



Fairness of artificial intelligence in healthcare: review and recommendations

Daiju Ueda¹ · Taichi Kakinuma² · Shohei Fujita³ · Koji Kamagata⁴ · Yasutaka Fushimi⁵ · Rintaro Ito⁶ · Yusuke Matsui⁷ · Taiki Nozaki⁸ · Takeshi Nakaura⁹ · Noriyuki Fujima¹⁰ · Fuminari Tatsugami¹¹ · Masahiro Yanagawa¹² · Kenji Hirata¹³ · Akira Yamada¹⁴ · Takahiro Tsuboyama¹² · Mariko Kawamura⁶ · Tomoyuki Fujioka¹⁵ · Shinji Naganawa⁶

Received: 14 June 2023 / Accepted: 17 July 2023 / Published online: 4 August 2023
© The Author(s) 2023

Abstract

In this review, we address the issue of fairness in the clinical integration of artificial intelligence (AI) in the medical field. As the clinical adoption of deep learning algorithms, a subfield of AI, progresses, concerns have arisen regarding the impact of AI biases and discrimination on patient health. This review aims to provide a comprehensive overview of concerns associated with AI fairness; discuss strategies to mitigate AI biases; and emphasize the need for cooperation among physicians, AI researchers, AI developers, policymakers, and patients to ensure equitable AI integration. First, we define and introduce the concept of fairness in AI applications in healthcare and radiology, emphasizing the benefits and challenges of incorporating AI into clinical practice. Next, we delve into concerns regarding fairness in healthcare, addressing the various causes of biases in AI and potential concerns such as misdiagnosis, unequal access to treatment, and ethical considerations. We then outline strategies for addressing fairness, such as the importance of diverse and representative data and algorithm audits. Additionally, we discuss ethical and legal considerations such as data privacy, responsibility, accountability, transparency, and explainability in AI. Finally, we present the Fairness of Artificial Intelligence Recommendations in healthcare (FAIR) statement to offer best practices. Through these efforts, we aim to provide a foundation for discussing the responsible and equitable implementation and deployment of AI in healthcare.

Keywords Fairness · Bias · Artificial intelligence · Healthcare · Medicine · Review

✉ Daiju Ueda
ai.labo.ocu@gmail.com

¹ Department of Diagnostic and Interventional Radiology, Graduate School of Medicine, Osaka Metropolitan University, 1-4-3 Asahi-Machi, Abeno-ku, Osaka 545-8585, Japan

² STORIA Law Office, Chuo-ku, Kobe, Hyogo, Japan

³ Department of Radiology, University of Tokyo, Bunkyo-ku, Tokyo, Japan

⁴ Department of Radiology, Juntendo University Graduate School of Medicine, Bunkyo-ku, Tokyo, Japan

⁵ Department of Diagnostic Imaging and Nuclear Medicine, Kyoto University Graduate School of Medicine, Sakyo-ku, Kyoto, Japan

⁶ Department of Radiology, Nagoya University Graduate School of Medicine, Nagoya, Aichi, Japan

⁷ Department of Radiology, Faculty of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Kita-ku, Okayama, Japan

⁸ Department of Radiology, Keio University School of Medicine, Shinjuku-ku, Tokyo, Japan

⁹ Department of Diagnostic Radiology, Kumamoto University Graduate School of Medicine, Chuo-ku, Kumamoto, Japan

¹⁰ Department of Diagnostic and Interventional Radiology, Hokkaido University Hospital, Sapporo, Japan

¹¹ Department of Diagnostic Radiology, Hiroshima University, Minami-ku, Hiroshima, Japan

¹² Department of Radiology, Osaka University Graduate School of Medicine, Suita City, Osaka, Japan

¹³ Department of Diagnostic Imaging, Graduate School of Medicine, Hokkaido University, Kita-ku, Sapporo, Hokkaido, Japan

¹⁴ Department of Radiology, Shinshu University School of Medicine, Matsumoto, Nagano, Japan

¹⁵ Department of Diagnostic Radiology, Tokyo Medical and Dental University, Bunkyo-ku, Tokyo, Japan

Introduction

Fairness is one of the core principles of artificial intelligence (AI) ethics [1–3], and in recent years, there has been an increase in efforts focusing on fairness in AI, with a growing number of publications highlighting the need for improvement [4–9]. Various biases are involved in developing and applying AI, and these biases can affect fairness by erroneously skewing AI results [10]. In the medical field, bias and discrimination in AI have been studied in various domains [11, 12]. The World Medical Association's Geneva Declaration cites factors such as “age, disease or disability, creed, ethnic origin, gender, nationality, political affiliation, race, sexual orientation, social standing or any other factor” as examples that should not influence a physician's obligation to their patients [13]. Therefore, fairness concerns arise if AI does not perform adequately for specific patients.

AI research in radiology is an active field in healthcare owing to its affinity for imaging [14], with the number of AI-related publications and medical device certifications increasing annually [15–17]. One important reason for this is the global shortage of radiologists [18–21]. In particular, Japan has numerous publications on AI use in the field of radiology, including X-ray [22–25], mammography [26–30], US [31], CT [32–40], MRI [41–50], and PET [51, 52]. This surge in Radiological AI publications in Japan could be related to Japan having both the lowest number of radiologists per capita and the highest number of CT and MRI machines per capita among the Organization for Economic Co-operation and Development (OECD) countries [53]. Furthermore, owing to the coronavirus disease 2019 (COVID-19) pandemic, the number of COVID-19-related studies from Japan has increased [54–61], with a significant increase in such research focusing on AI [62–64].

As physicians in this era of AI clinical practice, we must be mindful of fairness concerns arising from AI bias in healthcare to provide better care to all patients. This review aims to provide a comprehensive overview of concerns related to AI fairness, discuss strategies to mitigate AI biases, and emphasize the need for collaboration among stakeholders to ensure equitable AI integration. In doing so, it lays the foundation for discussing the responsible and equitable implementation and deployment of AI in healthcare.

First, fairness in healthcare is discussed. We then discuss the issue of bias in AI systems used in healthcare. Next, we suggest strategies to reduce bias such as using diverse data, validating algorithms, and educating clinicians and patients regarding AI. We then discuss ethical and legal issues such as patient consent, data privacy, accountability, and the need for transparency in AI

systems. Collaboration is key in this context; therefore, we explore the roles of various stakeholders, including physicians, AI researchers, policymakers, regulatory authorities, patients, advocacy groups, and professional associations. We include the best practices recommended for fairness in AI and the areas where more research is needed. Finally, we conclude the paper with a summary of our main findings.

Fairness concerns in healthcare

Defining fairness in healthcare

Fairness in healthcare is a multidimensional concept that includes the equitable distribution of resources, opportunities, and outcomes among diverse patient populations [65]. The concept of fairness is based on the fundamental ethical principles of justice, beneficence, and non-maleficence. Healthcare systems must provide access to high-quality care for all individuals without discrimination. In the context of radiology, fairness in AI refers to the development and deployment of unbiased AI that provides accurate diagnoses and treatments for all patients regardless of their social status or ethnic differences. Achieving this fairness requires a comprehensive understanding of the potential causes of bias in AI and development of strategies to mitigate these biases [66].

Biases of AI in healthcare

Generally, biases in AI can emerge from two main sources: the data used for algorithm training (data bias) and inherent design or learning mechanisms of the algorithm itself (algorithmic bias). However, in the healthcare context, additional biases may arise because of the complex nature of human interactions and decision-making processes. These additional biases can be classified into two types: those that originate from AI–clinician interactions and those that originate from AI–patient interactions [11]. An overview of these biases is shown in Fig. 1.

Data biases

Data bias refers to problems arising from the collection and organization of data used in AI training that can potentially have harmful effects on fairness and accuracy [67]. The types of data biases include minority bias, missing data bias, informativeness bias, and training–serving skew [11]. Minority bias occurs when the number of protected group members in the dataset is insufficient for AI to learn accurate statistical patterns. This can lead to decreased

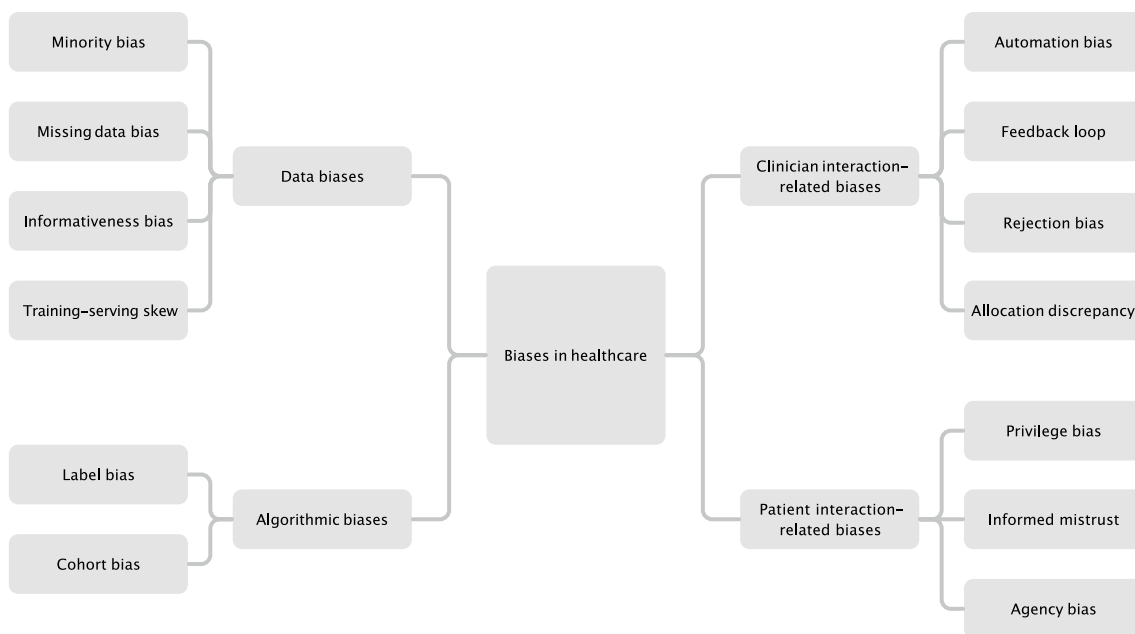


Fig. 1 Biases in healthcare of AI

performance and biased results when the algorithm is applied to these underrepresented groups. For example, many cardiovascular risk prediction algorithms have a history of being trained primarily on male patient data [68, 69]. This has led to an inaccurate risk assessment in female patients with different symptoms and risk factors. Missing data bias occurs when data from protected groups are missing nonrandomly, making it difficult for AI to generate accurate predictions. For example, if patients in contact isolation have fewer vital sign records than other patients, the algorithm may struggle to identify clinical deterioration. Informativeness bias occurs when the features used for detection are not as apparent for certain protected groups, lowering their informativeness when predictions are made. For example, identifying melanoma from images of patients with dark skin is more challenging than those with light skin [70, 71]. Training–serving skew refers to the mismatch between the data used for AI training and those used during deployment. This can arise from non-representative training data due to selection bias or from the deployment of the model on patients with a population prevalence different from that of the training data. In a study training AI to diagnose pneumonia from chest X-rays, the performance on unseen data from the institution where the training data were collected was significantly higher than its performance on data collected from external hospitals [72]. This common scenario means that estimations of AI performance based on internal test data

may overestimate its real-world performance on external data [73–75].

Algorithmic biases

Algorithmic bias refers to problems arising from the development and implementation of AI, which can negatively affect fairness and effectiveness. Even with representative data without data bias, AI can exhibit bias because of its inherent design or learning mechanisms. Algorithmic biases include label and cohort bias [11]. Label bias is a broad concept that includes test referral and interpretation bias. This occurs when AI training uses inconsistent labels, which may be influenced by healthcare disparities rather than universally accepted truths. This can lead to biased decision-making based on inaccurate or inconsistent information in the AI algorithms. For example, significant racial bias has been observed in commercially available algorithms used to predict patients' healthcare needs. Although several biases were affected, a major contributing factor to this algorithm's bias was its design, which used cost as a proxy for healthcare needs, leading to an underestimation of the needs of Black patients compared with those who are White with similar conditions [76]. Cohort bias occurs when AI is developed based on traditional or easily measurable groups without considering other potentially protected groups or varying levels of granularity. For example, mental health disorders have been underdiagnosed or misdiagnosed within lesbian, gay, bisexual, transgender, queer or questioning, intersex,

asexual, and other (LGBTQ+) populations [77]. One reason for this is that algorithms often do not take the granularity of the LGBTQ+ population into account and rely only on information about biological males and females. AI trained on such data may continue to overlook or misdiagnose mental health issues in these populations, potentially perpetuating existing disparities in mental healthcare.

Clinician interaction-related biases

When healthcare professionals interact with AI, biases related to interactions can occur, potentially affecting the algorithm's performance, fairness, and adoption [11]. One such bias is automation bias, which refers to the tendency to overly rely on AI when tasks are transferred from healthcare professionals to AI programs [78]. Overconfidence in algorithms can result in inappropriate actions based on inaccurate predictions. One study found that incorrect AI advice negatively affected radiologists' mammogram reading performance across all expertise levels. Inexperienced radiologists are more likely to follow incorrect AI suggestions [79]. Another bias related to interactions is the feedback loop [80]. This occurs when clinicians accept AI recommendations even if they are incorrect, leading the algorithm to relearn and perpetuate the same mistakes. Rejection bias refers to the conscious or unconscious desensitization to excessive alerts. Alert fatigue is a manifestation of this bias, as clinicians may ignore important alerts owing to an overwhelming number of false alarms [81, 82]. Finally, an allocation discrepancy occurs when the positive predictive values for protected groups are disproportionately low, leading the AI to withhold necessary resources, such as clinical attention or social services. Such resource allocation discrepancies can exacerbate disparities in care and outcomes among the affected groups.

Patient interaction-related biases

Biases related to interactions between patients and AI or the systems that incorporate them include privilege bias, informed mistrust, and agency bias [11]. Privilege bias occurs when certain populations cannot access AI in care settings or when these algorithms require technology or sensors that are not available to all populations [83]. This can lead to an unequal distribution of AI-driven healthcare benefits, potentially exacerbating existing healthcare disparities. Informed mistrust refers to the skepticism protected groups may have toward AI owing to historical exploitation and unethical practices in healthcare [84, 85]. This mistrust may lead these patients to avoid care or intentionally conceal information from clinicians or systems using AI. Agency bias arises when protected groups lack a voice in the development, use, and evaluation of AI [86]. These groups may

lack the access, resources, education, or political influence necessary to detect AI biases, voice concerns, and affect change. This lack of agency can result in AI inadequate at considering the needs and perspectives of protected groups, potentially leading to biases and disparities in healthcare outcomes.

Strategies to mitigate bias

Diverse and representative data

One of the most effective methods of mitigating AI biases is to ensure the use of diverse and representative datasets during AI development and training [67, 87]. This process entails carefully collecting and incorporating data from a wide range of sources to accurately reflect the demographics, characteristics, healthcare needs, and potential disparities in the target population. This diversity is not only critical for developing AI systems capable of catering to a multitude of patient requirements but also for fostering trust and confidence in AI-driven healthcare solutions. By incorporating data from various patient populations, age groups, disease stages, cultural and socioeconomic backgrounds, and healthcare settings, AI can learn to recognize, diagnose, and treat a broad spectrum of patient conditions with greater precision and contextual understanding. This comprehensive approach to data collection and curation prevents potential biases from occurring in AI systems, resulting in a reduction of disparities and promotion of equity in healthcare outcomes [76]. Furthermore, a diverse and representative dataset ensures that AI algorithms are rigorously tested across different scenarios, thereby enhancing their overall performance and utility. This enables healthcare providers to rely on AI-driven diagnostics and treatment recommendations, leading to improved patient care and reduced clinician workload.

Algorithm auditing and validation

Regular audits and AI validation play crucial roles in identifying and addressing potential biases and ensuring that AI systems remain fair, accurate, and effective in diverse healthcare settings. Independent audits by external experts or organizations can be conducted to evaluate the fairness, accuracy, and performance of AI, with adjustments made to the algorithms to correct identified biases [88]. The healthcare landscape is constantly changing; therefore, there is no guarantee that an AI algorithm with high performance will maintain its high performance in the future [89]. Validation studies are essential for verifying the effectiveness of AI in different patient populations and conditions [72]. The establishment of a dedicated department within hospitals for algorithm quality control has been advocated [90]. This department should be

responsible for continuously monitoring AI performance, identifying potential biases, and making the necessary updates to algorithms. This proactive approach to quality control would ensure that AI systems are held accountable and maintain their effectiveness in providing accurate and equitable care for all patients. Considering the growing prevalence of medical AI, practitioners must remain vigilant and evaluate key indicators, such as underdiagnosis rates and other health disparities, during the algorithm development process and after deployment. This ongoing evaluation will help identify and rectify emerging issues, ensuring that AI systems continue to serve patients effectively and equitably.

Education to both clinicians and patients

Educating clinicians and patients on the biases inherent in AI is crucial for fostering a shared understanding and promoting fairness in healthcare [1]. This educational process involves raising awareness of potential biases, sharing best practices to address them, and encouraging open discussions on the implications of AI in healthcare decision-making. Clinicians aware of AI biases can avoid overreliance on AI-generated results and make decisions based on more accurate information [91]. This increased awareness enables healthcare professionals to critically evaluate AI recommendations, weigh potential risks and benefits, and consider alternative sources of information when making patient care decisions. Additionally, clinicians can advocate for, and participate in, the development and evaluation of AI systems to ensure that their expertise and experience are incorporated into the models, further enhancing their accuracy and reliability. Patients who understand AI biases can make more informed and satisfactory decisions [92, 93]. By being aware of the potential limitations and biases of AI-generated recommendations, patients can engage in more meaningful conversations with their healthcare providers regarding treatment options and play a more active role in their care. This empowerment promotes patient-centered care and ensures that individual preferences, values, and circumstances are considered when making healthcare decisions. To foster a culture of continuous learning and improvement of AI, creating channels for feedback and collaboration among healthcare professionals and patients is essential. This can be achieved through workshops, conferences, online forums, or interdisciplinary collaborations that bring together diverse perspectives and experiences. By sharing knowledge, insights, and best practices, they can work together to identify and address biases and continuously refine AI systems to better serve the needs of all patients.

Ethical and legal considerations

Data privacy and security

Ensuring data privacy is an important ethical and legal consideration for AI fairness, as it has a significant impact on patient autonomy, trust in AI, and compliance with legal frameworks. Respecting patient autonomy and protecting confidential medical information is the foundation of ethical AI implementation, which can only be achieved by addressing important issues related to data privacy [94, 95]. One such issue is obtaining informed consent for data use [96, 97]. Patients must fully understand how their data are used, shared, and stored by AI. To achieve this, transparent communication regarding the purpose, risks, and benefits of data sharing is required to enable patients to make informed decisions regarding participating in AI-driven healthcare initiatives. Protecting data storage and transmission is an important aspect of data privacy [98]. Robust security measures, such as encryption and anonymization techniques, are required to protect patient data from unauthorized access, data breaches, and other cybersecurity threats. Moreover, strict access controls and audit mechanisms must be implemented to monitor and track data use, ensure accountability, and prevent data misuse. Compliance with privacy regulations, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States and the General Data Protection Regulation (GDPR) in the European Union, is essential for legally and ethically sound AI practice [99]. These regulations provide strict guidelines for the collection, storage, and processing of personal health information, and AI researchers and healthcare professionals must adhere to standardized data protection protocols. By addressing these challenges and ensuring data privacy, AI developers and healthcare professionals can foster trust in AI, maintain patient autonomy, and adhere to ethical and legal standards. This promotes the development and implementation of fair and equitable AI-driven healthcare solutions that respect the privacy and dignity of all patients.

Liability and accountability

To address the potential errors, harmful outcomes, and biases in the predictions generated by AI, clear guidelines for responsibility and accountability in healthcare AI must be established. This includes determining the roles and responsibilities of various stakeholders, such as physicians, AI developers, and healthcare institutions, in cases where misdiagnoses or other patient harm occur [100, 101]. Physicians should be responsible for verifying AI-generated diagnoses and integrating them into the clinical decision-making process. This may involve critically evaluating the AI outputs, considering them along

with other relevant clinical information, and making informed decisions regarding patient care. Conversely, AI developers have a responsibility to ensure the accuracy, reliability, and fairness of their algorithms. This includes addressing biases and continuously improving the algorithms based on feedback from the clinical community. Developers also need to provide clear guidance on the intended use and limitations of AI solutions, enabling physicians to make informed judgments regarding their application in patient care. Healthcare institutions play a crucial role in overseeing the integration of AI solutions into clinical workflows [90]. They must ensure that the necessary infrastructure, training, and support are available for the safe and effective use of AI. This includes developing policies and procedures for managing potential risks and harmful outcomes as well as monitoring and evaluating AI performance to ensure continuous quality improvement. A robust framework for accountability and responsibility enables AI stakeholders to address potential ethical and legal issues more effectively. As a result, trust in AI-driven healthcare solutions can increase, fostering responsible use and improving patient outcomes and overall care quality.

Transparency and explainability

Transparency and accountability are essential elements of ethical AI as they enable healthcare professionals and patients to understand the basis of AI-generated predictions and foster trust in AI [102]. To enhance the accountability of AI, developing interpretable algorithms, visualizing decision-making processes, and providing comprehensible explanations for AI predictions are important [103]. By improving transparency and accountability in AI in healthcare, both healthcare professionals and patients can be supported in making informed decisions, and ethical and legal concerns associated with the use of AI in healthcare can be addressed [104]. However, recognizing the limitations of explainability is important [104–107]. Even with saliency maps that visualize the areas of an image contributing to AI judgment, humans must decipher the meaning behind the explanation. When people favor meanings that confirm their beliefs or hypotheses, this is called confirmation bias. In other words, humans tend to interpret explanations positively even if the AI is not accurate or trustworthy. Recognizing the limitations of explainable AI is important for maintaining a realistic perspective of its potential benefits and drawbacks. By striking a balance among transparency, accountability, and understanding the limitations of explainable AI, healthcare professionals can address the ethical and legal concerns associated with the use of AI in healthcare.

Collaboration among stakeholders

Physicians, AI researchers, and AI developers

Collaboration among physicians, AI researchers, and AI developers is essential to address fairness concerns in AI [108, 110]. Physician participation can provide valuable domain expertise and insights for AI researchers. Recent AI, developed to utilize images [14, 111], and radiologists are particularly well matched. A cycle of improvement can be achieved through communicating expertise in the field with physicians and sharing their experience in using AI in actual medical practice. By working together, stakeholders can identify potential biases and develop effective strategies to mitigate them, ensuring that AI is fair, equitable, and effective. Additionally, empirical research on AI biases is often difficult for independent researchers to analyze as large-scale deployed algorithms are generally proprietary [112]. This makes it difficult to adequately assess bias, and active collaboration of not only AI researchers but also AI developers from companies is essential.

Policymakers and regulatory authorities

Policymakers and regulatory authorities play a crucial role in ensuring AI fairness through establishing comprehensive guidelines, standards, and regulations that govern the development and deployment of AI in healthcare [95, 113–115]. Through proactively shaping policies, they can promote the development of frameworks for AI design, training, and validation, ensuring that AI is built with fairness and inclusivity in mind. Fostering transparency and accountability in AI is also an important aspect of their responsibilities [5, 116]. Policymakers and regulatory authorities can implement requirements for AI developers to disclose their methodologies, data sources, and performance metrics, allowing for a better evaluation and comparison of AI. Furthermore, policymakers and regulatory authorities can allocate resources and funding towards AI innovation and research as well as towards addressing issues on fairness and equity in AI-driven healthcare. Through formulating policies that encourage the development of AI technologies focused on health equity, policymakers and regulators can minimize bias and ensure that all patients benefit from AI, regardless of their background or circumstances. This will contribute to a more equitable healthcare system, in which AI-driven solutions can improve patient outcomes and reduce disparities in access to high-quality care.

Patients and advocacy groups

Patients and advocacy groups serve a crucial function in advancing AI fairness as they contribute valuable insights

and firsthand experiences to the conversation. They can provide valuable insights into the needs and preferences of diverse patient populations, ensuring that AI addresses the specific challenges faced by various communities [86, 117, 118]. As patients directly affected by the AI output, they have a vested interest in identifying areas in which AI may be subject to potential biases and disparities in healthcare outcomes. By collaborating with patients and advocacy groups, physicians, AI researchers, and AI developers can gain a deeper understanding of the unique challenges and concerns faced by various patient populations and promote the development of more equitable and effective AI solutions tailored to individual needs [119]. This can also help build trust in AI-driven healthcare [84, 85]. By giving them a voice in the design, implementation, and evaluation of AI, organizations can demonstrate their commitment to patients by addressing their concerns and enhancing transparency.

Professional associations

Professional associations are pivotal in steering the development and implementation of AI-driven healthcare solutions, addressing ethical challenges, and promoting best practices. Establishing guidelines, standards, and ethical frameworks, fostering interdisciplinary collaborations, and facilitating open dialogue among all stakeholders will bridge this gap. Their unique position allows them to contribute to the development of fair and transparent policies and practices while ensuring that AI technologies are developed and deployed responsibly, equitably, and in the best interests of patients.

Recommendations and future directions

Best practices in healthcare for fairness of AI

To promote AI equity in healthcare and ensure fair and accurate care for all patients, developing a comprehensive strategy that addresses biases at multiple levels as well as ethical, legal, and practical concerns is essential. This approach should foster collaboration among key stakeholders to achieve equitable AI-driven healthcare solutions. We present the following recommendations, called the FAIR (Fairness of Artificial Intelligence Recommendations in healthcare) principles, which aim to ensure fair and equitable AI-driven healthcare solutions (see Table 1):

1. Ensuring diverse and representative data in AI development

Utilize diverse and representative data during AI development and training. This ensures that AI systems can better recognize, diagnose, and treat a wide range of

patient conditions, reduce disparities, and promote equity in healthcare outcomes.

2. Independent audits and validation of AI algorithms

Implement regular audits and validate AI algorithms by independent experts or organizations. This ensures objectivity and transparency in the evaluation process and helps identify potential biases, leading to necessary adjustments in the algorithms. Establish a dedicated system within hospitals for algorithm quality control to monitor AI performance continuously, identify potential biases, and update algorithms accordingly.

3. Education on AI biases for clinicians and patients

Educate clinicians and patients on the biases inherent in AI with ongoing education as needed. This will promote a shared understanding and encourage open discussions on the implications of AI in healthcare decision-making by creating channels for feedback and collaboration among healthcare professionals and patients. This can be achieved through workshops, conferences, online forums, and interdisciplinary collaborations.

4. Strengthening data privacy and security measures

Strengthen data privacy and security measures, ensuring compliance with existing legal frameworks such as HIPAA and GDPR. Develop transparent communication protocols to educate patients regarding data usage, storage, and sharing, allowing them to make informed decisions regarding participating in AI-driven healthcare initiatives.

5. Establishing liability and accountability frameworks

A robust framework for liability and accountability should be established, clearly defining the roles and responsibilities of physicians, AI developers, and healthcare institutions. Encourage continuous feedback and improvement of AI algorithms while maintaining transparency and providing guidance on the intended use and limitations of AI solutions.

6. Enhancing AI transparency and explainability

Enhance transparency and explainability in AI by developing interpretable algorithms, visualizing the model's decision-making processes, and providing explanations for AI predictions. Recognize the limitations of explainable AI and address potential biases to prevent overreliance on AI-generated outputs.

Table 1 FAIR (Fairness of Artificial Intelligence Recommendations in healthcare) statement

1. Ensuring diverse and representative data in AI development	Utilize diverse and representative data during AI development and training. This will ensure that AI systems can better recognize, diagnose, and treat a wide range of patient conditions, reducing disparities and promoting equity in healthcare outcomes
2. Independent audits and validation of AI algorithms	Implement regular audits and validation of AI algorithms by independent experts or organizations. This will ensure objectivity and transparency in the evaluation process and help identify potential biases, leading to necessary adjustments in the algorithms. Establish a dedicated system within hospitals for algorithm quality control to continuously monitor AI performance, identify potential biases, and update algorithms accordingly
3. Education on AI biases for clinicians and patients	Educate clinicians and patients on the biases inherent in AI with ongoing education as needed. This will promote a shared understanding and encourage open discussions on the implications of AI in healthcare decision-making by creating channels for feedback and collaboration among healthcare professionals and patients. This can be achieved through workshops, conferences, online forums, and interdisciplinary collaborations
4. Strengthening data privacy and security measures	Strengthen data privacy and security measures, ensuring compliance with existing legal frameworks such as HIPAA and GDPR. Develop transparent communication protocols to educate patients regarding data usage, storage, and sharing, allowing them to make informed decisions regarding participating in AI-driven healthcare initiatives
5. Establishing liability and accountability frameworks	Establish a robust framework for liability and accountability, clearly defining the roles and responsibilities of physicians, AI developers, and healthcare institutions. Encourage continuous feedback and improvement of AI algorithms, while maintaining transparency and providing guidance on AI solutions' intended use and limitations
6. Enhancing AI transparency and explainability	Enhance transparency and explainability in AI by developing interpretable algorithms, visualizing decision-making processes, and providing understandable explanations for AI predictions. Recognize the limitations of explainable AI and address potential biases to prevent overreliance on AI-generated outputs
7. Collaboration between physicians, AI researchers, and developers	Foster collaboration between physicians, AI researchers, and developers to share expertise, identify potential biases, and develop strategies to mitigate them. Encourage active participation of AI companies to support independent research on AI biases and improve algorithm fairness
8. Policymaker and regulatory authority involvement	Engage policymakers and regulatory authorities in developing comprehensive guidelines, standards, and regulations to ensure AI fairness, promote transparency and accountability, and allocate resources to support research and innovation in AI-driven healthcare
9. Patient and advocacy group participation in AI development and evaluation	Involve patients and advocacy groups in the design, implementation, and evaluation of AI solutions, giving them a voice in the decision-making process. Leverage their insights and experiences to address unique challenges and promote the development of equitable AI solutions tailored to individual needs
10. Professional association support	Professional associations help establish guidelines, standards, and ethical frameworks, and promote interdisciplinary collaborations and open discussions among all stakeholders. Their unique position enables them to aid in creating fair and transparent policies and practices

AI artificial intelligence, *HIPAA* the health insurance portability and accountability act, *GDPR* the general data protection regulation

7. Collaboration between physicians, AI researchers, and developers	accountability; and allocate resources to support research and innovation in AI-driven healthcare.
Foster collaboration among physicians, AI researchers, and developers to share expertise, identify potential biases, and develop strategies to mitigate them. Active participation of AI companies should be encouraged to support independent research on AI biases and improve algorithm fairness.	9. Patient and advocacy group participation in AI development and evaluation
8. Policymaker and regulatory authority involvement	Involve patients and advocacy groups in the design, implementation, and evaluation of AI solutions, giving them a voice in the decision-making process. Leverage their insights and experiences to address unique challenges and promote the development of equitable AI solutions tailored to individual needs.
Engage policymakers and regulatory authorities in developing comprehensive guidelines, standards, and regulations to ensure AI fairness; promote transparency and	10. Professional association support

Professional associations help establish guidelines, standards, and ethical frameworks, and promote interdisciplinary collaborations and open discussions among all stakeholders. Their unique position enables them to create fair and transparent policies and practices.

By implementing these recommendations and addressing biases in data and algorithms, stakeholders in the AI-driven healthcare sector can foster trust, transparency, and inclusivity. This will ensure that AI technologies are developed and deployed ethically, responsibly, and equitably for the benefit of all patients regardless of their differences. Ultimately, this approach contributes to a more equitable healthcare system and improves patient outcomes.

Research gaps and future work

Several research gaps and opportunities for future research to address concerns regarding AI bias and fairness exist. Randomized controlled trials should be conducted to explore the potential of AI in improving patient care and outcomes. These trials should include diverse populations and AI that are tailored to the specific needs of different demographic groups. The long-term impact of AI adoption in healthcare on patient treatment, outcomes, and physician workload should be investigated. The models should be monitored regularly to address biases that may emerge over time [89]. Developing new technologies for explainability and transparency is necessary to enable healthcare professionals and patients to better understand AI-generated predictions, foster trust in AI, and ensure its ethical deployment.

Conclusion

In this review, we first defined fairness in AI in the healthcare domain, introduced various biases with examples and potential countermeasures, and emphasized the importance of collaborating with stakeholders. Subsequently, we discussed important ethical and legal issues. As a result, we summarized the best practices into the FAIR statement. This includes preparing diverse and representative data, continuously validating AI, educating physicians and patients, and emphasizing the importance of interdisciplinary collaboration. Although implementing each best practice is difficult, these efforts have become increasingly important as AI integration advances in the medical field. Furthermore, AI technology is still evolving, and the situation is constantly changing, with new challenges emerging one after another. The emergence of tools like Chat Generative Pre-trained Transformer (ChatGPT) is expected to greatly change white-collar jobs, and physicians are no exception

[120–122]. We are currently in an era in which physicians require flexible thinking and the ability to respond quickly to new technologies.

Since its inception, AI has influenced several aspects of modern society, leading to notable advancements. The medical field has not remained untouched by this wave of change, with radiology particularly poised to harness the power of AI [123]. Given this unique position, the radiology community has a vital responsibility to share its experience in actively integrating AI into medicine [124–126], providing invaluable guidance and insight for other medical specialties. As pioneers in the implementation of AI, radiologists should champion AI equity in healthcare. Our early experience navigating the complex landscape of AI adoption and overcoming the challenges associated with its deployment can serve as a roadmap for other medical professionals. By doing so, we can ensure that AI benefits all patients, regardless of their backgrounds, and contributes to the greater good of society.

Acknowledgements We extend our gratitude to ChatGPT, an AI language model developed by OpenAI, for its invaluable assistance in providing insights, suggestions, and guidance throughout this study.

Funding There was no funding for this review.

Declarations

Conflict of interest There are no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* 2016;3:2053951716679679. SAGE Publications Ltd. <https://doi.org/10.1177/2053951716679679>
2. Jobin A, Ienca M, Vayena E. Artificial intelligence: the global landscape of ethics guidelines. *Nat Mach Intell.* 2019;1:389–99. <https://doi.org/10.1038/s42256-019-0088-2>.
3. Tsamados A, Aggarwal N, Cowlis J, Morley J, Roberts H, Taddeo M, et al. The ethics of algorithms: key problems and solutions. *AI & Soc.* 2022;37:215–30. <https://doi.org/10.1007/s00146-021-01154-8>.

4. Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S. Human decisions and machine predictions. *Q J Econ*. 2018;133:237–93. <https://doi.org/10.1093/qje/qjx032>.
5. Edwards V. Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke Law Technol Rev*. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/dltr16§ion=3.
6. Binns R. Fairness in machine learning: Lessons from political philosophy. In: Friedler SA, Wilson C, editors. *Proceedings of the 1st conference on fairness, accountability and transparency*. PMLR; 2018. p. 149–59
7. Selbst AD, Boyd D, Friedler SA, Venkatasubramanian S, Vertesi J. Fairness and abstraction in sociotechnical systems. In: *Proceedings of the conference on fairness, accountability, and transparency*. New York, USA. Association for Computing Machinery; 2019. <https://doi.org/10.1145/3287560.3287598>
8. Wong P-H. Democratizing algorithmic fairness. *Philos Technol*. 2020;33:225–44. <https://doi.org/10.1007/s13347-019-00355-w>.
9. Abebe R, Barocas S, Kleinberg J, Levy K, Raghavan M, Robinson DG. Roles for computing in social change. In: *Proceedings of the 2020 conference on fairness, accountability, and transparency*. New York, USA. Association for Computing Machinery; 2020. <https://doi.org/10.1145/3351095.3372871>
10. Bærøe K, Gundersen T, Henden E, Rommetveit K. Can medical algorithms be fair? Three ethical quandaries and one dilemma. *BMJ Health Care Inform*. 2022. <https://doi.org/10.1136/bmjhci-2021-100445>.
11. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring fairness in machine learning to advance health equity. *Ann Intern Med*. 2018;169:866–72. <https://doi.org/10.7326/M18-1990>.
12. Mehrabi N, Morstatter F, Saxena N, Lerman K, Galstyan A. A survey on bias and fairness in machine learning. *ACM Comput Surv*. 2021;54:1–35. <https://doi.org/10.1145/3457607>.
13. World Medical Association. Declaration of Geneva. World Medical Association; 1983
14. Ueda D, Shimazaki A, Miki Y. Technical and clinical overview of deep learning in radiology. *Jpn J Radiol*. 2019;37:15–33. <https://doi.org/10.1007/s11604-018-0795-3>.
15. Yuba M, Iwasaki K. Systematic analysis of the test design and performance of AI/ML-based medical devices approved for triage/detection/diagnosis in the USA and Japan. *Sci Rep*. 2022;12:16874. <https://doi.org/10.1038/s41598-022-21426-7>.
16. Zhu S, Gilbert M, Chetty I. The SF landscape of FDA-approved artificial intelligence/machine learning-enabled medical devices: an analysis of the characteristics and intended use. *Int J Med Inform*. 2021. <https://doi.org/10.1016/j.ijmedinf.2022.104828>.
17. Kelly BS, Judge C, Bollard SM, Clifford SM, Healy GM, Aziz A, et al. Radiology artificial intelligence: a systematic review and evaluation of methods (RAISE). *Eur Radiol*. 2022;32:7998–8007. <https://doi.org/10.1007/s00330-022-08784-6>.
18. Rimmer A. Radiologist shortage leaves patient care at risk, warns royal college. *BMJ*. 2017. <https://doi.org/10.1136/bmj.j4683>.
19. Nakajima Y, Yamada K, Imamura K, Kobayashi K. Radiologist supply and workload: international comparison—working group of Japanese college of radiology. *Radiat Med*. 2008;26:455–65. <https://doi.org/10.1007/s11604-008-0259-2>.
20. Gourad EE. UK radiologist staffing crisis reaches critical levels. *Lancet Oncol*. 2017. [https://doi.org/10.1016/S1470-2045\(17\)30806-9](https://doi.org/10.1016/S1470-2045(17)30806-9).
21. Mollura DJ, Culp MP, Pollack E, Battino G, Scheel JR, Mango VL, et al. Artificial intelligence in low- and middle-income countries: innovating global health radiology. *Radiology*. 2020;297:513–20. <https://doi.org/10.1148/radiol.2020201434>.
22. Toda N, Hashimoto M, Iwabuchi Y, Nagasaka M, Takeshita R, Yamada M, et al. Validation of deep learning-based computer-aided detection software use for interpretation of pulmonary abnormalities on chest radiographs and examination of factors that influence readers' performance and final diagnosis. *Jpn J Radiol*. 2023;41:38–44. <https://doi.org/10.1007/s11604-022-01330-w>.
23. Ueda D, Ehara S, Yamamoto A, Iwata S, Abo K, Walston SL, et al. Development and validation of artificial intelligence-based method for diagnosis of mitral regurgitation from chest radiographs. *Radiol Artif Intell*. 2022. <https://doi.org/10.1148/ryai.210221>.
24. Ueda D, Yamamoto A, Ehara S, Iwata S, Abo K, Walston SL, et al. Artificial intelligence-based detection of aortic stenosis from chest radiographs. *Eur Heart J Digit Health*. 2022;3:20–8. <https://doi.org/10.1093/ehjdh/ztab102>.
25. Matsumoto T, Ehara S, Walston SL, Mitsuyama Y, Miki Y, Ueda D. Artificial intelligence-based detection of atrial fibrillation from chest radiographs. *Eur Radiol*. 2022;32:5890–7. <https://doi.org/10.1007/s00330-022-08752-0>.
26. Ueda D, Yamamoto A, Takashima T, Onoda N, Noda S, Kashiwagi S, et al. Visualizing “featureless” regions on mammograms classified as invasive ductal carcinomas by a deep learning algorithm: the promise of AI support in radiology. *Jpn J Radiol*. 2021;39:333–40. <https://doi.org/10.1007/s11604-020-01070-9>.
27. Uematsu T, Nakashima K, Harada TL, Nasu H, Igarashi T. Comparisons between artificial intelligence computer-aided detection synthesized mammograms and digital mammograms when used alone and in combination with tomosynthesis images in a virtual screening setting. *Jpn J Radiol*. 2023;41:63–70. <https://doi.org/10.1007/s11604-022-01327-5>.
28. Ueda D, Yamamoto A, Onoda N, Takashima T, Noda S, Kashiwagi S, et al. Development and validation of a deep learning model for detection of breast cancers in mammography from multi-institutional datasets. *PLoS ONE*. 2022. <https://doi.org/10.1371/journal.pone.0265751>.
29. Honjo T, Ueda D, Katayama Y, Shimazaki A, Jogo A, Kageyama K, et al. Visual and quantitative evaluation of microcalcifications in mammograms with deep learning-based super-resolution. *Eur J Radiol*. 2022. <https://doi.org/10.1016/j.ejrad.2022.110433>.
30. Ueda D, Yamamoto A, Takashima T, Onoda N, Noda S, Kashiwagi S, et al. Training, validation, and test of deep learning models for classification of receptor expressions in breast cancers from mammograms. *JCO Precis Oncol*. 2021;5:543–51. <https://doi.org/10.1200/PO.20.00176>.
31. Ozaki J, Fujioka T, Yamaga E, Hayashi A, Kujiraoka Y, Imokawa T, et al. Deep learning method with a convolutional neural network for image classification of normal and metastatic axillary lymph nodes on breast ultrasonography. *Jpn J Radiol*. 2022;40:814–22. <https://doi.org/10.1007/s11604-022-01261-6>.
32. Ichikawa Y, Kanii Y, Yamazaki A, Nagasawa N, Nagata M, Ishida M, et al. Deep learning image reconstruction for improvement of image quality of abdominal computed tomography: comparison with hybrid iterative reconstruction. *Jpn J Radiol*. 2021;39:598–604. <https://doi.org/10.1007/s11604-021-01089-6>.
33. Nakai H, Fujimoto K, Yamashita R, Sato T, Someya Y, Taura K, et al. Convolutional neural network for classifying primary liver cancer based on triple-phase CT and tumor marker information: a pilot study. *Jpn J Radiol*. 2021;39:690–702. <https://doi.org/10.1007/s11604-021-01106-8>.
34. Okuma T, Hamamoto S, Maebayashi T, Taniguchi A, Hirakawa K, Matsushita S, et al. Quantitative evaluation of COVID-19 pneumonia severity by CT pneumonia analysis algorithm using deep learning technology and blood test results. *Jpn J Radiol*. 2021;39:956–65. <https://doi.org/10.1007/s11604-021-01134-4>.

35. Kitahara H, Nagatani Y, Otani H, Nakayama R, Kida Y, Sonoda A, et al. A novel strategy to develop deep learning for image super-resolution using original ultra-high-resolution computed tomography images of lung as training dataset. *Jpn J Radiol.* 2022;40:38–47. <https://doi.org/10.1007/s11604-021-01184-8>.
36. Kaga T, Noda Y, Mori T, Kawai N, Miyoshi T, Hyodo F, et al. Unenhanced abdominal low-dose CT reconstructed with deep learning-based image reconstruction: Image quality and anatomical structure depiction. *Jpn J Radiol.* 2022;40:703–11. <https://doi.org/10.1007/s11604-022-01259-0>.
37. Ohno Y, Aoyagi K, Arakita K, Doi Y, Kondo M, Banno S, et al. Newly developed artificial intelligence algorithm for COVID-19 pneumonia: Utility of quantitative CT texture analysis for prediction of favipiravir treatment effect. *Jpn J Radiol.* 2022;40:800–13. <https://doi.org/10.1007/s11604-022-01270-5>.
38. Matsukiyo R, Ohno Y, Matsuyama T, Nagata H, Kimata H, Ito Y, et al. Deep learning-based and hybrid-type iterative reconstructions for CT: Comparison of capability for quantitative and qualitative image quality improvements and small vessel evaluation at dynamic CE-abdominal CT with ultra-high and standard resolutions. *Jpn J Radiol.* 2021;39:186–97. <https://doi.org/10.1007/s11604-020-01045-w>.
39. Koretsune Y, Sone M, Sugawara S, Wakatsuki Y, Ishihara T, Hattori C, et al. Validation of a convolutional neural network for the automated creation of curved planar reconstruction images along the main pancreatic duct. *Jpn J Radiol.* 2023;41:228–34. <https://doi.org/10.1007/s11604-022-01339-1>.
40. Anai K, Hayashida Y, Ueda I, Hozuki E, Yoshimatsu Y, Tsukamoto J, et al. The effect of CT texture-based analysis using machine learning approaches on radiologists' performance in differentiating focal-type autoimmune pancreatitis and pancreatic duct carcinoma. *Jpn J Radiol.* 2022;40:1156–65. <https://doi.org/10.1007/s11604-022-01298-7>.
41. Cay N, Mendi BAR, Batur H, Erdogan F. Discrimination of lipoma from atypical lipomatous tumor/well-differentiated liposarcoma using magnetic resonance imaging radiomics combined with machine learning. *Jpn J Radiol.* 2022;40:951–60. <https://doi.org/10.1007/s11604-022-01278-x>.
42. Wong LM, Ai QYH, Mo FKF, Poon DMC, King AD. Convolutional neural network in nasopharyngeal carcinoma: How good is automatic delineation for primary tumor on a non-contrast-enhanced fat-suppressed T2-weighted MRI? *Jpn J Radiol.* 2021;39:571–9. <https://doi.org/10.1007/s11604-021-01092-x>.
43. Yasaka K, Akai H, Sugawara H, Tajima T, Akahane M, Yoshioka N, et al. Impact of deep learning reconstruction on intracranial 1.5 T magnetic resonance angiography. *Jpn J Radiol.* 2022. <https://doi.org/10.1007/s11604-021-01225-2>.
44. Nomura Y, Hanaoka S, Nakao T, Hayashi N, Yoshikawa T, Miki S, et al. Performance changes due to differences in training data for cerebral aneurysm detection in head MR angiography images. *Jpn J Radiol.* 2021;39:1039–48. <https://doi.org/10.1007/s11604-021-01153-1>.
45. Ishihara M, Shiiba M, Maruno H, Kato M, Ohmoto-Sekine Y, Antoine C, et al. Detection of intracranial aneurysms using deep learning-based CAD system: usefulness of the scores of CNN's final layer for distinguishing between aneurysm and infundibular dilatation. *Jpn J Radiol.* 2023;41:131–41. <https://doi.org/10.1007/s11604-022-01341-7>.
46. Miki S, Nakao T, Nomura Y, Okimoto N, Nyunoya K, Nakamura Y, et al. Computer-aided detection of cerebral aneurysms with magnetic resonance angiography: usefulness of volume rendering to display lesion candidates. *Jpn J Radiol.* 2021;39:652–8. <https://doi.org/10.1007/s11604-021-01099-4>.
47. Nakaura T, Kobayashi N, Yoshida N, Shiraishi K, Uetani H, Nagayama Y, et al. Update on the use of artificial intelligence in hepatobiliary MR imaging. *Magn Reson Med Sci.* 2023;22:147–56. <https://doi.org/10.2463/mrms.rev.2022-0102>.
48. Naganawa S, Ito R, Kawai H, Kawamura M, Taoka T, Sakai M, et al. MR imaging of endolymphatic hydrops in five minutes. *Magn Reson Med Sci.* 2022;21:401–5. <https://doi.org/10.2463/mrms.ici.2021-0022>.
49. Kabasawa H, Kiryu S. Pulse sequences and reconstruction in Fast MR imaging of the liver. *Magn Reson Med Sci.* 2023;22:176–90. <https://doi.org/10.2463/mrms.rev.2022-0114>.
50. Iwamura M, Ide S, Sato K, Kakuta A, Tatsuo S, Nozaki A, et al. Thin-slice two-dimensional T2-weighted imaging with deep learning-based reconstruction: Improved lesion detection in the brain of patients with multiple sclerosis. *Magn Reson Med Sci.* 2023. <https://doi.org/10.2463/mrms.mp.2022-0112>.
51. Nai YH, Loi HY, O'Doherty S, Tan TH, Reilhac A. Comparison of the performances of machine learning and deep learning in improving the quality of low dose lung cancer PET images. *Jpn J Radiol.* 2022;40:1290–9. <https://doi.org/10.1007/s11604-022-01311-z>.
52. Nakao T, Hanaoka S, Nomura Y, Hayashi N, Abe O. Anomaly detection in chest 18F-FDG PET/CT by bayesian deep learning. *Jpn J Radiol.* 2022;40:730–9. <https://doi.org/10.1007/s11604-022-01249-2>.
53. Kumamaru KK, Machitori A, Koba R, Ijichi S, Nakajima Y, Aoki S. Global and Japanese regional variations in radiologist potential workload for computed tomography and magnetic resonance imaging examinations. *Jpn J Radiol.* 2018;36:273–81. <https://doi.org/10.1007/s11604-018-0724-5>.
54. Cozzi D, Cavigli E, Moroni C, Smorchkova O, Zantonelli G, Pradella S, et al. Ground-glass opacity (GGO): a review of the differential diagnosis in the era of COVID-19. *Jpn J Radiol.* 2021;39:721–32. <https://doi.org/10.1007/s11604-021-01120-w>.
55. Aoki R, Iwasawa T, Hagiwara E, Komatsu S, Utsunomiya D, Ogura T. Pulmonary vascular enlargement and lesion extent on computed tomography are correlated with COVID-19 disease severity. *Jpn J Radiol.* 2021;39:451–8. <https://doi.org/10.1007/s11604-020-01085-2>.
56. Zhu QQ, Gong T, Huang GQ, Niu ZF, Yue T, Xu FY, et al. Pulmonary artery trunk enlargement on admission as a predictor of mortality in in-hospital patients with COVID-19. *Jpn J Radiol.* 2021;39:589–97. <https://doi.org/10.1007/s11604-021-01094-9>.
57. Fukuda A, Yanagawa N, Sekiya N, Ohyama K, Yomota M, Inui T, et al. An analysis of the radiological factors associated with respiratory failure in COVID-19 pneumonia and the CT features among different age categories. *Jpn J Radiol.* 2021;39:783–90. <https://doi.org/10.1007/s11604-021-01118-4>.
58. Özer H, Kılınçer A, Uysal E, Yormaz B, Cebeci H, Durmaz MS, et al. Diagnostic performance of radiological society of North America structured reporting language for chest computed tomography findings in patients with COVID-19. *Jpn J Radiol.* 2021;39:877–88. <https://doi.org/10.1007/s11604-021-01128-2>.
59. Zhuang Y, Lin L, Xu X, Xia T, Yu H, Fu G, et al. Dynamic changes on chest CT of COVID-19 patients with solitary pulmonary lesion in initial CT. *Jpn J Radiol.* 2021;39:32–9. <https://doi.org/10.1007/s11604-020-01037-w>.
60. Kanayama A, Tsuchihashi Y, Otomi Y, Enomoto H, Arima Y, Takahashi T, et al. Association of severe COVID-19 outcomes with radiological scoring and cardiomegaly: Findings from the COVID-19 inpatients database, Japan. *Jpn J Radiol.* 2022. <https://doi.org/10.1007/s11604-022-01300-2>.
61. Inui S, Fujikawa A, Gonoï W, Kawano S, Sakurai K, Uchida Y, et al. Comparison of CT findings of coronavirus disease 2019 (COVID-19) pneumonia caused by different major variants. *Jpn J Radiol.* 2022;40:1246–56. <https://doi.org/10.1007/s11604-022-01301-1>.

62. Walston SL, Matsumoto T, Miki Y, Ueda D. Artificial intelligence-based model for COVID-19 prognosis incorporating chest radiographs and clinical data; a retrospective model development and validation study. *Br J Radiol*. 2022;95:20220058. <https://doi.org/10.1259/bjr.20220058>.
63. Matsumoto T, Walston SL, Walston M, Kabata D, Miki Y, Shiba M, et al. Deep learning-based time-to-death prediction model for COVID-19 patients using clinical data and chest radiographs. *J Digit Imaging*. 2023;36:178–88. <https://doi.org/10.1007/s10278-022-00691-y>.
64. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of Covid-19: Systematic review and critical appraisal. *BMJ*. 2020;369:m1328. <https://doi.org/10.1136/bmj.m1328>.
65. Marmot M, Bell R. Fair society, healthy lives. *Public Health*. 2012;126(Suppl 1):S4–10. <https://doi.org/10.1016/j.puhe.2012.05.014>.
66. Ricci Lara MA, Echeveste R, Ferrante E. Addressing fairness in artificial intelligence for medical imaging. *Nat Commun*. 2022;13:4581. <https://doi.org/10.1038/s41467-022-32186-3>.
67. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H 3rd, Iii HD, et al. Datasheets for datasets. *Commun ACM*. 2021. New York: Association for Computing Machinery;64:86–92. <https://doi.org/10.1145/3458723>
68. Wenger NK. Women and coronary heart disease: A century after Herrick: Understudied, underdiagnosed, and undertreated. *Circulation*. 2012;126:604–11. <https://doi.org/10.1161/CIRCULATIONAHA.111.086892>.
69. Appelman Y, van Rijn BB, Ten Haaf ME, Boersma E, Peters SAE. Sex differences in cardiovascular risk factors and disease prevention. *Atherosclerosis*. 2015;241:211–8. <https://doi.org/10.1016/j.atherosclerosis.2015.01.027>.
70. Adamson AS, Smith A. Machine learning and health care disparities in dermatology. *JAMA Dermatol*. 2018;154:1247–8. <https://doi.org/10.1001/jamadermatol.2018.2348>.
71. Navarrete-Dechent C, Dusza SW, Liopyris K, Marghoob AA, Halpern AC, Marchetti MA. Automated dermatological diagnosis: hype or reality? *J Invest Dermatol*. 2018;138:2277–9. <https://doi.org/10.1016/j.jid.2018.04.040>.
72. Zech JR, Badgeley MA, Liu M, Costa AB, Titano JJ, Oermann EK. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study. *PLOS Med*. 2018. <https://doi.org/10.1371/journal.pmed.1002683>.
73. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170:51–8. <https://doi.org/10.7326/M18-1376>.
74. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286:800–9. <https://doi.org/10.1148/radiol.2017171920>.
75. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing radiology research on artificial intelligence: a brief guide for authors, reviewers, and readers—from the radiology editorial board. *Radiology*. 2020;294:487–9. <https://doi.org/10.1148/radiol.2019192515>.
76. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–53. <https://doi.org/10.1126/science.aax2342>.
77. Meyer IH. Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychol Bull*. 2003;129:674–97. <https://doi.org/10.1037/0033-2909.129.5.674>.
78. Anderson M, Anderson SL. How should AI be developed, validated, and implemented in patient care? *AMA J Ethics*. *Am Med Assoc*. 2019;21:E125–130. <https://doi.org/10.1001/amajethics.2019.125>.
79. Dratsch T, Chen X, Rezazade Mehrizi M, Kloeckner R, Mähringer-Kunz A, Püsken M, et al. Automation bias in mammography: The impact of artificial intelligence BI-RADS suggestions on reader performance. *Radiology*. 2023. <https://doi.org/10.1148/radiol.222176>.
80. Walsh CG, Chaudhry B, Dua P, Goodman KW, Kaplan B, Kavuluru R, et al. Stigma, biomarkers, and algorithmic bias: recommendations for precision behavioral health with artificial intelligence. *JAMIA Open*. 2020;3:9–15. <https://doi.org/10.1093/jamiaopen/ooz054>.
81. Lehman CD, Wellman RD, Buist DSM, Kerlikowske K, Tosteson ANA, Miglioretti DL, et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA Intern Med*. 2015;175:1828–37. <https://doi.org/10.1001/jamainternmed.2015.5231>.
82. Phansalkar S, van der Sijs H, Tucker AD, Desai AA, Bell DS, Teich JM, et al. Drug—drug interactions that should be non-interruptive in order to reduce alert fatigue in electronic health records. *J Am Med Inform Assoc Oxford*. 2012. Academic Press;20:489–93
83. Fiscella K, Williams DR. Health disparities based on socioeconomic inequities: implications for urban health care. *Acad Med*. 2004;79:1139–47. <https://doi.org/10.1097/00001888-200412000-00004>.
84. Gamble VN. Under the shadow of Tuskegee: African Americans and health care. *Am J Public Health*. 1997;87:1773–8. <https://doi.org/10.2105/ajph.87.11.1773>.
85. Boulware LE, Cooper LA, Ratner LE, LaVeist TA, Powe NR. Race and trust in the health care system. *Public Health Rep*. 2003;118:358–65. <https://doi.org/10.1093/phr/118.4.358>.
86. Minkler M, Wallerstein N. Community-based participatory research for health: From process to outcomes. John Wiley & Sons; 2011
87. Haibe-Kains B, Adam GA, Hosny A, Khodakarami F, Massive Analysis Quality Control (MAQC) Society Board of Directors, Waldron L, et al. Transparency and reproducibility in artificial intelligence. *Nature*. 2020. p. E14–6
88. Buolamwini J, Gebru T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In: Friedler SA, Wilson C, editors. Proceedings of the 1st conference on fairness, accountability and transparency. PMLR; 2018. p. 77–91
89. Finlayson SG, Subbaswamy A, Singh K, Bowers J, Kupke A, Zittrain J, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. 2021;385:283–6. <https://doi.org/10.1056/NEJMc2104626>.
90. Feng J, Phillips RV, Malenica I, Bishara A, Hubbard AE, Celi LA, et al. Clinical artificial intelligence quality improvement: Towards continual monitoring and updating of AI algorithms in healthcare. *npj Digit Med*. 2022. <https://doi.org/10.1038/s41746-022-00611-y>.
91. Cummings ML. Automation bias in intelligent time critical decision support systems. Decision making in aviation Routledge. 2004. <https://doi.org/10.2514/6.2004-6313>.
92. Gafni A, Charles C, McMaster University. Centre for health economics and policy analysis, Whelan T. Shared decision-making in the medical encounter: What does it mean?, or, it takes at least two to tango. Centre for Health Economics and Policy Analysis, McMaster University; 1994.
93. Coulter A, Collins A. Making shared decision-making a reality: No decision about me, without me. London: The King's Fund; 2011;621.

94. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLOS Med*. 2018. <https://doi.org/10.1371/journal.pmed.1002689>.
95. Price WN 2nd, Cohen IG. Privacy in the age of medical big data. *Nat Med*. 2019;25:37–43. <https://doi.org/10.1038/s41591-018-0272-7>.
96. Grady C. Enduring and emerging challenges of informed consent. *N Engl J Med*. 2015;372:2172. <https://doi.org/10.1056/NEJMc1503813>.
97. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? *JAMA*. 2000;283:2701–11. <https://doi.org/10.1001/jama.283.20.2701>.
98. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: Preserving security and privacy. *J Big Data*. 2018;5:1–18. <https://doi.org/10.1186/s40537-017-0110-7>.
99. Taylor L, Floridi L, van der Sloot B. Group privacy: New challenges of data technologies. Springer; 2016
100. Neri E, Coppola F, Miele V, Bibbolino C, Grassi R. Artificial intelligence: who is responsible for the diagnosis? *Radiol Med*. 2020;125:517–21. <https://doi.org/10.1007/s11547-020-01135-9>.
101. Price WN 2nd, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. *JAMA*. 2019;322:1765–6. <https://doi.org/10.1001/jama.2019.15064>.
102. van der Velden BHM, Kuijff HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal*. 2022. <https://doi.org/10.1016/j.media.2022.102470>.
103. Goebel R, Chander A, Holzinger K, Lecue F, Akata Z, Stumpf S, et al. Explainable AI: The new 42? Machine Learning and Knowledge Extraction. Springer International Publishing. 2018;295–303. https://doi.org/10.1007/978-3-319-99740-7_21
104. Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining explanations: An overview of interpretability of machine learning [Internet]; 2018. cs.AI. <http://arxiv.org/abs/1806.00069>. <https://doi.org/10.1109/DSAA.2018.00018>
105. Rudin C. Stop explaining Black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*. 2019;1:206–15. <https://doi.org/10.1038/s42256-019-0048-x>.
106. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artif Intell*. 2019;267:1–38. <https://doi.org/10.1016/j.artint.2018.07.007>.
107. Lipton ZC. The mythos of model interpretability. In: machine learning, the concept of interpretability is both important and slippery. *Queueing Syst*. 2018. Association for Computing Machinery;16:31–57. <https://doi.org/10.1145/3236386.3241340>
108. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med*. 2019;25:44–56. <https://doi.org/10.1038/s41591-018-0300-7>.
109. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380:1347–58. <https://doi.org/10.1056/NEJMr1814259>.
110. Obermeyer Z, Emanuel EJ. Predicting the future - Big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216–9. <https://doi.org/10.1056/NEJMp1606181>.
111. Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Deep learning with convolutional neural network in radiology. *Jpn J Radiol*. 2018;36:257–72. <https://doi.org/10.1007/s11604-018-0726-3>.
112. Burrell J. How the machine “thinks”: Understanding opacity in machine learning algorithms. *Big Data Soc*. 2016;3:2053951715622512. SAGE Publications Ltd;3. <https://doi.org/10.1177/2053951715622512>
113. Char DS, Shah NH, Magnus D. Implementing machine learning in health care - Addressing ethical challenges. *N Engl J Med*. 2018;378:981–3. <https://doi.org/10.1056/NEJMp1714229>.
114. Morley J, Machado CCV, Burr C, Cows J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. *Soc Sci Med*. 2020. <https://doi.org/10.1016/j.socscimed.2020.113172>.
115. Epstein RM, Fiscella K, Lesser CS, Stange KC. Why the nation needs a policy push on patient-centered health care. *Health Aff (Millwood)*. 2010;29:1489–95. <https://doi.org/10.1377/hlthaff.2009.0888>.
116. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion*. 2020;58:82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>.
117. Vayena E, Dzenowagis J, Brownstein JS, Sheikh A. Policy implications of big data in the health sector. *Bull World Health Organ*. 2018;96:66–8. <https://doi.org/10.2471/BLT.17.197426>.
118. Mittelstadt BD, Floridi L. The ethics of big data: current and foreseeable issues in biomedical contexts. *Sci Eng Ethics*. 2016;22:303–41. <https://doi.org/10.1007/s11948-015-9652-2>.
119. Coulter A, Ellins J. Effectiveness of strategies for informing, educating, and involving patients. *BMJ*. 2007;335:24–7. <https://doi.org/10.1136/bmj.39246.581169.80>.
120. Open AI. GPT-4 technical report [Internet]. arXiv [cs.CL]; 2023. <http://arxiv.org/abs/2303.08774>
121. Eloundou T, Manning S, Mishkin P, Rock D. GPTs are GPTs: An early look at the labor market impact potential of large language models [Internet]. arXiv; 2023. econ.GN. <http://arxiv.org/abs/2303.10130>
122. Ueda D, Walston SL, Matsumoto T, Deguchi R, Tatekawa H, Miki Y. Evaluating GPT-4-based ChatGPT’s clinical potential on the NEJM quiz [Internet]; 2023. medRxiv. <https://www.medrxiv.org/content/10.1101/2023.05.04.23289493v1>
123. Davis MA, Lim N, Jordan J, Yee J, Gichoya JW, Lee R. Imaging artificial intelligence: A framework for radiologists to address health equity, from the AJR special series on DEI. *AJR Am J Roentgenol*. 2023. American Roentgen Ray Society. <https://doi.org/10.2214/AJR.22.28802>
124. Shimazaki A, Ueda D, Choppin A, Yamamoto A, Honjo T, Shimahara Y, et al. Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method. *Sci Rep*. 2022;12:727. <https://doi.org/10.1038/s41598-021-04667-w>.
125. Ueda D, Yamamoto A, Nishimori M, Shimono T, Doishita S, Shimazaki A, et al. Deep learning for MR angiography: automated detection of cerebral aneurysms. *Radiology*. 2019;290:187–94. <https://doi.org/10.1148/radiol.2018180901>.
126. Ueda D, Yamamoto A, Shimazaki A, Walston SL, Matsumoto T, Izumi N, et al. Artificial intelligence-supported lung cancer detection by multi-institutional readers with multi-vendor chest radiographs: a retrospective clinical validation study. *BMC Cancer*. 2021;21:1120. <https://doi.org/10.1186/s12885-021-08847-9>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.