Chauhan *et al. BMC Medical Ethics* (2025) 26:140
https://doi.org/10.1186/s12910-025-01265-7

BMC Medical Ethics

RESEARCH Open Access

Evaluating trustworthiness in AI-Based diabetic retinopathy screening: addressing transparency, consent, and privacy challenges

Anshul Chauhan[4†], Debarati Sarkar[2†], Garima Singh Verma[4], Harsh Rastogi[4], Karthik Adapa[3] and Mona Duggal[4,1*]

Abstract

Background Artificial intelligence (AI) offers significant potential to drive advancements in healthcare; however, the development and implementation of AI models present complex ethical, legal, social, and technical challenges, as data practices often undermine regulatory frameworks in various regions worldwide. This study explores stakeholder perspectives on the development and deployment of AI algorithms for diabetic retinopathy (DR) screening, with a focus on ethical risks, data practices, governance, and emerging shortcomings in the Global South AI discourse.

Methods Fifteen semi-structured interviews were conducted with ophthalmologists, program officers, AI developers, bioethics experts, and legal professionals. Thematic analysis was guided by OECD principles for responsible AI stewardship. Interviews were analyzed using MAXQDA software to identify themes related to AI trustworthiness and ethical governance.

Results Six key themes emerged regarding the perceived trustworthiness of AI: algorithmic effectiveness, responsible data collection, ethical approval processes, explainability, implementation challenges, and accountability. Participants reported critical shortcomings in AI companies' data collection practices, including a lack of transparency, inadequate consent processes, and limited patient awareness about data ownership. These findings highlight how unchecked data collection and curation practices may reinforce data colonialism in low and middle-income healthcare systems.

Conclusion Ensuring trustworthy AI requires transparent and accountable data practices, robust patient consent mechanisms, and regulatory frameworks aligned with ethical and privacy standards. Addressing these issues is vital to safeguarding patient rights, preventing data misuse, and fostering responsible AI ecosystems in the Global South.

Keywords Artificial intelligence, Diabetic retinopathy screening, Trustworthiness

[†]Anshul Chauhan and Debarati Sarkar are co-first authors of the article.

*Correspondence:
Mona Duggal
monaduggal2@gmail.com
[1]Present address: Indian Council of Medical Research (ICMR) - National Institute of Research in Digital Health and Data Science (NIRDHDS), Ansari Nagar East 110029, New Delhi, India
[2]Interdisciplinary Research Programme, Indian Institute of Technology, Jodhpur, India
[3]Department of Radiation Oncology, UNC School of Medicine, Chapel Hill, USA
[4]Postgraduate Institute of Medical Research and Education, Chandigarh, India

Introduction

The rapid growth of health data curation, advancements in data storage, computational power, and intelligence analytics are reshaping healthcare technologies that utilize Artificial Intelligence (AI) and Machine Learning (ML) [1, 2]. With the advent of big data, large-scale health data sources, including social media platforms, web search engines, online forums, mobile applications, and wearable devices, have significantly expanded, providing more explicit insights into health determinants than traditional sources [3, 4]. This large and complex data has created new opportunities for AI to improve healthcare, including diagnosis, outcomes prediction, and patient care management [5, 6]. Rapid advancements in AI-based algorithms are already enhancing diagnostic accuracy and efficiency across multiple specialties [7]. However, AI ethics discussions are primarily led by high-income countries, with nearly 80% of studies from these regions. At the same time, perspectives from the Global South, including India, are critically underrepresented [8, 9]. In recent years, the ophthalmology specialty has leveraged advances in computing and big data to develop AI tools that support disease screening, treatment, and improve access to care [10, 11].

This study presents diabetic retinopathy (DR) as a case study, a severe microvascular complication of diabetes and a leading cause of preventable blindness globally [12]. AI algorithms are increasingly deployed for DR screening (DRS) to meet the rising demand driven by the growing diabetic population worldwide [13, 14]. AI has great potential to improve healthcare, but it also raises significant ethical and epistemic challenges, including data reliability, bias, and opaque decision-making. These issues create trust and transparency problems, making AI systems more complex to use responsibly [15]. Epistemic challenges arise because AI models, particularly deep learning models, are often opaque, making it challenging to ensure explainability and accountability [15–18]. Developing these algorithms requires large sets of digital fundus images for training and validation [10, 11]. This raises privacy concerns about data processing and anonymization, and patients perceptions of privacy and security in AI systems [19].

Recent studies on AI in DRS highlight key bioethical concerns, including unclear data ownership, inadequate consent processes, biased AI decisions that compromise patient autonomy, and a lack of accountability and transparency [20–22]. These issues pose significant risks to achieving equitable healthcare outcomes [20–22]. Recent research highlights the risk of data colonialism, in which tech companies collect data from low and middle-income countries (LMIC) without clear applications or benefits to the local population [23]. Therefore, as AI technologies are increasingly integrated into medical practice, strong governance and transparency mechanisms are essential due to their wide-ranging ethical, regulatory, and privacy impacts. The responsible adoption of AI in healthcare has raised concerns among researchers, healthcare practitioners, and regulatory bodies regarding data acquisition, ethical clearance, and participant consent [22, 24]. Well-documented AI training datasets often have small sample sizes and biased demographics, resulting in models that lack generalizability across diverse populations [22, 25]. Epistemological concerns arise as many AI models, particularly deep learning-based systems, function as "black boxes" with opaque decision-making processes [15]. Establishing regulatory and accreditation systems could promote the development of safe and sustainable AI models in healthcare [26]. Addressing ethical issues like accountability, transparency, fairness, and bias in AI requires interdisciplinary collaboration to build legal and ethical frameworks that reduce risks and support responsible use [27].

While some studies have focused on AI explainability and interpretability to improve trustworthiness, the ethical role of stakeholder involvement in its development and deployment remains underexplored [16–18]. Given the complexity of AI ethics, applying ethical principles throughout the AI lifecycle from design to deployment while involving diverse stakeholders, including providers, industry, and patients, is essential. Such perspectives are crucial for addressing ethical risks, improving patient outcomes, and ensuring the equitable, effective, and responsible adoption of AI in healthcare [25, 28, 29]. Despite growing interest, comprehensive studies on stakeholders' concerns, expectations, and insights across AI lifecycle stages, particularly within the Indian context, remain unexplored. Aligning perspectives from various stakeholders is key to successfully integrating AI into clinical practice, making implementations more effective and inclusive. Diverse perspectives are crucial for uncovering unique challenges, addressing ethical risks, enhancing patient outcomes, and protecting patient rights [28, 30].

This study synthesizes multifaceted ethical concerns and provides empirical evidence to inform the responsible adoption of AI-driven DRS. Guided by the Trustworthy AI Framework, the study explores how stakeholder perspectives on transparency, consent, and privacy influence the ethical development and deployment of AI-based DRS. It places particular focus on the adoption of these technologies within the Indian healthcare landscape.

Methods

This manuscript has been prepared using the Consolidated Criteria for Reporting Qualitative (COREQ) Research guideline for qualitative studies [31]. (see checklist supplementary file 1: Table 1).

Study design and recruitment

The study was conducted from November 2022 to December 2022. The study employed qualitative research methods to explore the stakeholders perspectives on AI development and deployment, professional decision-making, policy and legal aspects, and management. We prioritized incorporating insights through qualitative, in-depth interviews with key stakeholders contemplating informational power [32].

We conducted 15 in-depth qualitative interviews with English-speaking health system stakeholders, including ophthalmologists (Opth), retina specialists, program officers (PO), legal experts, bioethics experts, and industry partners (AI developers). To gather valuable insights, participants were selected using a convenience sampling strategy rather than a statistically representative sample [33]. The stakeholders involved ophthalmologists and retina specialists involved in AI deployment for DR screening and workflow optimization within their medical practice; program officers who help to ensure the integration of AI-enabled screening programs in the health system; legal professionals to understand the use of AI tools comply with legal and regulatory requirements; bioethics experts, role to address concerns about ethical and the potential biases inherent in AI systems; AI developers to understand the role of designing, developing, and refining AI algorithms, ensuring the tools are effective and meet clinical requirements. The ophthalmologists were posted at four district hospitals [34] of Punjab (Mohali, Moga, Faridkot, Amritsar), and a retina specialist working at the Advanced Eye Centre of Postgraduate Institute of Medical Education and Research (PGIMER), Chandigarh. The PO belonged to the National Program for Control of Blindness and Visual Impairment (NPCBVI) [35], the National Program for Control of Non-Communicable Diseases (NP-NCD) [25] at the state National Health Mission (NHM), and the Ministry of Health and Family Welfare (MoHFW) nationally. The NPNCD program officer was included because the program refers diabetic patients from diabetes clinics for DR screening. Similarly, the NPCBVI program officer was involved because DR screening is integrated into the program's guidelines. The industry partners (IP) were associated with commercially available DR screening companies with AI-building experience, legal experts in human rights and digital health, bioethics, and an independent bio-design expert.

Data collection

The in-depth interview guide was pilot-tested with five researchers (excluding authors) for clarity and length and revised accordingly, but the data was not included in the analysis. To ensure rich data from a range of participants with knowledge and experience of AI, no limitation on eligibility was imposed based on demographic characteristics (e.g., gender, age, etc.). We approached six ophthalmologists, of whom five participated; four out of five program officers were interviewed; two legal experts participated; two failed to respond; one bioethics person participated out of two; and three of five industry partners consented to join. Scheduling appointments and a lack of time were the major reasons for non-participation.

The qualitative data were collected using seven face-to-face and eight online (Zoom) interviews through semi-structured interviews, using the same interview guide for participants in each group. The data were collected at a time and place convenient for the participant. The participants were led to a quiet corner within the hospital and office to conduct the interviews. After explaining the study procedures, informed consent was obtained. Invited participants were contacted to arrange a time to conduct an online interview via the Zoom meeting platform (https://app.zoom.us/wc). The participants pr ovided verbal consent for the interview. The interviews lasted, on average, 45 min (50–60 min). The first author (AC) conducted all the interviews in English. The open-ended interview questions (supplementary file 2: A-E) encouraged participants to describe their views and share practice experiences in their own words. The interviews were recorded and transcribed (AC, HR, and GS), and all identifying information was removed. No repeat interviews were conducted, and the audio recordings were destroyed after the analysis.

Research team

AC and DS are research scholars with public and digital health backgrounds and significant social research experience. GS and HR, research associates with public health expertise, were also part of the team. The study was guided by two senior authors, MD and KA, who have extensive experience in responsible AI and digital health, offering valuable insights throughout the study's implementation and manuscript development.

Data analysis

The interview data were analyzed using inductive thematic analysis [33]. We used the MAXQDA Analytics Pro(24.5.1) Scientific Software Development VERBI GmbH (https://www.maxqda.com/) (AI-powered coding recently available for the software was not used in this study) to reflect on views in the interview transcripts to identify, retrieve, and clustering codes in an

iterative three-phase coding procedure. Several coding phases [36] combined with constant comparison techniques, embodied two researchers (DS, AC), both PhD scholars, who independently read the transcripts and reread them several times to foster familiarity with the data set, coded interviews, and codes were refined until a consensus was reached. The codebook was discussed with an expert (research experience over 20 years) for input (MD), and new codes were refined. The codes and themes were mapped and guided by the Organisation for Economic Co-operation and Development (OECD) principles for responsible stewardship of trustworthy AI, including inclusivity, fairness, privacy, transparency, explainability, robustness, security, safety, accountability, data governance, sustainability, and inclusivity. These principles were selected for their ability to encompass both AI actors and stakeholders. AI actors are entities directly involved in the AI lifecycle, like those deploying or operating AI. Stakeholders include all who are directly or indirectly impacted by AI systems. This model was selected to comprehensively capture the varied roles and impacts within the AI ecosystem [37]. The participants are denoted in italics with a unique ID number: Opth 1–5 (Ophthalmologists); PO 1–4 (Program Officer); LE 1–2 (Legal experts); IP 1–3 (Industry partner- AI developers); BE (Bio-Ethics Design), to preserve anonymity. Quotes were edited for readability to retain their original meaning; ellipses (…) show removed text.

Results

The stakeholder group ($n = 15$) consisted of 9 (60%) males and 6 (40%) females, with the majority aged 31–40 years (53.3%), followed by 41–50 years (33.3%) and 51–60 years (13.3%). The participants had an average professional experience of 12.3 ± 5.2 years during the interview. (see supplementary file 1: Table 2)

Data from all transcripts were mapped and classified to generate the final code matrix. An independent researcher coded a random number of selected interview transcripts to generate an intercoder reliability statistic [38]. The percentage agreement of intercoder reliability was 96.5. Figure 1 provides insights into the distribution of codes across different transcripts, and large nodes indicate the most frequent code. This highlights key concepts, such as accountability, ethical approval, data privacy, image quality, and data standardization, that are dominant in different parts of the dataset. The codes were transformed into themes and sub-themes. The analysis of interviews resulted in six theoretical themes: effectiveness of AI algorithm; responsible AI with data collection, quality, annotation, data privacy, and colonialism; ethical design and consideration; model training and classification; challenges of AI implementation; accountability and liability. Stakeholders discussed the situated nature of implementing AI tools in DR screening and care. Each theme is demonstrated and supported by quote(s) from the interviews.

We summarized the main raw quotes from the representative transcripts, as illustrated in Tables 1, 2 and 3.

Effectiveness of AI algorithms

All stakeholders felt that AI is critical in screening DR, managing up to 80% of cases at the community level, and reducing over-referrals, enabling eye specialists to concentrate on more severe conditions. However, concerns about AI's diagnostic accuracy persist, with some stakeholders expressing a preference for care by qualified eye surgeons, reflecting resistance to delegating clinical tasks to AI or less-trained personnel.

Responsible AI in relation to data

A few key sub-themes emerge to guide responsible AI use, particularly across the AI lifecycle from development to deployment, offering new perspectives on responsible approaches to data handling practices.

Data collection - The industry partners explained that data collection for AI development involves aggregating data from multiple (online and offline) sources. They focus on collecting data from clinical settings, partners, or community camps. An example of data collection involves gathering image data through large-scale screening efforts, with over 94,000 individuals screened to improve the model's performance. However, it unfolded that developers follow conducive and flexible data collection practices with private clinics' help. At the same time, in the public sector, they must navigate through more rigorous checks and balances and a longer approval process. The bioethics expert proposed that AI systems must transparently convey their functioning, data collection, and usage in clear, simple language, ensuring comprehensive understanding for users with varying levels of technological literacy.

Data privacy - All stakeholders emphasized the importance of data privacy and ownership in data-driven digital health realms surrounding ethical issues. However, the understanding of data privacy amongst patients and policymakers remains meager. The bioethics expert emphasized that collecting only necessary data and following data anonymization principles and compliance are crucial for protecting people from harm in case of a data breach (unauthorized access and misuse). The program officer emphasized that in the push to integrate AI into public health workflows, the primary concern is ensuring data privacy, preventing unauthorized access, obtaining informed consent, and maintaining data integrity.

Data anonymity - Data anonymization must be carefully considered to protect patient privacy rights. Anonymized data involves removing personally identifiable

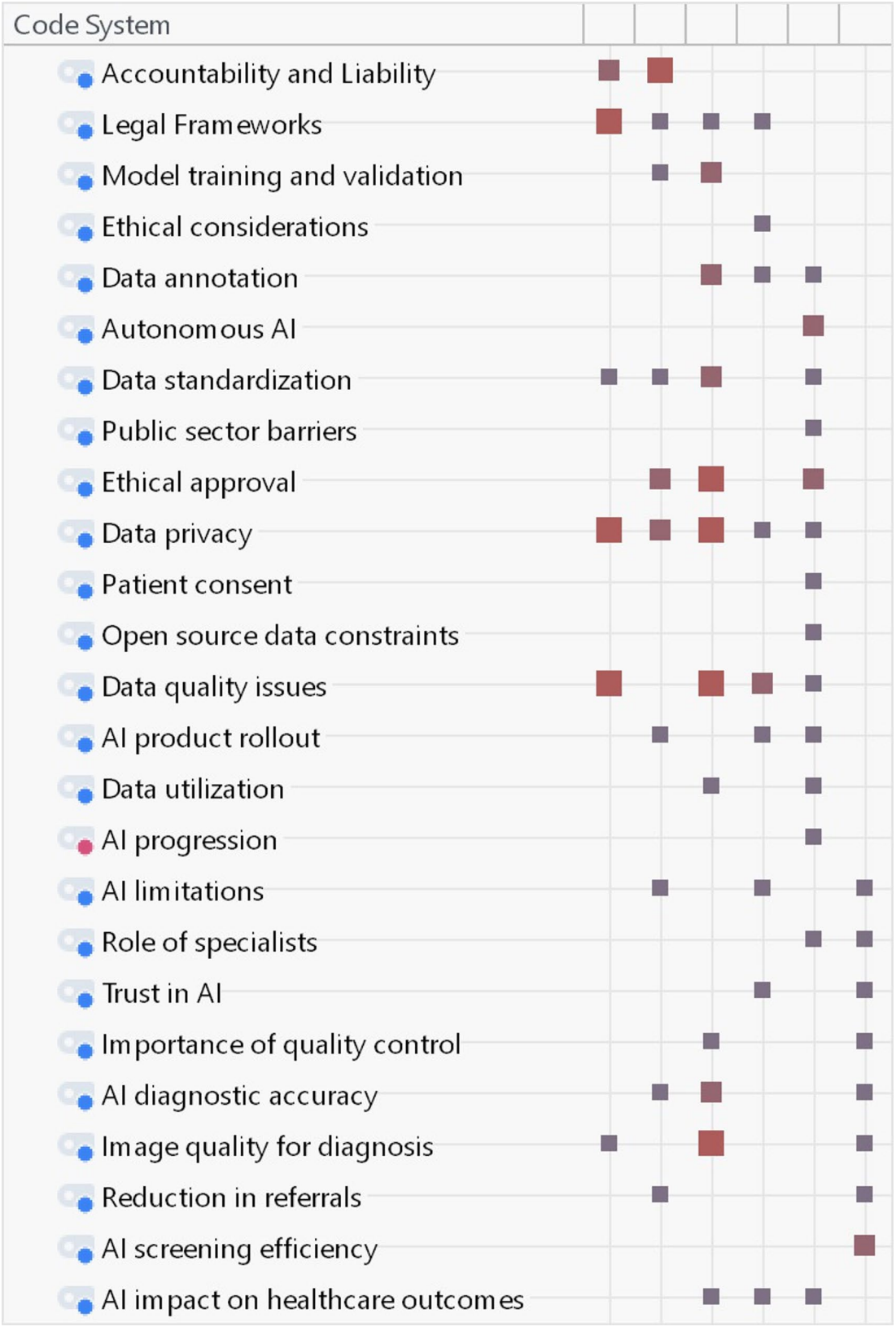

Fig. 1 Frequency of codes across dataset

Table 1 Quotes representing the effectiveness of AI algorithms and responsible AI themes

Theme	Sub-themes	Sample of illustrative quotes
Effectiveness of AI algorithms	**Efficiency of AI**	*"The primary purpose of AI is its use for screening purposes, and it can cover up to 80% of all community diabetic retinopathy-related problems without a needless referral." Opth – 1*
	Optimizing specialist resources	*"If we introduce a person of a lower caliber than an eye surgeon, be it AI, then I will not trust them." PO – 3*
Responsible AI in relation to data	**Data collection**	*"We started conducting camps on our own, and during the entire process, we have screened more than 94,000 people to date. This helps improve our AI model performance". IP-1* *"AI systems must clearly explain how they work, what data they collect, and how they are used in simple, non-technical language. It is essential to present this information in a way that resonates with a broad range of users, including those with limited technological literacy." BE*
	Data privacy	*"The meaning of data privacy is unclear to people. If I ask about privacy, they will say, what privacy? What will I do with privacy? I just want my good health". LE – 1* *"Doctors say "acha data chahiye de do isme kya hai" (If you need the data, take it; what's the big deal), and they have no issue with their data taken away." LE – 2*
	Data anonymity	*"Purpose limitation and data anonymization act as vital safeguards, ensuring that AI only collects and uses essential data and that individuals are protected from harm in the event of data breaches". BE* *"It is very well known that it is not possible to anonymize completely. It is very easy to reidentify and to bring it back away from anonymity". LE – 1*
	Data colonialism	*"AI Companies are quickly creating large data sets. They are self-regulating themselves and saying we are anonymizing, we are not revealing it into the public domain, so some amount of self-regulation is going on such that they don't get into a lot of trouble". LE – 1* *"Whatever data we could lay our hands on, we wanted to get that machine learning keeps improving. Hence, we did not have any exclusion criteria". IP – 3*

AI: Artificial intelligence, BE: Bioethics, IP: Industry partner, LE: Legal expert, Opth: Ophthalmologist, PO: Program officer

Table 2 Quotes representing the ethical consideration and approval and explainability themes

Theme	Sub-themes	Sample of illustrative quotes
Ethical consideration and approval	**Scope of AI inclusion in medical ethics**	*"The available ethical framework in India does not impose any compulsion to have any ethics clearances for AI training datasets" IP – 2* *"There is no need for them to worry about consent for AI development if there is already consent from the patient for his diagnosis or treatment." IP – 2*
	Ethical literacy	*"If you have to translate consent into a local language, and many translate the legal language into a local language, it becomes even more complicated." IP – 1* *"The developers say that no one understands the language of consent and that consent fatigue is present in the patient, so we should not worry about consent." LE – 2*
Explainability	**Validation process**	*"Classifying an algorithm depends on the problem space you are addressing. For example, an algorithm used in critical intensive care must undergo rigorous scrutiny, while one for screening purposes will have a different level of scrutiny." IP – 1* *"We use a technique, often called a 'heat map' to understand the parameters the AI bases its decisions on. This helps clarify how the algorithm operates and the factors driving its conclusions." IP – 3*

IP: Industry partner, LE: Legal expert

Table 3 Quotes representing the challenges of AI implementation and accountability and liability

Challenges of AI implementation	**Regulatory frameworks**	**"Currently, there is no ethical or regulatory framework in India for using AI for pure development purposes". IP – 2** **"Many digital healthcare platforms operate in undefined regulatory areas. They often piggyback on existing laws, which may not fully apply to them, allowing them to bypass rigorous testing and legal scrutiny." LE – 1**
Accountability and liability	**Misdiagnosis attribution**	*"AI is just a tool for clinicians; however, patient safety is the collective responsibility of everyone involved, from developers to healthcare providers. We need policies and education on appropriately using AI in practice." IP – 2* *"No one less than an ophthalmologist should be doing the DR grading. First, many people have DM, but very few have DR; those requiring intervention are few, and intervention is exact. To understand that interventions like laser or anti-VEGF should be prerogatives of eye specialists in tertiary care." PO – 3*

AI: Artificial intelligence, DM: Diabetes mellitus, DR: Diabetic retinopathy, IP: Industry partner, LE: Legal expert, PO: Program officer, VEGF: Vascular endothelial growth factor

information, protecting privacy, and reducing disclosure risks during data transfer. The main concern, however, is the risk of reidentification when proper anonymization is not achieved. Furthermore, AI companies are assertive in adhering to standard data collection, anonymization, and sharing principles, with a strong focus on ensuring the safety of patient data.

Data colonialism - Data colonialism refers to the appropriation of data from different sources, often needing adequate benefit or control for the data producers [39, 40]. However, a high volume of data sets is required and being created at scale to meet the growing need for training the machine learning algorithm. Noteworthy is the unaccounted accumulation of data without adhering to ethical considerations (ethical approval, patient consent).

In this context, the World Health Organization (WHO) highlights the necessity for ethical data governance frameworks [41] to combat data exploitation in the digital health sector, which often occurs without the consent of data owners [42]. Respondents indicated that AI companies prioritize collecting data from diverse environments to enhance diagnostic performance.

Ethical consideration and approval

The industry partner pointed out that institutional ethical approval is essential for research studies but not for developing AI algorithms. They noted that collecting patient data to develop machine-learning algorithms currently falls outside the purview of ethics, arguing that ethical boundaries will stifle technological innovation. The ophthalmologists argued that patient consent is unnecessary for algorithm development, claiming that obtaining consent will slow down the established clinical workflow.

Obtaining consent from individuals with low literacy is challenging because the language of consent forms often uses complex language, and locally translated versions can be difficult to understand. Ethical and legal experts note that AI developers sometimes bypass the consent process, claiming that obtaining multiple consents causes patients to experience *"consent fatigue"*.

Explainability

An industry partner shared insights into the trained model's validation process and asserted the importance of retrospective and prospective data for validation. Most participants accounted for their mixed viewpoints regarding collaboration with clinicians, integration of AI in clinical pathways, and explainability of AI decisions. The quotes (Table 2) reflect key themes around AI classification, clinical deployment, and explainability.

Challenges of AI implementation

The current regulatory landscape in India needs to adequately address the specific challenges and considerations related to AI applications in the healthcare industry. Many digital healthcare platforms function in ambiguous regulatory environments, often relying on existing laws that only partially pertain to them. This enables them to circumvent thorough testing and legal oversight. The other respondent acknowledged the importance of working in the AI regulatory space to ensure performance standards. (Table 3)

Accountability and liability

Current accountability and safety practices have yet to adapt to the potential patient harm arising from decisions made by AI-based clinical tools [43]. To address the theme of accountability and liability, the results pointed out that AI developers should not be held accountable for misdiagnoses made using their algorithms; instead, responsibility should lie with the healthcare providers using the tools. Additionally, regarding responsibility, the data inferred that. The program officer did not trust the AI-generated diagnosis, as treatment outcomes rely heavily on the accuracy of the diagnosis, which they questioned. (Table 3)

Discussion

Globally, 80% of AI-related studies, including those on stakeholder perspectives, originate from high-income countries, with limited contributions from low and middle-income countries (LMICs), like India, which are now emerging in this context [44, 45]. This gap highlights the underrepresentation of Global South perspectives in AI ethics discourse, despite the rapid adoption of AI in low- and middle-income country (LMIC) healthcare settings. This is the first study from India to explore stakeholder perspectives on the AI life cycle aligned with trustworthy AI principles [37]. Participants emphasized the importance of ethical frameworks, regulatory compliance, and transparency in informed consent and data privacy to protect personal data [24]. Concerns about image data collection and analysis biases highlight the need for greater transparency and fairness in AI algorithms to ensure equitable and responsible AI deployment [22]. It is crucial to protect patient data, address data ownership, and reduce bias in training data to safeguard privacy and ensure equitable diagnostic performance [22].

With the growing adoption of AI in healthcare, there is an increasing focus on developing ethical and trustworthy systems [46, 47]. While stakeholders acknowledged the importance of trustworthiness in AI design and deployment, the definition of ethics in the context of AI applications was less clear [46, 47]. Conventional informed consent forms were considered insufficient in

AI-based health research, as they often fail to capture the complexities of data aggregation, algorithmic processing, and secondary use focused on AI technologies [2]. Data protection, privacy, and unauthorized data sharing have become critical concerns in the evolving digital landscape [48]. Our study reflects the industry's prevailing *"take first, justify later"* approach to data collection, where the urgency to acquire data often overrides ethical concerns as critically articulated by a legal expert: *"If you need the data, take it; what's the big deal."* Data anonymization and de-identification are recommended to protect individuals' privacy [49]. However, stakeholders highlighted concerns about inadequate data anonymization practices, where retina fundus images could be easily re-identified by linking them with metadata, posing risks of privacy infringements. The stakeholders highlighted the need for clear responsibility in AI integration in healthcare, with defined moral accountability to ensure ethical deployment [50].

Institutional review boards should align their ethical evaluations to address AI-specific challenges, such as algorithmic bias and opacity. The processes for adverse event reporting, accountability, transparency in data collection, and effective risk management in health data research must be embedded in the ethical evaluations [51]. Stakeholders raised concerns about unscrupulous data collection practices in AI development, with one industry partner stating, *"We started conducting camps on our own," and "We collect the data first and see what can be done with it."* This approach, often called as data colonialism, prioritizes technological advancement without clear applications and heightens the risk of data privacy breaches and the exploitation of personal data [52, 53]. However, many ophthalmologists, program officers, industry experts, and patients in our study were unaware of data ownership, resulting in a limited understanding of the data fiduciary and principal obligations outlined in the Digital Personal Data Protection Act (DPDP) 2023 [54]. This Act in India establishes rules for handling digital personal data, aiming to strike a balance between individuals' right to privacy and the need to use data for lawful and legitimate purposes [54].

Respondents unanimously acknowledged the importance of ethical considerations in AI development. However, the lack of a common ethical approach led to differing views with ophthalmologists and industry partners perceiving it as unnecessary due to *"consent fatigue,"* and *"We cannot have separate consent for AI."* Legal and ethical experts regarded ethics as the cornerstone of the AI life cycle. Failing to inform patients about using their data and technology in care can undermine their autonomy, which depends on having sufficient information to make informed decisions [21, 50]. Empowering individuals to control their data requires supportive regulations, advanced technological standards, and public-private collaboration [48, 55]. Meanwhile, the industry partner stated, *"Do not stifle innovation with ethics,"* emphasizing the need for flexible ethical standards and an adaptable framework to keep pace with the evolving AI landscape.

The current AI ethics landscape shows a limited understanding of trustworthiness and should be validated through robustness and expert evaluation, despite [56]. Consistent with our findings, another study highlights uncertainty about trust in AI developers or deployment agencies. It also stresses that trustworthiness is often overused and lacks a clear and universally accepted definition [57]. Our study found that the lack of clear explanations for AI decisions is a key challenge. Although AI models demonstrate high accuracy in laboratory settings, they struggle in clinical settings due to limited real-world validation, biased training data, and opaque algorithms [58]. A clearly outlined framework should define responsibilities among developers, healthcare providers, regulatory authorities, and other stakeholders in designing, deploying, and using AI-powered tools [59].

A critical aspect of clinical application involves concerns about responsibility, accountability, and liability in the event of AI-induced diagnostic errors [20, 21, 58]. Stakeholders expressed mixed views on the accountability of AI in misdiagnosis. Industry partners held physicians responsible, while ophthalmologists placed the blame on developers. In contrast, legal and ethical experts advocated for a shared responsibility between clinicians and developers in the use of unvalidated AI systems. This aligns with the view expressed by one participant, who was skeptical about using AI for diagnosis: *"No one less than an ophthalmologist should be doing the DR grading."* Despite persistent fears that AI will replace doctors, its primary role is to augment and assist, rather than substitute for, clinical judgment [60–62]. Limited AI experience and understanding hinder its adoption in healthcare, emphasizing the need for contextual awareness and continuous learning before deployment in highly specialized, patient-centric settings [44, 63].

Regulatory concerns persist around the safety and efficacy of AI algorithms, especially when they do not align with existing care models [64]. Participants noted that *"Currently, no ethical or regulatory framework"* allows them to *"bypass rigorous testing and legal scrutiny."* Global AI governance initiatives emphasize the ethical adoption of AI, as outlined in frameworks established by the UN, OECD, and G20 [65, 66]. India's AI for India-specific regulatory framework fosters inclusive governance, while DEPA and NAIRP enhance data accessibility and collaboration [48, 67]. NITI Aayog's Responsible AI Guidelines (2020) further advocate for transparency, accountability, and fairness in the development and deployment of AI within the healthcare sector [68].

This article provides a comprehensive overview of the various issues, gaps, and challenges associated with the development and deployment of AI. However, it lacks an exploration of the social implications of AI in diagnosis and treatment, and the inclusion of only one ethics expert may have limited bioethics perspectives. Patient perspectives were not included, which limits the understanding of patient-centric ethical concerns. Future research with a larger and more diverse sample could facilitate the extrapolation of this theme.

Conclusion

The study appraised contributes a comprehensive understanding of potential challenges for AI development and deployment in healthcare. The most prominent themes that emerge are those of trust, privacy, consent, and data quality, each highlighting very complex problems. From this qualitative analysis, several concerns have been raised by stakeholders, primarily regarding the practicality of the data collection process and its quality, as well as the regulatory implications at multiple levels. This complexity surrounding trust, patient privacy and consent, ethics, and data integration presents avenues for further research. There is no question that AI holds great promise for improving healthcare outcomes. However, the integration of AI into healthcare systems is also loaded with trustworthiness and ethical considerations challenges. The findings of the study underscore the need to formulate strategies for the implementation of AI in healthcare to effectively harness AI's potential. For responsible adoption of AI technologies in the healthcare sector of India, the effort should involve reviewing and updating current organizational data and analytics governance and infrastructure, adequate training of healthcare providers in AI and data science, and fostering strategic collaborative partnerships to create an ethical AI ecosystem in the Indian healthcare sector. Besides the concentration of AI algorithms development for clinical diagnostics, there is a need for exploration of the ethical, legal, and governance dimensions. By embracing diverse perspectives, we can advance the development of trustworthy AI that upholds ethical principles and benefits society.

Abbreviations

AEC	Advanced eye centre
AI	Artificial intelligence
BE	Bioethics
COREQ	Consolidated Criteria for Reporting Qualitative
DR	Diabetic intelligence
GDPR	General Data Protection Regulation (GDPR)
IP	Industry partner
LE	Legal expert
MoHFW	Ministry of Health and Family Welfare
NHM	National Health Mission
NPCBVI	National Program for Control of Blindness and Visual Impairment
NPNCD	National Program for Control of Non-Communicable Diseases
Opth	Ophthalmologist
PO	Program officer
PGIMER	Postgraduate Institute of Medical Education and Research
VEGF	Vascular endothelial growth factor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12910-025-01265-7.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We extend our gratitude to all participants who contributed to this research.

Author contributions

AC and MD conceptualized the study. AC collected the data. AC, HR, GS, and DS conducted data analysis, with additional input from MD. AC and DS prepared the initial manuscript draft, which MD and KA reviewed. AC and DS revised the manuscript, and all authors read and approved the final version.

Funding

This research was funded by the National Institute of Transforming India (NITI) Aayog, India (PGI/IEC/2020/001342).

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

The study received approval from the Postgraduate Institute of Medical Education and Research (PGIMER) Institutional Ethics Committee (PGI/IEC/2020/001342) and followed the recommendations of the Declaration of Helsinki. Informed consent was obtained from all the study participants. The study has been registered under the Clinical Trials Registry India (CTRI/2022/10/046185).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 16 November 2024 / Accepted: 4 July 2025
Published online: 17 October 2025

References

1. Kurzweil R. The age of intelligent machines. Cambridge, Mass.: MIT Press. xiii; 1990. p. 565.
2. Mittelstadt BD, Floridi L. The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts. Sci Eng Ethics [Internet]. 2016;22(2):303–41. Available from: https://doi.org/10.1007/s11948-015-9652-2
3. Shin EK, Mahajan R, Akbilgic O, Shaban-Nejad A. Sociomarkers and biomarkers: predictive modeling in identifying pediatric asthma patients at risk of hospital revisits. NPJ Digit Med. 2018;1(1).
4. Mooney SJ, Pejaver V. Big data in public health: terminology, machine learning, and privacy. Annual Review of Public Health. Volume 39. Annual Reviews Inc.; 2018. pp. 95–112.
5. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. Vol. 22, BMC Public Health. BioMed Central Ltd; 2022.
6. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. Journal of the American Medical Informatics Association. Volume 27. Oxford University Press; 2020. pp. 491–7.

7. Ahuja AS. The impact of artificial intelligence in medicine on the future role of the physician. PeerJ. 2019;2019(10).
8. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell [Internet]. 2019;1(9):389–99. Available from: https://doi.org/10.1038/s42256-019-0088-2
9. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol [Internet]. 2005;8(1):19–32. Available from: https://doi.org/10.1080/1364557032000119616
10. Vujosevic S, Aldington SJ, Silva P, Hernández C, Scanlon P, Peto T et al. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol [Internet]. 2020;8(4):337–47. Available from: https://doi.org/10.1016/S2213-8587(19)30411-5
11. Ting DSW, Peng L, Varadarajan AV, Keane PA, Burlina PM, Chiang MF et al. Deep learning in ophthalmology: The technical and clinical considerations. Prog Retin Eye Res [Internet]. 2019;72:100759. Available from: https://www.sciencedirect.com/science/article/pii/S1350946218300909
12. Rajalakshmi R. The impact of artificial intelligence in screening for diabetic retinopathy in India. Eye. 2020;34(3):420–1.
13. Lim JI, Regillo CD, Sadda SVR, Ipp E, Bhaskaranand M, Ramachandra C et al. Artificial intelligence detection of diabetic retinopathy: subgroup comparison of the eyeart system with ophthalmologists' dilated examinations. Ophthalmol Sci. 2023;3(1).
14. Grzybowski A, Brona P, Lim G, Ruamviboonsuk P, Tan GSW, Abramoff M, et al. Artificial intelligence for diabetic retinopathy screening: a review. Eye (Basingstoke. Volume 34. Springer Nature; 2020. pp. 451–60.
15. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. J Med Ethics. 2021;47(5):329–35.
16. Fuhrman JD, Gorre N, Hu Q, Li H, El Naqa I, Giger ML. A review of explainable and interpretable AI with applications in COVID-19 imaging. Medical Physics. Volume 49. John Wiley and Sons Ltd; 2022. pp. 1–14.
17. Hasani N, Morris MA, Rhamim A, Summers RM, Jones E, Siegel E, et al. Trustworthy artificial intelligence in medical imaging. Vol. 17, PET clinics. W.B. Saunders; 2022. pp. 1–12.
18. Zhang Z, Genc Y, Wang D, Ahsen ME, Fan X. Effect of AI explanations on human perceptions of Patient-Facing AI-Powered healthcare systems. J Med Syst. 2021;45(6):64.
19. Williamson SM, Prybutok V. Balancing privacy and progress: A review of privacy challenges, systemic oversight, and patient perceptions in AI-Driven healthcare. Applied Sciences (Switzerland. Volume 14. Multidisciplinary Digital Publishing Institute (MDPI); 2024.
20. Abdullah YI, Schuman JS, Shabsigh R, Caplan A, Al-Aswad LA. Ethics of artificial intelligence in medicine and ophthalmology. Asia-Pacific Journal of Ophthalmology. Volume 10. Lippincott Williams and Wilkins; 2021. pp. 289–98.
21. Ursin F, Timmermann C, Orzechowski M, Steger F. Diagnosing diabetic retinopathy with artificial intelligence: what information should be included to ensure ethical informed consent?? Front Med (Lausanne). 2021;8.
22. Crew A, Reidy C, van der Westhuizen HM, Graham M. A narrative review of ethical issues in the use of artificial intelligence enabled diagnostics for diabetic retinopathy. Journal of evaluation in clinical practice. John Wiley and Sons Inc; 2024.
23. Couldry N, Mejias UA. Data Colonialism: Rethinking Big Data's Relation to the Contemporary Subject., Television. & New Media [Internet]. 2018;20(4):336–49. Available from: https://doi.org/10.1177/1527476418796632
24. Duggal M, Chauhan A, Kankaria A, Gupta V, Roy A, Verma P et al. Responsible Adoption of Cloud-Based Artificial Intelligence in Health Care: A Validation Case Study of Multiple Artificial Intelligence Algorithms for Diabetic Retinopathy Screening in Public Health Settings. Taylor and Francis. 2024 (In press). 2024.
25. Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. Transl Vis Sci Technol. 2020;9(2):1–6.
26. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application WHO guidance. Journal of the American Medical Informatics Association. Volume 27. Oxford University Press; 2020. pp. 491–7.
27. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: mapping the debate. Big Data Soc. 2016;3(2).
28. Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M et al. The ethics of AI in health care: A mapping review. Soc Sci Med [Internet]. 2020;260:113172. Available from: https://www.sciencedirect.com/science/article/pii/S0277953620303919
29. Coeckelbergh M. AI Ethics [Internet]. 2020. Available from: https://www.researchgate.net/publication/339103412
30. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C et al. Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res. 2017;19(11).
31. Tong A, Sainsbury P, Craig J. Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. Int J Qual Health Care. 2007;19(6):349–57.
32. Malterud K, Siersma VD, Guassora AD. Sample Size in Qualitative Interview Studies: Guided by Information Power. Qual Health Res [Internet]. 2015;26(13):1753–60. Available from: https://doi.org/10.1177/1049732315617444
33. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2):77–101.
34. Ministry of Health. & Family Welfare G of I. Indian Public Health Standards, Sub District Hospital and District Hospital, 2022.
35. Ministry of Health &. Family Welfare G of I. National Programme for Control of Blindness & Visual Impairment(NPCBVI) [Internet]. [cited 2024 Oct 26]. Available from: https://npcbvi.mohfw.gov.in/Home
36. Green J, Thorogood N. Qualitative Methods for Health Research [Internet]. London: SAGE Publications Ltd; 2018. (Introducing Qualitative Methods series). Available from: http://digital.casalini.it/9781526448804
37. Organization for Economic Co-operation and Development. Organization for Economic Co-operation and Development, Legal Instruments, Recommendation of the Council on Artificial Intelligence [Internet]. 2024 [cited 2024 Nov 15]. Available from: https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449
38. Lombard M, Snyder-Duch J, Bracken CC. Content Analysis in Mass Communication: Assessment and Reporting of Intercoder Reliability. Hum Commun Res [Internet]. 2002;28(4):587–604. Available from: https://doi.org/10.1111/j.1468-2958.2002.tb00826.x
39. Obia V. The costs of connection: how data is colonizing human life and appropriating it for capitalism. Inf Commun Soc [Internet]. 2023;26(9):1908–10. Available from: https://doi.org/10.1080/1369118X.2022.2062254
40. Nick Couldry, Ulises A. Mejias preface. The costs of connection: how data is colonizing human life and appropriating it for capitalism. Stanford University Press; 2019.
41. Thatcher J, O'Sullivan D, Mahmoudi D. Data colonialism through accumulation by dispossession: new metaphors for daily data. Environ Plan D. 2016;34(6):990–1006.
42. Sekalala S, Chatikobo T. Colonialism in the new digital health agenda. BMJ Global Health. Volume 9. BMJ Publishing Group; 2024.
43. Habli I, Lawton T, Porter Z. Artificial intelligence in health care: accountability and safety. Bull World Health Organ. 2020;98(4):251–6.
44. Vo V, Chen G, Aquino YSJ, Carter SM, Do QN, Woode ME. Multi-stakeholder preferences for the use of artificial intelligence in healthcare: A systematic review and thematic analysis. Volume 338. Social Science and Medicine. Elsevier Ltd; 2023.
45. Thenral M, Annamalai A. Challenges of building, deploying, and using AI-Enabled telepsychiatry platforms for clinical practice among urban indians: A qualitative study. Indian J Psychol Med. 2021;43(4):336–42.
46. European Union. On Artificial Intelligence-A European approach to excellence and trust White Paper on Artificial Intelligence A European approach to excellence and trust [Internet]. 2020. Available from: https://ec.europa.eu/commission/sites/beta-political/files/political-guidelines-next-commission_en.pdf
47. The European Union. Ethics Guidelines For Trustworthy AI A High-Level Expert Group on Artificial Intelligence [Internet]. Brussels. 2019. Available from: https://ec.europa.eu/digital-
48. NITI Aayog. Data Empowerment And Protection Architecture-Draft for Discussion Data Empowerment And Protection Architecture Draft for Discussion, August 2020 [Internet]. [cited 2024 Oct 26]. Available from: https://www.niti.gov.in/sites/default/files/2023-03/Data-Empowerment-and-Protection-Architecture-A-Secure-Consent-Based.pdf
49. Ministry of Electronics and Information Technology (MeitY) Government of India. Report Of Committee– D On Cyber Security, Safety, Legal And Ethical Issues [Internet]. 2018 [cited 2024 Oct 27]. Available from: https://www.meity.gov.in/artificial-intelligence-committees-reports
50. Tigard DW. Artificial Moral Responsibility: How We Can and Cannot Hold Machines Responsible. Cambridge Quarterly of Healthcare Ethics [Internet].

2021/06/10. 2021;30(3):435–47. Available from: https://www.cambridge.org/core/product/1FB82B1728EB240D059DF318B39FA13A
51. Vu T, Throne R. Current trends in AI ethics for software as a medical device (SaMD). IRB, human research protections, and data ethics for researchers. IGI Global Scientific Publishing; 2025. pp. 145–64.
52. European Parliament, European Council. General Data Protection Regulation. (2016). [Internet]. [cited 2024 Oct 26]. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/ PDF/?uri = CELEX:32016R0679&from = EN.
53. Approach Document. for India Part 1-Principles for Responsible AI. 2021.
54. Ministry of Electronics and Information Technology (MeitY). Digital Personal Data Protection Act (DPDP) 2023. 2023; Available from: https://www.meity.gov.in
55. Marda V. Artificial intelligence policy in india: A framework for engaging the limits of data-driven decision-making. Philosophical Trans Royal Soc A: Math Phys Eng Sci. 2018;376(2133).
56. Durán JM, Formanek N. Grounds for trust: essential epistemic opacity and computational reliabilism. Minds Mach (Dordr). 2018;28(4):645–66.
57. Reinhardt K. Trust and trustworthiness in AI ethics. AI Ethics. 2023;3(3):735–44.
58. Zafar S, Mahjoub H, Mehta N, Domalpally A, Channa R. Artificial intelligence algorithms in diabetic retinopathy screening. Current Diabetes Reports. Volume 22. Springer; 2022. pp. 267–74.
59. Gupta A, Raj A, Puri M, Gangrade J. Ethical considerations in the deployment of AI. Volume 45. Tuijin Jishu/Journal of Propulsion Technology; 2024.
60. Fogel AL, Kvedar JC. Artificial intelligence powers digital medicine. NPJ Digit Med [Internet]. 2018;1(1):5. Available from: https://doi.org/10.1038/s41746-017-0012-2
61. Jha S, Topol EJ. Adapting to Artificial Intelligence: Radiologists and Pathologists as Information Specialists. JAMA [Internet]. 2016;316(22):2353–4. Available from: https://doi.org/10.1001/jama.2016.17438
62. Mintz Y, Brodie R. Introduction to artificial intelligence in medicine. Minimally Invasive Therapy & Allied Technologies [Internet]. 2019;28(2):73–81. Available from: https://doi.org/10.1080/13645706.2019.1575882
63. Yang J, Blount Y, Amrollahi A. Artificial intelligence adoption in a professional service industry: A multiple case study. Technol Forecast Soc Change [Internet]. 2024;201:123251. Available from: https://www.sciencedirect.com/science/article/pii/S0040162524000477
64. Roberts H, Cowls J, Hine E, Morley J, Wang V, Taddeo M et al. Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. The Information Society [Internet]. 2023;39(2):79–97. Available from: https://doi.org/10.1080/01972243.2022.2124565
65. Roberts H, Hine E, Taddeo M, Floridi L. Global AI governance: barriers and pathways forward. Int Aff. 2024;100(3):1275–86.
66. UNESCO. Artificial Intelligence: UNESCO calls on all Governments to implement Global Ethical Framework without delay.
67. Ministry of Electronics and Information Technology (MeitY) G of I. Report Of Committee. - A On Platforms And Data On Artificial Intelligence [Internet]. 2018 [cited 2024 Oct 27]. Available from: https://www.meity.gov.in/artificial-intelligence-committees-reports
68. Adopting the Framework. A Use Case Approach on Facial Recognition Technology RESPONSIBLE AI #AIForAll.

Publisher's note