AI in Healthcare: Impacts, Risks and Regulation to Mitigate Adverse Impacts

Retno Larasati

Knowledge Media Institute, The Open University. Walton Hall, Milton Keynes, United Kingdom

Abstract

The rapid advancements in artificial intelligence (AI) have revolutionised various industries, including healthcare. AI systems in healthcare hold immense potential for improving patient outcomes, accelerating diagnosis, and enhancing overall healthcare delivery. However, the introduction of AI in healthcare also brings forth significant ethical, legal, and societal implications that necessitate a robust regulatory framework. This position paper aims to discuss the use and the risk of AI in healthcare, and also how the regulations or guidelines aiming to mitigate their potential adverse impact on individuals and societies corresponds to it.

Keywords

AI Healthcare, AI Harms, AI Regulation, Ethics of AI

1. Introduction

Artificial Intelligence (AI) holds transformative promise in reshaping healthcare delivery, enhancing patient outcomes, and expediting diagnostic processes. Within the healthcare domain, AI systems play a pivotal role in aiding medical practitioners across diverse responsibilities, encompassing tasks like the interpretation of medical imaging, drug discovery endeavors, and continuous patient monitoring. The proficiency of AI in processing extensive and intricate datasets, coupled with its capacity to furnish insights and predictions with heightened precision and celerity compared to human capabilities, positions it as a valuable instrument for the advancement of healthcare. The assimilation of AI into the healthcare landscape has catalyzed the evolution of innovative techniques and methodologies, profoundly elevating the overall quality of healthcare services.

However, the adoption of AI in healthcare also raises significant ethical, legal, and societal concerns. The potential risks associated with the use of AI in healthcare include privacy breaches, discrimination, misdiagnosis, and the creation of new health inequalities. These risks highlight the importance of developing a robust regulatory framework that can effectively mitigate any adverse impacts on individuals and societies.

This position paper aims to critically examine the use and risks of AI in healthcare and to analyse how regulations or guidelines can mitigate their potential adverse impacts. The paper will provide an overview of the impacts of AI on healthcare, discussing the various applications of AI in healthcare, their potential benefits, and the challenges and risks associated with their adoption. It will examine the existing regulatory frameworks that aim to mitigate the risks associated with the use of AI in healthcare and identify the gaps and limitations of these frameworks.

2. Contextualising the transformative potential of AI in healthcare

The field of healthcare is witnessing the transformative potential of artificial intelligence (AI), a technology that holds promise for revolutionizing patient care, clinical decision-making, and healthcare delivery. This section aims to provide a contextual understanding of the transformative potential of AI in healthcare by exploring its key applications, benefits, and implications..

The utilization of Artificial Intelligence (AI) applications in healthcare has the potential to significantly enhance the medical profession's capacity to diagnose and treat patients. Given the voluminous, intricate, and extensive nature of medical data, AI applications are well-equipped to better exploit complex data for disease detection through image analysis [1, 2], drug discovery using pattern recognition [3], and surgery via medical robots [4], as well as patient monitoring.

The transformative potential of AI in healthcare offers several notable benefits. Firstly, it enhances diagnostic accuracy by augmenting healthcare professionals' capabilities, reducing errors, and improving accuracy in interpreting complex medical data. For instance, IBM Watson was compared to human experts for 1,000 cancer diagnoses and identified treatment options missed by doctors in 30% of cases [1]. Additionally, previous research

---

[1]https://www.nytimes.com/2016/10/17/technology/ibm-is-

has demonstrated that Google's AI system can identify breast cancers and reduce the number of missed cases by 9.4% in a US sample and 2.7% in a UK sample compared to the original radiologist diagnoses [5]. Secondly, AI improves efficiency and workflow by automating administrative tasks and optimizing resource allocation, allowing healthcare professionals to focus more on patient care. Lastly, AI enables personalized medicine by analyzing diverse patient data, leading to tailored treatment plans that consider individual characteristics and genetic profiles. For example, a systematic review of the use of AI in medical imaging found that the technology has the potential to improve diagnostic accuracy and reduce variability among radiologists [6]. Another study found that the use of AI in predicting heart failure mortality could potentially save lives and reduce healthcare costs [7].

The future potential of AI in healthcare is based on the premise that it can aid healthcare professionals in delivering care that is more accessible, affordable, and easier to administer. Moreover, AI has the potential to enhance the performance of healthcare professionals as it has been shown to outperform human healthcare professionals in some tasks. With this level of capability, AI has great potential to enhance healthcare and enable clinicians/doctors to allocate more time with their patients.

While integrating AI into healthcare brings substantial benefits, it is essential to address ethical and societal implications. Data privacy and security become critical considerations as AI involves collecting and analyzing sensitive patient data. Explainable AI has emerged as a potential solution to some of these concerns by making AI systems more transparent and understandable to humans. The question of interpretability is not a mere matter of intellectual curiosity, it is both a risk and a responsibility [8]. Due to the possible "high stakes", medical professionals and patients may not be comfortable to straightly follow a non-transparent system's decision. Many AI researchers are working toward developing interpretable methods that are suitable for medical cases [9, 10, 11, 12, 13].

Robust data governance frameworks and cybersecurity measures are necessary to protect patient privacy. Bias and fairness are also significant concerns, as AI algorithms may inherit biases from biased training data, requiring efforts to ensure fairness and diversity. Research has found that explainable AI can help to identify and mitigate biases in AI systems, improving fairness and reducing the risk of harmful outcomes [14]. These are still considered as potential solution as there is no real-world impact recorded yet. Moreover, the integration of AI should emphasize human-AI collaboration, where healthcare professionals are trained to effectively

interpret AI-generated insights while retaining the final responsibility for decision-making.

It is important for researchers to consider the broader impacts of their work on these various aspects of our ecosystem, and to work towards the development of AI systems that maximise benefits and minimise risks. Understanding and navigating the contextual landscape of AI in healthcare are essential to maximize its transformative potential while mitigating potential risks.

3. Assessing Risks of AI Systems in Healthcare

However, the use of AI in healthcare also raises concerns about the potential risks and unintended consequences of this technology. This paper identified four possible risks, such as, bias, safety, interpretability, and privacy.

One of the primary risks in AI healthcare applications is the quality and bias of the underlying data. AI algorithms rely heavily on comprehensive and accurate data to generate reliable insights. However, if the training data is incomplete, biased, or unrepresentative, it can lead to skewed outcomes, misdiagnoses, or unequal treatment. For example, outside of healthcare, cases such as Google Photos algorithm falsely classified black people as gorillas [2], or when the Microsoft chat-bot turned racist in a day [3], caused by unrepresentative data. This can have severe consequences for patients and can perpetuate existing health disparities. Unfortunate example, a study published in the journal Science found that a popular algorithm used to prioritise patients for extra healthcare services was biased against Black patients. The algorithm was more likely to prioritise white patients over Black patients who had the same level of health needs [15]. Therefore, ensuring the collection of diverse and unbiased data is crucial to mitigate this risk. This can be achieved through measures such as increasing the diversity of the data sets used to train AI algorithms and ensuring the data is representative of the population being served, especially the minority groups.

The safety and reliability of AI systems in healthcare are paramount. Errors, malfunctions, or vulnerabilities in AI algorithms can have severe consequences for patients. It is essential to ensure rigorous testing, validation, and ongoing monitoring of AI systems to minimize the risk of adverse events and maximize patient safety. This can be achieved through regulatory bodies and agencies setting up standards and guidelines for the development and deployment of AI systems in healthcare.

---

counting-on-its-bet-on-watson-and-paying -big-money-for-it.html

[2]https://mashable.com/2015/07/01/google-photos-black-people-gorillas/

[3]https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

AI models often operate as black boxes, making it challenging to understand how they arrive at specific decisions or predictions. With the advent of deep learning and commercialised AI, the issue of using black-box methods has become increasingly relevant. This lack of interpretability and explainability raises concerns about trust, accountability, and potential errors in critical healthcare decisions [16]. It is vital to develop methods that enhance the interpretability of AI systems, enabling healthcare professionals to understand the reasoning behind AI-generated outcomes. This can be achieved through approaches such as transparent decision-making models, which provide explanations for the recommendations made by the AI systems.

Lastly, AI in healthcare raises complex ethical and legal considerations. Privacy and data protection, consent, and the responsible use of patient data are critical aspects to address. The potential for AI to inadvertently discriminate or violate ethical principles, such as patient autonomy or confidentiality, necessitates clear guidelines and ethical frameworks. This can be achieved through the establishment of ethical committees and regulatory authorities that oversee the ethical and legal implications of AI in healthcare.

4. Mitigating Adverse Impacts: Frameworks and Regulations

4.1. Existing regulations, guidelines, and frameworks for AI in healthcare

Existing regulations, guidelines, and frameworks play a critical role in mitigating the potential adverse impacts of AI in healthcare. In the European Union, the Medical Device Regulation (MDR) provides regulatory requirements for medical devices, including AI systems used in healthcare [17]. The MDR aims to ensure the safety and performance of AI systems in healthcare. The National Health Service (NHS) in the UK also has developed an AI Code of Conduct to guide the safe and ethical use of AI technologies in healthcare. It covers aspects such as transparency, accountability, and patient privacy. Similarly, in a more general scope, the European Commission's regulatory framework for AI aims to establish clear guidelines and rules to promote the development and deployment of trustworthy AI systems while ensuring the protection of individuals and fundamental rights. The International Organization for Standardization (ISO) has published a series of standards on AI that include requirements for safety, transparency, and accountability. The ISO/IEC 27701 standard provides a framework for privacy management that applies to AI systems that process personal data.

The General Data Protection Regulation (GDPR) is an-

other significant regulation that affects AI in healthcare. The GDPR mandates the protection of personal data and the right to privacy of individuals. AI systems in healthcare that process personal data must comply with the GDPR's requirements for data protection, including data minimization, transparency, and accountability.

In the U.S., Food and Drug Administration (FDA) has issued "Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan" [18]. They offer recommendations for pre-market submissions, including considerations for algorithm changes and ongoing monitoring of AI systems.

Several international organizations have also developed guidelines and frameworks for risk assessment in AI healthcare. The World Health Organization (WHO) has developed guidelines on the use of AI in healthcare, focusing on the ethical and governance implications. The guidelines emphasize the importance of transparency, accountability, and human oversight in the development and deployment of AI in healthcare.

While these regulations, guidelines, and frameworks provide important guidance on risk assessment in AI healthcare, some critics argue that they are too general and lack teeth. They contend that existing regulations are often insufficient to address the complexities of AI systems and their potential risks. Moreover, there is a lack of standardization in the application of AI in healthcare, making it challenging to develop uniform regulations and guidelines.

4.2. Discussion

Despite the existence of several regulations, guidelines, and frameworks for AI in healthcare, criticisms and concerns about their adequacy and effectiveness continue to be raised. The first criticisms is that the regulations are too general and do not provide specific guidance on how to ensure the safety, efficacy, and ethical use of AI systems in healthcare. For instance, the GDPR and the UK GDPR do not offer explicit guidance on how to address the interpretability and explainability of AI algorithms [19]. This lack of specificity can result in confusion and inconsistencies in the implementation and enforcement of the regulations.

Moreover, some argue that the existing regulations are too weak and fail to adequately protect individuals' privacy and data rights. The GDPR has been criticized for being too vague and not going far enough to protect individuals' privacy rights [20]. Some argue that the GDPR's provisions on consent, purpose limitation, and data minimization are not enough to address the risks posed by AI systems, such as those related to profiling and discrimination.

Not only deemed too weak, one of the criticisms of the current regulations, guidelines, and frameworks for

AI in healthcare is their perceived lack of efficacy or "toothless" [21]. This is due to the absence of a clear accountability mechanism or regulatory body to oversee the adherence to regulations. As a result, some stakeholders have expressed concerns about the ineffectiveness of these regulations in preventing any potential negative impacts of AI in healthcare. Another critique of the regulations is that they may stifle innovation and hinder the adoption of beneficial AI healthcare applications. Regulations that are too restrictive or burdensome may discourage companies from investing in the development and deployment of AI systems in healthcare, which could limit the potential benefits to patients and healthcare providers.

Another major critique of existing regulations and frameworks for AI in healthcare is that they are often too general and lack specificity [22]. This can make it challenging to effectively regulate and ensure the safe and ethical use of AI in healthcare. However, it is important to note that creating specific regulations for every possible application of AI in healthcare is a daunting task, as AI is constantly evolving and being implemented in new and innovative ways. Additionally, the lack of clarity and specificity in existing regulations can result in confusion and inconsistency in their application. This can lead to variations in how AI systems are regulated across different jurisdictions and countries, making it difficult for companies to comply with regulations and for patients to have confidence in the safety and effectiveness of AI systems.

Nonetheless, it is important to acknowledge that the development of specific regulations and enforcement mechanisms for AI in healthcare is a challenging task, given the diverse range of applications and the rapid advancement of technology. Thus, it is crucial to engage a diverse group of stakeholders, including AI experts, healthcare professionals, policymakers, and patients, in the development and implementation of regulations to ensure their effectiveness and relevance.

Furthermore, it is important to note that despite the critiques, the existing regulations, guidelines, and frameworks have undoubtedly played an important role in promoting the responsible development and deployment of AI systems in healthcare. They have helped to establish a baseline of expectations and standards for the industry, and have provided a starting point for discussions around the ethical, legal, and social implications of AI in healthcare. Moreover, they have helped to raise awareness and understanding of the potential risks and benefits of AI in healthcare among stakeholders, including patients, healthcare providers, regulators, and policymakers.

4.3. Consideration for regulations, guidelines, and frameworks

Moving forward, it is essential to address the weaknesses and critiques of the existing regulations and frameworks to ensure that they remain relevant and effective in the face of rapidly evolving technologies and emerging risks. One approach could be to develop more specific and detailed guidance on how to address the interpretability and explainability of AI algorithms, and how to mitigate the risks related to bias and discrimination. Additionally, it may be necessary to revisit and update the existing regulations to ensure that they are fit for purpose in the current technological landscape and take into account the latest developments and best practices in AI in healthcare.

The regulation need to cover the whole AI development process. From the first stage of development: data collection, regulatory frameworks should address potential biases and discriminatory practices in AI algorithms and data sources. They should promote fair access to healthcare services and ensure that AI systems do not perpetuate or amplify existing disparities in healthcare delivery. Integrating ethical considerations related to fairness and equity fosters equitable access, reduces disparities, and promotes social justice in AI healthcare applications.

To after the development stage: patient interaction, regulatory frameworks should establish guidelines that prioritise patient autonomy by ensuring informed consent and transparent communication. Patients should be made aware of the use of AI systems in their care, the potential risks and benefits, and their right to consent or refuse AI-driven interventions. These ethical considerations promote patient empowerment and protect individual rights within AI healthcare.

Protecting patient privacy and maintaining confidentiality are paramount ethical obligations. Regulatory frameworks must incorporate stringent guidelines for data privacy, ensuring proper anonymisation, secure storage, and controlled access to sensitive healthcare information. The integration of ethical considerations related to data privacy safeguards patient trust and upholds the ethical obligations of healthcare providers and AI developers. Even though the European Union's GDPR and U.S. Office for Civil Rights' guidance on the use of AI in healthcare includes provisions that require organizations to obtain and store personal data,the specific requirements and implementation of informed consent in AI healthcare may vary depending on the jurisdiction and the particular application of AI technology.

Ethical AI in healthcare requires transparency and explainability to build trust, promote accountability, and facilitate proper decision-making. Regulatory frameworks emphasised the transparency of AI algorithms, data sources, and decision processes. They should also

mandate explainable AI models that provide **understandable** and **justifiable** explanations for their outputs [23]. Integrating ethical considerations related to transparency and explainability fosters responsible AI practices and supports the responsible use of AI systems in healthcare. As previously mentioned, the regulation should be specific and not too general. For example, the law and applicable regulations require certain information to be included in the explanation for AI systems that uses personal data to train, test or deploy. According to the UK Information Commissioner's Office (ICO), the two subcategories of explanations required by data protection laws are process-based explanations, which relate to information about the governance of the AI system throughout its design and deployment, and outcome-based explanations, which concern what happens in the case of a particular decision [24]. In contrast with these regulations, most explainable AI approaches only cover outcome-based explanation [25, 26], which is why regulations should specify what should be included in the explanation.

While AI systems can assist healthcare professionals, human oversight and responsibility remain essential. Regulatory frameworks should emphasise the role of healthcare professionals in the decision-making process and clarify the extent to which AI systems should augment or support human judgment. The integration of ethical considerations related to human oversight ensures that the ultimate responsibility for patient care lies with trained healthcare professionals, while AI systems serve as valuable tools. Ultimately, a collaborative and multi-stakeholder approach will be required to develop a comprehensive and effective regulatory framework that balances the potential benefits of AI in healthcare with the need to protect individuals' rights and interests.

5. Conclusion

Despite the critiques, the existing regulations, guidelines, and frameworks have undoubtedly played an important role in promoting the responsible development and deployment of AI systems in healthcare. They have helped to establish a baseline of expectations and standards for the industry, and have provided a starting point for discussions around the ethical, legal, and social implications of AI in healthcare. Moreover, they have helped to raise awareness and understanding of the potential risks and benefits of AI in healthcare among stakeholders, including patients, healthcare providers, regulators, and policymakers.

Overall, this paper contributes to the ongoing discussion on AI in healthcare and regulatory framework that ensures that AI is used in a way that is ethical, transparent, and responsible. Such a framework would ensure

that the benefits of AI in healthcare are fully realised while minimising any negative impacts on individuals and societies.

Acknowledgement

References

[1] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. H. Beck, Deep learning for identifying metastatic breast cancer, arXiv preprint arXiv:1606.05718 (2016).

[2] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature 542 (2017) 115.

[3] Y.-C. Lo, S. E. Rensi, W. Torng, R. B. Altman, Machine learning in chemoinformatics and drug discovery, Drug discovery today 23 (2018) 1538–1546.

[4] Y. Kassahun, B. Yu, A. T. Tibebu, D. Stoyanov, S. Giannarou, J. H. Metzen, E. Vander Poorten, Surgical robotics beyond enhanced dexterity instrumentation: a survey of machine learning techniques and their role in intelligent and autonomous surgical actions, International journal of computer assisted radiology and surgery 11 (2016) 553–568.

[5] E. D. Pisano, Ai shows promise for breast cancer screening, 2020.

[6] R. Aggarwal, V. Sounderajah, G. Martin, D. S. Ting, A. Karthikesalingam, D. King, H. Ashrafian, A. Darzi, Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis, NPJ digital medicine 4 (2021) 65.

[7] C. R. Olsen, R. J. Mentz, K. J. Anstrom, D. Page, P. A. Patel, Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure, American Heart Journal 229 (2020) 1–17.

[8] Y. Xie, G. Gao, X. Chen, Outlining the design space of explainable intelligent systems for medical diagnosis, arXiv preprint arXiv:1902.06019 (2019).

[9] P. Nguyen, T. Tran, N. Wickramasinghe, S. Venkatesh, Deepr: a convolutional net for medical records, IEEE journal of biomedical and health informatics 21 (2017) 22–30.

[10] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, J. Gao, Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks, in: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, 2017, pp. 1903–1911.

[11] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, et al., Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning, arXiv preprint arXiv:1711.05225 (2017).

[12] G. Zhao, B. Zhou, K. Wang, R. Jiang, M. Xu, Respond-cam: Analyzing deep models for 3d imaging data by visualizations, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2018, pp. 485–492.

[13] J. Wu, D. Peck, S. Hsieh, V. Dialani, C. D. Lehman, B. Zhou, V. Syrgkanis, L. Mackey, G. Patterson, Expert identification of visual primitives used by cnns during mammogram classification, in: Medical Imaging 2018: Computer-Aided Diagnosis, volume 10575, International Society for Optics and Photonics, 2018, p. 105752T.

[14] K. Alikhademi, E. Drobina, D. Prioleau, B. Richardson, D. Purves, J. E. Gilbert, A review of predictive policing from the perspective of fairness, Artificial Intelligence and Law (2022) 1–17.

[15] Z. Obermeyer, R. Nissan, M. Stern, S. Eaneff, E. J. Bembeneck, S. Mullainathan, Algorithmic bias playbook, Center for Applied AI at Chicago Booth (2021).

[16] G. Marcus, Deep learning: A critical appraisal, arXiv preprint arXiv:1801.00631 (2018).

[17] E. Parliament, the Council of the European Union, Regulation (eu) 2020/561 of the european parliament and of the council of 23 april 2020 amending regulation (eu) 2017/745 on medical devices, as regards the dates of application of certain of its provisions, Official Journal of the European Union 63 (2020) 18–22.

[18] Food, D. Administration, et al., Proposed regulatory framework for modifications to artificial intelligence/machine learning (ai/ml)-based software as a medical device (samd) (2019).

[19] S. Wachter, B. Mittelstadt, L. Floridi, Why a right to explanation of automated decision-making does not exist in the general data protection regulation, International Data Privacy Law 7 (2017) 76–99.

[20] C. Castets-Renard, Accountability of algorithms in the gdpr and beyond: A european legal framework on automated decision-making, Fordham Intell. Prop. Media & Ent. LJ 30 (2019) 91.

[21] A. Rességuier, R. Rodrigues, Ai ethics should not remain toothless! a call to bring back the teeth of ethics, Big Data & Society 7 (2020) 2053951720942541.

[22] L. Floridi, Establishing the rules for building trustworthy ai, Nature Machine Intelligence 1 (2019) 261–262.

[23] R. Larasati, A. De Liddo, E. Motta, The effect of explanation styles on user's trust, in: Proceedings of the Workshop on Explainable Smart Systems for Algorithmic Transparency in Emerging Technologies co-located with IUI 2020, 2020.

[24] E. Kazim, A. Koshiyama, Explaining decisions made with ai: a review of the co-badged guidance by the ico and the turing institute, Available at SSRN 3656269 (2020).

[25] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (xai), IEEE Access 6 (2018) 52138–52160.

[26] A. Adadi, M. Berrada, Explainable ai for healthcare: from black box to interpretable models, in: Embedded Systems and Artificial Intelligence: Proceedings of ESAI 2019, Fez, Morocco, Springer, 2020, pp. 327–337.