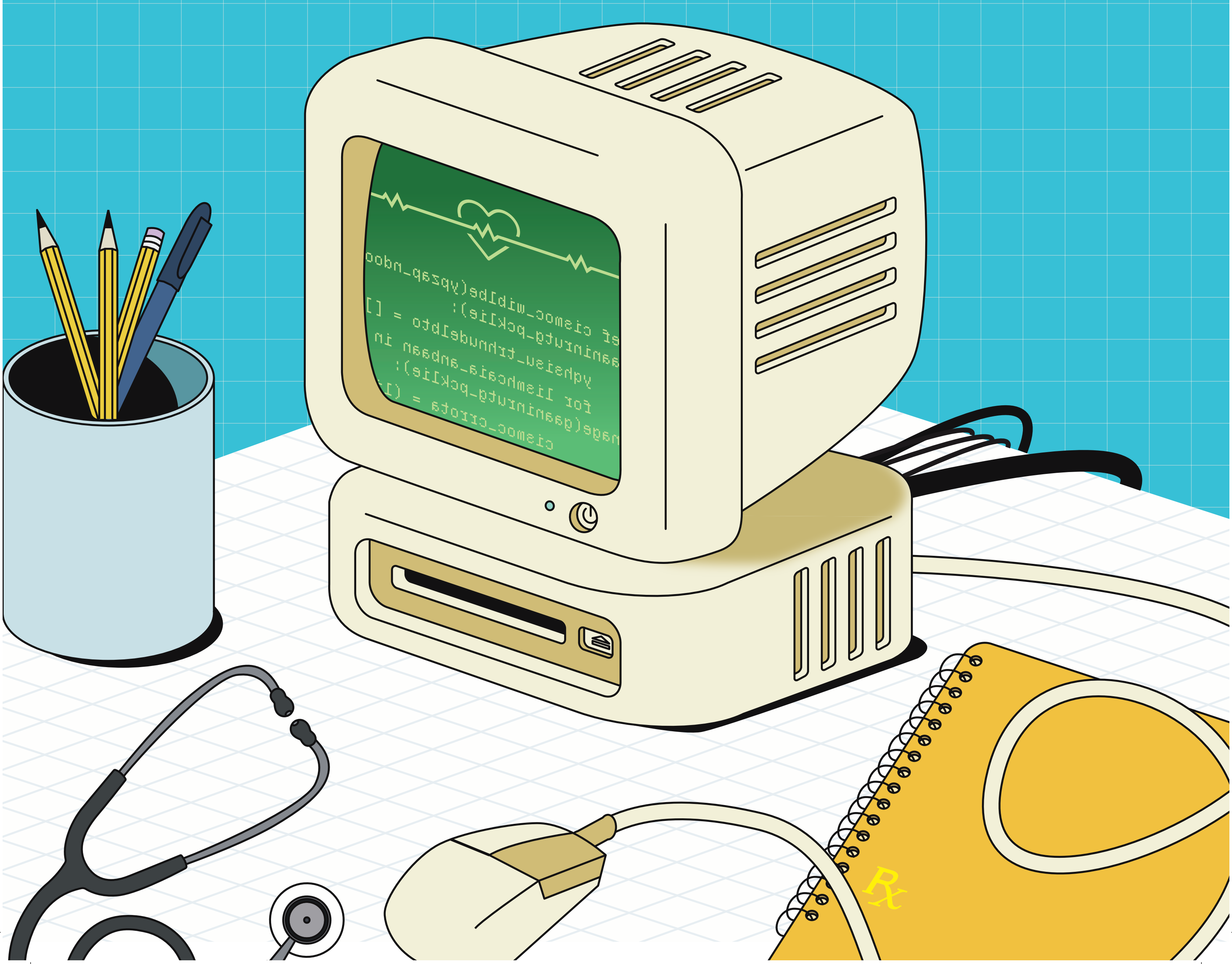
Centre for Internet & Society

AI for Healthcare

Understanding Data Supply Chain and Auditability in India

NOVEMBER 2024



Centre for Internet & Society

AI for Healthcare

Understanding Data Supply Chain and Auditability in India

NOVEMBER 2024

Research Team

Research and Writing | Amrita Sengupta (PI), Shweta Mohandas (Co-PI), *(In alphabetical order)* Abhineet Nayyar, Chetna VM, Puthiya Purayil Sneha, Yatharth

Research Support | Anjana R Nair and Pallavi Bedi

Data Collection | INJ Partners

Institutional Review Board | Sigma Research and Consulting Pvt Ltd

Report Review | *(In alphabetical order)* Divij Joshi (Doctoral Researcher, University College London), Pallavi Bedi (Senior Researcher, Centre for Internet and Society), Sunita Sheel Bandewar (PhD), (Secretary General, Forum for Medical Ethics Society; Working Editor, Indian Journal of Medical Ethics)

Design | Sharath Ravishankar

Copyediting | The Clean Copy

Acknowledgements

The CIS team would like to thank the study participants, working across healthcare institutions, technology companies and startups, academia, civil society organisations and policymaking for their time and inputs on the topic of AI auditing and healthcare in India. We are grateful to Dr. Bhaskar Rajakumar, Sneha Sai and Rohit Satish from ARTPARK@IISc and Sunita Sheel Bandewar from Forum for Medical Ethics Society for their continued collaboration on this project. The team would also like to thank Chiara Furtado (CIS) for initial support and guidance through the IRB process.

This material is based upon work supported in whole or in part by The Notre Dame-IBM Tech Ethics Lab. Such support does not constitute endorsement by the sponsor of the views expressed in this publication.

Contents

Executive Summary ... 3
Table of Figures ... 7
List of Abbreviations ... 8
1. INTRODUCTION ... 9
1.1. Scope of the study ... 10
Defining AI and AI in healthcare ... 11
1.2. Structure of this report ... 12
2. METHODS ... 13
2.1. Study design ... 13
2.2. Sampling ... 13
2.3. Recruitment and data collection ... 15
2.3.1. Desk-based secondary research ... 16
2.3.2. Quantitative data ... 16
2.3.3. Qualitative data ... 17
2.4. Data analysis ... 17
2.5. Study ethics ... 18
2.6. Study limitations ... 18
3. AI IN HEALTHCARE: A BACKGROUND ... 19
3.1. Use cases for AI in healthcare ... 19
3.2. Governance of AI in healthcare ... 21
3.2.1. Regulatory and policy developments in the background of AI and healthcare ... 21
3.2.2. Ethical principles and practice ... 23
3.3. AI audits ... 24
3.3.1. The internal vs external debate ... 25
3.3.2. Auditing frameworks and contents of AI audits ... 26
3.4. The data supply chain framework ... 28
3.4.1. Data sourcing ... 28
3.4.2. Data processing ... 29
3.4.3. Data learning and model development ... 29
3.4.4. Model deployment ... 30

4. MAIN FINDINGS AND DISCUSSION: DATA SUPPLY CHAIN 31
4.1. Data sourcing 31
4.1.1. Reliance on both open and proprietary data sources 31
4.1.2. Reliance on datasets from the Global North 32
4.2. Data processing 35
4.2.1. Data quality checks, while in operation, bring a significant burden 35
4.2.2. Anonymisation and removal of personally identifiable information are key priorities for all stakeholders 37
4.3. Data learning and model development 40
4.3.1. Collaboration between medical professionals and AI developers remains limited 40
4.3.2. Feedback sharing between the technology developers and their users is often irregular and indirect 42
4.4. Model deployment 44
4.4.1. Reliance on external vendors 44
4.4.2. Medical professionals hesitate to adopt AI 45
4.4.3. Gaps in training and education among professionals on the use of AI 46
5. MAIN FINDINGS AND DISCUSSION: AI AUDITING AS A RESPONSE 48
5.1. Current state of AI audits 48
5.1.1. Understanding and applications of AI audits remain scattered and incoherent among stakeholders 48
5.1.2. AI audits prioritise mainstream and intersectional concepts such as data privacy and security in their scope 50
5.2. Audit as a governance tool for AI systems 51
5.2.1. Theoretically, AI audits can increase accountability, aid in standardisation, and improve algorithmic performance 51
5.2.2. Organisational and systemic constraints inhibit the effectiveness and success of AI audits 52
6. RECOMMENDATIONS 53
6.1. Improve data management across the AI data supply chain 54
6.1.1. Adopt standardised data-sharing policies 54
6.1.2. Emphasise not just data quantity but also data quality 56
6.2. Streamline AI auditing as a form of governance 57
6.2.1. Standardise the practice of AI auditing 57
6.2.2. Build organisational knowledge and inter-stakeholder collaboration 58
6.2.3 Prioritise transparency and public accountability in auditing standards 58
6.3. Centre public good in India's AI industrial policy 59
6.3.1. Adopt focused and transparent approaches to investing in and financing AI projects 59
6.3.2. Strengthen regulatory checks and balances for AI governance 61
7. CONCLUSION 62

Executive summary

The use of artificial intelligence (AI) technologies constitutes a significant development in the Indian healthcare sector, with industry and government actors showing keen interest in designing and deploying these technologies. Even as key stakeholders explore ways to incorporate AI systems into their products and workflows, a growing debate on the accessibility, success, and potential harms of these technologies continues, along with several concerns over their large-scale adoption. A recurring question in India and the world over is whether these technologies serve a wider interest in public health. For example, the discourse on ethical and responsible AI in the context of emerging technologies and their impact on marginalised populations, climate change, and labour practices has been especially contentious.

For the purposes of this study, we define AI in healthcare as the use of artificial intelligence and related technologies to support healthcare research and delivery. The use cases include assisted imaging and diagnosis, disease prediction, robotic surgery, automated patient monitoring, medical chatbots, hospital management, drug discovery, and epidemiology. The emergence of AI auditing mechanisms is an essential development in this context, with several stakeholders ranging from big-tech to smaller startups adopting various checks and balances while developing and deploying their products. While auditing as a practice is neither uniform nor widespread within healthcare or other sectors in India, it is one of the few available mechanisms that can act as guardrails in using AI systems.

This report aims to understand the prevalence and use of AI auditing practices in the healthcare sector. By mapping the data supply chain underlying AI technologies, the study aims to unpack i) how AI systems are developed and deployed to achieve healthcare outcomes and, more importantly, ii) how AI audits are perceived and implemented by key stakeholders in the healthcare ecosystem.

Our primary research questions are as follows:

What is the current data supply chain infrastructure for organisations operating in the healthcare ecosystem in India?

What auditing practices, if any, are being followed by technology companies and healthcare institutions?

What best practices can organisations based in India adopt to improve AI auditability?

This was a mixed methods study, comprising a review of available literature in the field, followed by quantitative and qualitative data collection through surveys and in-depth interviews. The findings from the study offer essential insights into the current use of AI in the healthcare sector, the operationalisation of the data supply chain, and policies and practices related to health data sourcing, collection, management, and use. It also discusses ethical and practical challenges related to privacy, data protection and informed consent, and the emerging role of auditing and other related practices in the field. Some of the key learnings related to the data supply chain and auditing include:

Based on these findings, this report offers a set of recommendations addressed to different stakeholders such as healthcare professionals and institutions, AI developers, technology companies, startups, academia, and civil society groups working in health and social welfare. These include:

Technology companies, medical institutions, and medical practitioners rely on an equal mix of proprietary and open sources of health data and there is significant reliance on datasets from the Global North.

Data quality checks are extant, but they are seen as an additional burden; with the removal of personally identifiable information being a priority during processing.

Collaboration between medical practitioners and AI developers remains limited, and feedback between users and developers of these technologies is limited.

There is a heavy reliance on external vendors to develop AI models, with many models replicated from existing systems in the Global North.

Healthcare professionals are hesitant to integrate AI systems into their workflows, with a significant gap stemming from a lack of training and infrastructure to integrate these systems successfully.

The understanding and application of audits are not uniform across the sector, with many stakeholders prioritising more mainstream and intersectional concepts such as data privacy and security in their scope.

Improve data management across the AI data supply chain

Adopt standardised data-sharing policies.
This would entail building a standardised policy that adopts an intersectional approach to include all stakeholders and areas where data is collected to ensure their participation in the process. This would also require robust feedback loops and better collaboration between the users, developers, and implementers of the policy (medical professionals and institutions), and technologists working in AI and healthcare.

Emphasise not just data quantity but also data quality.
Given that the limited quantity and quality of Indian healthcare datasets present significant challenges, institutions engaged in data collection must consider their interoperability to make them available to diverse stakeholders and ensure their security. This would include recruiting additional support staff for digitisation to ensure accuracy and safety and maintain data quality.

Streamline AI auditing as a form of governance

Standardise the practice of AI auditing.
A certain level of standardisation in AI auditing would contribute to the growth and contextualisation of these practices in the Indian healthcare sector. Similarly, it would also aid in decision-making among implementing institutions.

Build organisational knowledge and inter-stakeholder collaboration.
It is imperative to build knowledge and capacity among technical experts, healthcare professionals, and auditors on the technical details of the underlying architecture and socioeconomic realities of public health. Hence, collaboration and feedback are essential to enhance model development and AI auditing.

Prioritise transparency and public accountability in auditing standards.
Given that most healthcare institutions procure externally developed AI systems, some form of internal or external AI audit would contribute to better public accountability and transparency of these technologies.

Centre public good in India’s AI industrial policy

Adopt focused and transparent approaches to investing in and financing AI projects.

An equitable distribution of AI spending and associated benefits is essential to guarantee that these investments and their applications extend beyond private healthcare, and that implementation approaches prioritise the public good. This would involve investing in entire AI life cycles instead of merely focusing on development and promoting transparent public–private partnerships.

Strengthen regulatory checks and balances for AI governance.

While an overarching law to regulate AI technologies may still be under debate, existing regulations may be amended to bring AI within their ambit. Furthermore, all regulations must be informed by stakeholder consultations to guarantee that the process is transparent, addresses the rights and concerns of all the parties involved, and prioritises the public good.

Table of figures

Figure No.	Figure title	Page no.
1	Sample size– Stakeholder-wise breakdown of survey respondents	**14**
2	Sample size – Stakeholder-wise breakdown of qualitative interviews	**15**
3	Use cases of AI systems in healthcare	**20**
4	An MAA framework that uses SMACTR and FMEA approaches	**27**
5	Visual representation of the Data Supply Chain framework	**28**
6	Percentage of respondents from healthcare institutions, technology companies, and medical professionals on the types of data sources used for AI design/deployment	**31**
7	Percentage of respondents from healthcare institutions, technology companies, and medical professionals on types of data sources used (top 3)	**32**
8	Percentage of respondents from technology companies and healthcare institutions on how they verify the accuracy and reliability of data sources	**35**
9	Percentage of respondents from healthcare institutions and technology companies using benchmark datasets or checking for imbalances in datasets	**35**
10	Percentage of respondents from healthcare institutions and technology companies, on the different practices they follow while cleaning medical data	**36**
11	Percentage of medical professionals on their practices related to the PII of patient	**37**
12	Percentage of respondents from healthcare institutions and technology companies, on the different practices they follow during data handling	**38**
13	Percentage of respondents from technology companies, healthcare institutions, and medical professionals who partner with other stakeholders in developing and deploying AI systems	**41**
14	Percentage of respondents from technology companies who work in AI development and testing with prior work experience in AI and healthcare	**41**
15	Percentage of medical professionals who provide feedback on AI systems, on their preferred methods of sharing feedback	**42**
16	Percentage of medical professionals, on the degree of challenges they face in integrating AI systems into clinical workflows	**44**
17	Percentage of respondents from healthcare institutions, on how they source AI systems	**45**
18	Types of AI-focused training across different stakeholders, either provided by organisations or undertaken by respondents individually	**46**
19	Percentage of respondents who prioritise policies, protocols, and compliance for reviews, evaluations, and audits	**49**
20	Percentage of respondents that prioritise different aspects of the AI data supply chain for reviews, evaluations, and audits	**50**
21	Recommendations for stakeholders in the AI and healthcare ecosystem	**53**

List of abbreviations

ABDM	Ayushman Bharat Digital Mission
AI	Artificial Intelligence
AI4SG	AI for Social Good
API	Application Programming Interfaces
ASHAs	Accredited Social Health Activists
CDC	Centers for Disease Control and Prevention
CDSCO	Central Drugs Standard Control Organisation
CIS	Center for Internet & Society
DIA	Digital India Act
DISHA	Digital Information Security in Healthcare Act 2018
DPDPA	Digital Personal Data Protection Act 2023
DSP	Data Supply Chain
EHR	Electronic Health Records
EU	European Union
FDA	Food and Drug Administration
GDP	Gross domestic product
NHS	National Health Service UK
NMDR	National Medical Device Rules
OECD	Organisation for Economic Co-operation and Development
PII	Personally Identifiable Information
GDPR	General Data Protection Regulation
GPUs	Graphics Processing Units
HCAI	Human Centred AI
HIIL	Human in the Loop
IAIC	IndiaAI Innovation Centre
ICMR	Indian Council of Medical Research
IDI	In-Depth Interviews
IRB	Institutional Review Board
ISO	International Organisation for Standardisation
IT-ES	Information Technology-Enabled Services
LLM	Large Language Model
MAA	Medical Algorithmic Audit
MEITY	Ministry of Electronics and Information Technology
NASSCOM	National Association of Software and Service Companies
NDHM	National Digital Health Mission
NHA	National Health Authority
PMJAY	Pradhan Mantri Jan Arogya Yojana
PPP	Public Private Partnerships
R&D	Research and Development
US	United States
WHO	World Health Organisation

1. Introduction

Despite serving a population of over 1.4 billion people as of 2022, India's healthcare system has significant shortcomings, a fact starkly highlighted by the COVID-19 pandemic.[1] Government expenditure on domestic health amounted to just 1.1% of India's GDP in 2021 – nearly a sixth of the global average of 6.5%.[2] The density of physicians, i.e., the number of physicians per 10,000 population, is estimated to be 7.3, compared to the global average of 17.2.[3] [4] Concerns about affordability further exacerbate these gaps in the system's accessibility and capacity. Over 75% of households who needed hospitalisation in 2017 financed their care through out-of-pocket expenses, and at least 80% of these cases were covered neither by private nor public insurance.[5]

To address the need for accessibility and affordability in India's healthcare system, state and market actors are increasingly turning to artificial intelligence (AI) technologies as a potential solution. However, it remains unclear whether AI technologies are well-suited to meeting these needs. NASSCOM, the country's leading industry association of technology and IT companies, mentions that "integration of technology with healthcare, especially AI, is increasingly becoming crucial in enabling anytime and anywhere care."[6] [7] Organisations such as the World Economic Forum and NITI Aayog have expressed similar expectations.[8] [9] In fact, the development and deployment of AI systems across sectors – including for healthcare – has seen extensive support by the state, as is best evidenced by the Indian government's recent investment of USD 1.2 billion in building computing infrastructure and boosting AI-related innovation in the country.[10]

At the same time, ethical and legal concerns regarding the use of these technologies – especially in a domain as safety-critical and context-dependent as public health – have received limited attention[11]. These include concerns such as the risk of algorithmic bias, the lack of transparency, and concerns around data privacy and security.[12] [13] Further, with increasing narratives of 'AI for good', both within India[14] and globally[15], it is critical to ask whether the problem that needs to be solved requires an AI intervention in the first place. Techno-solutionist approaches are progressively starting to dominate the tech-policy landscape. Before developing these AI-based solutions, it is vital to determine whether AI is being used as an enabler to solve the problem or if it is being force-fitted to a problem it may or may not be able to solve. Moreover, given the black-box nature of today's dominant AI systems, there is little information on how and where these technologies source their training data and how they make decisions for and about people in different contexts.[16] Further, legislative and regulatory interventions to resolve most of these concerns are yet to be introduced.

1. *"India", World Health Organization (WHO), accessed 25 October 2024*
2. *"Domestic General Government Health Expenditure (% of GDP)", World Bank Group, accessed 25 October 2024*
3. *Ibid*
4. *"Density of Physicians (per 10,000 Population)", WHO, accessed 25 October 2024*
5. *"Health and Family Welfare Statistics in India 2019-20", Ministry of Health and Family Welfare, Government of India, 2020, 138.*
6. *"About Us", Nasscom, accessed 25 October 2024.*
7. *"How AI Is Transforming the Future of Healthcare in India", Nasscom, accessed 25 October 2024*
8. *"AI in Healthcare: India's Trillion-dollar Opportunity", World Economic Forum, 18 October 2022*
9. *"National Strategy for Artificial Intelligence. #AIForAll", NITI Aayog, Government of India, 2018, 24*
10. *"India Announces $1.2 Bln Investment in AI Projects", Reuters, 7 March 2024*
11. *Mary Cummings, "Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings", AI Magazine 42, no. 1 (2021): 6–15*
12. *Daniel Schönberger, "Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications", International Journal of Law and Information Technology 27, no. 2 (2019): 171-203*
13. *Nithesh Naik et al., "Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility?", Frontiers in Surgery 9 (2022)*
14. *National Strategy for Artificial Intelligence #AIForAll", NITI Aayog, Government of India.*
15. *"AI for Good", AII for Good, accessed 25 October 2024*
16. *Hanhui Xu and Kyle Michael James Shuttleworth, "Medical Artificial Intelligence and the Black Box Problem: A View Based on the Ethical Principle of 'Do No Harm'", Intelligent Medicine 4, no. 1 (2023): 52–57*

Partly to bridge this governance gap, and provide procedures for more robust checks on AI systems, several white papers and reports have prescribed frameworks for conducting audits to mitigate these and other concerns.[17] However, research on such AI audits, let alone their efficacy, remains limited. Further, within the Indian AI and healthcare landscape, several questions remain unanswered, such as what is the current state of design, development, and deployment of AI in healthcare? What data sources are being used to develop these AI systems? What checks and balances are being implemented to reduce harm and increase transparency? It is against this context of increasing investments in AI for the public good, and a lack of information and clarity on how these AI systems are being developed and deployed for healthcare that we situate our research. The key research questions and objectives of our study are described in the following section.

1.1. Scope of the study

By mapping the data supply chains that underlie AI technologies in the health sector, our study investigates i) how AI systems are being developed and deployed to achieve healthcare outcomes and, more importantly, ii) how AI audits are perceived and implemented by key stakeholders in the ecosystem. More specifically, the research focuses on three broad questions:

- **What is the current data supply chain infrastructure for organisations operating in the healthcare ecosystem in India?**
- **What auditing practices, if any, are being followed by tech companies and healthcare institutions?**
- **What best practices can organisations based in India adopt to improve AI auditability?**

Nevertheless, the study's scope is determined by certain limitations, as we attempt to contextualise and present the findings from our research. First, the report focuses primarily on use cases where the deployment of AI systems is most relevant from the perspective of the Indian healthcare system. For instance, AI-based technologies used for ancillary purposes, such as insurance disbursal, are not included in this study's scope. Similarly, the study does not focus on the use of AI in drug discovery due to its vast scope, the different stakeholders involved, and the differences in the AI lifecycle.[18]

Second, the study prioritises AI systems designed for healthcare providers and professionals over independent patient use, given the former's centrality to the healthcare system. In that sense, medical practitioners are seen as the end user for the purposes of this study.

17. Xiaoxuan Liu et al., "The Medical Algorithmic Audit", The Lancet Digital Health 4, no. 5 (2022): e384–97

18. The OECD's Framework for the Classification of AI Systems defines the "life cycle" as "planning and design; collecting and processing data; building and using the model; verifying and validating; deployment; and operating and monitoring."OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, accessed 29 October, 2024.

Defining AI and AI in healthcare

It can be difficult to accurately define AI, as it is primarily used as a catch-all marketing term to brand a broad and often disparate set of computing techniques to simulate human intelligence. AI technologies can range from traditional analysis techniques, such as decision trees, to novel technologies, such as large-language models.[19]

UNESCO describes AI as systems "built from data, hardware and connectivity, allowing machines to mimic human intelligence such as perception, problem-solving, linguistic interaction or creativity".[20] Building upon this definition, we define AI as systems that use technologies such as machine learning, computer vision, neural networks, and language models to assist, automate, or replicate tasks traditionally associated with human intelligence. The tasks performed by AI include learning, reasoning, problem-solving, processing language, and perception.[21]

In this study, we define AI in healthcare as the use of artificial intelligence and related technologies to support healthcare research and service delivery. The use cases include assisted imaging and diagnosis, disease prediction, robotic surgery, automated patient monitoring, medical chatbots, hospital management, drug discovery, epidemiology, etc.

The landscape of AI systems is vast and rapidly evolving. Instead of being exhaustive in its scope, this study adopts a purpose-based classification approach, explicitly focusing on AI systems that are used to:

- **aid in early disease detection, personalised treatment plans, and predictive analytics;**
- **assist in radiographic evaluation to detect anomalies and diagnose diseases more accurately;**
- **provide round-the-clock assistance for basic medical inquiries and improve patient engagement;**
- **analyse vast amounts of patient data, including medical records, imaging, and genetic information;**
- **extract valuable information from unstructured clinical notes and electronic health records; and**
- **support on-field diagnosis, especially in low-resource areas.**

The objective behind choosing a purpose-based classification approach is to focus on a small but highly popular set of tools that rely on varying types of technological foundations, including but not limited to neural networks, computer vision, and natural language processing.

19. *Will Douglas Heaven, "What Is AI?", MIT Technology Review, 12 October 2024*
20. *"Artificial Intelligence", UNESCO, 28 June 2024*
21. *"What Is Artificial Intelligence (AI)?", ISO, accessed on 25 October 2024*

1.2. **Structure of this report**

This section provides a brief overview on how the report is structured.

Chapter 2 (Methods) provides an in-depth introduction to the study's data collection and analysis methods. In **Chapter 3** (AI in Healthcare: A Background), we use available research and literature to present an overall picture of the landscape of AI systems in healthcare – including relevant use-cases, the ethical principles surrounding their implementation, the existing regulatory landscape, and the concept of 'AI auditing'. It is also in this chapter that we introduce and discuss the 'Data Supply Chain' (DSC) framework and its four constituent stages – Data Sourcing, Data Processing, Model Development, and, lastly, Model Deployment.

In **Chapter 4** (Main findings and discussion: Data Supply Chain), we use this DSC framework to present evidence from our quantitative and qualitative research across all the four stages. Instead of focusing purely on the technical details of developing and deploying AI systems in healthcare, our findings also expand on the relationships between the multiple stakeholders that form part of the data supply chain, including technology companies, startups, healthcare institutions, and medical professionals. We also use chapter 4 to discuss the implications of these findings for the practice of AI auditing, which we elaborate in much more detail in the subsequent chapter.

Continuing this thread, we begin **Chapter 5** (Main findings and discussion: AI auditing as a response) by highlighting the current state of AI auditing, as understood by the many stakeholder groups covered under this study. We follow this landscape view with a deep-dive into the notion of AI auditing as a governance mechanism, to underline not only the possibilities attached to the practice but also the gaps and risks that are likely to resist its application.

Subsequently, **Chapter 6** (Recommendations) distils these findings into actionable suggestions for key decision-making authorities across the landscape of healthcare AI systems – focusing on improving the underlying data supply chains as well as streamlining the practice of AI auditing for the purposes of governance. This is finally followed by **Chapter 7** (Conclusion), which summarises the study's key takeaways and implications for the introduction of AI systems in India's healthcare sector.

2. Methods

Given the limited literature available on the prevalence of AI systems in the Indian healthcare sector, this study is exploratory in nature and intended as an initial research intervention to understand the topic better. As AI auditing frameworks are not yet widely prevalent in India, the data supply chain is used as an entry point to understand the algorithmic infrastructures that underlie AI systems currently being developed and deployed in India. The study design, therefore, is based on a mixed methods approach to arrive at a fuller, more comprehensive picture of the prevalence of AI in Indian healthcare systems and the use of auditing mechanisms, if any.

2.1. Study design

The discourse on AI applications is currently technological, with limited interventions aimed at understanding the socio-technical aspects of these systems. A combination of qualitative and quantitative research was used to gain insights into the larger socio-cultural factors shaping the perception and adoption of AI in healthcare in India. In its first phase, the study leveraged a scoping literature review and stakeholder mapping to i) map the existing state of knowledge on the topic, ii) develop and refine the core hypotheses, and iii) identify important organisations involved in building healthcare-related AI systems in India.

The research questions and the refined hypotheses were subsequently coded into detailed questionnaires. Following this, the research team undertook primary data collection through three quantitative surveys of 500 respondents, and 18 in-depth interviews with representatives from domains of relevance to the topic of research and enquiry. These constituencies included medical professionals, AI technology professionals, academicians, policymakers, and civil society groups working on AI and healthcare, data privacy, and regulation. The study setting is India, with a study population that comprised the aforementioned domains, all situated in India or an Indian branch of an international healthcare institution, IT/ITES[22] company, or startup.

2.2. Sampling

For the quantitative surveys, the total sample size included 500 respondents, comprising 150 medical professionals, 175 representatives from healthcare institutions, and 175 respondents from technology companies. The surveys employed purposive sampling methods. Within purposive sampling, an expert sampling approach was used to identify individuals with expertise in AI and healthcare in India.

22. *This is an acronym for Information technology/ Information technology-enabled services.*

Figure 1 presents the stakeholder-wise breakdown of the 500 survey respondents.

Figure 1: **Sample size- Stakeholder-wise breakdown of survey respondents**

Stakeholder group 1: **Medical professionals**

Total count	150
Median age	41 years old
Gender breakup	50% men, 50% women
Current role	→ 31% practise medicine in a public healthcare facility → 74% practise medicine in a private healthcare facility → 45% conduct research in an academic or medical institution → 9% conduct research with healthcare-focused start-ups, consulting firms and technology companies

Stakeholder group 2: **Respondents from healthcare institutions**

Total count	175
Median age	39 years old
Gender breakup	51% men, 49% women
Primary organisational affiliation	→ 33% work at standalone private hospitals and 11% work at large private hospital chains → 23% work at private medical research institutes → 16% work at public hospitals → 16% work at public medical research institutes → 1% work at community health centres or clinics

Stakeholder group 3: **Respondents from technology companies**

Total count	175
Median age	34 years old
Gender breakup	50% men, 50% women
Primary organisational affiliation	→ 56% work at healthcare-focused consulting/IT/ITes companies (33% at a global organisation and 24% at an Indian organisations) → 26% work at HealthTech startups (4% at a global organisations and 22% at an Indian organisations) → 18% work at a BigTech company (9% at a global organisations and 9% at an Indian organisations)

Source: CIS survey of professionals in AI and healthcare, January-April 2024.

Figure 2: **Sample size – Stakeholder-wise breakdown of qualitative interviews**

Stakeholder groups	
Tech Companies/IT/ITES services (that have worked in AI healthcare space)	4
AI in Healthcare startups/ developers	2
Academics working on AI and healthcare, data privacy and regulation	3
Policy makers/government officials working on areas of health and digitisation	2
Medical Professionals: doctors as well as professionals involved in various roles in healthcare institutions	5
Civil society/patient advocacy groups	2
Total count	**18**

Source: CIS interview of professionals in AI and healthcare, January-April 2024

2.3. Recruitment and data collection

For the quantitative surveys, the respondents were onboarded with the aid of a third-party data collection agency that used its internal database – supplemented with licensed databases such as ZoomInfo and Linkedin Sales Navigator – to identify and source the requisite respondents. The third-party data collection agency was provided with a series of screening questions to determine the respondents' qualifications for the survey. After ensuring adequate data quality and conducting hygiene checks, the agency shared the raw datasets for each stakeholder with the research team, which then analysed these outputs internally to identify findings and insights.

Data collection for the surveys was implemented by the agency between January and April 2024 through self-administered web surveys. The research team at CIS only received anonymised datasets of the survey responses from the data collection agency. The anonymised datasets were stored on a secure internal cloud-based server, with access restricted to only the research team.

Throughout the data collection process, the research team conducted several quality checks. The data cleaning and verification process involved flags to remove or revalidate observations. The flags identified inconsistencies in observations, combination of responses that lead to unlikely or impossible outcomes, and discrepancies in text responses. When observations received beyond a certain number of flags, the respondent was either re-contacted by the survey agency or removed entirely from the sample for analyses. The survey agency conducted re-surveys to replace these observations.

For the in-depth interviews, research participants were identified by the data collection agency based on the requirements shared by the research team at CIS. In addition, the research team also used referential or snowball sampling to identify people working on the broader themes of AI and ethics and data privacy, management, and regulation across the six key stakeholder groups listed in Figure 2. As with the quantitative data, we obtained informed consent from all the participants to record, store, use, and delete data. All personally identifiable information (PII) was separated from the actual study dataset/interview recordings and stored on a secure, internal cloud server.

2.3.1. **Desk-based secondary research**

The first phase of this study consisted of a review of literature to map the existing state of knowledge on the research topic. This comprised a scoping review to understand better the available literature on auditing AI in healthcare and the key stakeholders in this space in the Indian context. The available literature on auditing AI is very limited for the Indian healthcare sector, given that AI development and deployment are still in the early stages. We therefore surveyed a variety of sources, including the academic literature, journalistic reportage, and policy developments on AI and healthcare, data management, and privacy. Key thematic areas that emerged were the discourse on ethics of AI use, the potential for algorithmic bias and its impact on marginalised populations, and the role of health digitisation and the data supply chain in determining the design of AI infrastructures when they are finally deployed in healthcare systems. Consequently, the study adopted the data supply chain framework as an entry point to understand the prevalence and use of auditing practices for AI systems in India.

2.3.2. **Quantitative data**

Quantitative data was collected using a self-administered web survey tool facilitated by a third-party data collection agency. Three separate surveys targeted different stakeholder groups – medical professionals, respondents from healthcare institutions, and respondents from technology companies. The surveys covered topics such as AI use in healthcare, data supply chains, audit practices, and challenges related to AI deployment in healthcare. The research team performed a thorough data cleaning process and analysed the survey data primarily using descriptive statistics.

Given the study's exploratory nature, convenience and purposive sampling approaches were used for quantitative data collection, albeit with a few integral constraints:

a. **Only individuals with a reasonable understanding of and involvement in researching, developing, or deploying healthcare-related AI were included**

b. **Only Indian nationals who are currently based in India – even if their parent organisation is not – were included**

c. **An equal distribution of men (including trans-men) and women (including trans-women) was ensured across all stakeholder groups**

The survey tools for different stakeholders included sections on the following themes:

- **Socio-demographic information**
- **Familiarity with AI use in healthcare**
- **AI audit and data supply chain**
- **Ethical challenges and concerns with the use of AI tools in healthcare and**
- **Opinions on the future of AI use in India.**

2.3.3. **Qualitative data**

In-depth interviews (IDIs) were conducted via virtual meetings with the recruited participants. In addition to their experience working in AI in healthcare, demographic parameters, such as the participants' job roles, designations, and gender, were considered during recruitment. Although CIS and the third-party data collection agency collaborated to identify and recruit IDI participants, the interviews themselves were conducted solely by the core research team. Minimal personal information (including name, designation, organisation and contact details) required to meet the research goals was collected by the implementing organisation and the research team. The core research team conducted the interviews after securing consent from the research participant with regards their preferences for recording, note-taking, and attribution.

The primary objective behind the IDIs was to gather descriptive and open-ended information from AI experts in healthcare based on themes derived from the research questions. The interviewees were drawn from the study's three core stakeholder groups and complementary domains, such as academia, policymaking, and civil society.

The interview guide for the in-depth interviews included the following themes:

- **Data collection and data quality**
- **Familiarity or hesitation with AI use in healthcare**
- **Audit and familiarity with auditing practices**
- **Ethical and regulatory challenges**
- **Deployment of AI systems**
- **Opinions on the future of AI use in India**

2.4. **Data analysis**

Data collection was followed by analysis for a period of three months, during which time the research team worked on collating insights from the quantitative and qualitative datasets. This comprised coding the collected data (quantitative and qualitative) based on key themes related to the data supply chain. These key themes were data sourcing, collection and management; data quality and security; the sourcing, development and deployment of AI systems; ethical and practical challenges; training and infrastructure; and auditing and regulation. The quantitative data was further disaggregated based on key stakeholders, such as healthcare professionals, hospitals, and tech companies.

The IDIs were analysed using the inductive approach, where two members of the research team read the transcripts and notes from each recording to identify the themes and categorise them. Once the categories were finalised, a framework was devised to summarise the raw data and examine the key emerging themes. Each category was coded in a unique colour, simplifying the analysis of the interview segments. It also helped in bundling categories into themes and flagging similarities and differences in opinion within the stakeholder groups. The topics that did not fit into any of the categories were added as an uncategorised or miscellaneous section.

The qualitative and quantitative datasets were also parsed for any patterns, unique findings, or contextual factors that inform the development and deployment of AI systems. They were also reviewed for the use of auditing practices, if any. These learnings were then consolidated through discussions to arrive at a set of recommendations aimed at some of the key stakeholders in the study.

2.5. **Study ethics**

The Institutional Review Board at SIGMA Research and Consulting approved the study in December 2023. Their review encompassed a detailed examination of the study's design, ethical considerations, a risk assessment matrix, and a data management plan. We had a data management plan to address ethical risks related to participant confidentiality and handling PII. It also included measures to safeguard the rights of research participants, including privacy and withdrawal from the study, informed consent, and transparency regarding the use of the study data. For the in-depth interviews, we established specific protocols for separating PII from study data and adopted mechanisms to ensure informed consent for recording and providing appropriate attribution. For the survey data, additional checks were implemented by the data collection agency to ensure the accuracy and verifiability of the datasets. Finally, we established protocols for storing, retaining, using, and deleting both data sets by the agency and CIS after the required time.

2.6. **Study limitations**

- **Desk-based secondary research:** The most significant limitation was the lack of sufficient secondary literature on the prevalence of auditing AI in the Indian healthcare sector. As a result, the team relied on a range of sources to generate an overview of the landscape, including but not limited to the use cases of AI in the Indian healthcare sector, potential risks of bias and harm, the management of health data, and the use of auditing practices if any.
- **Survey:** The survey used convenience and purposive sampling approaches to identify experts within the field of AI and healthcare in India. Therefore, the survey data is not representative of the population, posing a limitation. The data analysis mainly employed descriptive statistics. The surveys were self-reported and conducted online which can lead to issues like poor interpretation of questions and participant fatigue that may affect data reliability. Additionally, self-reported interviews can introduce response bias, where participants might provide socially desirable answers over honest ones.[23] To address these concerns, the research team implemented a rigorous data cleaning process to eliminate inconsistencies. We have ensured that all data presented in this report reflects the respondents' views accurately.
- **In-depth interviews:** Qualitative methods have limitations, particularly concerning gender representation in startups and IT/ITES sectors. Efforts to engage with patient advocacy groups were unsuccessful, as our attempts to connect through established networks did not lead to identifying organisations focused on AI in healthcare or health digitisation. There were also concerns about recording, storing, and retaining PII and attribution. To mitigate these, our consent form included measures such as anonymisation and allowed participants to decline or stop recording at any time. As mentioned earlier, protocols were also implemented for the effective management, use, and disposal of data by the data collection agency and CIS.

23. *F. Kreuter, S. Presser, and R. Tourangeau, "Social Desirability Bias in CATI, IVR, and Web Surveys: The Effects of Mode and Question Sensitivity," Public Opinion Quarterly 72, no. 5 (December 1, 2008): 847–65*

3. **AI in Healthcare:** A Background

In line with global trends, industry estimates expect the use of AI tools in India's healthcare system to grow exponentially in the near future. A 2023 NASSCOM study, for example, claimed that the use of data and AI in healthcare has "the potential to add $25-$30 billion to India's GDP by 2025."[24] This belief is clearly visible in the country's healthtech sector, which includes a range of startups as well as big-tech entities such as Microsoft and Google – partnering with large-scale private hospitals, such as the Apollo chain of hospitals.[25] Although private investments and public initiatives continue to incentivise the use of AI systems in healthcare, governance of these systems remains sparse.

In this chapter, we aim to investigate this lacuna in governance processes and build a shared understanding of AI for healthcare in India. Relying primarily on existing literature, this chapter presents a background reading of this ecosystem, including the various use cases where AI systems are deployed, existing regulatory and governance mechanisms, and the role of AI audits in the process.

3.1. **Use cases for AI in healthcare**

The use of AI in healthcare is slowly gaining momentum, with a range of areas and sub-disciplines receiving increased interest from researchers and technologists. These include, but are not limited to, disease diagnosis, health and wellness monitoring, virtual care consulting, drug discovery, and even medication management.[26] In fact, large-scale AI systems, including IBM's Watson and Google's Deep Mind, are used in several healthcare-related activities, such as detecting non-communicable diseases, including diabetes and certain cancers, providing medical assistance, and monitoring patients remotely.[27]

In India, the use of AI systems in healthcare has also seen increased interest from technology companies and various state governments. For instance, the Karnataka government recently partnered with AstraZeneca to deploy AI-based solutions for screening for lung cancer in the state.[28] Similarly, the Andhra Pradesh government has collaborated with Hellokidney.ai, a startup, to develop and roll out a mobile application that could perform screenings for kidney-related diseases.[29] In addition to these public-private partnerships (PPPs), major Indian hospital chains – such as Manipal, Max, and Apollo – claim to use AI-based tools in many of their departments.[30] [31] [32]

24. *"How AI Is Transforming the Future of Healthcare in India", Nasscom.*

25. *Shashank Saini, "Healthtech Startup Landscape in India", siliconindia, accessed 25 October 2024*

26. *Junaid Bajwa et al., "Artificial Intelligence in Healthcare: Transforming the Practice of Medicine", Future Healthcare Journal 8, no. 2 (2021): e188–94*

27. *Ibid.*

28. *PTI, "K'taka Govt Joins Hands with AstraZeneca to Deploy AI-Based Lung Cancer Screening Technology", ETHealthworld, 13 October 2023*

29. *"State Government to Screen Kidney Diseases with AI-Powered Mobile App", INDIAai, 20 March 2024*

30. *"Elevate Radiology with Artificial Intelligence: Transforming Images into Insightful Reports, Seamlessly", Manipal Hospitals Radiology Group, accessed 25 October 2024*

31. *"Artificial Intelligence-Based Echocardiography", Max Healthcare, accessed 25 October 2024*

32. *BL Chennai Bureau, "Apollo Hospitals Launches AI-Powered Clinical Intelligence Engine for Doctors", BusinessLine, 7 February 2023*

In Figure 3, we present a non-exhaustive list of healthcare-related areas that have integrated key AI systems, along with illustrative examples of these solutions.

Figure 3: **Use cases of AI systems in healthcare**

Healthcare-related areas	Popular purpose(s) for using AI	Illustrative examples/use cases
Imaging-based diagnosis	Processing images and scans to diagnose diseases and/or identify problem zones[33]	Qure.ai's 'qLC-Suite' product studies chest X-rays to help early-stage lung cancer detection[34]
Primary health consultations	Facilitating doctor-patient communication (for instance, through chatbots) and designing personalised treatment plans[35]	Kommunicate's 'Healthcare AI' chatbot covers services like telemedicine and preventive care, among others[36]
Administration and management	Predicting maintenance requirements and improving decision-making in non-clinical areas[37]	Microsoft's AI services automate operational and administrative tasks, such as documenting electronic medical records[38]
Automated care services	Guiding robotic arms during high-risk surgeries and enabling personalised care robots, especially for the elderly[39] [40]	Meril's 'Cuvis Joint Robotic System' is a fully automated surgical robot for artificial knee joint replacements[41]
Drug discovery	Predicting potential chemical compounds and enabling drug-repurposing[42] [43]	Petris's platform technology uses neural network models to predict protein-molecule interactions[44]
Epidemiology	Facilitating disease surveillance, especially in high-contingency scenarios[45]	Wadhwani AI's 'Event-Based Disease Outbreak Monitoring' system scans digital media to track events that signal an impending disease outbreak[46]
Medical research	Informing the process of designing, conducting, and reviewing clinical studies[47]	Oncoshot's 'Recommend Tx' is a large-language model that supports cancer patients in reviewing and finding matching clinical trials[48]

33. Filippo Amato et al., "Artificial Neural Networks in Medical Diagnosis", Journal of Applied Biomedicine, January 11 (2013); 45–58

34. " qLC-Suite : AI for Early Detection of Lung Cancer", qureAI accessed 25 October 2024

35. Kay T. Pham, Amir Nabizadeh, and Salih Selek, "Artificial Intelligence and Chatbots in Psychiatry", Psychiatric Quarterly 93, no. 1 (2022): 249–53

36. "AI-Powered Chatbot for Healthcare | Deliver Superior Patient Experience", Kommunicate, accessed 25 October 2024

37. Abeer Malik and Barry Solaiman, "AI in Hospital Administration and Management: Ethical and Legal Implications", in Research Handbook on Health, AI and the Law, edited by Barry Solaiman and I. Glenn Cohen (Cheltenham, UK: Edward Elgar Publishing, 2024), 21–40.

38. Shritama Saha, "How Microsoft HoloLens Bridges the Healthcare Gap in Rural India", AIM, 28 May 2024

39. Amit Gupta et al., "Training and Credentialing in Robotic Surgery in India: Current Perspectives", Journal of Minimal Access Surgery 18, no. 4 (2022): 497

40. James Wright, "Inside Japan's Long Experiment in Automating Elder Care", MIT Technology Review, 9 January 2023

41. "CUVIS Joint Robot System for Artificial Joint Surgery", Meril, accessed 25 October 2024

42. Debleena Paul et al., "Artificial Intelligence in Drug Discovery and Development", Drug Discovery Today 26, no. 1 (2020): 80–93

43. Yadi Zhou et al., "Artificial Intelligence in COVID-19 Drug Repurposing", The Lancet Digital Health 2, no. 12 (2020): e667-76

44. "Peptris: AI Plays Tetris with Proteins", Peptris, accessed 25 October 2024

45. Pranav Anjaria et al., "Artificial Intelligence in Public Health: Revolutionising Epidemiological Surveillance for Pandemic Preparedness and Equitable Vaccine Access", Vaccines 11, no. 7 (2023): 1154

46. "E-Health: AI Solutions", Wadhwani AI, accessed 25 October 2024

47. Matthew Hutson, "Cutting to the Chase", Nature, vol. 627, 14 March 2024

48. TNN, "AI to Match Indian Cancer Patients with Most Effective Clinical Trials", The Times of India, 18 June 2023

Although integrating AI systems in these areas is a more developed phenomenon in the Global North, findings from our survey confirm that many of them are also popular use cases in the Indian context. For instance, over 40% of the surveyed medical professionals listed early disease detection and drug discovery as two prominent areas where AI systems are already being used. However, this increasing integration of AI in healthcare is not without risks. In fact, harms such as algorithmic bias and gaps in accountability are crucial challenges, especially given the safety-critical and consent-based nature of healthcare delivery.[49]

In the next section, we discuss existing regulations, policies, and frameworks that serve as guardrails for the use of AI in the context of India's healthcare system.

3.2. **Governance of AI in healthcare**

Conversations on the governance of AI systems have gained prominence globally. From global measures such as the UN AI advisory body's report to more regional regulations such as the EU's AI Act, there have been attempts to regulate AI solutions to mitigate the risks and harms that these systems pose.[50] [51] Additionally, the last few years have witnessed significant global research on ethical codes, principles, and frameworks for using AI systems in healthcare. Some of this research has been translated into guiding documents, such as the World Health Organisation's *Ethics and Governance of Artificial Intelligence for Health*.[52]

As we highlight in the subsequent sections, the governance of AI systems in healthcare is relatively nascent in India.

3.2.1. **Regulatory and policy developments in the background of AI and healthcare**

Even before the introduction of modern and multi-dimensional technologies such as AI, the regulation of medical devices, drugs, and healthcare services in India has had a long history. To begin with, health, being a state subject under the Indian Constitution, is the responsibility of the respective state government, and the union government has only a secondary role. During the COVID-19 pandemic, the differences between state laws on healthcare became clearer as states adopted different measures to contain the pandemic.[53]

Secondly, the significant role played by private actors and market forces in manufacturing medical equipment and drugs often means that incentives to prioritise public health are often weighed against economic priorities such as trade and foreign investment.[54] This is also reflected in certain legislative efforts, such as the Drugs and Cosmetics Act 1940 and the Indian Medical Council Act 1956 (now replaced by the National Medical Commission Act, 2019), both of which govern manufacturers as well as medical professionals, where the patient or the consumer is outside the purview of the laws.

49. Mary Cummings, "Rethinking the Maturity of Artificial Intelligence in Safety-Critical Settings".

50. "Governing AI for Humanity", United Nations, 2024

51. Artificial Intelligence Act, 2024

52. "The AI Act Explorer", Future of Life Institute, accessed 25 October 2024

53. Kiran Kumar Gowd, Donthagani Veerababu, and Veeraiahgari Revanth Reddy, "COVID-19 and the Legislative Response in India: The Need for a Comprehensive Health Care Law", Journal of Public Affairs 21, no. 4 (2021)

54. Dinesh Singh Thakur and Prashant Reddy Thikkavarapu, The Truth Pill (Simon & Schuster, 2022).

More recent and more relevant to the nature of our discussion is the dual phenomenon of digitisation of medical records and digitalisation of healthcare through efforts such as telemedicine. Though there was a shift to tech-centric, data-driven healthcare service delivery models following COVID-19, many of these measures were introduced prior, and the pandemic only accelerated their adoption.[55] Since 2017, the National Health Authority (NHA) has spearheaded the creation of ecosystems, policies, and guidelines for collecting and managing health data as well as for creating health IDs. Named Ayushman Bharat Health Account numbers or ABHA numbers, these health IDs are unique 14-digit numbers that allow people to access and share their health records through established digital methods.[56]

Although the NHA has released several data management policies, they remain scattered and do not have any legislative backing, removing them from any judicial oversight.[57] In fact, despite some initial traction in the form of the Digital Information Security in Healthcare Act (DISHA), 2017, the only legislation currently applicable to data protection is the recently passed Digital Personal Data Protection Act, 2023 (DPDPA).[58] [59] However, many DPDPA clauses do not cover healthcare data. For example, the Act no longer categorises health data as 'sensitive personal data' – a categorisation that existed in the earlier drafts of the law – thereby reducing the much-needed extra protection for purposes such as training AI models.[60]

Another effort relevant to AI systems in healthcare is the National Medical Device Rules (NMDR), 2017 – created by the Central Drugs Standard Control Organisation under the Drug and Cosmetics Act.[61] Through a 2020 notification, the NMDR was amended to cover a range of devices, including but not limited to "a software or an accessory" designed to serve the purposes of disease diagnosis, patient monitoring, and injury assistance, among others.[62] However, whether this definition includes the vast gamut of AI systems remains unclear, creating yet another ambiguity in the governance of these solutions. Similar attempts to govern digitally mediated healthcare can also be noticed in developments such as the 2020 Telemedicine Guidelines, which cover new communication modes between medical professionals and patients.[63] Despite their clear focus on informed consent and personal data security, these guidelines suffer from concerns of jurisdictional overlap between the state and the centre, as well as the lack of a dedicated governance mechanism.[64]

Without appropriate, timely, and informed regulatory interventions, the existing policy apparatus will be unable to address the apparent and emerging risks of implementing AI in healthcare.. Consequently, one popular approach to fill this policy gap – best exemplified through global discourses on 'ethical AI', 'responsible AI', and 'AI for social good' – can be found in creating and adopting ethics-based principles and guidelines.[65] [66] [67]

We unpack some of these discourses briefly in the following subsection.

55. *Dipika Jain, "Regulation of Digital Healthcare in India: Ethical and Legal Challenges", Healthcare 11, no. 6 (2023): 911*

56. *"ABHA– Ayushman Bharat Health Account", National Health Authority, accessed 10 September 2024*

57. *Tejasi Panjiar, "Second Time's Not the Charm: Health Data Management Policy Misses the Mark Again", Internet Freedom Foundation, 23 May 2022*

58. *"Data Transfer of Digital Health Records", Press Information Bureau, 16 July 2019*

59. *The Digital Personal Data Protection Act, 2023*

60. *Akshayy S. Nanda, "The Impact of the DPDP Act, 2023 on the Healthcare Industry: A Detailed Exploration", ETLegalWorld, 23 June 2024*

61. *The Medical Devices Rules, 2017,*

62. *Ibid.*

63. *"Telemedicine Practice Guidelines – Enabling Registered Medical Practitioners to Provide Healthcare Using Telemedicine", Medical Council of India, 25 March 2020*

64. *U. Venkatesh, Gandhi P. Aravind, and Anbu Ananthan Velmurugan, "Telemedicine Practice Guidelines in India: Global Implications in the Wake of the COVID-19 Pandemic", World Medical Health Policy 14, no. 3 (2022): 589–99*

65. *"Global AI Ethics and Governance Observatory", UNESCO, accessed 25 October 2024*

66. *Nenad Tomašev et al., "AI For Social Good: Unlocking the Opportunity for Positive Impact", Nature Communications 11, no. 1 (2020)*

67. *Trishan Panch, Heather Mattie, and Rifat Atun, "Artificial Intelligence and Algorithmic Bias: Implications for Health Systems", Journal of Global Health 9, no. 2 (2019); Michael J. Rigby, "Ethical Dimensions of Using Artificial Intelligence in Health Care", The AMA Journal of Ethics 21, no. 2 (2019): E121-24*

3.2.2. **Ethical principles and practice**

The fast-changing nature of AI technologies, juxtaposed against the slow speed of regulations and concerns around regulations potentially stifling innovation, has led to the establishment of ethical codes, principles, guidelines, or frameworks for AI systems in healthcare as the preferred mode to introduce some form of checks on AI systems.

In fact, there is now a considerable body of literature on the ethical considerations and dilemmas in using AI, including principle-based guidance for data collection and management.[68] For example, a prominent application of this idea is evident in the Responsible AI for All approach document released by NITI Aayog in 2021.[69] Although the document remains sector-agnostic in its scope, it explicitly relies on ethical considerations while prescribing guiding principles.[70]

On the other hand, more pertinent to healthcare are principles espoused by the field of bioethics – 'autonomy', 'beneficence', 'non-maleficence', and 'justice' – which have also been adopted in many ways to govern AI systems for healthcare.[71] [72] Building on these four principles, for instance, in 2023, the Indian Council for Medical Research (ICMR) released the Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare (ICMR Guidelines). The document lists ten principles that ICMR considers essential for developing and deploying AI systems for healthcare.[73] These include i) autonomy, ii) safety and risk minimisation, iii) trustworthiness, iv) data privacy, v) accountability and liability, vi) optimisation of data quality, vii) accessibility, equity, and inclusiveness, viii) collaboration, ix) non-discrimination and fairness, and x) validity.[74] The document also expands on guidelines for academics, technology companies, clinicians, and the government, among others, to govern AI systems at different phases of its lifecycle.[75]

Despite their comprehensiveness and coverage of multi-stakeholder perspectives, the ICMR guidelines are just that – guidelines. For instance, the document empowers ethics committees in hospitals and medical research institutions to check for these principles and assess potential AI-related proposals for "data source quality, safety, anonymisation, data selection biases, participant protection, and the possibility of stigmatisation", among others.[76] However, as reported by the National Ethical Guidelines for Biomedical and Health Research Involving Human Participants, 2017, ethics committees usually focus on using AI systems in research and not as much in practice.[77]

Currently, the 2023 ICMR guidelines have the most detailed mapping of ethical principles guiding the use of AI in India's healthcare system, while the 2021 guidelines by NITI Aayog provide a more general framework. But without an overarching (or a healthcare-specific) law to govern AI systems, or even without a mandate to enforce these guidelines, their application in healthcare delivery remains unevaluated and, most likely, disparate. This has, inevitably, led to a situation where governance of AI systems has come to depend primarily on self-regulation – which allows private stakeholders, especially technology companies, to "escape democratic oversight" and indulge in "ethics-washing".[78] [79] [80]

68. *Jessica Morley et al., "The Ethics of AI in HealthCare: A Mapping Review", Social Science & Medicine 260 (2020): 113172*

69. *"Responsible AI, #AI for All – Approach Document for India. Part 1 – Principles for Responsible AI", NITI Aayog, 2021*

70. *Ibid.*

71. *"Post #9: How Bioethics Can Inform Ethical AI Governance", Edmond & Lily Safra Center for Ethics, 15 March 2024*

72. *Anna Jobin, Marcello Ienca, and Effy Vayena, "The Global Landscape of AI Ethics Guidelines", Nature Machine Intelligence 1, no. 9 (2019): 389–99*

73. *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare, 2023, ICMR*

74. *Ibid.*

75. *Ibid.*

76. *Ibid.*

77. *National Ethical Guidelines for Biomedical and Health Research Involving Human Participants, 2017, ICMR*

78. *Alyssa Wong, "Regulatory Gaps and Democratic Oversight: On AI and Self-Regulation", Schwartz Reisman Institute, 21 September 2023*

79. *Ben Wagner, "Ethics as an Escape from Regulation. From 'Ethics-Washing' to Ethics-Shopping?", in Being Profiled: Cogitas Ergo Sum: 10 Years of Profiling the European Citizen, edited by Emre Bayamlıoğlu, Irina Baraliuc, Liisa Janssens, and Mireille Hildebrandt (Amsterdam University Press, 2018), 84–89*

80. *'Ethics-washing' is a neologism that is used "to describe the phenomenon of instrumentalising ethics by misleading communication, creating the impression of ethical artificial intelligence (AI), while no substantive ethical theory, argument, or application is in place or ethicists involved"*

As elaborated in this section, the governance of AI systems in India's healthcare sector is highly fragmented, with many laws, policies, and regulations attempting to create a 'responsible' AI ecosystem. This scattered approach, along with a lack of standardisation, has also led to opaqueness – best exemplified through the use of contracts and non-disclosure agreements between parties engaged in deploying AI in healthcare, which allows them to determine their own set of rules, stipulations and processes – combined with a growing trend of self-regulation. One way these issues of governance (or a lack thereof) and transparency intersect can be seen in the emerging discourse on 'AI auditing', a concept we elaborate on in the subsequent section.

3.3. **AI audits**

Traditionally, auditing has been integral in evaluating complex systems and processes to determine if they comply with organisational, industrial, and regulatory goals and standards.[81] In India, they are usually associated with financial and monetary policymaking. For instance, the Office of the Comptroller and Auditor General is a constitutionally appointed body that monitors and reports on the collection and expenditure of public funds to "promote accountability, transparency and good governance".[82] Similarly, many regulations and policies rely on statutory auditing to introduce an element of financial transparency. A prominent example of this is found in the Companies Act, 2013, which not only mandates certain companies to conduct internal audits but also lays down detailed clauses on who should conduct them and how.[83]

On the other hand, the use of audits to evaluate AI systems, let alone those designed specifically for healthcare purposes, is relatively nascent and does not yet rely on a consistent framework. For starters, auditing an AI system depends on factors such as its intended purpose, the structure of its data supply chain, and the regulatory landscape of the relevant sector. For example, if the purpose is to validate an AI system, then ex-ante audits (i.e., those used before the deployment of an AI system) or in media res audits (i.e., those used during an iterative design process) are preferred.[84] [85] Despite this, some common features have emerged in recent years, often in the context of the Global North.

Consequently, we use the following two subsections to unpack the existing discourse around two integral aspects of AI audits for healthcare – i) who conducts them, i.e., the question of accountability and independence, and ii) what they include, i.e., the question of underlying frameworks and auditing parameters.

81. *Deborah Raji et al., "Closing the AI Accountability Gap", FACCT Proceedings, 27 January 2020*

82. *"Our Vision, Mission and Values", Comptroller and Auditor General of India, accessed 25 October 2024*

83. *The Companies Act, 2013. Government of India*

84. *Hidde Lycklama, et al. 2024. "Holding Secrets Accountable: Auditing Privacy-Preserving Machine Learning", arXiv:2402.15780*

85. *Ibid.*

3.3.1. The internal vs external debate

India's healthcare system consists of a diverse set of stakeholder groups, including many whose incentives are not always aligned with each other. Given the fragmented nature of policies around AI auditing, this misalignment is also visible in the relationship between the auditor and the stakeholder being audited – most prominently in the decision between an internal or an external auditing process. Although existing principles – such as the 2023 ICMR guidelines (which suggest both internal and external audits at varying stages) – play a specific role, the choice is also influenced by other factors.[86]

Internally regulated audits, for example, allow the audited stakeholder not just to select the auditor but, in the absence of an external mandate, also to decide the scope, depth, and outcomes of the auditing process.[87] Consequently, they can include or exclude the entire development cycle of an AI solution from the scope of the audit without caring about confidentiality risks or adverse public perceptions. Along similar lines, their greater access to the algorithm and direct communication with developers can enable internal auditors to build ethical awareness within the organisation.[88] More importantly, a stronger alignment with organisational leadership makes it more likely that an internally regulated audit's findings will be translated into tangible changes.[89]

On the other hand, externally regulated audits – often conducted by third-party organisations – may not have the same visibility regarding the algorithm as its internal counterparts. Thus, their influence on the development of the algorithm could be limited, especially when enforcement of AI audits remains voluntary. However, third-party auditors also have the liberty to deploy more inclusive auditing frameworks, thereby allowing them to inspect AI systems on a broad range of parameters, not just those that are considered integral by the stakeholders being audited.[90] [91] If linked with legal and statutory guidelines, externally regulated audits can also aid in maintaining public accountability and informing policy action.[92]

Nevertheless, both types of audits also come with their own challenges. Internally regulated audits are often accompanied by fears of corporate interference and retaliation. In contrast, externally regulated audits, even when bound by law, can become mere checklists and enable a culture of audit-washing.[93] Although enforceable guidelines and laws can counter some of these risks, much also depends on the purpose underlying the audit and, more importantly, what the audit includes.

We turn to this latter question in the following subsection.

86. *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare, 2023, ICMR.*

87. *Birhane Abeba, et.al. "AI Auditing: The Broken Bus on the Road to AI Accountability". In 2024 IEEE Conference on Secure and Trustworthy Machine Learning (SaTML), 9-11 April 2024, Toronto, Canada, 612–43*

88. *Inioluwa Deborah Raji et al., "Closing the AI Accountability Gap".*

89. *Birhane Abeba et.al. "AI Auditing: The Broken Bus on the Road to AI Accountability".*

90. *Ibid.*

91. *Inioluwa Deborah Raji et al. "Outsider Oversight: Designing a Third Party Audit Ecosystem for -AI Governance."*

92. *Ibid 89.*

93. *Ibid 89.*

3.3.2. **Auditing frameworks and contents of AI audits**

As discussed earlier, no single approach to auditing AI systems in healthcare is considered appropriate or mandatory. Instead, we have a broad spectrum of auditing frameworks, such as clinical testing, pharmacovigilance, and peer review.[94] [95] The landscape is further complicated by the fact that in the absence of established frameworks, each of these approaches can be deployed varyingly.

Consequently, the objective of this subsection is not to provide an exhaustive list of existing frameworks but instead highlight a few prominent ones as illustrative examples. Many of these frameworks also overlap with each other's scope and focus areas, and some are often used in conjunction with others.

1. **Algorithmovigilance** is a framework that recognises that "algorithms have the potential for both great benefit and harm and, therefore, require study."[96] More specifically, it includes a set of "scientific methods and activities relating to the evaluation, monitoring, understanding, and prevention of adverse effects of algorithms in health care."[97] Instead of relying solely on automated statistical tests to monitor the performance of AI systems in specific contexts and with particular datasets, algorithmovigilance incorporates a range of human factors that affect the development and deployment of AI systems.[98] These include but are not limited to their impact on medical decision-making and clinical workflows, transparent reporting of algorithmic vulnerabilities and bugs, and standardisation and harmonisation of data collection and reporting practices across healthcare institutions.[99]
2. **Ethical AI**, or the EAI framework, alternatively, is an auditing approach focused on "operationalising ethics, grounded in existing guidelines that provide actionable solutions"[100]. Structured around the AI lifecycle of i) data management, ii) model development, and ii) deployment and monitoring, the framework assesses AI solutions against ten ethical principles, including but not limited to the four principles of bioethics (autonomy, beneficence, non-maleficence, and justice).[101] In fact, by further expanding these four principles to embrace "sustainability" (the ecological impact of AI systems) and "solidarity" (the impact on social cohesion and the lives of the marginalised), the EAI framework also includes concerns that are critical for a developing country such as India.
3. **Medical Algorithmic Audit**, or MAA, is a tool to "better understand the weaknesses of an artificial intelligence system and put in place mechanisms to mitigate their impact."[102] At its core, the MAA framework builds on two existing AI audit approaches – SMACTR (scoping, mapping, artefact collection, testing, and reflection) and FMEA (failure mode and effects analysis).[103] Most prominently, the MAA framework allows internal developers to pre-empt errors and biases during the development process while recommending clinical actions, such as introducing human oversight. Furthermore, integrating robust feedback loops also lends itself well to a dynamic where the responsibility to audit is shared between AI developers, healthcare providers, and end-users.

---

94. Pharmacovigilance is the process and science of monitoring the safety of medicines and taking action to reduce the risks and increase the benefits of medicines

95. "Pharmacovigilance", European Commission, accessed 25 October 2024

96. Peter J. Embi, "Algorithmovigilance – Advancing Methods to Analyze and Monitor Artificial Intelligence-Driven Health Care for Effectiveness and Equity", JAMA Network Open 4, no. 4 (2021): e214622, p 2

97. Ibid.

98. Alan Balendran et al., "Algorithmovigilance, Lessons from Pharmacovigilance", Npj Digital Medicine 7, no. 1 (2024)

99. Alan Balendran et al., "Algorithmovigilance, Lessons from Pharmacovigilance".

100. Pravik Solanki, John Grundy, and Waqar Hussain, "Operationalising Ethics in Artificial Intelligence for Healthcare: A Framework for AI Developers", AI And Ethics, no. 3 (2022): 223–40

101. Ibid.

102. Xiaoxuan Liu et al., "The Medical Algorithmic Audit", The Lancet Digital Health 4, no. 5 (2022): e384–97

103. Inioluwa Deborah Raji et al., "Closing the AI accountability Gap".

Figure 4 presents a detailed description of the MAA framework and its associated activities.

Figure 4: **An MAA framework that uses SMACTR and FMEA approaches**

Scoping	**Mapping**	**Artifact collection**	**Testing**	**Reflection**	**Post audit**
Define audit scope	Map artificial intelligence system	Audit checklist ∘ Intended use statement ∘ Intended impact statement ∘ FMEA clinical pathway mapping ∘ FMEA clinical task risk analysis ∘ FMEA risk priority number document ∘ Datasets ∘ Data description ∘ Data, including explainability artifacts ∘ Data flow diagram ∘ The artificial intelligence model itself, if available ∘ Model summary ∘ Previous evaluation materials	Exploratory error analysis	Risk mitigation measures	Algorithmic audit summary report
Understand intended use	Map health-care task		Subgroup testing	Developer actions	Plan re-audit
Define intended impact	Identify personnel and resources		Adversarial testing	Clinical actions	
	Identify and prioritise risks				
	FMEA				

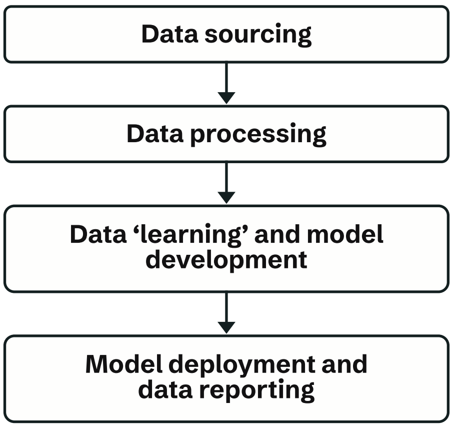
Source: The Medical Algorithmic Audit, The Lancet Digital Health 4, no. 5 (April 5, 2022)[104]

As we saw in previous sections, the landscape of AI systems in India's healthcare system, while nascent, is expected to grow rapidly in the coming years. In contrast, the regulatory and ethical regimes governing these technologies remain fragmented, voluntary, and without any legal mandate attached to them. In this context, AI auditing has emerged as a prominent – yet largely self-regulatory – practice for organisations to create and use AI systems 'ethically'.

In the following and concluding section of this chapter, we introduce the concept of the data supply chain (DSC) framework and briefly discuss the typical lifecycle of a healthcare-focused AI solution. The objective of this section is not only to provide readers with a common framework to capture the study's findings but, more importantly, to stress the criticality of data to today's AI systems.

104. *Xiaoxuan Liu et al., "The Medical Algorithmic Audit".*

Figure 5: **Visual representation of the Data Supply Chain framework and the four key steps it encompasses**



Source: CIS representation of data supply chain for AI

3.4. **The data supply chain framework**

Given the diverse spectrum of AI systems today, especially in healthcare, no single framework can offer an exhaustive, end-to-end understanding of every AI solution. However, the AI lifecycle approach is often used to discuss more generalised findings about a broad array of AI technologies, algorithms, and solutions. Usually consisting of three phases, the approach imagines an AI solution as a combination of i) design, ii) development, and iii) deployment practices, which can be, in many cases, further broken down into specific steps and activities. For example, data collection and preparation are usually undertaken during the design phase, whereas the creation of actual models usually occurs at the development stage.[105]

Building on the foundations of this lifecycle approach, we adopt a modified version commonly used in production research. Instead of just tracking the activities underlying an AI solution, the DSC framework also focuses on the data flows that underpin these technologies.[106] This framing (illustrated in Figure 5) is particularly helpful in India's multi-stakeholder context, where medical records are often digitised at the healthcare delivery site or by frontline workers. In contrast, the training of AI systems is usually led by technology companies and startups, thereby necessitating data flow between multiple stakeholders.

The prevailing practices and processes at each stage are elaborated in the following subsections, with a specific focus on India's healthcare system. The subsequent chapter on the study's findings goes into further detail and presents the gaps and challenges across the DSC.

3.4.1. **Data sourcing**

To work accurately and maintain their validity, AI systems require large quantities of good-quality data that have a few distinct features. At the data sourcing stage, for instance, the representativeness of the training data across various geographic and demographic parameters, such as gender, genetic history, dietary lifestyle, and economic background, is integral. Given the highly contextual nature of healthcare delivery, this representativeness is inevitably tied to the data collection process and the sources from which this data is collected.

In India, significant amounts of data collection are undertaken by community health workers, such as accredited social health activists (ASHAs) – mainly women who provide basic healthcare assistance in rural areas while gathering data on personal and community health indicators. As a result of this dual responsibility, their work is usually built on a negotiation between the trust they build and the invasive nature of data collection.[107] Collected through digital and physical forms in conversation with community members, this data is forwarded to appropriate data repositories, often serving a broad set of purposes.[108]

105. *Daswin De Silva and Damminda Alahakoon, "An Artificial Intelligence Life Cycle: From Conception to Production", Patterns 3, no. 6 (2022): 100489*

106. *Konstantina Spanaki et al., "Data Supply Chain (DSC): Research Synthesis and Future Directions", International Journal of Production Research 56, no. 13 (2017): 4447–66*

107. *"About Accredited Social Health Activist (ASHA)", National Health Mission, n.d.*

108. *Divya Thakkar et al., "When Is Machine Learning Data Good?: Valuing in Public Health Datafication", Association for Computing Machinery, 322 (2022): 1–16*

With the advent of digitisation, these repositories have also come to include electronic healthcare registries and databases, such as those being maintained under the National Digital Health Mission.[109] Additionally, in the absence of a single repository of health-related data, many AI developers also rely on private databases such as those managed by clinics, hospital chains, retail pharmacies, and diagnostic centres.[110] Furthermore, many AI systems use open data sources such as TensorFlow or Google Images for their training datasets, leading to concerns of geographic and demographic bias (we will return to this point in Chapter 5).[111] [112]

3.4.2. **Data processing**

Once the training data has been sourced from the identified repositories, it usually undergoes a process of inspection and pre-preparation at the data processing stage. The central objective of this step in the DSC is to mitigate data quality issues in the raw datasets, particularly in the form of unjustified gaps, missing documentation, and duplication of values, among others. This 'cleaning' of datasets to make them complete and legible for algorithmic training is often undertaken by appointed data stewards, usually those with data entry, data collection, and project management experience.[113]

Another essential part of data processing is data annotation, where identifying features and labels are added to the prepared datasets to enable the training of an AI solution.[114] [115] In fact, high-quality annotated data is essential for training machine learning models for the use cases of imaging, radiology, and diagnostics, and, therefore, annotators with some form of medical expertise are required.[116] Combined with the overall labour-intensive nature of data annotation, many technology companies and AI developers often rely on in-house teams and medical professionals to annotate data or verify existing annotations before proceeding to the next stage.[117]

3.4.3. **Data learning and model development**

After the collection and processing of raw data to acquire clean and well-annotated datasets, model development follows. At this stage, an AI model – developed by technology companies, independent researchers, or healthcare institutions – is trained using requisite machine-learning techniques and processed datasets to perform a specific task. Depending on the intended purpose, this union of data with an AI algorithm involves one or more of two activities.[118]

1. **Training or re-training:** In use cases where AI systems are needed to perform a specific task, applied statistical techniques such as deep learning are used to train these models on large, annotated datasets. For example, Apollo Hospitals recently collaborated with Microsoft to launch the Clinical Intelligence Engine – an AI tool trained on Indian data that uses natural language processing to "emulate a doctor's decision-making capabilities".[119] Alternatively, the resource-intensive process of creating new AI models has given way to the process of re-training, which involves adapting pre-trained clinical language models to deliver specific outcomes.[120] A prominent example is RETFound, a foundation model for interpreting retinal images that can be adapted to perform disease detection tasks more efficiently than its conventionally trained counterparts.[121]

109. *Amit Mishra et al., "Mapping Healthcare Data Sources in India", Journal of Health Management 24, no. 1 (2022): 146–59*

110. *Samaneh Madanian et al., "mHealth and Big-Data Integration: Promises for Healthcare System in India", BMJ Health & Care Informatics 26, no. 1 (2019): e100071*

111. *"Models & Datasets", TensorFlow, accessed 25 October 2024*

112. *Candace Makeda Moore, "The Challenges of Health Inequities and AI", Intelligence-Based Medicine 6 (2022): 100067*

113. *Divya Thakkar et al., "When Is Machine Learning Data Good?: Valuing in Public Health Datafication".*

114. *Bárbara C. Benato et al., "Semi-Automatic Data Annotation Guided by Feature Space Projection", Pattern Recognition 109 (2020): 107612*

115. *Ibid.*

116. *Ayomide Owoyemi et al., "Artificial Intelligence for Healthcare in Africa", Frontiers in Digital Health 2 (2020)*

117. *Rina Chandran, Adam Smith, and Mariejo Ramos, "AI Boom Is Dream and Nightmare for Workers in Global South", Context, 14 March 2023*

118. *Divy Thakkar et al., "When Is Machine Learning Data Good?: Valuing in Public Health Datafication".*

119. *Suneeta Reddy, "How Apollo Hospitals Is Using AI to Detect and Treat Range of Diseases", The Week, 10 December 2023*

120. *Jason Fries et al., "How Foundation Models Can Advance AI in Healthcare", Stanford University Human-Centered Artificial Intelligence, 15 December 2022*

121. *Yukun Zhou et al., "A Foundation Model for Generalizable Disease Detection From Retinal Images", Nature 622 (2023): 156–63*

2. **Clinical validation and monitoring:** Besides training and re-training AI models, geographically and demographically suitable datasets are often used to validate and monitor existing tools. For instance, calculating and benchmarking the efficacy of AI algorithms using a sample dataset can allow regulators and policymakers to test for issues of bias, appropriateness, and effectiveness. Similarly, hospitals and clinics that deploy AI systems can also adopt validation and monitoring systems to ensure that these technologies are not generating inaccurate or unethical results.[122] [123]

Although differing in their overall purpose, all these processes rely on a substantial amount of digital and physical infrastructure to conduct the requisite training, re-training, or monitoring. This infrastructure, which is broadly referred to as 'compute', includes not just large-scale data centres and warehouses – to store the datasets, their respective libraries, or the AI model's parametric data – but also hardware technologies such as graphics processing units (GPUs) and the requisite software that enable these GPUs, all of which are integral to model development.[124] Industry estimates (as of August 2023) suggest that with 151 data centres spanning over 11 million square feet, India ranks 14th globally in terms of its data centre inventory.[125]

3.4.4. **Model deployment**

At the model deployment and data reporting stage, the trained/re-trained and validated AI solution is introduced and monitored in real-life use cases. As highlighted earlier in the report, these use cases range from smart chatbots that enable patient–doctor communication to more specialised solutions that perform specific tasks such as diagnosing lung cancer in potential patients.

The multi-stakeholder nature of healthcare delivery in India means that AI systems are often deployed in clinical scenarios through a partnership between the stakeholder owning the technology and the stakeholder looking to deploy it. An illustrative example of such a relationship to diagnose diabetic retinopathy can be seen in the 2018 collaboration between Microsoft and ForusHealth – a manufacturer of advanced medical devices used for various aspects of eye care.[126] Through the partnership, Microsoft's AI-based retinal-imaging APIs (or application programming interfaces) were integrated into the latter's 3nethra devices, a range of portable tools that can be used to conduct retinal scans in more inaccessible regions of the country.[127] [128] Similarly, Google recently collaborated with Apollo Radiology International to provide them with AI-based imaging solutions to detect diseases such as lung cancer, breast cancer, and tuberculosis.[129] [130]

However, in the absence of an overarching law governing the use of these solutions, the terms of engagement underlying such partnerships remain unclear. As we elaborate in the subsequent chapter, deploying an AI solution in a clinical setting is not always straightforward, and it often faces both infrastructural and capacity-related challenges.

122. *Tim Showalter. "Unlocking the Power of Health Care AI Tools Through Clinical Validation", MedicalEconomics, 2 August 2023*

123. *"Validation", Ferrum Health, accessed 25 October 2024*

124. *Jai Vipra, "Computational Power and AI", AI Now Institute, 19 April 2024*

125. *"India Data Centers – Entering Quantum Growth Phase", Colliers and Confederation of Indian Industry, 2023, 6*

126. *"650 Partners to Drive AI for All in India: Microsoft", Microsoft Stories India, 29 March 2018*

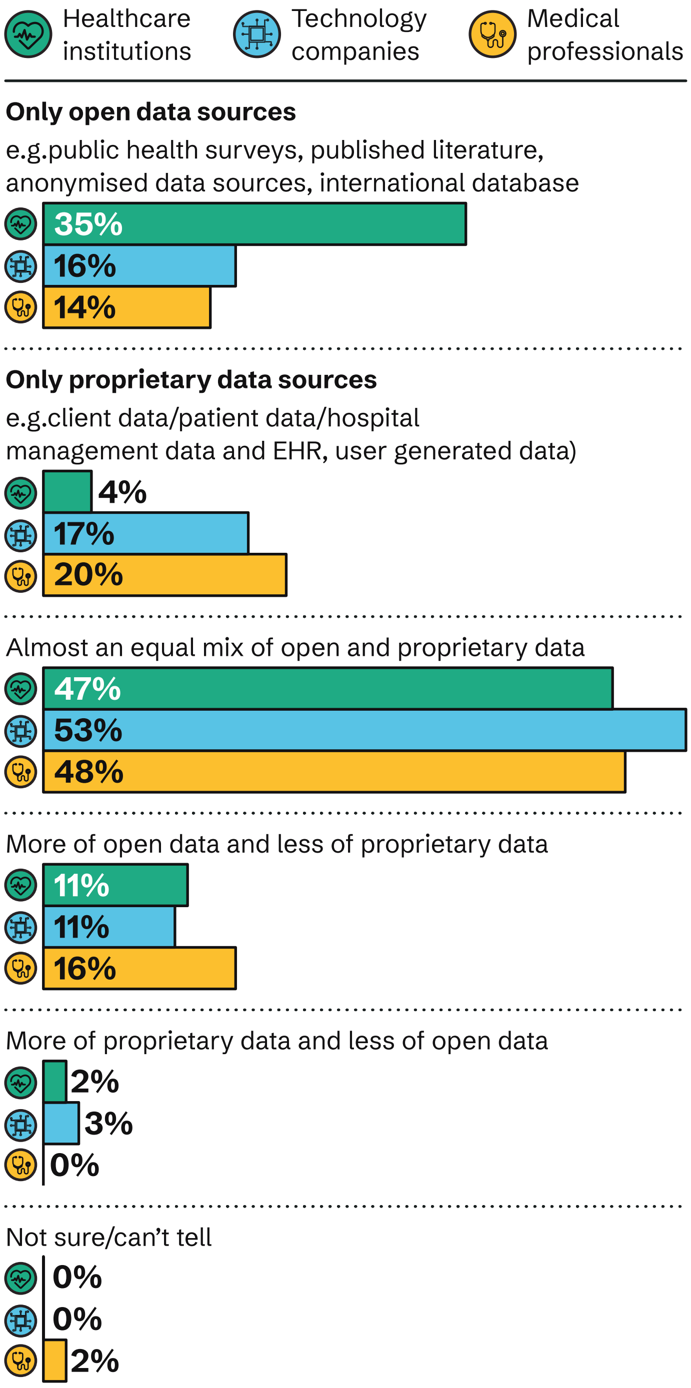
127. *"650 Partners to Drive AI for All in India: Microsoft ", Microsoft Stories India*

128. *"The Future of Digital Eye Health", Forus Health, accessed 25 October 2024*

129. *Shravya Shetty, "How AI Supports Early Disease Detection in India", Google, 19 March 2024*

130. *Shritama Saha, "Google Partners with Apollo Radiology for Early Disease Detection in India", AIM, 22 March 2024*

Fig 6: **Percentage of respondents from healthcare institutions, technology companies, and medical professionals on the types of data sources used for AI design/deployment**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Medical professionals (n = 133); healthcare institutions (n = 162); technology companies (n = 171)

4. **Main findings and discussion:** Data supply chain

In this section, we discuss the findings around the data supply chain for AI. This chapter has four subsections: data sourcing, data processing, data learning and finally model deployment.In each subsection, we first present the quantitative findings from the survey, followed by the qualitative findings from the interviews. At the end of each subsection, we also have a short discussion and audit implications section.

4.1. **Data sourcing**

Our survey and interview findings emphasise that the availability of health data, especially Indian datasets, as well as open and interoperable data, remains a persistent issue when developing AI systems for healthcare in India. Further, challenges associated with data collection, cleaning, digitisation, integration across systems/ institutions, and access for diverse stakeholders persist in developing robust Indian datasets for Indian AI models.

4.1.1. **Reliance on both open and proprietary data sources**

As per our survey, nearly 50% of the respondents from the medical field (including those working in healthcare institutions and medical professionals) and 53% of respondents from technology companies shared that they use a mix of open and proprietary data sources to develop and deploy AI systems in their fields (see Figure 6). Significantly, 35% of respondents from healthcare institutions said that they use only open data sources, compared to only 16% from technology companies and 14% of medical professionals.

This was further corroborated by interviewees. For instance, an academic working on machine learning in healthcare stated, Fig 7: **Percentage of respondents from healthcare institutions, technology companies, and medical professionals on types of data sources used**

* Highlighted entries indicates dependence on global north data sources

Top 3 data sources: healthcare institutions, medical professionals, technology companies	**% agreement**
Healthcare institutions	
Historical data from medical service providers (eg. hospitals, nursing home and medical research institutions) (Proprietary data source)	55%
* Open-sourced data from other countries (eg. US, Europe) (Open data source)	44%
Clinical data from research and medical trails (Proprietary data source)	43%
Medical professionals	
* Data and datasets available publicly such as Tensor Flow, Google images, SSRN, Google Scholar etc. (Open data source)	50%
* Open-sourced data from other countries (eg. US, Europe) (Open data source)	44%
Clinical data from research and medical trails (Proprietary data source)	43%
Technology companies	
* Data and datasets available publicly such as Tensor Flow, Google images, SSRN, Google Scholar etc. (open data source)	48%
* Proprietary data from technology companies (eg. Google, Amazon, IBM etc.) (proprietary data source)	38%
Clinical data from research and medical trails (proprietary data source)	37%

Source: CIS survey of professionals in AI and healthcare, January-April 2024. Medical professionals (n = 133); healthcare institutions (n = 162); technology companies (n = 171)

> “We take two approaches. One is publicly available datasets. We use those datasets. For our own datasets, we collaborate with hospitals with all appropriate approvals.
>
> As of now, we use mostly publicly available datasets, but we try to source more indigenous datasets from our collaborators. The public datasets are mostly non-Indian datasets – mostly from European and American institutions.”

In our interviews there were also some indications of tie ups with hospitals. For instance, big technology companies that provide software and hardware relating to AI and healthcare mostly rely on data from hospitals through contracts, data collection agencies, and public data. Significantly, most startups relied on various data sources to train their systems such as data from the Global North, data collected by themselves, open public data etc. On the other hand, larger companies had the market power to engage with large hospitals to share data, including government hospitals.

According to a senior executive at a startup working on AI diagnosis,

> “Teleradiology companies are the largest sources of data for [name redacted] and any other AI company as well. Most of the training data comes from India, some from Europe or the UK.”

4.1.2. **Reliance on datasets from the Global North**

Our survey revealed that proprietary data sources, particularly clinical data from research and medical trials, ranked among the top three data sources for all stakeholders – medical professionals, technology companies, and healthcare institutions. Historical data from medical service providers was the top data source for healthcare institutions (see Figure 7).

It is important to note that besides the notion of proprietary and open source data, there is a heavy reliance on Global North data; approximately 44% of healthcare institutions and medical professionals rely on open-source data from the US or Europe. Further, 50% of medical professionals and 48% of technology companies rely on open sources such as TensorFlow and SSRN, many of which have a disproportionate amount of data sets from Global North contexts.[131]

131. Candace Makeda Moore, “The Challenges of Health Inequities and AI”.

During our interviews with startups across various sectors of AI and healthcare in India, including cancer detection and mental health, we discovered that their data sources were diverse, spanning from the US and EU to information gathered by their own systems. There was also little clarity on sources of public data in India – our interviews with policymakers indicated that while health data policies have highlighted the need to share anonymous health data with researchers and businesses, this was yet to be implemented. This corroborated the information from interviews with academics who stated that they relied on publicly available international data for their research.

A common observation from different stakeholders during our interviews was that there was insufficient health data from India to effectively train AI systems. During discussions on data collection, medical professionals highlighted that India was falling behind in digitising health records, with many still preferring handwritten notes and prescriptions. They also mentioned the need to familiarise and train medical students in digital note-taking and data recording. According to one of our interviewees working closely with Ayushman Bharat Digital Mission:

> “We have data, but not enough digitised data available in the market. We should have data at source digitised, and then it should be made available for academia and the tech sector to consume.”

The lack of open-source datasets specific to the Indian population has led to a reliance on large public databases from the National Health Service (NHS) (UK), Center for Disease Control and Prevention CDC (US), and the Organisation for Economic Co-operation and Development (OECD). These datasets are from the Global North and, as a result, have been developed and tested on data drawn from populations in these geographies. Some of our interviewees commented that the lack of Indian datasets implies that many AI models were perhaps developed outside India and used data based on non-Indian populations. In such instances, a primary concern is the introduction of bias at the level of diagnosis, which also affects treatment and further care.

> “In any new project, we check whether it is possible to detect something in public datasets; most of these are from Europe and the US. Indians are different from Caucasians, so when algorithms are trained on datasets from outside, the algorithms are trained to look for abnormalities in a certain way, which may not work for Indian bodies and healthcare.”
>
> **—An academic working on community health and AI.**

As noted by some interviewees, bias may manifest at different stages, either within the dataset or at the level of the system. Unfortunately, the lack of diversity within Indian datasets further complicates this problem. For instance, data collection among marginalised populations is lower, and specific indicators such as caste and gender do not show up adequately within available datasets. Despite ongoing efforts, certain pockets of the population get left out of data collection efforts.

> "Currently, we are introducing a lot of bias since we are relying on external datasets. For example, if I'm a pulmonary patient, 90% saturation is normal for me. So, if you consider 90% to be high and if you prepare a model on that basis, it will be biased. So the marriage of medical and collected knowledge and clinical knowledge has not happened."
>
> **—A clinical doctor studying AI applications in critical care monitoring**

Another important observation according to some of the interviewees was the need for more integration between various data collection sources and systems and for creating a centralised system of health data management. This would facilitate better access to health data and aid in building more robust AI models.

Discussion and audit implications

Data quality and sourcing can impact the performance of AI applications on ground – they may not work well or exhibit bias when applied to specific populations. Datasets for building healthcare AI systems are often sourced from the US, Europe, and China. Consequently, when this data is used to inform healthcare-related decisions in India, it often fails to perform well and sometimes outrightly racially excludes groups from healthcare.[132] [133] Further, health data is often concentrated in specific medical fields such as radiology and imaging, thus leading to more healthcare-centric AI applications in these fields.[134]

The scarcity of good quality Indian data could be one reason for the reliance on international data, which is often readily available in the form of public datasets (for example, models and datasets available on Tensorflow).[135] [136] A study of datasets and research in AI in the clinical setting revealed that the US and China are disproportionately overrepresented, and high-income countries take up the top ten spots of the ranking. In the scoping review, the US accounted for 48% of datasets and authors, while India contributed only 1.6%.[137]

Furthermore, there are multiple reasons why international data, primarily from the US and Europe, are preferred, such as the easy availability of data, the quality checks that go into ensuring the data is of good quality, the standardised format of the data (e.g., having imaging data in Digital Imaging and Communications in Medicine standard).[138]

When the data source is unknown or placed in different contexts or geographies, it might be harder to audit it. As discussed earlier, healthcare institutions rely overwhelmingly on private players for proprietary AI-based solutions. Exposing such systems, specifically their data sourcing process, to an aptly designed AI audit can potentially identify and address such biases. However, conducting AI audits may be difficult without access to the datasets and information regarding the circumstances under which they were collected.[139] Tracing the sources of data is important for understanding and preempting the challenges with the data during an AI audit.[140] An auditor may examine the training for potential over-representation or under-representation of certain populations. While access to only training data may be insufficient for a robust audit, it can help identify data management practices carried out by the organisation in question.[141]

132. *Azra Ismail and Neha Kumar, "AI in Global Health: The View from the Front Lines", In CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, Article 598 (2021): 1-26*

133. *Ziad Obermeyer et al., "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations", Science 366, no. 6464 (2019): 447–53*

134. *Leo Anthony Celi et al., "Sources of Bias in Artificial Intelligence That Perpetuate Healthcare Disparities – A Global Review", PLOS Digital Health 1, no. 3 (2022): e0000022*

135. *"Models & Datasets", TensorFlow, accessed 30 October 2024*

136. *Shweta Mohandas, "AI and Healthcare in India: Looking Forward", The Centre for Internet and Society (2017)*

137. *Leo Anthony Celi et al., "Sources of Bias in Artificial Intelligence That Perpetuate Healthcare Disparities"*

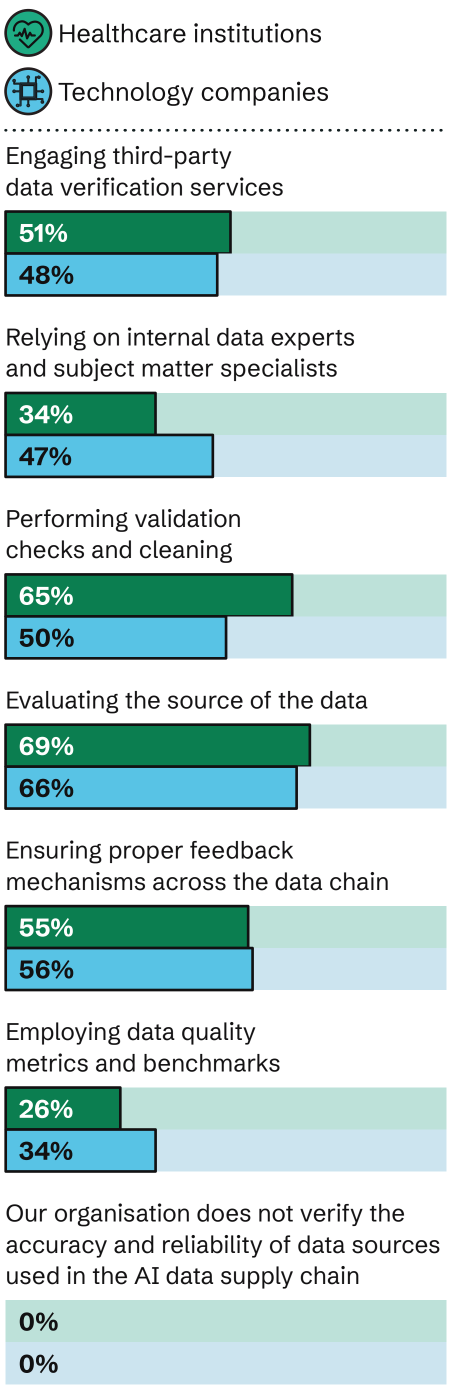
138. *Ibid.*

139. *Nithya Sambasivan et al., "'Everyone Wants to Do the Model Work, Not the Data Work': Data Cascades in High-Stakes AI," CHI '21: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, May 6, 2021*

140. *Emma Galdon Cavell, "Checklist for AI Auditing | European Data Protection Board"*

141. *Sarah H. Cen and Rohan Alur, "From Transparency to Accountability and Back: A Discussion of Access and Evidence in AI Auditing," arXiv (Cornell University), October 7, 2024*

Fig 8: **Percentage of respondents from technology companies and healthcare institutions on how they verify the accuracy and reliability of data sources**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Healthcare institutions (n = 162); technology companies (n = 171)

4.2. **Data processing**

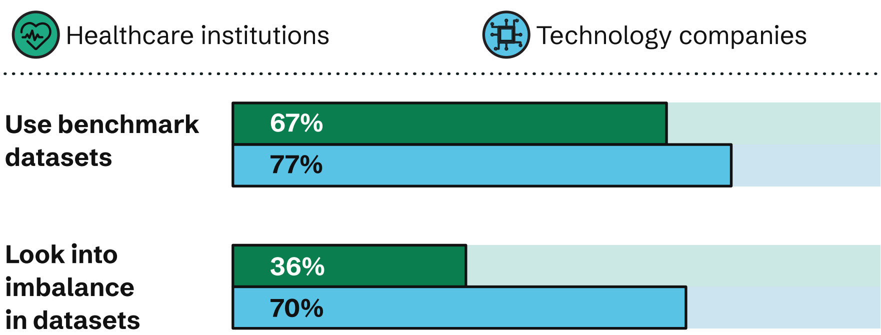
Our interviews revealed challenges in data quality as well as subsequent issues in handling, cleaning, and processing data. Meanwhile, our surveys pointed to certain checks organisations have in place during data processing. (For questions on data processing, we relied on the framework for AI developers proposed by Solanki et al, 2022)

4.2.1. **Data quality checks, while in operation, bring a significant burden**

During our survey, technology companies and healthcare institutions reported conducting several data quality-related checks while collecting medical data. Evaluating data sources remained the most commonly conducted check among technology companies (66%) and healthcare institutions (69%). In contrast, employing data quality metrics and benchmarks was the least conducted check; just 26% of respondents from healthcare institutions and 34% from technology companies selected this check. Comparatively, more technology companies (47%) than healthcare institutions (34%) reported relying on internal data and subject matter experts (see Figure 8).

Our survey showed a vast difference between respondents when it came to looking into imbalance in datasets. Imbalanced class data is a data issue that arises due to the non-uniform distribution of samples among classes – training the model on such imbalanced data results in outcomes that are biased to certain categories.[142] 70% of respondents from technology companies reported checking for imbalances in datasets, while only 36% of respondents from healthcare institutions said so. However, the difference was not very significant between medical institutions (67%) and technology companies (77%) in their use of benchmark datasets (see Figure 9). Benchmark datasets refer to resources published explicitly as datasets that can be used for evaluation, are publicly available or accessible on request, and have clearly defined evaluation methods. These are standard datasets used to evaluate and compare the performance of machine learning models, algorithms, or systems.[143]

Fig 9: **Percentage of respondents from healthcare institutions and technology companies using benchmark datasets or checking for imbalances in datasets**



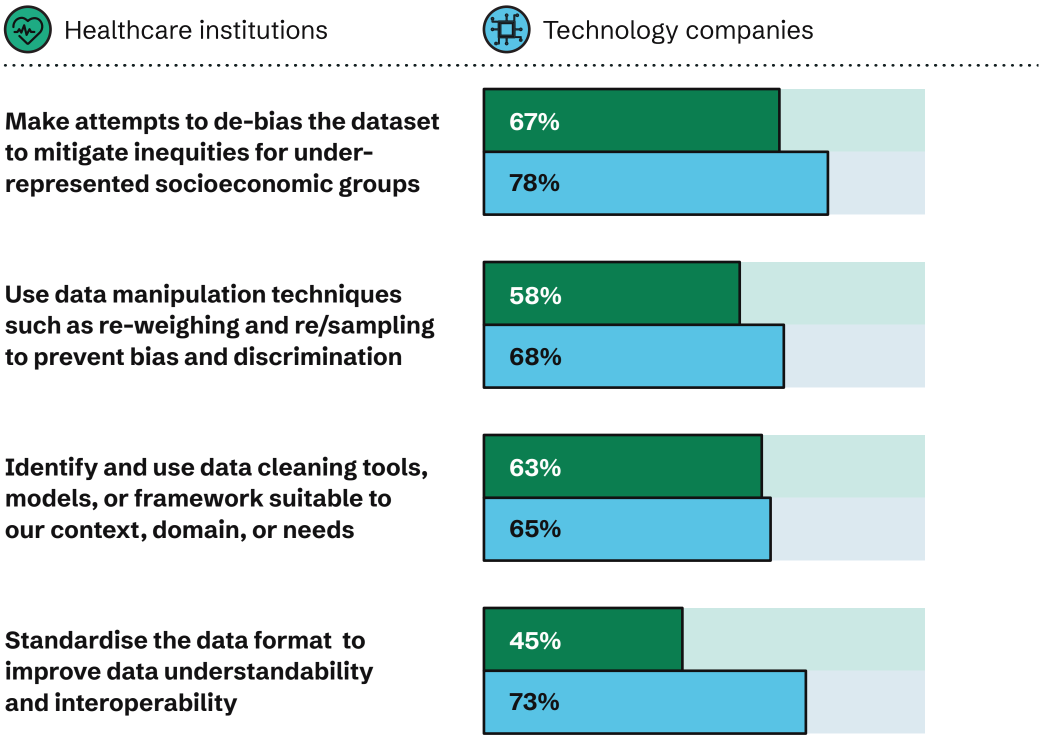
Source: CIS survey of professionals in AI and healthcare, January-April 2024. Healthcare institutions (n = 162); technology companies (n = 171)

142. *Khansa Rasheed et al., "Explainable, Trustworthy, and Ethical Machine Learning for Healthcare: A Survey", Computers in Biology and Medicine 149 (2022)*

143. *Kathrin Blagec et al., "Benchmark Datasets Driving Artificial Intelligence Development Fail to Capture the Needs of Medical Professionals", Journal of Biomedical Informatics 137 (2022): 104274*

As Figure 10 suggests, with regard to cleaning medical data, 3 out of 4 respondents from technology companies mentioned attempts to de-bias datasets; similarly, nearly 7 out of 10 respondents from healthcare institutions reported de-biasing datasets. However, standardising these datasets for interoperability is more prevalent among technology companies (73%) than healthcare institutions (45%).

Fig 10: **Percentage of respondents from healthcare institutions and technology companies, on the different practices they follow while cleaning medical data**



Source: CIS survey of ptofessionals in AI and healthcare, January-April 2024. Healthcare institutions (n = 162); technology companies (n = 171)

Our interviews with medical professionals revealed that data to be used for medical applications require extensive processing and cleaning to be useful. A respondent from one of the technology companies cautioned that proper steps had to be taken before deploying AI models – the respondent emphasised that the companies approaching medical institutions should be transparent about the shortcomings of the AI system regarding bias and not attempt to conceal them. A few startups reported that they ensure data quality by addressing it during the data capture and collection workflows. This included standardising equipment and processes, verifying data at each step through expert-led annotation and review of data, and minimising known low-quality data sources.

> “We have developed certain quality checks – we check if the image is focused properly. We capture multiple images as a part of screening for cancers. These images may have improper labelling, e.g., the right breast is labelled as the left breast, and the labelling could incorrectly say malignant tumour on the right side instead of the left side.”
>
> **—A senior executive at a startup working on AI and cancer diagnosis**

Interviewees also pointed to the labour and time needed to make the data useful for training models. A senior executive at a startup working on AI diagnosis remarked,

> “We need a lot of annotators to double-check the quality of the image. We need a lot of expertise to ensure that the image is of good quality. Data quality is the primary concern for the algorithm.”

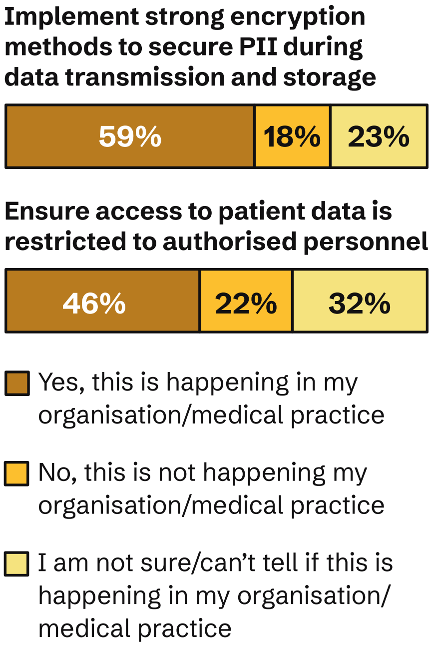
Similarly, an orthopaedic surgeon working with robotic joint replacement confirmed,

> “Data for medical use needs a lot of processing and cleaning before it can be useful.”

A few of the academics we interviewed also emphasised that data quality was crucial for better research outcomes. They observed that data from private hospitals were of better quality than public health data, which, according to them, was “deeply flawed” and was not of the required standard for developing models. Regardless, data quality checks were also undertaken in hospitals, though only one interviewee mentioned the checks undertaken in hospitals:

> “We have a data quality check team, from time to time, bi-weekly/monthly data quality check. Besides computers, it is done manually through trained personnel. Data quality tools are also available to check that the data is complete and the database is well maintained.”
>
> **—A doctor working in a premier public hospital in India**

Fig 11: **Percentage of medical professionals on their practices related to the PII of patients**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Medical professionals (n = 150)

Furthermore, interviewees from startups and academia stressed the importance of having standardised datasets. For instance, due to higher standardisation and digitisation, data quality is much higher in some medical fields, such as diagnostics and imaging.

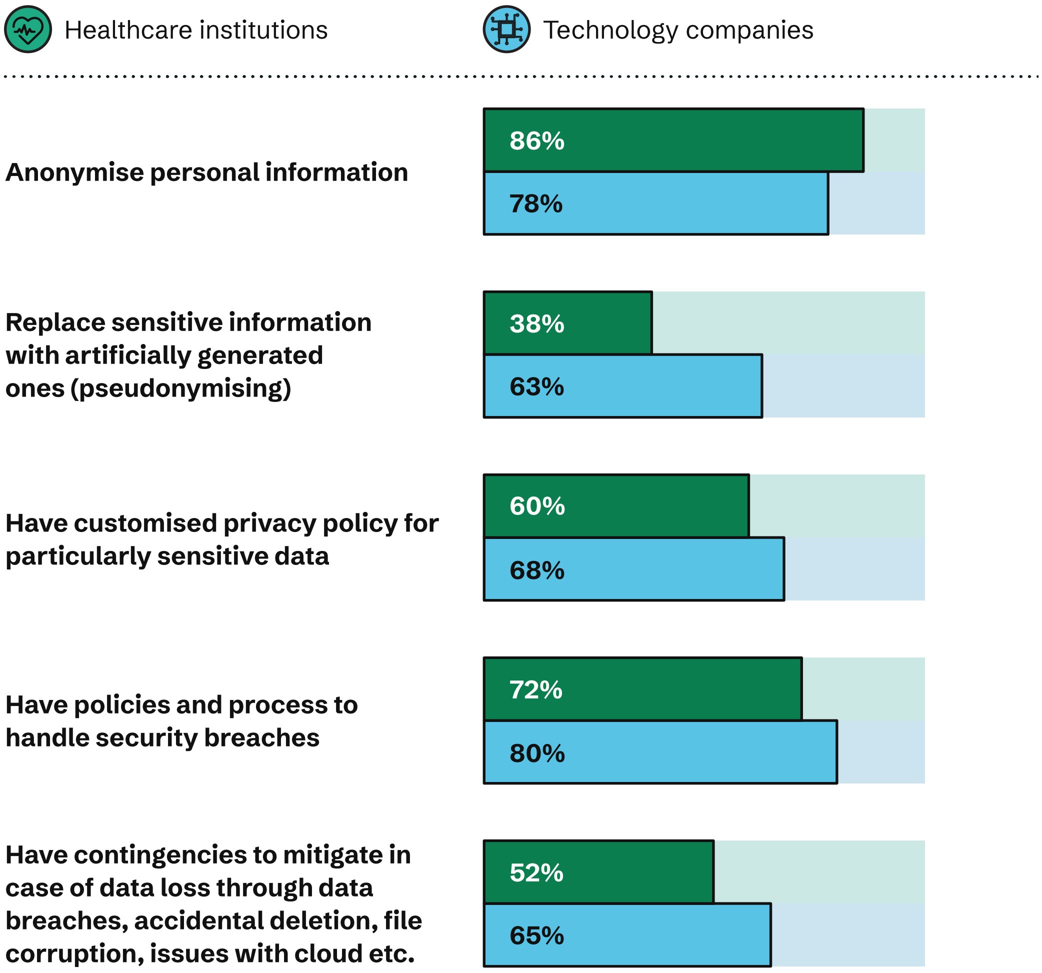
4.2.2. **Anonymisation and removal of personally identifiable information are key priorities for all stakeholders**

To a certain extent, all stakeholders understood the importance of privacy and security of patient data, although there is room for improvement in conducting these practices.

As we see in Figure 11, in our survey, when asked about their practices to protect personally identifiable information (PII) of patients, 59% of medical professionals confirmed that they implement robust encryption methods to secure PII during data transmission and storage, and 46% said they ensure access to patient data is restricted to authorised personnel. Despite this, 32% of medical professionals expressed uncertainty about the robustness of restrictions on patient data access, indicating concerns regarding control and security of patient data.

Meanwhile, when it comes to data handling, anonymisation of personal information (78% of respondents from technology companies and 86% respondents from healthcare institutions) and establishing policies to handle security breaches (80% of respondents from technology companies and 72% of respondents from healthcare institutions) were the most common practices. However, pseudonymisation was prevalent more among technology companies (63%) than in healthcare institutions (38%) (see Figure 12).

Fig 12: **Percentage of respondents from healthcare institutions and technology companies, on the different practices they follow during data handling**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Healthcare institutions (n = 162); technology companies (n = 171)

Consistent with the survey findings, in our interviews with startups, we found that one of the common data processing steps was removing PII from the dataset. While this was considered a data security measure, one interviewee clarified that the data was also anonymised to remove biases. Additionally, startups mentioned that they restricted access to raw data with PII by implementing data access controls for staff and enabling multi-factor authentication.

> “For all data that we collect, consent is mandatory, and it is explicitly mentioned that the data will be used for research purposes. We store the collected data in cloud storage, and the identity of the patient is masked before being used for algorithm development.”
>
> **—A doctor working in diagnostics and patient care using wearable technologies.**

> “We store some metadata, say gender, or the name of the manufacturer, etc. So, we prepare a database with the image path and metadata, and we get only pixel data and store that… Access restrictions are in place, including two-factor authentication; only R&D staff can access this data.”
>
> **—A senior executive at a startup working on AI diagnosis**

Discussion and audit implications

Even as of January 2022, a staggering 60–70% of mid-sized and large hospitals in India lacked adequate systems for collecting and storing digital health records or maintaining access controls.[144] Similarly, sources of data released by the government are often incomplete due to various reasons; for example, in the cancer registry, a significant number of patients are not included in the official statistics until their death;[145] in the Birth and Death Registry, data on the cause of death continues to remain unreliable due to poor compliance with the International Classification of Diseases guidelines.[146] All these issues impact data quality and necessitate stringent data-cleaning processes that are often burdensome for many stakeholders. While the ABDM has been heralded as a way to digitise and collect health data, the 2023 report on the data errors in the Ayushman Bharat - Pradhan Mantri Jan Arogya Yojana (PM-JAY),[147] further shed light on the issues in data quality and necessitate stringent data cleaning requirements, which are often burdensome for various stakeholders involved, as our findings revealed.

Similarly, although our findings point to stakeholders taking data security and privacy concerns seriously – such as by taking steps to anonymise health data – much of this action is limited to fulfilling compliance and liability requirements. This is all the more important in light of the many health data breaches that have surfaced in India over the past three years – affecting some of India’s premier healthcare institutions,[148] government data platforms,[149] and even private insurance companies.[150]

144. *Abhijit Ahaskar and Moumita Deb Choudhury, “Indian Hospitals, Clinics, Labs Selling Data without Consent”, Mint, 18 January 2022*

145. *Arvind Pandey et al., “Health Information System in India: Issues of Data Availability and Quality”, Demography India 39, no. 1 (2010): 111–128.*

146. *Ibid.*

147. *Garima Sadhwani, “‘Dummy Numbers & Mismanagement’: 5 Takeaways from CAG Report on Ayushman Bharat,” TheQuint, August 16, 2023,*

148. *ET HealthWorld and www.ETHealthworld.com, “From AIIMS Delhi to ICMR, Data Breaches Haunt Crores of Indians,” ETHealthworld.Com, November 13, 2023,*

149. *John Xavier, “Explained | What Does the Alleged CoWIN Data Leak Reveal?,” The Hindu, June 18, 2023*

150. *Toi Tech Desk, “Star Health Hacked: Name, Address, Phone Numbers, Medical Reports and Other Data of 31 Million Customers Available for Free on Telegram,” The Times of India, September 23, 2024*

Lastly, while the stakeholders did not necessarily use the word audit in the interviews, a few data quality check techniques mentioned during the interviews and surveys may be considered to serve the same purpose as AI audits. For example, using benchmark datasets can help stakeholders understand the performance, consistency, and limitations of their AI models.[151] However, AI benchmarks of direct clinical relevance are scarce and fail to cover most activities that clinicians want to see addressed globally.[152] Therefore, it is unclear what kind of benchmark datasets were used in the Indian context, given the limited availability of such open benchmark datasets. Yet, not all practices are equally widespread across stakeholders or are comprehensive on their own. For instance, de-identifying data can help address data leaks and biases; but studies show that de-identification techniques can themselves be biased.[153]

4.3. **Data learning and model development**

Developing AI systems requires not just sourcing and processing training datasets but also their subsequent use to train the underlying model. The technical details of this training process vary significantly based on a range of factors, including but not limited to the intended use case of the AI system. In this section, however, we examine the model development process in the multi-stakeholder context of India's healthcare system and, more specifically, the importance of collaboration and feedback between medical professionals and technology companies.

4.3.1. **Collaboration between medical professionals and AI developers remains limited**

Model development is considered more challenging and rewarding than data collection and processing, since a contextual understanding of datasets is fundamental for developers to train models effectively.[154] However, the lack of multi-stakeholder collaboration remains a dominant concern when it comes to knowledge and data sharing between various stakeholders during this stage. For instance, only 26% of respondents from the surveyed technology companies collaborate with doctors, and only 31% collaborate with medical researchers for AI-related discussions. Comparatively, 78% collaborate with other technology companies, and 66% collaborate with AI researchers and academics.

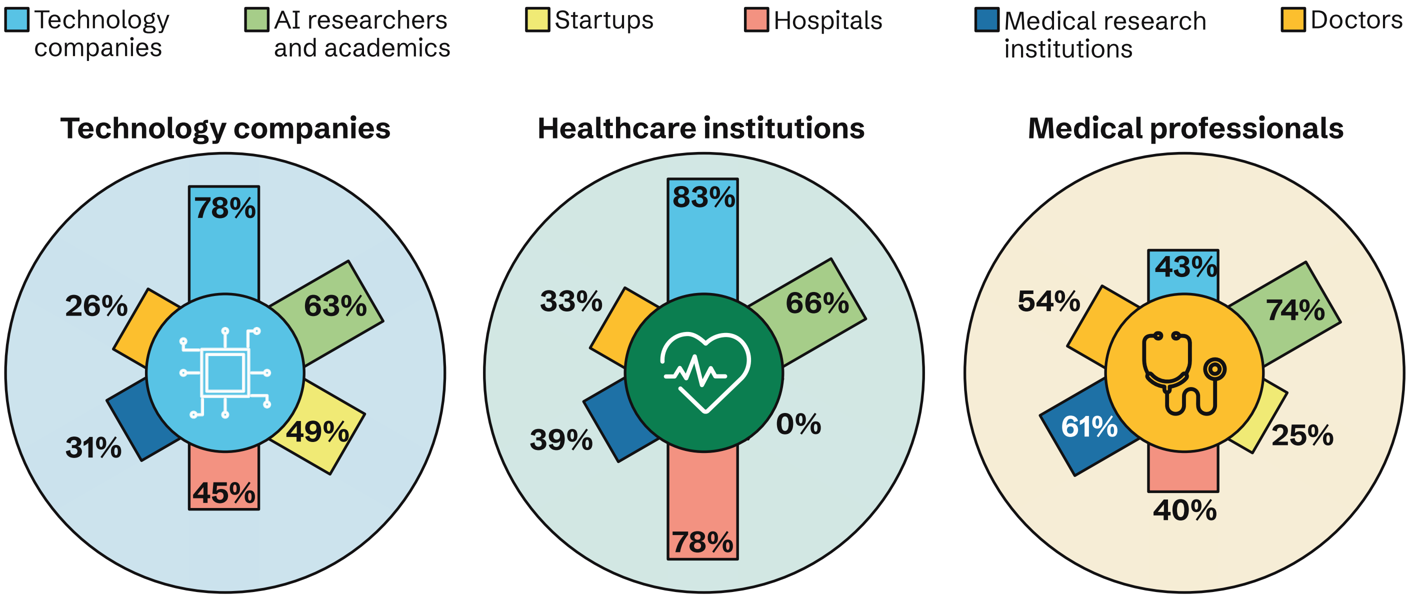
151. *Nikos Sourlos et al., "Recommendations for the Creation of Benchmark Datasets for Reproducible Artificial Intelligence in Radiology", Insights Into Imaging 15, no. 1 (2024)*

152. *Kathrin Blagec et al., "Benchmark Datasets Driving Artificial Intelligence Development Fail to Capture the Needs of Medical Professionals"*

153. *Yuxin Xiao et al., "In the Name of Fairness: Assessing the Bias in Clinical Record De-Identification", arXiv.org (2023)*

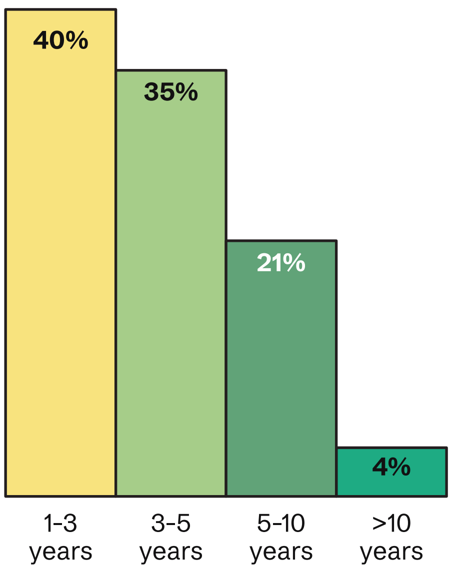
154. *Divy Thakkar et al., "When Is Machine Learning Data Good? Valuing in Public Health Datafication", In CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, Article 322 (2022): 1-16*

Fig 13: **Percentage of respondents from technology companies, healthcare institutions, and medical professionals who partner with other stakeholders in developing and deploying AI systems**



Source: CIS survey of professionals in AI and healthcare, January- April 2024. Medical professionals (n = 150); healthcare institutions (n = 175); technology companies (n = 175)

Fig 14: **Percentage of respondents from technology companies who work in AI development and testing with prior work experience in AI and healthcare**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Technology companies (n = 175)

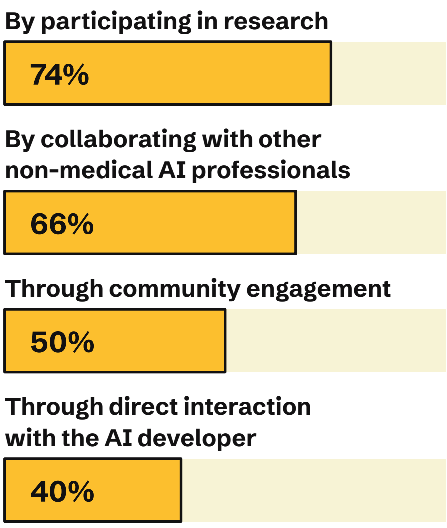
As Figure 13 suggests, this trend is also apparent in the collaboration patterns of medical professionals. While 74% of them collaborate with AI academics and 61% work with medical research institutions, only 43% have partnered with technology companies, and just 25% have collaborated with startups for discussions and debates related to AI systems.

The gap in domain-specific knowledge is made even wider by the generally low level of healthcare-related expertise in technology companies. About 41% of technology company respondents we surveyed reported AI development and testing as one of their areas of work. Yet, as Figure 14 indicates, only 25% had worked at the intersection of AI and healthcare for more than five years, given the relative novelty of AI in India's healthcare industry.

A civil society interviewee noted that many AI systems are being developed by companies lacking healthcare experience and without input from medical professionals, highlighting the lack of collaboration between stakeholders.

> "...healthcare is an ongoing process that cannot be reduced to stats and numbers. Organisations with no experience in healthcare are entering the space through 'AI', these organisations risk packaging pseudoscience/discredited science into the AI model...most often they don't have doctors or medical expertise on their teams."
>
> **—A researcher who works on AI audits and human rights**

Fig 15: **Percentage of medical professionals who provide feedback on AI systems, on their preferred methods of sharing feedback**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Medical professionals (n = 82)

4.3.2. **Feedback sharing between the technology developers and their users is often irregular and indirect**

A similar disconnect between different stakeholders is also evident in the inadequate sharing of feedback between medical professionals (one of the models' purported beneficiaries) and AI developers. To begin, only 55% of medical professionals we surveyed provide regular feedback on the AI systems they use. Among those who do, as seen in figure 15, only 40% interact directly with the developer of the AI model in question, whereas 74% share their feedback indirectly by "participating in research."

This weak feedback loop potentially isolates stakeholders' priorities of model development. For example, while 46% of medical professionals and 49% of respondents from healthcare institutions believe transparency and explainability are important ethical considerations in using AI systems, only 29% of respondents from technology companies did so.

A practising clinician who worked with AI companies stressed the need for feedback from medical professionals while developing products, saying:

> "I have used all these systems from [name redacted], and the documentation process is absolutely outdated. There is no way to take an image of the blood report, and it gets automatically uploaded."

Similarly, emphasising the need for feedback and how feedback from doctors helped them retrain the algorithm, a senior executive at a startup working on AI and cancer diagnosis said,

> "We get feedback from docs on false positives/negatives. They share the datasets, and we retrain the algorithms."

Discussion and audit implications

Developing AI systems to address healthcare needs calls for expertise in the underlying technologies and a rigorous understanding of basic medical principles.

While technology companies in charge of this process frequently allay the medical fraternity's fears over AI-based decision-making and risks to their jobs, their inclusion in the development process remains limited.[155] [156] Our findings indicate that this is evident in how medical professionals share feedback with AI developers, i.e., primarily through published research and much less through direct interactions. Given the overall low level of sectoral expertise among our respondents from technology companies, this lack of collaboration further hinders the successful integration of AI systems in clinical workflows.

At present, however, the popular discourse on model development primarily focuses on the problem of data training infrastructure in India. A range of recent state-led initiatives – such as the 2020 draft National Data Centre Policy and the newly launched IndiaAI Mission – aim to build AI infrastructure in the country.[157] [158] Although the creation of such infrastructure can, theoretically, incentivise the private industry to innovate, it is imperative that the returns from this level of public investment benefit all. This is especially important since creating large-scale data centres requires significant amounts of land, energy, and water – all of which can lead to adverse climate outcomes.[159] [160]

The increased interest in the uses of AI systems across the healthcare sector has brought to light several challenges in their design and deployment. Some of these challenges, such as the absence of a secure and private data pipeline, not only prevent AI from scaling effectively but can also lead to skewed training datasets, thereby compromising the algorithm's decisional accuracy.[161] The current AI for the healthcare market has, at the same time, also relied on increased purchase and implementation costs, making them unaffordable to most people.[162] Although there is a push to create health data for Indian people and healthcare institutions, the high costs of buying and integrating the resulting AI system can limit their adoption in under-funded settings, thereby risking a form of one-way value extraction through data collection.[163]

More urgently, in a safety-critical domain such as public health, algorithmic bias and data security risks further skewed outcomes while deploying AI systems. Overcoming these risks requires developing AI models that are inclusive for all the stakeholders in the data supply chain. Given the decentralised nature of existing data supply chains, cross-sectoral collaboration is both a necessity for AI audits and a potential outcome for auditing processes to aspire for. For instance, modular IT audits allow stakeholders with their requisite expertise and exposure to audit various aspects of an AI system, thereby sharing the cost of conducting an end-to-end audit.[164] This approach can be beneficial when procuring or using these technologies (due to unaffordability) is already challenging for healthcare providers.

155. *Medicine Man: How AI Is Bringing Humanity Back into Healthcare", Microsoft News Centre Europe, 10 June 2019*

156. *"Will AI Eventually Replace Doctors?", Kellogg Insight, Northwestern University, 31 January 2023*

157. *"Data Centre Policy 2020 (Draft for Discussion)", Ministry of Electronics and Information Technology, Government of India (2020)*

158. *"Cabinet Approves Ambitious IndiaAI Mission to Strengthen the AI Innovation Ecosystem", PIB, 7 March 2024*

159. *Lynn H Kaack et al., "Aligning Artificial Intelligence with Climate Change Mitigation", HAL (2021): hal-03368037*

160. *Jude Coleman, "AI's Climate Impact Goes beyond Its Emissions", Scientific American, 20 February 2024*

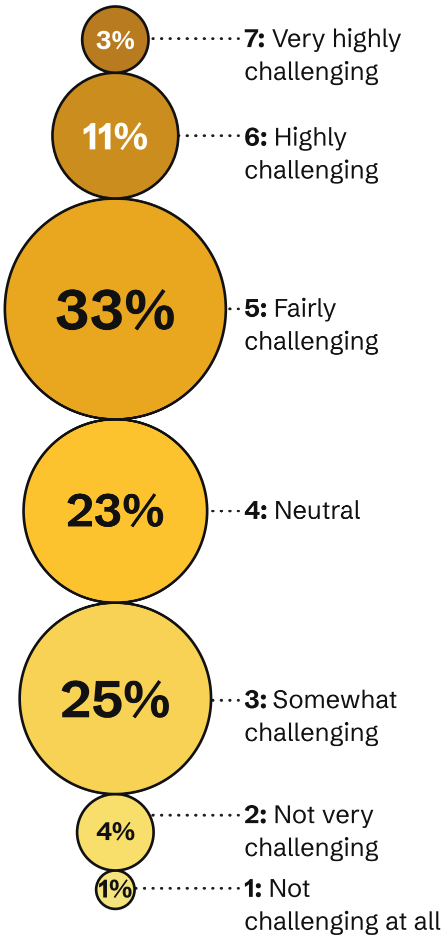
161. *Shivaram Kalyanakrishnan et al., "Opportunities and Challenges for Artificial Intelligence in India", In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society (2018): 164–170*

162. *Ibid.*

163. *Madhumita Murgia, Code Dependent: Living in the Shadow of AI (Picador, 2024), 126–127*

164. *Syed Rizvi et al., "A Modular Framework for Auditing IoT Devices and Networks", Computers & Security 132 (2023): 103327*

Fig 16: **Percentage of medical professionals, on the degree of challenges they face in integrating AI systems into clinical workflows**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Medical professionals (n = 150)

4.4. **Model deployment**

After its development and validation (as per requisite standards, if any), the AI system is usually deployed in a clinical setting, often as a pilot. Our survey findings indicate that 47% of medical professionals find integrating AI systems into their clinical workflows challenging, albeit to varying degrees (Figure 16). Separately, this finding was also echoed by our respondents from healthcare institutions, 44% of whom listed the integration of AI systems in clinical workflows as a challenge.

Our interview with a respondent from a technology company also shed light on the challenges they face when selling AI systems to hospitals. They stated that hospitals were still unsure how to account for the expense of procuring AI and identify the person responsible for its purchase, implementation, and maintenance – for instance, whether it would be the chief administrative officer or the chief technology officer. They also stated that the hospital had to do top-down pricing and determine how to recover the cost of these expensive AI systems from patients. According to a senior executive working in product management in the oncology department of an international health technology company:

> "For well-known products such as a CT or ultrasound machine, it is very clear for hospitals where they should go to buy it; there are multiple vendors and trade shows, and the hospitals can plan their budget cycle to accommodate the expense of the machines. With AI vendors [AI developers], there is no single place to capture their customers [hospitals and medical professionals] and sell to them. In addition, the ability to integrate these AI systems into existing hospital ecosystems, without demanding additional IT infrastructure and costs, is something (small) AI vendors struggle with."

While highlighting the infrastructure issues and scarcity of medical professionals in the context of public hospitals, a clinical doctor studying AI application in critical care monitoring also stated,

> "As of now, not a single AI tool is applied in our current setup. Only digitisation is happening. AI application is only happening in a private capacity for their ease of practice."

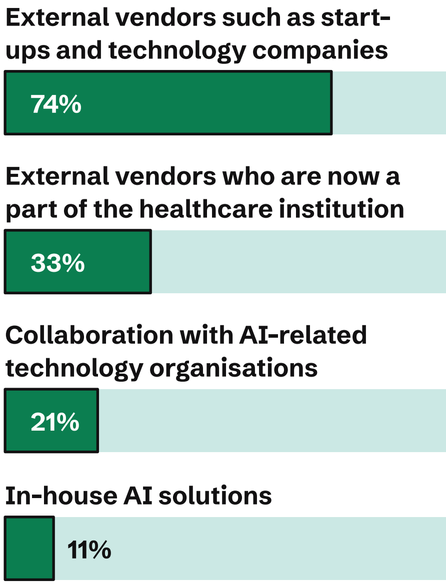
In the following subsections, we further discuss this challenge of AI integration and a few of its contributing factors, particularly the overreliance of medical professionals on external vendors for AI systems, hesitation and a lack of trust in adopting these systems, and limited training and education in the use of AI in healthcare.

4.4.1. **Reliance on external vendors**

There are significant gaps in resources and funding to facilitate AI deployment. Our surveys show that healthcare providers lack the resources and infrastructure to introduce AI systems into healthcare institutions. For instance, 38% of medical professionals and 27% of healthcare institutions face significant challenges in terms of inadequate resources and infrastructure for deploying AI.

In-depth interviews show that while AI technologies can enhance efficiency in healthcare systems, implementing them requires funding and resources.

Fig 17: **Percentage of respondents from healthcare institutions, on how they source AI systems**



Source: CIS survey of professionals in AI and healthcare, January-April 2024. Healthcare institutions (n = 175)

> “In the long run, investing in AI technologies can reduce time and manpower, giving AI better credibility for the Indian healthcare system. The funding/ investment needs to come from the Indian Medical Council or the government.”
>
> **—A practising clinician who has worked with AI companies**

As Figure 17 suggests, healthcare providers who source AI systems primarily depend on AI systems from technology companies and start-ups, with only 11% reporting using in-house systems. This can create dependencies on external vendors in addition to a lack of healthcare expertise among technology companies and difficulties in integrating AI into clinical workflows.

4.4.2. **Medical professionals hesitate to adopt AI**

Our interviews indicate a need to build trust among healthcare providers in using AI systems. In our interviews with startups, they stated that medical professionals were still sceptical about integrating AI into their workflows. It was only after they were convinced of the clinical results and the performance of the AI system that they were ready to use AI applications. Another way to build trust in AI for medical professionals, such as doctors, is by having explainable results from AI.

> “...we have also seen that there needs to be more trust-building in AI systems. For instance, you must have heard that for radiologists, it is becoming a threat. But, in fact, many radiologists today use their own judgement, so there is a bit of lack of trust in these technologies.”
>
> **—A researcher examining health-related law and policy issues**

In the same vein, a few policymakers also confirmed that medical professionals’ trust in AI systems improved when the system demonstrated that their accuracy was as good as that of doctors and when the medical professionals understood how the decisions were made.

Our interviews with medical professionals revealed that various factors contributed to the hesitation displayed by the medical fraternity, including the additional workload associated with learning AI systems and the fear of AI systems replacing doctors.

> “Most of the healthcare providers are not confident in the technology, and they may have a fear of being replaced. Most doctors believe that it has to be completely evidence-based, and many are sceptical to use it with confidence. For the time being, the acceptance is very, very low because of this lack of confidence in the tools.”
>
> **—An orthopaedic surgeon working with robotic joint replacement**

A doctor working in diagnostics and patient care using wearable technologies noted:

> “Most of us are taught to practice traditional medicine, and it appears like the machines have come in between, so there is reluctance on both doctors and patients.”

Fig 18: **Types of AI-focused training across different stakeholders, either provided by organisations or undertaken by respondents individually**

Nature of trainings provided by healthcare institutions

Healthcare institutions

1. Training on data, privacy, and security (65%)
2. Training to increase general awareness of AI tools (51%)
3. Training on AI ethics (45%)

Nature of training undertaken by respondents individually

Technology companies

1. Participating in online AI communities (63%)
2. Speaking with professionals from other fields (58%)
3. Attending conferences and workshops (58%)

Medical professionals

1. Attending conferences and workshops (65%)
2. Reading research papers and publications (59%)
3. Participating in online AI communities (47%)

Source: CIS survey of professionals in AI and healthcare, January- April 2024. Medical professionals (n = 150); healthcare institutions (n = 175); technology companies (n = 175)

Academics and medical professionals also flagged the issue of liability. For instance, who would be liable for an error in diagnosis made by an AI application that aids medical professionals? A common concern we heard from doctors, startups, and academics was that AI was meant to assist doctors and should not take their place.

4.4.3. **Gaps in training and education among professionals on the use of AI**

Healthcare providers need to be trained and educated on using AI systems for the seamless integration of AI into clinical workflows. Similarly, it is also vital for technology companies to have healthcare expertise when developing AI systems for healthcare. Furthermore, educating all stakeholders about different facets of AI and healthcare enhances collaboration and communication during implementation. Our surveys reveal key challenges in training and professional development of different stakeholders. Notably, 57% of medical professionals cite the lack of AI related training and education as an important challenge. Meanwhile, 59% of respondents from technology companies struggle to effectively communicate AI-decision-making processes to healthcare professionals and patients.

Training and professional development in AI within healthcare are currently fragmented, with varying levels of engagement across different stakeholders. As seen in Figure 18, healthcare institutions provide training to their staff with 65% offering courses on data, privacy, and security, and 51% focusing on raising awareness of AI tools. Medical professionals actively engage with AI research, with 65% staying updated through conferences and 59% by reading research publications. Among technology companies, 63% follow AI advancements via online communities, 58% interact with professionals from other disciplines, and 58% attend AI-related conferences and workshops.

In our interviews, medical professionals either using AI or working on developing an AI system for healthcare stated that healthcare institutions had the added responsibility of training doctors to integrate digitisation and AI into their workflows. An academic trained in medical ethics and bioethics shared that sometimes medical professionals had to educate themselves through self-learning and peer-to-peer learning:

> “...healthcare providers are looking to each other and articles, papers, and speakers [experts] to enlighten them on what this new world of AI in healthcare will look like. There is an exploratory mode of educating people since there is no streamlined approach for the same.”

Interviewees pointed out that most medical colleges in India still train doctors to take handwritten notes, creating a learning curve for medical graduates who must adapt to a healthcare system that requires digitised observations. They also highlighted that healthcare institutions did not have updated curriculums that included new technologies such as AI. One interviewee also stated that there were not enough computers in some medical colleges to train doctors in using AI tools. Due to insufficient training from institutions, medical professionals have to acquire new skills independently, learn alongside their courses, or pick them up on the job.

As a result of the gap between the need for training and its availability, medical professionals invest their personal resources in understanding how to use AI in healthcare. An academic trained in medical ethics and bioethics admitted,

> “We are in this stage of learning and understanding, and I don’t think many AI systems have been deployed.”

Similarly, a researcher who works on AI audits and human rights shared a concern:

> “Medical professionals can end up hating the [AI] systems because they don’t understand how the system works. More nuanced training on how these systems work with their limitations is needed.”

According to an academic trained in medical ethics and bioethics,

> “We have a national medical commission, and ideally, from time to time, there should be an effort in educating healthcare professionals, especially on the latest understandings of diagnostics, technology, etc, but there is not a concrete system for this... Healthcare professionals rely on online resources and connect with specialist societies, which might have documents on this as it is a new topic for the industry. And they may speculate on how it will change things.”

Discussion and audit implications

While the AI models might work well in a lab setting during their development, their real test is when they are deployed and used by the end users. The integration of AI as a “product”, especially in a medical setting, requires regulatory approvals, large amounts of data for it to work in a real-world setting, staff trained in working with AI systems, integration of the AI system into new and existing workflows, and training or reskilling of medical staff.[165] In India, there is a need to assess existing health infrastructure to determine if it is ready to integrate AI in public and private medical institutions. Ad hoc AI deployment also means that there is a lack of a common understanding of liability, especially who would be liable for errors made by AI systems; the doctors using it and the developers making it are equally concerned about who would be responsible for potential harms caused by AI.[166] There is also the added uncertainty of regulating AI as a product and whether it would come under the Medical Devices Rules. This uncertainty is more prevalent in AI-enabled devices such as smartwatches.[167] This also ties into the earlier point about the reliance on off the shelf AI systems, versus systems made inhouse that are unique to the hospital or medical practice.

Auditing could be used to check how these systems are deployed and integrated into the workflow of medical institutions or professionals. They could be undertaken by the developers building the AI system and the institution using the AI system. Audits can play a crucial role in assessing the accuracy of these systems, identifying biases and errors, and fostering greater confidence and trust among medical professionals.[168] For medical institutions, audits could also serve the purpose of checking the number of professionals trained in using AI systems and how many among them use these systems.[169] Feedback from these users could be used to assess the need to integrate more AI into the workflow.

165. *Jianxing He et al., “The Practical Implementation of Artificial Intelligence Technologies in Medicine”, Nature Medicine 25, no. 1 (2018): 30–36*

166. *Shweta Mohandas, “AI and Healthcare in India: Looking Forward”*

167. *“Fitness Wearables and the Law in India – Part 3”, Ronin Legal, 3 October 2024*

168. *Thomas P Quinn et al., “Trust and Medical AI: The Challenges We Face and the Expertise Needed to Overcome Them”, Journal of the American Medical Informatics Association 28, no. 4 (2020): 890–94*

169. *Chih-Hsien Hsu and Marcia Y. Sakai, “Auditing Program Evaluation Audits: Executive Training Exercise for Assessing Management Thinking, Planning, and Actions”, Journal of Business Research 62, no. 7 (2008): 680–89*

5. **Main findings and discussion:** AI auditing as a response

The deployment of AI in India's healthcare system is relatively new, and as a result, there are limited documented precedents for AI audits; most of them are largely voluntary and left to the discretion of the technology companies and healthcare institutions deploying these systems. For instance, only 7% of respondents from technology companies surveyed in our research reported "regular security audits and vulnerability assessments" for their AI systems.

However, there have been sporadic and fragmented evaluations and reviews (that were highlighted in the data processing section), which we present further evidence of in this chapter. We also explore how audits and aligned methods are commonly understood and examine the practices and processes prioritised by different stakeholders.

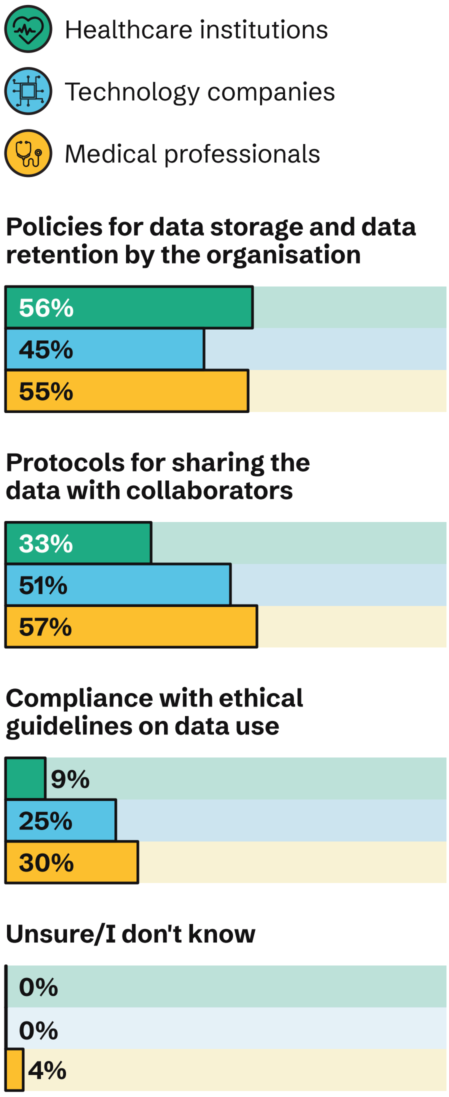
5.1. **Current state of AI audits**

5.1.1. **Understanding and applications of AI audits remain scattered and incoherent among stakeholders**

Our interviews reveal the diversity of AI audits among the wide range of stakeholders in the data supply chain. Often, it is used interchangeably with internal or external review processes, regulatory and compliance-related mechanisms, and checks for the quality, accuracy, security, and privacy of AI systems. Several interviewees expressed that it is still too early for auditing systems to be implemented, considering the difficulties associated with health data collection, digitisation, and implementation of AI models. However, it is possible these would be set up with the increased use of these systems.

A few medical professionals and those working in technology companies and startups also felt that current audit practices were not completely standardised.

Fig 19: **Percentage of respondents who prioritise policies, protocols, and compliance for reviews, evaluations, and audits**



Source: CIS survey of professionals in AI and healthcare, January- April 2024. Medical professionals (n = 133); healthcare institutions (n = 162); technology companies (n = 171)

> "To me, I wouldn't use the word 'audit', but 'validate'. So, what is the process that someone uses to validate an AI model before it is put into practice with real patients and real healthcare professionals? Who signed off on the model and the process they followed are questions that would really require some form of standardisation to build long-term trust."
>
> **—A person working closely with the central government and some state governments on digital health**

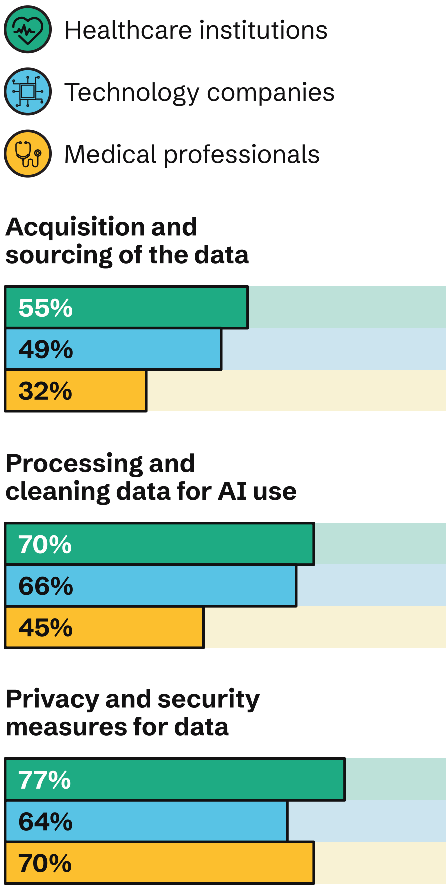
Furthermore, our interviews indicated a lack of consensus regarding whether an audit should serve as an internal or external accountability mechanism. For instance, among the technology companies we interviewed, a few use a range of internal audit/review processes, such as establishing data quality checks and evaluating internal policies and protocols for data storage and sharing (Figure 19). Meanwhile, a few startups and AI researchers we interviewed stated the importance of external audits, especially in the context of external compliance, such as by the International Organisation for Standardisation (ISO), the National Health Authority (NHA), and the Central Drugs Standard Control Organisation (CDSCO). Some companies that provide global services also mentioned undertaking audits to maintain compliance with the regulatory requirements of the United States Food and Drug Administration and the EU's General Data Protection Regulation.

> "The auditors who come actually have no idea about the technology that we are building. Regulatory authorities like the NHA and CDSCO oversee your AI system's compliance with regulatory requirements. They do regulatory audits and access to the AI systems, their conformity to safety, and they mainly check on the efficacy and quality standards."
>
> **—A senior executive heading the health and AI team at a big Indian technology company**

> "There should also be regular audits; for example, if it is a diagnostic machine and it is coming up with a diagnosis, then this judgement of the machine has to be audited. This is also true for treatments that should be audited against outcomes and levels of patient satisfaction. It depends on what the patients are experiencing. The gold standard for now is the doctor, so against that, you audit what the machine is doing. That is how you see whether it is doing what it is supposed to do."
>
> **—An academic trained in medical ethics and bioethics**

170. Although reviews, evaluations, and audits differ from each other in purpose and scope, we chose to use them together in our surveys mainly because these words are often used interchangeably.

Fig 20: **Percentage of respondents that prioritise different aspects of the AI data supply chain for reviews, evaluations, and audits**



Source: CIS survey of professionals in AI and healthcare, January- April 2024. Medical professionals (n = 133); healthcare institutions (n = 162); technology companies (n = 171)

5.1.2. AI audits prioritise mainstream and intersectional concepts such as data privacy and security in their scope

As Figure 20 highlights, the main focus of current AI audits and review processes is on privacy and security, although other aspects also hold some value for stakeholders.

According to a senior executive working for an AI startup in mental health,

> "We don't collect any data without privacy notices. We immediately redact the PII, anonymise, and randomise the data. We don't store any transcript-level data; for instance, we cannot read the entire chat."

> "Not everyone on the team has access to the names and other details of patients. Further, we anonymise and delink patient details from the images. If the patient asks for deletion of data from servers, we do that as well...The regulatory team oversees these aspects; earlier, we also had a privacy officer looking into these issues."
>
> **—A senior executive at a startup working on AI and cancer diagnosis**

Discussion and audit implications

The limited and scattered adoption of audits as an assessment framework can largely be attributed to the absence of a legal mandate. Notwithstanding, many government bodies, such as the Supreme Audit Institution, have proposed various auditing checklists for the responsible use of AI systems.[171] [172] [173] In fact, even the ICMR-prescribed guidelines (discussed earlier in this report) encourage audits at multiple stages to minimise risk, facilitate accountability, and reduce bias.[174]

The absence of an overarching and agreed-upon auditing framework implies that even if technology companies or healthcare institutions were to conduct them, the content of these audits would vary substantially. Similarly, depending on the priorities of the audit-authorising body, there is no guarantee that the findings of these audits will be made public or even acted upon. Recent studies on the impact of AI audits have raised critical concerns about power asymmetries within the audit process, particularly in scenarios lacking third-party oversight.[175]

---

171. *Shashidharan, K.P. "Using AI for Audit Techniques", The Hindu, 4 October 2023*

172. *Husanjot Chahal and Samanvya Hooda, "Auditing AI: What Is It and Why Does It Matter for India?", ORF, 30 January 2024*

173. *KR Srivats, "ICAI, MeitY Will Join Hands to Develop 'AI Audit Tool' to Oversee India Inc", BusinessLine, 8 July 2024*

174. *Ethical Guidelines for Application of Artificial Intelligence in Biomedical Research and Healthcare, 2023. ICMR.*

175. *Birhane Abeba et.al. "AI Auditing: The Broken Bus on the Road to AI Accountability".*

5.2. **Audit as a governance tool for AI systems**

As we have seen, many organisations engage in various practices related to AI auditing despite not performing a comprehensive audit or labelling it as such. In addition to being a standalone finding of this study, it should also be noted that the same inconsistency and heterogeneity in audit practices made it difficult for us to gauge the actual effectiveness of an AI audit. Nevertheless, auditing – like other governance tools – is constrained by particular challenges, and the lack of evidence does not necessarily indicate evidence is absent.

We use this section to present and discuss the potential and challenges associated with AI audits. Given the lack of information from our primary sources, we predominantly use existing evidence to undertake this analysis.

5.2.1. **Theoretically, AI audits can increase accountability, aid in standardisation, and improve algorithmic performance**

An audit's outcomes can vary substantially depending on the choices made during the auditing process. For instance, an AI audit that uses indicators and parameters in line with ethical guidelines is likely to find more credibility than its counterparts. Nevertheless, there are some overarching benefits that AI audits offer for medical professionals, technology companies, healthcare institutions, and policymakers, among others:

- **It enables transparency and accountability.** In the present scenario, where AI systems are procured via external vendors, who, in turn, develop them using data sources and foundational models from the Global North, an AI audit can act as a strong lever for promoting accountability. A societal audit of such systems – one that accounts not just for regulatory obligations or technical mandates but also for social expectations, such as environmental sustainability and ethical labour practices – can improve transparency, inform policymaking, and build people's overall confidence in these technologies.[176]
- **It improves current healthcare-based AI systems.** A design-focused audit, on the other hand, can reveal crucial technical gaps in the design of an AI system. For instance, an audit focusing on algorithmic bias can be integral in identifying the source of the bias and simplifying the process of upgradation. Similarly, by indicating critical junctures in the functioning of the system, the audit can also enable human-in-the-loop and society-in-the-loop interventions, which can take the form of external oversight on some elements of the DSC.[177]
- **It provides a common comparative benchmark.** Although sector-level policies, such as the ICMR guidelines, provide a holistic view of the underlying principles, their interpretation in technical applications remains varied. In this context, audits using these guidelines as their core framework can help normalise their prevalence and translation into technical choices, such as the data sources used to train the AI models. Without an overarching law, this can provide stakeholders – such as healthcare institutions that procure AI systems, investors who fund them, or technologists who design them – with an acceptable benchmark.

176. Kasia Chmielinski et al., "How to Make AI Audits Societal, Not Just Technical", Mozilla, 10 January 2024
177. Shayne Longpre et al., "A Safe Harbor for AI Evaluation and Red Teaming", arXiv.org (2024)

5.2.2. **Organisational and systemic constraints inhibit the effectiveness and success of AI audits**

A senior executive working in product management in the oncology department of an international health technology company stated:

> "A lot of auditing happens during the development of the AI, and post-release, it follows a regular reporting cycle. Auditing post-release is very hard... fairly well-oiled for non-healthcare sectors, like within tech giants like Google etc. But in healthcare, it's still early."

The success of an audit hinges on various structural factors, such as the organisation's auditing capacity and the level of public oversight of such audits. As a consequence, some limitations emanating from these factors can also impede the success of an AI audit, including

- **Risk of audit-washing[178]:** AI audits assess a wide range of variables, and given the current lack of regulatory oversight, there is no consistent standard that AI audits must adopt in India. In this context, audits might just cover the bare minimum and be used as a "smoke screen for corporate responsibility".[179] This is particularly relevant for audits that are internally regulated as they rely on proprietary and confidential information. These audits often disclose minimal information about the underlying principles, which leads to concerns regarding independence.[180]
- **Increased workload and administrative burden:** Although strict compliance with established auditing standards can, in theory, reduce the prevalence of audit-washing, it can also have a counter-effect on particular stakeholders in the data supply chain. For instance, healthcare institutions and medical professionals already struggling to afford privately developed AI systems may find it substantially more burdensome to conduct an end-to-end audit of their AI systems. Auditing-related regulations must, therefore, acknowledge and account for organisation-level barriers.
- **Inconsistency in design and adoption:** Ultimately, the effectiveness of AI audits as a governance tool relies significantly on local contexts, capabilities, and choices. While some of these variables can be addressed through context-specific guidelines – especially for domains such as healthcare – our study indicates that the current state of affairs does not reflect this. Whether an AI audit is even conducted and, if so, by whom, remain largely two unanswered questions about the implementation process followed, something that is imperative to the integrity and credibility of any AI audit.

178. *A firm that has audited itself or submitted to inadequate audit can provide false assurance that it is complying with norms and laws, possibly "audit-washing" problematic or illegal practices. A poorly designed or executed audit is at best meaningless. At worst, it can deflect attention from or even excuse harms that the audits are supposed to mitigate. Audit washing is a cousin of "green washing" and "ethics washing" – the acquisition of sustainability or ethical credibility through cosmetic or trivial steps: Goodman and Trehu. "AI Audit-washing and Accountability". September 2022*

179. *Birhane Abeba et.al. "AI Auditing: The Broken Bus on the Road to AI Accountability"*

180. *Ibid.*

6. Recommendations

The findings in this report focus primarily on the current state of the data supply chains underlying AI systems in India's healthcare sector and the practice of auditing such systems for governance purposes. More particularly, we have flagged a range of challenges – such as the absence of representative datasets, the unaffordability of AI systems, and the lack of collaboration between different stakeholders. These challenges not only affect AI systems' functioning but also hinder the ability of internal and external auditors to guide decision-making processes across the different stages of the DSC.

In this chapter, we propose some recommendations to address these challenges to inform policy and industry action. Figure 21 summarises these proposals, followed by a more detailed discussion of our recommendations and their rationale.

Figure 21: **Recommendations for stakeholders in the AI and healthcare ecosystem**

6.1. Improve data management across the AI Data Supply Chain	Build standardised data sharing policies	Emphasise on not just data quantity but also data quality	
6.2. Streamline AI auditing as a form of governance	Standardise the practice of AI auditing	Build organisational knowledge and inter-stakeholder collaboration	Prioritise transparency and public accountability in auditing standards
6.3. Centre public good in India's AI industrial policy	Adopt focused and transparent approaches to investments and financing of AI projects	Strengthen regulatory checks and balances	

6.1. Improve data management across the AI data supply chain

Sourcing, processing and use of good quality data to train underlying models is at the heart of any AI system. More so for a safety-critical sector like healthcare, which relies on streamlined collaboration between many different stakeholders, such as technology companies, healthcare institutions, and medical practitioners. However, policies related to the management of the DSC, such as the Health Data Management policy[181], often do not complement each other, thereby creating redundancies and ignoring vulnerabilities at the same time.

Consequently, with AI systems becoming commonplace in healthcare, we provide below a few recommendations to strengthen the process of data management. These would include comprehensive policy actions to address challenges of non-standardised data practices as well as the lack of multi-stakeholder interactions. In addition to having robust policies, it is also critical to focus on data quality as much as quantity. These questions are vital, especially at this stage where health data digitisation in India is picking up alongside the development and deployment of AI systems.

6.1.1. Adopt standardised data-sharing policies

This research highlighted the challenges in obtaining India-specific healthcare data, flagging the dependency on international data sources or partnerships with large-sized Indian hospitals to train AI systems. Meanwhile, the increased digitisation of healthcare data through apps (e.g., telemedicine) and the push for a universal health ID in India have led to the collection of health data, albeit scattered across silos and in various formats. While existing health data policies flag the importance of making anonymised health data available to researchers and other organisations, these policies have not been implemented.

A standardised data-sharing policy must be adopted across all sectors collecting health data to guarantee data security while enabling access for researchers and organisations. Nonetheless, a standardised policy alone cannot guarantee the rights of individuals; to achieve this, it is essential to incorporate intersectionality among frameworks and stakeholders while making policies.

Take an intersectional approach for frameworks and stakeholders. Policies must consider the views of multiple stakeholders, and involve public participation, to ensure they are effective and do not harm any stakeholder group. We have already discussed the need for better collaboration between medical professionals and AI developers. Similarly, there is a need for multiple stakeholders, such as medical professionals, academics, technology companies, civil society, and patient advocacy groups, to participate in developing policies on health data, and AI and healthcare. Frameworks from other disciplines must be considered to develop a policy that can respond to a continually evolving technology. One example is the inclusion of the "privacy by design" concept in the earlier drafts of the Data Protection Bill[182] and, more recently, the Guidelines for the Prevention of Dark Patterns, released by the Department of Consumer Affairs which was published to safeguard consumers from unfair trade practices.[183]

181. Ministry of Health and Family Welfare, "National Digital Health Mission: Health Data Management Policy," policy, National Digital Health Mission, 2021

182. "The PDP Bill 2019 Through the Lens of Privacy by Design", Centre for Internet and Society, 12 November 2020

183. Guidelines for Prevention and Regulation of Dark Patterns, 2023. Government of India

Further, adopting a human-centred AI design approach, which involves diverse stakeholders across the AI cycle, can mitigate potential harms. Additionally, co-opting feminist bioethics and decolonial approaches to data governance can foreground the needs of marginalised and vulnerable populations while developing and regulating these technologies so that they may benefit public health systems at large.

Collaboration between professionals from different fields at the intersection of AI and healthcare is essential for gaining better insights and overcoming isolation in AI research. While the development of AI has immensely benefited from academic research, greater collaboration between academics and technology companies in India is needed, especially in AI and healthcare. Partnerships between medical research institutions and technology companies could help identify interventions where AI is needed while facilitating the flow of feedback from end-users to companies on implementation challenges. This could offer opportunities for young medical professionals to understand and engage with AI at the start of their careers.

According to the collaborative model for ethics in medical AI, for effective collaboration between AI designers and developers, medical experts, and other stakeholders: (i) medical doctors should play an active role in model development; (ii) AI designers should engage with doctors to better understand and interpret the outputs of medical AI; and (iii) medical doctors should undertake continuous evaluation of AI in day-to-day clinical practice and share it with AI developers.[184] For healthcare institutions and medical professionals, engaging with professionals outside of medicine, such as technologists, data scientists, and AI ethicists, can offer a broader perspective on AI tools and their applications.[185] Our research shows that medical professionals, including doctors, predominantly engage with AI use in medicine through academic research. While this is necessary at the current stage of AI deployment in India, it is equally vital for doctors, especially those in clinical practice, to receive practical training and experience with AI tools to collaborate effectively with technologists and adopt a more interdisciplinary approach to AI research.

Facilitate continuous feedback loops. Effective and continuous communication is needed to ensure better collaboration between stakeholders. For a system based on data and subject to changes and updates, AI applications in healthcare need to consider regular feedback and collaboration, and not just at the development stage. There is a need for robust feedback mechanisms involving all stakeholders at every stage of AI deployment to boost intersectionality and strengthen collaboration.

Healthcare providers and technology companies need to maintain active communication with each other through established partnerships between technology companies/startups and large healthcare organisations as well as smaller clinics and individual medical professionals. Apart from partnerships, contracts between technology firms and hospitals should provide for regular feedback collection and include methods to assess the system at specific checkpoints.

Provide guidelines for the implementers and their capacity to adopt AI. It is important to note that medical institutions – the actual implementers of the AI policies – need to update themselves with the different policies to implement them correctly, in addition to their existing care or administrative work.[186] A scattered policy landscape creates space for non-compliance due to a lack of knowledge or over-compliance due to fear of liability. A standardised policy would benefit medical professionals and healthcare organisations by providing guidelines for storing data securely and maintaining proper anonymisation before sharing the data with third parties. For example, the Electronic Health Records Standards of 2016[187] should be updated, included in current regulatory conversations, linked to existing health data programmes, and referred to while making new health data policies.

184. *Torbjørn Gundersen and Kristine Bærøe, "The Future Ethics of Artificial Intelligence in Medicine: Making Sense of Collaborative Models", Science and Engineering Ethics 28, no. 2 (2022): 1–16*

185. *Daiju Ueda et al., "Fairness of Artificial Intelligence in Healthcare: Review and Recommendations", Japanese Journal of Radiology 42, no. 1 (2023): 3–15*

186. *Shweta Mohandas, "Privacy, Policy and Preparedness and the Road towards India's Digital Health Ecosystem", Panorama (2024): 105–20*

187. *"Electronic Health Record (EHR) Standards for India 2016 Standards Set Recommendations v2.0", Ministry of Health and Family Welfare (2016)*

6.1.2. **Emphasise not just data quantity but also data quality**

While the functioning of AI systems depends largely on the quantity of data, their success also depends on its quality. As highlighted in this report's findings, various stakeholders in the healthcare and AI ecosystems face challenges related to gaps in data quantity and quality, which they have attempted to address by using datasets from the Global North. However, using these datasets to diagnose and treat a population as diverse as India is bound to create issues of bias and harm, including misdiagnosis.

Maintain interoperability along with data collection. The Ayushman Bharat Digital Mission (ABDM) and the plan to share anonymised health data could make such data from India available to multiple stakeholders. A press release by the Ministry of Health also acknowledged that creating Electronic Health Records (EHRs) would create avenues to integrate them into emerging technologies such as AI.[188] Similarly, the Digital Health Incentive Scheme implemented by the NHA incentivises hospitals, diagnostic labs, and other institutions to contribute to healthcare digitisation efforts; the incentive allows these institutions to earn up to INR 4 crore.[189] In tandem, healthcare institutions appear to be competing to register more patients under the ABDM. According to one report, a district hospital in Uttar Pradesh created over 2000 health IDs (each with a unique Ayushman Bharat Health Account [ABHA] Number) in a day and registered 148,000 patients over three months.[190] While these recent efforts are well-intentioned, there is a need to consider who bears the burden of data collection and management. Currently, the federated approach to health data means that this data is collected and stored by hospitals and clinics. However, better support systems are needed to assist them in managing the added responsibility of collecting, digitising, and storing data.[191] Moreover, there is an urgent need for greater standardisation of data formats and improved interoperability[192] of data. This will help minimise the duplication of data and increase access to health data that had previously been trapped in silos.

Have systems in place to ensure data quality during collection. In addition to focusing on the quantity of data being collected, there is also a need to look at how this data is being collected and if it is being recorded accurately. In the same report on the records of health ID creation, the Chief Medical Superintendent stated that the hospital initially did not have data entry operators for creating the ABHA ID for patients, and the hospital staff undertook this task.[193] In an already overburdened health system, an additional task that requires patience along with attention to detail and speed is not only challenging but also impossible to provision for. Therefore, it is essential to review and examine why health data is being collected and who stands to benefit from it. Poor data quality neither helps the patient, who might end up with an incorrect diagnosis, nor benefits the companies that use it to train their AI models.

Before emphasising the collection of health data, it is essential to prioritise accurate data over speed and quantity. The captured data must add value and not just add to the burden of healthcare institutions and patients. While policies can mandate recording data accurately, they will not work until they are implemented effectively. Public and private hospitals should be able to demonstrate that they have an administrative system in place and self-assess their capacity before collecting and digitising data. There should ideally be a separate set of staff based on the number of patients to ensure data that is being collected is accurate and mistakes are minimised. Healthcare institutions should also update their privacy policies as well as internal data policies based on the Digital Personal Data Protection Act 2023 (DPDPA), including facilitating the right of the patient to correct, complete, update, and erase their personal data.[194]

188. *"Update on Artificial Intelligence (AI) in Health Sector", PIB, 17 March 2023*

189. *ANI, "Revolutionising Healthcare and Empowering Patients with Ayushman Bharat Digital Mission", NDTV-Dettol Banega Swasth India, 22 December 2023*

190. *Ibid.*

191. *"Ayushman Bharat Digital Mission: Boon or Bane?", Accountability Initiative: Responsive Governance, 30 May 2023*

192. *Shivam Soni, "H1: Healthcare", Aapti Institute, 30 May 2023*

193. *ANI "Revolutionising Healthcare and Empowering Patients with Ayushman Bharat Digital Mission"*

194. *The Digital Personal Data Protection Act, 2023. Government of India*

6.2. **Streamline AI auditing as a form of governance**

India's healthcare system suffers from low capacity, poor physical and digital infrastructure, inequitable access, and unaffordability. At the same time, public health is a domain that requires not just well-funded and ethical institutions but also an empowered civil society and, most importantly, proactive regulators. Theoretically, an AI audit can be effective in governing AI systems; however, as discussed in the previous chapter, its success depends on various factors.

Given the relative nascency of AI auditing as a governance response, enhancing its reliability will require specific interventions from policymakers and the many organisations across the DSC. To this end, we list a few recommendations. These suggestions are informed primarily by our findings on the current state of AI audits in India's healthcare system, supported by existing research on auditing AI systems, wherever relevant.

6.2.1. **Standardise the practice of AI auditing**

One of the findings of this study is that there is no common understanding of AI auditing across medical professionals, technology companies, and healthcare institutions. In the absence of overarching regulations to govern the practice, our respondents not only interpreted the word 'auditing' differently but also labelled various processes as AI auditing.

Even institutionally developed guidelines that refer to AI auditing – such as the 2023 ICMR Guidelines – do not refer to particular frameworks or standards. Although some of this heterogeneity stems from the diversity of AI-related use cases, it hinders AI auditing from becoming a scalable practice.

Consequently, some level of standardisation – led by state actors, the medical community, or even an independent audits standards board – would allow for substantial progress on this front.[195] Alternatively, in consultation with medical professionals, auditing experts, and AI developers, civil society organisations can develop an actionable framework for the audit community. While contextualised for the Indian landscape, these efforts can borrow from existing interventions, such as the Artificial Intelligence Auditing Framework (developed by the Institute of Internal Auditors) or the CRISP-DM Framework (developed by the Information Systems Audit and Control Association).[196] [197]

Some form of uniformisation in AI auditing could also simplify decision-making for many organisations across the data supply chain. For instance, whether an AI audit should be conducted by internally appointed auditors or their external counterparts is an important decision with vastly different potentials, implications, and risks.[198] As discussed in Chapter 4, while internal audits allow for adaptiveness and versatility, they often face a higher risk of corporate interference. In contrast, external audits will likely remain toothless without an appropriate enforcement mandate. By collaborating and developing informed standards for questions such as these, policymakers and similarly situated stakeholders can create a common benchmark for AI auditing and set the agenda for a more robust policy instrument in the future.

195. *David Manheim et al., "The Necessity of AI Audit Standards Boards", arXiv (2024)*

196. *"Global Perspectives and Insights: The IIA's Artificial Intelligence Auditing Framework", The Institute of Internal Auditors, Global, accessed 4 November 2024*

197. *Andrew Clark, "The Machine Learning Audit: CRISP-DM Framework", ISACA (2018)*

198. *Andika Pramukti, "Internal Audit versus External Audit: A Qualitative Perspective", Golden Ratio of Auditing Research 4, no. 2 (2024): 78–88*

6.2.2. **Build organisational knowledge and inter-stakeholder collaboration**

Similar to financial or environmental audits, which are relatively more mature tools of governance, reliable auditing of AI systems also requires participants to be well aware of sectoral nuances.[199]

Develop multi-sectoral knowledge and build capacity for AI audits. In the case of India's healthcare sector, these nuances include not just the technical knowledge of the underlying architecture, but also the socioeconomic realities of public health. It is all the more important for the audit team to develop an interdisciplinary understanding in situations where internal audits are used[200]. For instance, technology companies attempting to audit their health AI systems inevitably require in-house technical experts, such as data scientists and engineers, to possess adequate knowledge of auditing processes and pitfalls. The same holds for the ecosystem of auditors, many of whom would require, at least, basic exposure to and training in AI systems and their interaction with the many medical use cases.[201] [202] Bodies such as the Institute of Internal Auditors and other capacity-building organisations can also provide potential trainers for such exercises.[203]

Enable collaboration and sharing of feedback for AI audits. The multi-stakeholder nature of the existing data supply chains necessitates the involvement of all the relevant parties in the auditing process.[204] As discussed in this report, collaboration and feedback sharing between technology companies and medical professionals remains inadequate and irregular. Addressing this gap through appropriate policy and market interventions is imperative to improve the process of model development and enhance the practice of AI auditing. For example, risks associated with vendor lock-in or partitioned knowledge, which are more relevant for healthcare institutions, can be substantially mitigated through collaboratively designed AI audits.[205]

6.2.3. **Prioritise transparency and public accountability in auditing standards**

At its core, an audit evaluates complex systems to determine their compliance with predetermined standards.[206] Whether these standards exist at the organisational, sectoral, or regulatory level depends on various factors. However, considering the particularly sensitive nature of healthcare, the value of public accountability and transparency cannot be overstated – a principle that both domestic and global standard-setting institutions have repeatedly articulated.[207]

Our study finds that most healthcare institutions procure externally developed AI systems, which remain unavailable for independent evaluations due to their largely closed-source nature. As a result, addressing this gap would require modification of existing technological and regulatory infrastructures. For instance, disclosure of certain information about the AI model or the underlying training data can allow policymakers and other stakeholders in the DSC to effectively leverage red teaming – a structured testing effort that relies on independent evaluation to identify flaws and vulnerabilities in an AI system.[208] [209]

Alternatively, in cases where real-time or public testing of an AI system is not feasible or practical, a retrospective internal AI audit can help make these systems more accountable to public representatives and their constituents.[210] However, to minimise the risk of audit washing, these internal audits must be accompanied by appropriate documentation and result in public reporting of their outcomes to enable some level of external decision-making.[211]

199. *"Auditing Artificial Intelligence", ISACA, accessed 4 November 2024*

200. *Jakob Mökander, "Auditing of AI: Legal, Ethical and Technical Approaches", Digital Society 2, no. 49 (2023): 1–32*

201. *"Global Perspectives and Insights: The IIA's Artificial Intelligence Auditing Framework", The Institute of Internal Auditors, Global.*

202. *"Auditing Artificial Intelligence", ISACA.*

203. *"About IIA", The Institute of Internal Auditors, India, accessed 4 November 2024*

204. *Ibid.*

205. *Ibid.*

206. *Inioluwa Deborah Raji et al., "Closing the AI Accountability Gap".*

207. *Jana Fehr et al., "A Trustworthy AI Reality-Check: The Lack of Transparency of Artificial Intelligence Products in Healthcare", Frontiers in Digital Health 6 (2024): 1–11*

208. *Shayne Longpre et al., "A Safe Harbor for AI Evaluation and Red Teaming"*

209. *Artificial Intelligence Index Report 2024, Stanford University HAI (2024), pg 234*

210. *Katharina Simbeck, "They Shall Be Fair, Transparent, and Robust: Auditing Learning Analytics Systems", AI And Ethics 4, no. 2 (2023): 555–71*

211. *Gemma Galdon Clavell, "AI Auditing - Checklist for AI Auditing", European Data Protection Board (2023)*

6.3. Centre public good in India's AI industrial policy

Increasingly, national governments have been strategising to increase public funding for AI through commitments such as increasing the R&D budget, setting up industrial and investment funds in AI startups, and investing in networks, infrastructure, and AI-related public procurements.[212] These strategies focus on a range of actions, including building skills and foregrounding governance, in addition to allocating financial investments. A collection of these investment, regulatory, and government-spending strategies aimed at building AI capabilities at a national level can be termed an AI industrial policy.[213] In this section, we recommend some critical AI regulatory and financing-related issues to consider as we move towards greater AI-related spending in the country.

6.3.1. Adopt focused and transparent approaches to investing in and financing AI projects

India committed in 2024 to investing upwards of USD 1.2 billion in AI projects, including but not limited to computing infrastructure.[214] In an official statement, the Ministry of Electronics and Information Technology announced that this investment is "poised to catalyse various components of the IndiaAI Mission, including pivotal initiatives like the IndiaAI Compute Capacity, IndiaAI Innovation Centre (IAIC), IndiaAI Datasets Platform, IndiaAI Application Development Initiative, IndiaAI FutureSkills, IndiaAI Startup Financing, and Safe & Trusted AI."[215] Clearly, there is a strong interest in enhancing India's AI capabilities and developing the necessary infrastructure to achieve this goal.

According to Stanford University's annual AI Index report, aside from public investments in AI, India ranked fifth in investments received by startups offering AI-based products and services from public and private entities in 2023.[216] While channelling robust investments into AI and AI infrastructure is a welcome move, several factors need consideration. These factors extend beyond the binaries of domestic vs global AI to critically assess who stands to benefit from these investments in the AI ecosystem.

Ensure equitable distribution of AI spending and associated benefits. It will be critical to ensure that, within healthcare, investments in AI and the subsequent use of these AI systems are not limited to the private healthcare ecosystem but are made available to public healthcare to fulfil its fundamental promise of improving affordability and access to healthcare through AI. Today, much of the discourse on AI in healthcare hinges on the promise of future benefits with limited evidence to that effect, except in certain use cases, as shared earlier in the report. Investments in AI should focus on enhancing benefits and avoid amplifying existing harms.

212. *National Strategy for Artificial Intelligence, NITI Aayog (2018)*
213. *Amba Kak, "AI Nationalism(s): Executive Summary", AI Now, 12 March 2024*
214. *"India Announces $1.2 Bln Investment in AI Projects", Reuters, 7 March 2024*
215. *Deborah Grey, "India Announces US$ 1.25 Billion Investment into AI", W.Media, 11 March 2024*
216. *Artificial Intelligence Index Report 2024, Stanford University HAI*

Despite an overall decline in AI private investment in 2023, funding for generative AI surged, "nearly octupling from 2022 to reach $25.2 billion."[217] In India, within the healthcare sector, investments are expected to rise in global healthcare-specific large language models (such as MedPaLM2, Meditron, and Hippocratic AI), vernacular language models (such as Sarvam AI, Krutrim), and newer models that can be trained using the government language repository (Bhashini) for the Indian context.[218] Even when it comes to investments in generative AI for healthcare, the principles of cautious implementation and prioritising public good (as may be possible through vernacular language models) need to be applied diligently and without discrimination.

Invest in AI life cycles instead of just at the point of initial development. The use of AI in healthcare will require initial capital investments and ongoing expenditure for maintenance, incorporating larger amounts of patient data, updating software algorithms, and ensuring hardware operability within healthcare institutions.[219] It may also mean large-scale system upgrades and setting up of infrastructure, especially in under-resourced healthcare setups (community health centres, public hospitals, etc). While making investment strategies, the last-mile delivery of such AI applications and their ongoing maintenance need to be seriously considered.

In addition, even when there are promises of economic benefits through the use of AI, India needs thoughtful and strategic direction in the development of AI products and services that are useful and do not follow short-term "hype cycles" such that "high investments in compute are linked to actual economically beneficial outcomes."[220]

Encourage public-private partnerships to optimise for public good over profits. Public private partnerships (PPPs) are often seen as a golden bullet for obtaining cost efficiency and optimising the private sector's expertise while using the public sector's regulatory and other infrastructural resources to address critical demands.[221] Not just in terms of efficiency, PPPs are also seen as a solution to the ethical challenges that the use of AI brings.[222] However, the litmus test for these AI-based PPPs should centre around public good and whether these investments are genuinely reaping benefits for people at large, particularly in the case of healthcare, by democratising access to it instead of focusing solely on economic or business interests.

Besides, there also needs to be more transparency in AI-related PPPs and AI investments. Recently, the government invited bids for the empanelment of vendors providing AI services. The Request for Empanelment sought entities such as data centres and cloud service providers to submit bids to provide access to high-speed AI infrastructure.[223] Such public calls allow for greater transparency on the various opportunities for investments in AI for different stakeholders and help create more trust in the ecosystem. For PPPs, regular reporting on such partnerships and financial disclosures can aid in creating greater transparency.

217. *Ibid.*

218. *Pankaj Jethwani, "Trends Shaping India's Healthcare Startup Landscape", Inc42 Media, 12nFebruary 2024*

219. *Jianxing He et al., "The Practical Implementation of Artificial Intelligence Technologies in Medicine"*

220. *Jai Vipra, "A Compute Agenda for India", CyberBRICS (2024)*

221. *K. Rajendra Prasad et al., "AI in Public-Private Partnership for IT Infrastructure Development"*

222. *Devroop Dhar, "Imperative of Public-Private Partnerships in Ethical AI Development in India", Techcircle, 26 August 2024*

223. *PTI, "Government Invites Bids to Empanel AI Infra Providers under Rs 10,372-Cr India AI Mission", The Economic Times, 17 August 2024*

6.3.2. **Strengthen regulatory checks and balances for AI governance**

In 2023, there were talks regarding a comprehensive law to oversee emerging technologies, specifically the Digital India Act, but these discussions appear to have lost momentum.[224] The absence of overarching legislation to regulate new technologies, such as AI, which are gaining significance daily, has resulted in a legal system that struggles to keep pace with technological advancements. While it is understood that lawmaking is an arduous process, requiring multiple stakeholders to agree on a common framework, there is a need to have more focused investments in AI-related regulations, especially when it is gaining prominence in critical areas such as healthcare and insurance. This was reiterated by one of our civil society interviewees, who noted that without regulations and redlines, it would be difficult to ensure that organisations follow ethical and human rights-preserving principles. While a new regulation takes time, another possible way could be to amend existing regulations to bring AI and similar new technologies into their ambit; the Medical Devices Rule 2017 and the Telemedicine Guidelines of 2020 prove how new technologies can be brought under existing laws. Another such opportunity is available with the DPDPA Rules[225], which are yet to be released. These Rules could include provisions governing the need for anonymisation of data, checks on automated decision-making, and adding companies that collect and process health data as significant data fiduciaries.

Alongside examining regulation, it is also essential to emphasise the need for transparency, not just as a principle in the regulation but also in the drafting process of the regulation. There is a need to have multiple stakeholders, including academics, civil society, and startups, involved in the drafting process. An example of the need for transparency and greater stakeholder involvement is the recent AI advisory published by the IT Ministry on 1 March 2024, requiring all AI companies to obtain permission from the government "to make their products available to users online in India" within a limited timeframe;[226] the advisory triggered a lot of panic and uncertainty among AI companies, both big and small. While the IT Ministry withdrew this advisory within a fortnight, it revealed the need for extensive multi-stakeholder consultations, before releasing such guidelines and hearing the opinions of the people most affected by the regulation.

One such step in the right direction is the recent talks about creating AI safety institutes[227] with a mandate to identify harms and set standards that could guide future regulations on AI. While it is not confirmed where these AI safety institutes will be housed, it would be important that these are not restricted to institutes of technology but also include those teaching medicine and social sciences.

224. *Das, "Questions Arise on Digital India Act, Other Tech Regulations in Coalition Govt", Mint, 6 June 2024*

225. *Ashutosh Mishra, "Balance Privacy & AI Innovation in DPDP Rules: Global Tech Body Urges Govt", Business Standard, 22 September 2024*

226. *The Hindu Bureau, "IT Ministry Replaces AI Advisory, Drops Requirement of Government's Permission", The Hindu, 16 March 2024*

227. *Aditi Agrawal, "Govt Mulls Setting up Artificial Intelligence Safety Institute", Hindustan Times, 13 October 2024*

7. Conclusion

In recent years, the Indian health sector has seen a notable rise in initiatives aimed at implementing AI in healthcare. However, there is a need to address the challenges that have emerged from introducing even simple digital solutions without accounting for ground realities. For instance, the unavailability of Aarogya Setu (the COVID-19 safety app) on feature phones resulted in some of the most vulnerable people being excluded from working, travelling, or seeking medical help during the pandemic.[228] Under these circumstances, rapid and extensive AI development is outpacing regulatory responses and has emerged as a critical area of concern.

One of the key findings of this study is the range of challenges prevalent in the DSC and key instruments in AI development and deployment. The reliance on data from the Global North affects not only academics and technology companies using such data but also medical professionals using AI systems trained on this data, along with patients who are subjected to the decisions of these systems. In contrast, efforts to build India-specific health data transfer the burden of data collection, quality maintenance, and security to an already overburdened healthcare system and healthcare professionals. Meanwhile, the lack of medical professionals in developing these AI systems has resulted in gaps in understanding how AI can be best integrated into their workflows. It raises questions of unfamiliarity, lack of trust, and the need for separate training on using AI systems. Furthermore, as in other parts of the Global South, India faces significant infrastructural constraints in implementing AI for healthcare.[229]

This study highlights that, although the integration of AI in healthcare is currently uneven, its presence is undeniable. We anticipate significant growth with the continuous development of innovative use cases and AI products. However, there is a need to critically assess which use cases of AI can improve the current healthcare system in India and which issues can be addressed with equally effective and more affordable non-AI interventions. When it comes to AI governance, implementing legislation to address the challenges posed by AI is a formidable task. The time and regulatory resources required to develop comprehensive legislation that meets the needs of all stakeholders are not just substantial but also hard to foresee.

Consequently, this study investigated auditing as a governing tool that exists in other sectors and is being practised by some technology companies and healthcare institutions. While auditing has its merits and limitations, it is a possible intervention to ensure that AI systems and their results are evaluated for biases and harms. Auditing also serves as institutional memory, documenting the rationale and processes involved in AI development and deployment, which can be reviewed in the event of legislation.

In conclusion, while India needs to embrace new technologies and explore ways to improve its healthcare system, it must assess its infrastructural capacity, especially in public health systems.

Additionally, understanding the readiness of people who will use and be impacted by these changes is crucial. Existing governance mechanisms, including data protection and data security, warrant careful consideration.

Finally, it is vital that we address the risks that large-scale implementation of these systems could bring, before doubling down on them and expanding without proper guardrails in place.

228. *Special Correspondent, "Cannot Deny Services for Not Installing Aarogya Setu App: HC", The Hindu, 19 October 2020*

229. *Vidushi Marda, "Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making", Philosophical Transactions of the Royal Society A 376 (2018): 20180087*

Centre
for Internet
& Society