

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IJLRA welcomes submissions from all branches of law, provided the work is original, unpublished, and submitted in accordance with the prescribed submission guidelines.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.

MANY HANDS, NO HANDS? THE PROBLEM OF MANY HANDS AND THE RESPONSIBILITY GAP IN ARTIFICIAL INTELLIGENCE-DRIVEN HEALTHCARE

AUTHORED BY - MS. NANCY

Ph.D. Scholar, Hidayatullah National Law University

Abstract

Artificial intelligence (AI) systems are moving from the laboratory into clinical practice, where they triage patients, read images, predict deterioration and recommend treatment. When such systems err, patients are harmed, yet the question of who must answer for the harm is far from settled. This paper examines two related difficulties that complicate that question: the problem of many hands, which arises because AI-enabled care is produced by dispersed developers, data custodians, regulators, hospitals and clinicians, and the responsibility gap, which arises because learning systems act in ways that no human fully predicts or controls. Drawing on the philosophical literature on moral responsibility and on legal developments in India, the United Kingdom and the European Union, the paper argues that the gap in healthcare is less a metaphysical void than an institutional failure, aggravated by opacity, algorithmic bias, automation bias and thin consent practices. It proposes a layered, role-based model of distributed responsibility, supported by documentation duties, calibrated liability rules, meaningful human oversight and compensation mechanisms, so that no injured patient is left without an answerable party.

Keywords: *Artificial intelligence; healthcare; responsibility gap; problem of many hands; medical negligence; algorithmic bias; explainability; accountability.*

1. Introduction

Medicine has long depended on trust in persons: the physician who examines, the nurse who monitors, and the institution that stands behind them. Artificial intelligence is altering that arrangement. Deep learning systems now classify skin lesions with accuracy comparable to dermatologists,¹ detect diabetic retinopathy from retinal photographs,² and, in at least one

¹Andre Esteva and others, 'Dermatologist-level classification of skin cancer with deep neural networks' (2017) 542 *Nature* 115.

²Varun Gulshan and others, 'Development and Validation of a Deep Learning Algorithm for Detection of Diabetic

regulated product, produce a diagnostic output without a clinician interpreting the image.³ Enthusiasts envisage a “high-performance medicine” in which human and machine intelligence combine to improve accuracy, efficiency and access.⁴ A more sober reading of the evidence cautions that most studies are retrospective, few have been tested in real clinical settings, and claims of superiority over clinicians are frequently overstated.⁵

Whatever the true performance of these tools, they will sometimes fail, and when they do the injured person will ask a simple question: who is answerable? Imagine a hospital that deploys a commercial imaging algorithm to prioritise chest radiographs. The model, trained on data from other institutions, under-detects a condition in a patient population it has rarely seen. A busy radiologist, trusting the prioritisation, reads the scan late; the patient deteriorates. The developer points to the hospital’s configuration choices, the hospital to the developer’s validation claims, the radiologist to institutional protocol, and the regulator to a clearance granted on the basis of limited evidence. Each account is plausible and none is wholly false. Yet the patient, who consented to none of these decisions, bears the harm alone.

This scenario captures two distinct but interacting problems that dominate current debate on the ethics of AI in healthcare. The first is the *problem of many hands*, first articulated in the context of public administration, which holds that when many individuals contribute to an outcome it becomes difficult to identify anyone who can be held responsible for it.⁶ The second is the *responsibility gap*, the claim that autonomous, learning machines act in ways that their designers and operators cannot foresee or control, so that no human can fairly be blamed for what they do.⁷

This paper argues that in healthcare these problems are real but tractable. The gap is not a metaphysical void that only a new legal personality for machines could fill; it is largely an institutional and regulatory failure that can be closed by deliberate design. Section 2 sets out the conceptual foundations. Section 3 maps the many hands involved in clinical AI. Section 4 examines four ethical challenges (opacity, bias, automation bias and consent) that widen the

Retinopathy in Retinal Fundus Photographs’ (2016) 316 *JAMA* 2402.

³Michael D Abràmoff and others, ‘Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices’ (2018) 1 *npj Digital Medicine* 39.

⁴Eric J Topol, ‘High-performance medicine: the convergence of human and artificial intelligence’ (2019) 25 *Nature Medicine* 44.

⁵Myura Nagendran and others, ‘Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies’ (2020) 368 *BMJ* m689.

⁶Dennis F Thompson, ‘Moral Responsibility of Public Officials: The Problem of Many Hands’ (1980) 74 *American Political Science Review* 905.

⁷Andreas Matthias, ‘The responsibility gap: Ascribing responsibility for the actions of learning automata’ (2004) 6 *Ethics and Information Technology* 175.

gap. Section 5 reviews how tort, product liability and regulatory law in India and comparable jurisdictions currently respond. Section 6 proposes a layered model of distributed responsibility, and Section 7 concludes.

2. Conceptual Foundations

2.1 *The Problem of Many Hands*

Thompson's original formulation concerned officials in bureaucratic organisations, but its logic travels readily to technological systems. Helen Nissenbaum observed in the 1990s that computerised society produces a "barrier to accountability" precisely because software is built by teams, embedded in institutions and used by others, so that harms emerge from a chain of contributions none of which is clearly culpable on its own.⁸ She identified further obstacles that are directly relevant to healthcare: the tendency to treat software bugs as inevitable, the habit of blaming the computer as a scapegoat, and the practice of claiming ownership of software while disclaiming liability for its consequences.

Van de Poel, Royakkers and Zwart have refined the concept. They distinguish between the problem as an epistemic matter (we cannot tell who did what) and as a normative one (even if we could, individual contributions may not warrant blame), and they argue that responsibility must be understood in both backward-looking and forward-looking senses.⁹ Backward-looking responsibility concerns accountability and blame after harm; forward-looking responsibility concerns the duties, roles and virtues that prevent harm in the first place. This distinction echoes H L A Hart's classic separation of role-responsibility, causal responsibility, liability-responsibility and capacity-responsibility,¹⁰ and it is helpful because the many-hands problem is usually at its worst for blame and at its most manageable for role-based duties.

2.2 *The Responsibility Gap*

Matthias argued that traditional ways of ascribing responsibility presuppose that the operator of a machine can control it and foresee its behaviour. Learning automata erode both conditions, since their behaviour depends on training data and adaptation that the manufacturer cannot fully specify in advance.¹¹ Robert Sparrow reached a similar conclusion in relation to autonomous weapons, suggesting that neither programmer, commander nor machine could

⁸Helen Nissenbaum, 'Accountability in a computerized society' (1996) 2 *Science and Engineering Ethics* 25.

⁹Ibo van de Poel, Lambèr Royakkers and Sjoerd D Zwart, *Moral Responsibility and the Problem of Many Hands* (Routledge 2015).

¹⁰H L A Hart, *Punishment and Responsibility* (Oxford University Press 1968) 211–230.

¹¹Matthias (n 7) 175–183.

justly be held responsible for an unpredictable killing.¹² Santoni de Sio and Mecacci have more recently shown that “the” responsibility gap is in fact a family of at least four gaps: in culpability, in moral accountability, in public accountability and in active responsibility, each requiring a different response.¹³

The thesis is contested. Critics argue that responsibility has always been attributed to humans in complex socio-technical settings, that the “gap” confuses unpredictability with the absence of control, and that answerability can be reassigned through institutional practice.¹⁴ John Danaher has argued that the real difficulty is a “retribution gap”: the public’s desire to see someone blamed and punished cannot be met when a machine has caused the harm.¹⁵ Mark Coeckelbergh adds a relational perspective, linking responsibility to two classical conditions, control and knowledge, and arguing that explainability matters because it allows people affected by AI to understand and contest what has been done to them.¹⁶

For present purposes, two points follow. First, the responsibility gap and the many-hands problem are mutually reinforcing: opacity in the machine makes it harder to disentangle the contributions of the humans who surround it. Second, the practical fact remains that injured patients today often cannot identify who should compensate them or why. The task is less to settle the philosophical dispute than to design institutions in which fair attribution is restored.

3. The Many Hands of Clinical AI

A clinical AI system is the product of a long supply chain of decisions, and every link carries distinct moral duties. At the origin lie *data custodians*: hospitals, laboratories, insurers and public health bodies that collect, label and release the data from which models learn. Their choices about what to record, whom to include and how to label outcomes embed assumptions that later surface as predictions. *Developers and data scientists* then select model architectures, define the target variable, tune thresholds, and decide how much testing is enough. Commercial *vendors and manufacturers* market the resulting product, make representations about its intended use and performance, and control updates after release.

Downstream, *regulators* decide whether a tool is cleared, under what classification, and with

¹²Robert Sparrow, ‘Killer Robots’ (2007) 24 *Journal of Applied Philosophy* 62.

¹³Filippo Santoni de Sio and Giulio Mecacci, ‘Four Responsibility Gaps with Artificial Intelligence: Why they Matter and How to Address them’ (2021) 34 *Philosophy & Technology* 1057.

¹⁴Daniel W Tigard, ‘There Is No Techno-Responsibility Gap’ (2021) 34 *Philosophy & Technology* 589; Deborah G Johnson, ‘Technology with No Human Responsibility?’ (2015) 127 *Journal of Business Ethics* 707.

¹⁵John Danaher, ‘Robots, law and the retribution gap’ (2016) 18 *Ethics and Information Technology* 299.

¹⁶Mark Coeckelbergh, ‘Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability’ (2020) 26 *Science and Engineering Ethics* 2051.

what post-market obligations. *Hospital administrators and procurement committees* decide whether to buy it, how to integrate it into workflows and what training to provide. *Clinicians* decide whether to rely on a given output, how to document that reliance and how to explain it to patients. Finally, *patients* supply data, often through routine consent that bears little relation to the eventual uses, and are the ones who bear the consequences.

Three features of this chain aggravate the many-hands problem. The first is *temporal dispersion*: the decisions that shape harm may be taken years apart, by actors who never meet. A labelling decision in one hospital in one year becomes a bias in another hospital's tool several years later. The second is *epistemic asymmetry*: developers know the model but not the clinical setting; clinicians know the setting but not the model. Neither has the full picture needed to anticipate failure, which weakens the claim that either could have foreseen and prevented it. The third is *contractual fragmentation*. Software licences routinely allocate risk away from vendors through disclaimers, indemnities and "decision support only" clauses, while hospital policies allocate it to the individual clinician. Responsibility is thereby redistributed by private ordering, typically towards the party with the least bargaining power and the least knowledge of the system.

4. Ethical Challenges that Widen the Gap

The ethical literature on AI in healthcare identifies a wide range of concerns, including privacy, fairness, transparency, safety and accountability.¹⁷ Four of these deserve closer attention because each directly affects the allocation of responsibility among the many hands.

4.1 Opacity and the Limits of Explainability

Many high-performing models, particularly deep neural networks, are opaque to developers and users alike. This opacity compromises the epistemic condition of responsibility: a clinician who cannot understand why a system flagged one patient and not another cannot meaningfully assess whether to follow its advice. Alex John London has argued that accuracy should take priority over explainability in medicine, pointing out that clinicians have always used interventions whose mechanisms were poorly understood so long as they were empirically effective.¹⁸ Cynthia Rudin responds that, in high-stakes domains, "explainable" post-hoc

¹⁷Jessica Morley and others, 'The ethics of AI in health care: A mapping review' (2020) 260 *Social Science & Medicine* 113172; Brent Daniel Mittelstadt and others, 'The ethics of algorithms: Mapping the debate' (2016) 3(2) *Big Data & Society* 1.

¹⁸Alex John London, 'Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability' (2019) 49(1) *Hastings Center Report* 15.

approximations of black-box models can be misleading and that inherently interpretable models should be preferred wherever they perform comparably.¹⁹

The responsibility perspective offers a way through this dispute. Even if a model cannot be fully explained, the *institutions* around it can be made explainable: what data was used, how it was validated, in which populations it was tested, what error rates were found and what the intended clinical role is. Opacity at the level of the algorithm need not entail opacity at the level of governance. Where a vendor or hospital cannot supply such institutional transparency, the claim that no one could have foreseen the harm loses much of its force. Explainability therefore matters less as a technical property than as a precondition for contestation, a point that also resonates with the debate on whether data protection law confers any meaningful right to an explanation of automated decisions.²⁰

4.2 Algorithmic Bias and Distributive Injustice

Healthcare AI inherits the biases of the data and systems on which it is built. In a widely cited study, Obermeyer and colleagues found that a commercial algorithm used to allocate care-management resources to millions of patients in the United States systematically underestimated the needs of Black patients, because it used past healthcare expenditure as a proxy for health need, and expenditure itself reflected unequal access.²¹ Similar concerns have been raised about race-based correction factors embedded in clinical algorithms.²² In India, where datasets are skewed toward urban, tertiary-care and English-language settings and where gender, caste, class and regional disparities are pronounced, a model validated on one population may perform poorly on another with consequences that are neither visible nor easily attributable.

Bias presents a many-hands problem in its purest form. The proxy variable was chosen by a developer; the data reflected structural inequality; the hospital adopted the tool without subgroup testing; and the clinician saw only a risk score. Each can claim that bias was not their intention, and none individually “caused” the injustice. Yet the principle of justice in

¹⁹Cynthia Rudin, ‘Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead’ (2019) 1 *Nature Machine Intelligence* 206.

²⁰Sandra Wachter, Brent Mittelstadt and Luciano Floridi, ‘Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation’ (2017) 7 *International Data Privacy Law* 76.

²¹Ziad Obermeyer and others, ‘Dissecting racial bias in an algorithm used to manage the health of populations’ (2019) 366 *Science* 447.

²²Darshali A Vyas, Leo G Eisenstein and David S Jones, ‘Hidden in Plain Sight: Reconsidering the Use of Race Correction in Clinical Algorithms’ (2020) 383 *New England Journal of Medicine* 874.

biomedical ethics requires fair distribution of benefits and burdens,²³ and a system that systematically disadvantages particular groups violates that principle regardless of whether any actor intended it. Responsibility here is best understood as forward-looking: someone must be assigned the duty to test, monitor and correct for disparities before and after deployment.

4.3 Automation Bias and the Moral Crumple Zone

A third challenge arises at the human–machine interface. Research on automation consistently shows that people tend to over-rely on automated recommendations, accepting them even when they are wrong, and to under-monitor systems that usually work.²⁴ Lisanne Bainbridge’s “ironies of automation” point out that automation leaves humans with the hardest residual tasks while depriving them of the practice needed to perform them.²⁵ In clinical settings this means that a clinician formally retained “in the loop” may in practice have little capacity to catch the system’s errors, particularly under time pressure.

Madeleine Clare Elish has described the resulting dynamic as the “moral crumple zone”: just as the crumple zone of a car absorbs the force of impact to protect the vehicle, the human operator in a complex automated system absorbs moral and legal responsibility for failures that originate elsewhere in the system.²⁶ Healthcare is especially vulnerable to this pattern. The clinician is the visible face of the system and the legally recognised duty-holder in most jurisdictions, so liability tends to flow to her regardless of whether she had adequate information, training or time. The consequence is a perverse double bind: a clinician who follows an erroneous recommendation may be blamed for negligent reliance, while one who overrides a correct recommendation may be blamed for ignoring evidence-based technology.

4.4 Informed Consent and Patient Autonomy

Finally, AI complicates informed consent. Modern medical law has moved from a physician-centred to a patient-centred standard of disclosure. In the United Kingdom, *Montgomery v Lanarkshire Health Board* requires disclosure of all material risks and reasonable

²³Tom L Beauchamp and James F Childress, *Principles of Biomedical Ethics* (8th edn, Oxford University Press 2019) ch 7.

²⁴Kate Goddard, Abdul Roudsari and Jeremy C Wyatt, ‘Automation bias: a systematic review of frequency, effect mediators, and mitigators’ (2012) 19 *Journal of the American Medical Informatics Association* 121; Raja Parasuraman and Dietrich H Manzey, ‘Complacency and Bias in Human Use of Automation: An Attentional Integration’ (2010) 52 *Human Factors* 381.

²⁵Lisanne Bainbridge, ‘Ironies of automation’ (1983) 19 *Automatica* 775.

²⁶Madeleine Clare Elish, ‘Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction’ (2019) 5 *Engaging Science, Technology, and Society* 40.

alternatives,²⁷ and the Supreme Court of India in *Samira Kohli v Dr Prabha Manchanda* likewise insisted on consent that is real and informed.²⁸ It is unclear what these standards require when a diagnosis or treatment plan is shaped by an algorithm. Should patients be told that AI was used, what its error rates are, whether it was validated on people like them, or whether they may opt for a human-only assessment? Scholars have argued that meaningful patient-centred decision-making requires at least some disclosure of the role and limitations of AI, but that the opacity of the tools and the pressure of clinical time make this difficult.²⁹ Where patients are not told, they cannot exercise the very autonomy that grounds their right to contest harm, and the responsibility gap widens from a legal problem into a failure of respect for persons.³⁰

5. Legal and Regulatory Responses

5.1 Medical Negligence

In most common-law systems, liability for medical harm is determined by the law of negligence. In England, the *Bolam* test asks whether the clinician acted in accordance with a practice accepted as proper by a responsible body of medical opinion, subject to the requirement in *Bolitho* that such opinion be logically defensible.³¹ The Supreme Court of India has adopted the *Bolam* standard and, in *Jacob Mathew v State of Punjab*, held that criminal liability of doctors requires gross negligence or recklessness rather than mere error of judgment.³² Since *Indian Medical Association v V P Shantha*, medical services have also fallen within consumer protection law, giving patients a relatively accessible forum for compensation.³³

These doctrines assume a human professional exercising judgment. They are poorly equipped for a situation in which the standard of care itself is in flux. If AI-assisted diagnosis becomes common, will a clinician who declines to use it be negligent? If she uses it and follows it, will she be protected by reliance on an approved tool? Price, Gerke and Cohen have noted that physicians may face liability both for following erroneous AI recommendations and for

²⁷*Montgomery v Lanarkshire Health Board* [2015] UKSC 11, [2015] AC 1430.

²⁸*Samira Kohli v Dr Prabha Manchanda* (2008) 2 SCC 1.

²⁹Jens Christian Bjerring and Jacob Busch, 'Artificial Intelligence and Patient-Centered Decision-Making' (2021) 34 *Philosophy & Technology* 349.

³⁰Beauchamp and Childress (n 23) ch 4.

³¹*Bolam v Friern Hospital Management Committee* [1957] 1 WLR 582; *Bolitho v City and Hackney Health Authority* [1998] AC 232.

³²*Jacob Mathew v State of Punjab* (2005) 6 SCC 1.

³³*Indian Medical Association v V P Shantha* (1995) 6 SCC 651.

departing from correct ones, and that the incentives created by this uncertainty are likely to push toward defensive conformity.³⁴

5.2 Product Liability

Product liability could in principle shift responsibility to developers and manufacturers, who are better placed to detect and prevent defects. In India, the Consumer Protection Act 2019 introduced a dedicated chapter on product liability, imposing liability on manufacturers, service providers and sellers for harm caused by defective products or deficient services. Whether standalone software falls within the statutory definition of “product” is, however, not free from doubt, and the burden of proving a defect in a self-learning system is formidable when the claimant lacks access to training data, code or validation records. Where software qualifies as a medical device, it may be regulated under the Medical Devices Rules 2017, though the regulatory treatment of adaptive and continuously learning software remains underdeveloped. The European Union has moved furthest in confronting these difficulties. The revised Product Liability Directive expressly treats software as a product, extends liability to harm caused by updates and the absence of required updates, and introduces rebuttable presumptions and disclosure powers designed to ease the claimant’s evidentiary burden.³⁵ These reforms recognise that the many-hands problem is partly an information problem and that shifting the burden of proof can be as important as defining the category of defendant.

5.3 Regulation, Soft Law and Data Protection

Regulatory frameworks are also beginning to allocate responsibilities prospectively. The EU Artificial Intelligence Act classifies AI systems used as safety components of medical devices as high-risk and imposes obligations on providers and deployers concerning risk management, data governance, technical documentation, transparency and human oversight.³⁶ The World Health Organization’s guidance on AI for health identifies principles including transparency, explainability, accountability and equity, and stresses that responsibility must be assigned clearly across the life cycle of a technology.³⁷ In India, the Indian Council of Medical Research

³⁴W Nicholson Price II, Sara Gerke and I Glenn Cohen, ‘Potential Liability for Physicians Using Artificial Intelligence’ (2019) 322 *JAMA* 1765.

³⁵Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products [2024] OJ L 2024/2853.

³⁶Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence [2024] OJ L 2024/1689, arts 6, 9–15 and 26.

³⁷World Health Organization, *Ethics and governance of artificial intelligence for health: WHO guidance* (WHO 2021).

has issued ethical guidelines for the application of AI in biomedical research and healthcare which emphasise accountability, human oversight and patient safety, but they are advisory rather than binding.³⁸

Data protection law supplies a further layer. The Digital Personal Data Protection Act 2023 recognises the rights of data principals and imposes duties on data fiduciaries, building on the constitutional right to privacy recognised in *Justice K S Puttaswamy v Union of India*.³⁹ However, it does not, in terms, create a general right to an explanation of AI-assisted clinical decisions, and it was not designed to allocate responsibility for algorithmic harm. India therefore currently relies on a patchwork of general tort principles, consumer protection, sector-specific device regulation and non-binding ethical guidance, none of which squarely addresses the allocation of responsibility among the many hands of clinical AI.

6. Towards a Model of Distributed Responsibility

If the foregoing analysis is correct, closing the responsibility gap in healthcare requires neither the fiction of machine responsibility nor the scapegoating of the clinician. Floridi and Sanders have argued that artificial agents may be “accountable” in a thin sense without being morally responsible,⁴⁰ and Nyholm has suggested that responsibility in human–machine collaborations should be located in the humans and institutions that direct and benefit from the system.⁴¹ Building on these insights and on the distinction between backward- and forward-looking responsibility, this paper proposes a layered model comprising five elements.

First, role-specific duties. Each actor in the chain should bear defined duties proportionate to its knowledge and control. Developers should owe duties of rigorous validation, representative data, subgroup testing, honest labelling of intended use and limitations, and post-market surveillance. Vendors should be responsible for the accuracy of their performance claims and for managing updates. Hospitals should owe duties of safe procurement, local validation, training, workflow design and continuing monitoring. Clinicians should owe duties of reasonable, informed reliance rather than blind deference or reflexive distrust. Defining these duties in advance converts a diffuse collective problem into a set of specific obligations against which conduct can later be measured, and it responds to the many-hands problem in its forward-

³⁸Indian Council of Medical Research, *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare* (ICMR 2023).

³⁹*Justice K S Puttaswamy (Retd) v Union of India* (2017) 10 SCC 1.

⁴⁰Luciano Floridi and J W Sanders, ‘On the Morality of Artificial Agents’ (2004) 14 *Minds and Machines* 349.

⁴¹Sebastian Nyholm, ‘Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci’ (2018) 24 *Science and Engineering Ethics* 1201.

looking form.

Second, documentation and traceability. Because the many-hands problem is partly epistemic, the law should require that the life cycle of clinical AI be documented in an auditable form: provenance of training data, model versions, validation results, deployment decisions, overrides and incident reports. Evidence from comparable real-world deployments shows why this is necessary. An external validation of a widely implemented proprietary sepsis prediction model, for instance, found that it performed substantially worse than its developer had reported, a discovery that was possible only because independent researchers could examine hospital data.⁴² Mandatory logging and access rights for regulators and claimants would make such scrutiny routine and would reverse the informational advantage that currently shields developers and institutions from challenge.

Third, calibrated liability. Liability rules should track control and capacity to prevent harm. Developers and vendors, who profit from the technology and can best identify and mitigate defects, should bear primary liability for design and validation failures, supported by presumptions of defect where documentation is missing or the claimant cannot reasonably access technical information. Hospitals should bear enterprise liability for deployment, procurement and oversight failures. Clinicians should be liable only where they departed from a reasonable standard of informed reliance, with legislative or judicial clarification that following a properly validated, regulator-approved tool within its intended use is, absent warning signs, evidence of reasonable care. This allocation resists the moral crumple zone by ensuring that the burden falls on those with the greatest power to prevent the harm.

Fourth, meaningful human oversight. Human oversight must be designed, not merely declared. A requirement that a human remain “in the loop” is hollow if the person has no time, training, information or authority to challenge the system. Institutions should therefore be required to ensure realistic workloads, training in the limitations of AI, interfaces that communicate uncertainty, and protection for clinicians who override recommendations in good faith. Oversight duties should rest on the institution that structures the clinician’s working conditions as much as on the individual clinician.

Fifth, no-fault compensation and collective mechanisms. Even a well-designed regime will leave some harms unattributable. To address Danaher’s retribution gap and the practical gap in compensation, jurisdictions should consider no-fault compensation schemes or mandatory insurance funded by developers, vendors and healthcare providers. Such schemes, which exist

⁴²Andrew Wong and others, ‘External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients’ (2021) 181 *JAMA Internal Medicine* 1065.

in some countries for vaccine injury and birth-related injury, ensure that injured patients are compensated without the need to trace a chain of causation through opaque systems, while a separate accountability track preserves incentives for safety and learning. Where the harm reveals systemic failure, regulators should have the power to investigate and publish findings in the manner of accident investigation bodies in aviation, thereby addressing the gaps in public accountability and active responsibility identified by Santoni de Sio and Mecacci.

Two objections deserve a brief reply. Heavy duties may be said to chill innovation, but clear and predictable allocation of responsibility is more conducive to responsible innovation than uncertainty. Documentation demands may also conflict with trade secrecy, but this can be managed through confidential regulatory access and court-supervised disclosure. For resource-constrained systems such as India's, duties should scale with risk, and public procurement can itself become a regulatory lever by making validation and monitoring conditions of purchase.

7. Conclusion

The problem of many hands and the responsibility gap describe a genuine danger in AI-enabled healthcare: that harms will occur for which no person or institution is held answerable. This danger is aggravated by opacity, bias, automation bias and thin consent, and it is presently unaddressed by legal frameworks designed for a world of individual professional judgment. Yet the danger is not inevitable. The many hands of clinical AI are identifiable, their roles are definable, and their capacity to prevent harm can be assessed. What is required is a deliberate choice to assign responsibility in advance, to make the system traceable, to match liability to control, and to protect patients who are harmed from the burden of proving what only the system's builders and operators can know.

For countries such as India, now rapidly adopting digital health infrastructure, the choices made in procurement rules, device regulation and negligence doctrine will determine whether AI deepens inequity or contributes to more accountable care. A healthcare system in which responsibility is shared should not become one in which responsibility is lost. The ethical task is to ensure that, however many hands contribute to a clinical decision, the patient always has someone to ask why, and someone obliged to answer.