics review, and informed consent as important components of responsible medical AI. India can build on these principles by making institutional AI governance a practical part of healthcare management rather than a voluntary afterthought.

The final purpose of such legislation should be accountability without fear. Patients should have a meaningful remedy when an AI system causes harm, but healthcare professionals should also be able to use useful technologies without believing that every machine error will expose them to automatic punishment. Developers, hospitals, and doctors should each have clear obligations which correspond to their actual control over the technology.

4. AI Medical Education Integration

In a country as populous and healthcare-dependent as India, the base of the healthcare sector are the students in the medical field. Since AI still has some time to catch up, we can make our country AI-ready so they aren't unprepared. The upcoming generation is crucial for the assimilation of AI in this sector, which seems inevitable in the near future.

There was a survey done in Kerala wherein around 20,000 medical students participated.³⁹ As per that research, many students showed a keen interest in AI, however, one interesting fact was that only a handful of them (39.8%) understood the nitty-gritty of AI, and more than half of them agreed that they lacked knowledge regarding the limitations of AI. This poses a serious question, how will the sector perform if the one who uses it has no idea about the complications of the device. Hence, medical education should be mandatorily included, especially in those medical fields wherein AI is already being used. This would result in more effective use of the machine as well as a growth in the trust factor between the machine and its operator which would ultimately benefit the patients and the market.

Medical education is therefore one of the most practical safeguards available to India. A doctor does not need to become a computer scientist, but a doctor using AI should understand the basic principles of how the system works, what kind of data it depends on, where bias can arise, and when the output may be unreliable. Training should also cover the difference between correlation and causation, the risk of automation bias, privacy responsibilities, cybersecurity,

³⁹ K, J. G., & Singh, P. (2024). Attitude towards artificial intelligence among medical students in Kerala. *E-Learning and Digital Media*, 0(0). <https://doi.org/10.1177/20427530241276138>

and the procedure for reporting an AI-related incident.

CONCLUSION

As India moves steadily toward the integration of AI in healthcare, it must do so with caution, foresight, and responsibility. While AI holds great promise ranging from enhanced diagnostics to life-saving surgeries, it is equally important to address the glaring challenges some of which are discussed here. By adopting international best practices, investing in indigenous data infrastructure, and equipping the next generation of medical professionals with AI literacy, India can not only safeguard its population but also emerge as a global leader in ethical and effective medical AI deployment.

The path forward is therefore not a choice between regulation and innovation. India needs both. AI can increase access to expertise, improve early diagnosis, support overloaded doctors, and create new forms of personalised healthcare. At the same time, these benefits will only be sustainable when patients can trust the systems being used on them. Trust will depend on whether the devices are properly approved, the data is representative, the reasoning can be meaningfully reviewed, consent is respected, and there is a clear mechanism to identify and remedy harm.

