LEGAL AND ETHICAL IMPLICATIONS OF ARTIFICIAL INTELLIGENCE IN CLINICAL DECISION-MAKING

Rohan Abraham Daniel, School of Law, Christ (Deemed to be University)

ABSTRACT

Artificial Intelligence (AI) is rapidly changing the landscape of modern healthcare, most profoundly in the area of clinical decision-making. From diagnostic imaging using deep learning technology to natural language processing tools that query electronic health records, AI systems are now actively participating in the field of healthcare, including the detection of diseases, recommendation of treatment pathways and prediction of patient outcomes. This technology has promised enhanced precision, reduced clinician burden, greater accessibility, and improved level of health outcomes. Yet their integration into the clinic generates equally substantial legal and ethical tensions. Some of the main concerns and questions raised are of liability allocation when AI contributes to a misdiagnosis, the adequacy of the informed consent where algorithmic logic is used, the systemic injustice embedded in biased training data, and the jurisdictional fragmentation of regulatory oversight remain unresolved in law and contested in ethics. This article undertakes a comprehensive examination of these tensions and analyses the existing legal frameworks across key jurisdictions including the United States, the European Union (EU), and India and evaluates their adequacy in addressing AI specific challenges such as adaptive learning and black box decision making architecture and looks at the foundational bioethical principles in the context of algorithmic medicine. The main argument is that AI has to play an auxiliary role rather than being an alternative to human clinical judgment, but that in turn calls for a recalibration of the doctrine of liability, informed consent, and accountability.

Keywords: Artificial Intelligence, Clinical Decision Making, Medical Liability, Informed Consent, Healthcare Regulation, Data Protection, Liability.

1. Introduction

The mixing of artificial intelligence and clinical medicine represents one of the most important technological developments of the twenty first century. AI systems are broadly defined as computational systems capable of performing tasks that ordinarily require human intelligence, including pattern recognition, inference, and prediction. These systems have been increasingly rooted in the operational infrastructure of modern healthcare. Machine learning algorithms can now inspect radiological images for early markers of malignancy. Natural language processing tools can now take out clinically relevant signals from unstructured patient records. Predictive analytics engines can calculate patients by risk of deterioration or the need for hospital readmission. The transformative potential of these technologies is not theoretical and large-scale clinical studies have demonstrated that AI models can match or even in some cases exceed specialist level performance in specific diagnostic tasks which includes the detection of diabetic retinopathy, skin cancer classification, and the identification of pneumonia from chest radiographs.[1]

Yet this integration is not merely a technical or clinical phenomenon, it is a legal and ethical event of significant magnitude. The introduction of a non-human decision making participant into the medical encounter disrupts legal frameworks that have, for centuries, promised medical liability on the individual professional's exercise of clinical judgment.[2] It challenges ethical structures which are based on the trust relationship between physician and patient and raises urgent questions about who bears responsibility when an AI influenced decision results in patient harm. These are questions that existing tort doctrine, product liability regimes and professional disciplinary frameworks are currently ill equipped to resolve.[3]

Compounding these doctrinal difficulties are features peculiar to advanced AI systems. Unlike conventional medical devices like a surgical instrument or an infusion pump, modern machine learning models are not static artefacts. They may evolve post their deployment through continued training on new data, meaning the system that receives regulatory approval may differ materially from the system operating at the point of patient harm.[4]Their internal logic

[1] Andre Esteva in, 'Dermatologist-level classification of skin cancer with deep neural networks' (2017) 542 Nature 115; Varun Gulshan in., 'Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs' (2016) 316 JAMA 2402.
[2] W. Nicholson Price II, 'Regulating Black-Box Medicine' (2017) 116 Michigan Law Review 421, 425–430.
[3] I. Glenn Cohen et al., 'The Legal and Ethical Concerns That Arise from Using Complex Predictive Analytics in Health Care' (2014) 33 Health Affairs 1139.
[4] I. Glenn Cohen et al., 'The Legal and Ethical Concerns That Arise from Using Complex Predictive Analytics in

may be effectively impenetrable even to their designers. This is the so-called 'black box' problem making it impossible to reconstruct the reasoning pathway behind the harmful output. They may encode, amplify, or generate discriminatory patterns derived from historical injustice in medical data, producing systematically inferior outcomes for already marginalised patient populations.

This article proceeds in eight substantive parts. Following this introduction, Part II talks about AI clinical decision support within a technical and taxonomic framework. Parts III and IV analyse the legal and ethical implications based on these issues, with a more focused attention to liability, privacy, regulatory gaps, justice, and transparency. Part V examines the impact of AI on the doctor and patient relationship. Part VI is a comparative analysis of the regulatory approaches across the United States, the European Union (EU), and India. Part VII deals with the challenge of innovation against patient protection. Part VIII concludes with concrete proposals in relation to these problems.

2. Understanding AI in Clinical Decision Making: A Technical and Taxonomic Framework

Before we go on to discuss the legal and ethical implications, it is necessary to understand with some degree of specificity as to what exactly AI in clinical decision making entails. The term 'artificial intelligence' is a very broad term and it includes a very large number of things and the legal and ethical importance of a particular AI application would depend on its architecture, function, and the level of autonomy it has.

2.1 Technical Architecture

The dominant pattern in using contemporary medical AI is supervised machine learning, and within that, deep learning using convolutional neural networks (CNNs) for image based tasks and transformer architectures for text based clinical applications[5]. These models are trained on large, labelled data sets to identify statistical associations between the input features and the designated outcomes. The quality, size and representativeness of training data are therefore determinative not only by the performance of the model but also the patterns of error and bias

Health Care' (2014) 33 Health Affairs 1139.

[5] Eric J. Topol, 'High-performance medicine: the convergence of human and artificial intelligence' (2019) 25 Nature Medicine 44.

the model will exhibit after its deployment.

A critical distinction for legal and regulatory purposes is between closed systems, especially those with fixed parameters post deployment and also adaptive or continuously learning systems, which update their parameters in response to the new data which it encounters during clinical use. Closed systems are more amenable to pre-market validation while adaptive systems, by contrast, present a moving regulatory target: the validated version and the deployed version may become substantially different over time, potentially in ways that introduce new failure modes or biases due to the new information which is being introduced to the machine.

2.2 Functional Taxonomy

AI clinical tools may be categorized in two types: the degree of automation (decision and support given versus autonomous decision) and the clinical function performed (diagnostic, prognostic, therapeutic, or monitoring). Most currently deployed systems occupy the decision to support the end of the automation spectrum i.e. they generate outputs, risk scores, diagnostic probabilities, treatment suggestions and then these are presented to a human clinician who retains the final decisional authority. However, the trend towards greater automation is clear and visible, and fully autonomous systems which can act on patients without clinician intermediation are emerging in areas such as autonomous screening programs and in closed loop treatment delivery.

This taxonomy is of extreme importance from the legal point of view as in a decision support model, the physician has the knowledge and the responsibility for the ultimate decision, and the AI is simply an input, not a determinant. In the autonomous model, that framework becomes deeply strained and the attribution of responsibility in the case of a bad outcome must be derived from the already existing principles. The legal frameworks discussed in this article are in a large part still calibrated towards the decision support model, even if the reality of deployment extends beyond this.

3. Legal Implications of AI in Clinical Settings

3.1 The Liability Problem: - Reframing Medical Malpractice

Medical malpractice law, as it has evolved across common law jurisdictions, rests on a tripartite structure which is: a duty of care which is owed by the clinician to the patient, a breach of that

duty measured against the standard of a reasonably competent practitioner, and attributable harm which is. This framework requires a human professional at the center of the clinical encounter, whose judgment, skill, and knowledge can be assessed against a set professional standard. The integration of AI into that encounter starts to de-establish each element of this structure in multiple distinct ways.

Consider first the standard of care, as AI diagnostic tools achieve levels of accuracy that match or in some cases exceed human specialists, there is a credible argument which is given by scholars such as Price and Cohen who say that the standard of care may evolve to incorporate AI assisted practice. If for example a reasonably competent radiologist would use a validated AI system for a mammography screening, a radiologist who declines to use such a system and misses a lesion which was detectable by AI may face enhanced liability for failing to employ the available technology. Similarly, a clinician who defers to an AI recommendation without exercising independent clinical judgment may equally face liability for not following professional responsibility.

The causation element in this presents difficulties, majorly in establishing that an AI system's output caused a patient's harm requires both attribution (that the physician had relied on the AI) and a factual analysis (that a correct AI output would have led to a different clinical decision and outcome). When dealing with complex clinical decisions and probabilities provided by the AI as output, this causal chain is hard to reconstruct and may even be disputed with expert opinions from multiple sides.[6] Courts have yet to develop a settled doctrine for this analysis, and the opacity of many AI systems which do not have a method to articulate the reasoning behind their outputs compounds the main evidential problem.

Liability may also be distributed across a broader set of actors than traditional malpractice usually follows. When an AI system contributes to a harmful outcome the responsibility may attach to the treating clinician who relied on the output, the healthcare institution that selected, deployed, and was maintaining the system, the developer or manufacturer of the AI software, the person or company which helped in the training data of the AI and in some regulatory frameworks, a certifying body that approved that the system as safe. This diffusion of the potential defendants and the absence of a clear allocation rule creates strategic incentives for

[6] See David Hyman and Charles Silver, 'Medical Malpractice Litigation and Tort Reform: It's the Incentives, Stupid' (2006) 59 Vanderbilt Law Review 1085.

each party to externalise the responsibility. This increases the risk that patients are left without an effective remedy.

3.2 Product Liability and the Challenge of Adaptive AI

Where a clinician has not been negligent but the AI system itself has produced a defective output, product liability doctrine offers an alternative avenue for a patient's redress. In many jurisdictions AI healthcare tools are classified as medical devices and are therefore subject to product liability principles.[7] Under both the U.S. products liability framework and the EU's Product Liability Directive it is held that manufacturers may be strictly liable for harms caused by defective products. A defect may arise in design, manufacturing or through a failure to provide sufficient warnings regarding the known risks.

AI systems introduce complications at every level of this analysis[8]. Firstly, the very concept of a defect can be contested when applied to a probabilistic system. An AI model that performs at 95% accuracy is not defective in any conventional engineering sense, yet it will generate harmful outputs in 5% of cases which can be considered as defective. Secondly, the adaptive learning problem means that a system may have been non-defective at the time of manufacture but may have evolved into a defective state through post deployment learning. Under the traditional product liability doctrine, manufacturers are not ordinarily liable for harms arising from subsequent modifications which are made to their products, but it is less clear as to whether a manufacturer can avoid liability for harms arising from the foreseeable evolution of an adaptive system which it deliberately designed to learn.

Thirdly, the black-box problem imposes serious evidential burdens on plaintiffs.[9] Demonstrating that a product is defective ordinarily requires showing what the product did and how it did it. Where an AI system's internal reasoning is effectively inaccessible in nature, establishing that a particular output was the product of a design defect rather than an inherent and acceptable probable error may be impossible without accessing the training data, model architecture, and validation records that the developers typically treat as proprietary.

[7] Council Directive 85/374/EEC of 25 July 1985 on liability for defective products [1985] OJ L210/29. See also Restatement (Third) of Torts: Products Liability § 2 (Am. Law Inst. 1998).
[8] Gary E. Marchetti, 'Artificial Intelligence: The Future of Medicine and the Challenge to the Practice of Law' (2019) 48 Stetson Law Review 465, 480–482.
[9] See generally W. Nicholson Price II and I. Glenn Cohen, 'Privacy in the Age of Medical Big Data' (2019) 25 Nature Medicine 37.

Legislative intervention to mandate meaningful disclosure of AI system characteristics and how they make a decision may therefore need to be a prerequisite for effective product liability enforcement within this domain.

3.3 Data Protection, Privacy, and the Architecture of Medical AI

Medical AI systems are at their foundation data systems. Their development requires large quantities of patient health information and their deployment generates additional streams of sensitive data. The legal regulation of personal health data therefore intersects with the governance of AI at multiple points.

In the European Union the General Data Protection Regulation (GDPR) establishes the foundational framework.[10] Health data is classified as a special category and is subject to much stricter protection, and processing is in principle prohibited unless one of the specified exceptions applies. Article 22 of the GDPR confers on individuals the right to not be subject to solely automated decision making that produces legal or significant effects, which has direct relevance to autonomous AI diagnostic systems. The right to explanation for automated decisions provides at minimum a normative basis for demanding that AI systems can be capable of providing intelligible accounts at their own accord.

For instance, the Digital Personal Data Protection Act, 2023 (DPDPA), which is a recent legislative development in India that establishes data protection requirements for the processing of personal data.[11] However, there is no specific provision in the DPDPA relating to health data and AI, and as far as secondary legislation is concerned, it is in its infancy. The lack of specific health data governance is a notable gap in patient protection, especially considering India's ambition to become a significant market for AI driven telemedicine services.

3.4 Regulatory Architecture and Its Gaps

The regulatory challenge which is posed by AI medical tools is acute. Traditional pre-market regulatory patterns for medical devices are poorly adapted to systems that learn, evolve and operate across multiple clinical environments. The U.S. Food and Drug Administration has acknowledged this challenge and has been developing its AI/ML-Based Software as a Medical

[10] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L119/1, arts 9, 22.
[11] Digital Personal Data Protection Act 2023 (India), s 4(1).

Device (SaMD) Action Plan, which contains a predetermined change control plan allowing adaptive AI systems to evolve within the pre-approved parameters.[12] However, this framework is still in development, and the current oversight of post market AI performance remains as a significant gap.

The European Union has advanced what is currently the most comprehensive AI-specific regulatory instrument, the EU AI Act which was adopted in 2024, which classifies medical AI systems as high risk and subjects them to multiple assessments, transparency requirements, human oversight obligations, and also post market monitoring mandates.[13] The Act's risk based approach as it distinguishes between unacceptable, high, limited, and minimal risk AI applications provides a very coherent structural framework, although its effective implementation will depend on the capacity of national assessment bodies and the resources available for post market surveillance.

4. Ethical Implications of AI in Clinical Decision Making

4.1 Autonomy, Informed Consent and the Disclosure Problem

Respect for patient autonomy is a cornerstone of bioethical thought and its legal counterpart as emphasized in the doctrine of informed consent. Both are based on the idea that patients are entitled to make decisions about their own medical care on the basis of adequate information, communicated in a form they can understand, and free from any undue influence. The integration of AI into clinical decision making creates new and unresolved tensions within this framework.

The threshold question is whether the use of AI constitutes information which is material to patient consent. In jurisdictions which apply a reasonable patient standard including the United Kingdom after Montgomery v Lanarkshire Health Board[14] a clinician is obliged to disclose any information that a reasonable patient in their position would consider significant. There are grounds for concluding that the involvement of an AI system in diagnosis or treatment recommendation crosses this threshold, Patients may hold views about algorithmic medicine

[12] FDA. See also FDA, 'Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device' (2019).
[13] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) [2024] OJ L1689/1, arts 6, 9–15.
[14] Montgomery v Lanarkshire Health Board [2015] UKSC 11, [87] (Lords Kerr and Reed).

that affect their choices and the identifiable limitations of specific AI systems known error rates, different demographic performance disparities and opacity of reasoning are features which can make a reasonable patient wish to consider.

Nevertheless, an effective disclosure of the characteristics of the AI system creates the problem of how to effectively communicate the information to patients who may not have the literacy skills to understand it. This makes the ethical duty not to simply provide the raw technical documentation, but to effectively communicate to patients the relevant material characteristics of the AI system, what it is used for, what it can do, what it cannot do, the level of confidence that is placed in it, and to what extent the doctor's judgment is being informed or shaped by what the AI recommends.

4.2 Beneficence, Non-Maleficence, and the Performance Validation Imperative

The twin principles of beneficence are acting in the patient's best interest, non-maleficence and doing no harm. This created a system which has an ongoing obligation of performance validation on all actors responsible in the AI clinical ecosystem. It is possible to design an AI system that has high average performance but is systematically harmful to certain sub-populations. It is equally possible to design an AI system that is accurate at the time of initial deployment but over time becomes less accurate as clinical practice, patient population, or disease prevalence change. The ethical obligation to do good and avoid harm is not met by a pre-market validation exercise but by ongoing surveillance, re-validation, and in the event of performance degradation, system modification or removal.

A morally distressing aspect of the present AI healthcare environment is the unbalanced distribution of incentives for the application of AI technology. The application of AI technology is accompanied by strong incentives for the developer and healthcare institution, whereas the incentive for safety surveillance after application is very low, especially when the cause of an adverse event cannot be attributed to the AI technology itself but rather to individual clinical decision making. The need for mandatory performance surveillance after application, including reporting obligations and access to adverse event data, is both ethical and legal in nature.

4.3 Algorithmic Bias, Structural Justice, and Health Equity

According to the Rawlsian sense, the principle of justice requires the fair distribution of social

goods[15]. This is placed under acute pressure by the phenomenon of algorithmic bias in AI's used for clinical purposes. Here, bias does not imply any preconceived prejudice, but rather a systematic and disproportionate error in certain predefined demographic groups due to shortcomings in the training data, model design, or evaluation approach.

The empirical evidence of significant algorithmic bias is substantial and alarming. The landmark study by Obermeyer et al in 2019 showed that a commonly used commercial differentiation algorithm consistently underestimated the health care need of Black patients in comparison with White patients with the same health status, as the algorithm used healthcare costs as a measure of need and healthcare costs are a reflection of differential healthcare access.[16]In a study conducted in 2018 by Buolamwini and Gebru, it was found that facial recognition technology was significantly poorer for darker-skinned females than for lighter-skinned males.[17] In the field of dermatology, AI diagnostic tools trained predominantly on images of lighter skin have demonstrated reduced diagnostic accuracy for patients with darker skin pigmentation.[18]

One should understand the importance of structural aspects of algorithmic bias. Algorithmic biases in AI systems do not stem from ill intent, they stem from data that reflects the biases of social and medical systems in which it was created. An AI system that is trained on historical clinical data without any specific intervention will reflect, replicate, and even intensify historical biases in how it approaches diagnosis and treatment, particularly among those who already suffer a disadvantage in accessing health care. The use of biased AI systems, therefore, has the potential not only to not improve health disparities but also to worsen them.

However, to respond meaningfully to algorithmic bias, it is necessary to implement interventions at different levels, including the diversification of training datasets, disaggregated evaluation of AI systems' performances across different demographic groups as a mandatory part of the approval process, representation of communities impacted by algorithmic bias in AI development and validation as well as structural mechanisms of accountability that recognize

[15] John Rawls, A Theory of Justice (Belknap Press 1971) 52–65; Norman Daniels, Just Health: Meeting Health Needs Fairly (Cambridge University Press 2008).
[16] Ziad Obermeyer et al., 'Dissecting racial bias in an algorithm used to manage the health of populations' (2019) 366 Science 447.
[17] Joy Buolamwini and Timnit Gebru, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' (2018) 81 Proceedings of Machine Learning Research 1.
[18] Adewole S. Adamson and Avery Smith, 'Machine Learning and Health Care Disparities in Dermatology' (2018) 154 JAMA Dermatology 1137.

algorithmic bias as a legally cognizable harm.

4.4 Transparency, Explainability, and the Black-Box Problem

The opacity of advanced AI systems, particularly deep learning system represents an unique challenge to the norms and expectations of professional accountability in medicine. The norms and expectations of professional accountability in medicine have traditionally been grounded in the assumption that a physician should be able to explain the rationale or reasons underlying a particular decision or action to the patient, to other physicians, and to the courts. A physician who is unable to explain why she or he made a particular diagnosis or prescribed a particular course of action is professionally and legally vulnerable under negligence. The challenge raised by black-box AI systems is whether the norms and expectations of professional accountability should or can be extended to such systems.

The emerging field of Explainable AI (XAI) addresses precisely this challenge. Techniques such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive explanations)[19] aim to generate human-intelligible explanations for AI outputs by identifying, for instance, which features of a radiological image most strongly influenced a diagnostic classification, or which patient variables most heavily weighted a risk stratification decision. These tools enable physicians to check AI outputs against their own clinical knowledge and to provide patients and courts with accounts of AI influenced decisions.

Additionally, it is not clear that the degree of explainability is not dependent on situation, with high level probability outputs being adequate in routine low stakes screening but life altering treatment decisions requiring more fine grained explanation.

5. The Doctor-Patient Relationship in the Age of Algorithmic Medicine

The relationship between physician and patient is created not only by the exchange of diagnostic and therapeutic information but by empathy, trust, and mutual recognition that are foundational to its function. Research in the sociology and psychology of medicine consistently demonstrates that the quality of the therapeutic alliance independently predicts health outcomes across a range of conditions above and beyond the technical quality of treatment.

[19] Scott M. Lundberg and Su-In Lee, 'A Unified Approach to Interpreting Model Predictions' (2017) 30 Advances in Neural Information Processing Systems 4765.

The incorporation of AI in clinical encounter raises several potential risks in terms of the relational aspects of the encounter. In the case of significant investment of cognitive and temporal space in the interface of AI by the physicians, the space for engaging the patient may be diminished. 'Alert fatigue' is a phenomenon which has been a recognized issue in the context of electronic health records and the incorporation of AI generated recommendations has the potential for diminishing the level of clinical vigilance when such recommendations are common and considered reliable.

There is also the risk of automation complacency which is documented as the tendency for human operators to reduce active oversight of processes they perceive as reliable and automated.[20] In a clinical setting, this can be evidenced by a decreased likelihood of questioning the output of an AI, a decreased level of engagement with the patient's history and clinical findings, and a decreased level of personal responsibility. This is particularly problematic because an AI will inevitably make errors in precisely those cases which it has encountered the least, the rare diagnosis, the patient whose characteristics differ from the norm; these are precisely when these cases need the most clinical vigilance.

The flip side of these risks is that, used with keen awareness in both relational and technological aspects, AI could potentially strengthen the doctor-patient relationship by freeing medical professionals from time consuming administrative and analytical work leaving them with more time to focus on patient engagement. AI tools that can assist with administrative tasks, data retrieval, and even basic risk assessment may allow medical professionals to commit themselves even more to the relational and empathic aspects of medical work. The impact of AI on the doctor-patient relationship will depend on how it is used.

6. Comparative Regulatory Approaches: United States, European Union, and India

6.1 The United States: FDA Oversight and the SaMD Framework

In the United States the primary regulatory authority over AI clinical tools is the Food and Drug Administration also known as the FDA and they classify AI based clinical decision support software meeting specified criteria as specified for a Software as a Medical Device

[20] Raja Parasuraman and Victor Riley, 'Humans and Automation: Use, Misuse, Disuse, Abuse' (1997) 39 Human Factors 230.

(SaMD). Under the 21st Century Cures Act and subsequent FDA guidance.[21] Clinical decision support software is subject to regulatory oversight where it is 'not intended to replace or inform clinical judgment' and where its underlying methodology is non-transparent to the intended user. This test has attracted criticism for its ambiguity and for potentially excluding from oversight some of the highest-risk AI applications.

In the proposed regulatory framework by the FDA for AI/ML-based SaMDs, the term "predetermined change control plan" is introduced, which would allow the manufacturer of adaptive AI systems to make changes to their systems even after approval within a predetermined range of changes to the algorithms used in the system, without requiring a new pre-market submission for each iteration. Although this approach accommodates the technological reality of continuously learning systems, some critics have argued that the envelope of allowable change may be defined too liberally, and that the post market surveillance approaches may not be sufficient to detect a degradation of system performance in a clinical setting.

6.2 The European Union: The AI Act and the MDR

The EU's regulatory architecture for clinical AI operates at the intersection of the Medical Devices Regulation (MDR) and the recently adopted AI Act. The MDR subjects AI-based medical devices to certain assessment requirements, including clinical investigation mandates, quality management system obligations, and post market clinical follow up requirements. The AI Act adds a horizontal layer of requirements specific to high risk AI systems, including mandatory transparency, human oversight, accuracy and robustness standards, and cybersecurity requirements.

Another interesting feature of the EU’s system is the focus on individual rights in relation to AI decision-making. The interplay between Article 22 of the GDPR’s prohibition on purely automated decision making and the AI Act’s provisions regarding human oversight creates a system in which completely autonomous AI clinical decision making is not permissible, but in which individual rights to challenge AI-influenced clinical decision making do exist.

[21] 21st Century Cures Act, Pub L 114-255 (2016), s 3060; FDA, 'Clinical Decision Support Software: Guidance for Industry and FDA Staff' (2022).

6.3 India: Emerging Frameworks and Legislative Gaps

India's regulatory landscape is unique as it is a critical market for AI driven diagnostic tools and a large producer of healthcare AI. It thus has a vested interest in developing a critical regulatory framework. While the DPDPA introduces the basic data protection principles, there is no dedicated AI regulation framework in India. Medical devices are governed by the Central Drugs Standard Control Organisation (CDSCO) under the Medical Devices Rules 2017,[22] but the classification criteria and the assessment processes for AI based clinical tools remain insufficiently developed.

The Indian government has released a National Health Stack policy framework and a National AI Strategy (NITI Aayog's AI for All)[23], which acknowledged the transformative potential of AI in healthcare, particularly for increased specialist level diagnostic capability in rural populations. However, these policy documents have yet to be applied to any specific AI clinical tool regulations. In the meantime, Indian patients do not enjoy the legal protection available to their counterparts in the EU and the United States. The development of a comprehensive, contextually relevant AI health regulation framework is an issue of urgency.

6.4 The Case for International Harmonisation

Clinical AI by its very nature is a transnational phenomenon because the data that is used to train the model is gathered from around the world. The model is developed by organisations that operate in many countries, and the model will be used in healthcare systems that have different regulatory requirements. Jurisdictional fragmentation creates opportunities for regulatory arbitrage and the deliberate deployment of AI systems in less stringent regulatory environments. A degree of international regulatory convergence, facilitated through the World Health Organization[24] or the International Medical Device Regulators Forum (IMDRF), would reduce these risks and provide a more coherent governance foundation for global health AI.

7. Balancing Innovation and Protection: Towards a Principled Regulatory Architecture

There is an ongoing balance in the regulation of emerging technologies between the danger of

[22] Medical Devices Rules 2017 (India), r 2(s), notified by the Ministry of Health and Family Welfare.
[23]NITI Aayog, 'National Strategy for Artificial Intelligence: #AIforAll' (2018) <https://www.niti.gov.in/sites/default/files/2023-03/NationalStrategy-for-AI-Discussion-Paper.pdf> accessed 1 March 2025.
[24] World Health Organization, 'Ethics and Governance of Artificial Intelligence for Health: WHO Guidance' (WHO Press 2021) 43–47.

over regulation that stifles useful innovation too early, versus the danger of under regulation that permits harm to develop before adequate regulation is put in place. With respect to clinical AI, the balance is particularly delicate, as the advantages of a AI assisted medicine, in terms of improved diagnostic accuracy, earlier disease detection, reduction of burnout among healthcare professionals, and expansion of specialist healthcare to underserved populations, are considerable and the cost of over caution is expressed in terms of preventable morbidity and mortality.

A risk proportionate approach for the regulatory framework provides a constructive solution for the tension between these two values. In a risk proportionate approach, the pre-market validation, transparency, explainability, and post-market regulatory requirements are aligned with the level of risk associated with the specific AI application. This is determined by a structured assessment of the consequence of error, the degree of human oversight retained, and the vulnerability of the affected patient population. Lower risk AI tools may appropriately be subject to lighter oversight. High risk tools, those influencing diagnostic or treatment decisions for serious conditions, in settings with limited clinician oversight, or for vulnerable populations should be subject to more stringent requirements.

Regulatory sandboxes, or environments in which new AI applications can be tested in real world clinical environments under the supervision of regulators, present an opportunity to balance innovation with regulation. Several regulatory environments, including the UK by way of its Medicines and Healthcare products Regulatory Agency (MHRA), are investigating the concept of sandboxes in relation to digital health technologies. If well designed, such mechanisms hold promise in speeding the development of real world data regarding the performance of AI technologies, as well as highlighting failure modes before widespread rollout and fostering the requisite regulatory expertise in an ever changing world.

8. Conclusion: Towards Responsible AI in Clinical Medicine

Artificial intelligence in clinical decision making represents a transformative development, one that carries profound implications not only for the technical practice of medicine but also for its ethical commitments, its legal structures, and its social function. This article has tried to map the principal legal and ethical tensions generated by this transformation, and to evaluate the adequacy of existing frameworks for addressing them. The analysis done in this article yields several overarching conclusions.

Firstly, the existing legal regimes in tort law, product liability, data protection law, and medical devices are not entirely suited to address the specificities of clinical AI, including its adaptivity, lack of transparency, and potential for bias, as well as the allocation of responsibility among several actors. Legislative and regulatory interventions are needed to address the identified gaps, and while the EU AI Act is imperfect, it can provide useful guidance for the design of a comprehensive high risk AI regulatory framework.

Secondly, the fundamental principles of bioethics, namely autonomy, beneficence, non maleficence and justice are not outdated in the era of algorithmic medicine, they simply require reinterpretation. Informed consent must be re-thought in terms of algorithmic disclosure. The monitoring of performance in the application of the principles of beneficence and non-maleficence must be continuous rather than in breaks. Justice requires structural intervention to identify and remediate algorithmic bias as a systemic and not merely technical problem.

Thirdly, the human clinician must be the moral and legal center of the clinical encounter. AI must be designed as a powerful tool that supplements the judgment of the clinician, and not as a substitute for the clinician. This has implications for the design of the AI system (which has override potential), the training of the clinician, the allocation of responsibility, and the regulatory structure (with the requirement of human oversight in high risk applications).

The future of artificial intelligence in healthcare will not be determined by the capabilities of the technology, but by the decisions that we make now in legislatures, regulatory bodies, professional organizations, healthcare organizations and in the courts. These decisions would decide whether we use artificial intelligence to ensure that the benefits of good medicine can be shared by all people in a way that is efficient and equitable, or whether we use it to cause new forms of medical harm and injustice. Making good decisions about artificial intelligence in healthcare requires the kind of careful, interdisciplinary, and patient centered thinking that this article has tried to contribute.