REGULATING ARTIFICIAL INTELLIGENCE IN CLINICAL MEDICINE: GOVERNANCE FRAMEWORK PRINCIPLES FOR SAFE AND ACCOUNTABLE IMPLEMENTATION

Nauman Ismat Butt[1]

1. Department of Medicine & Allied, Azra Naheed Medial College, Superior University Lahore Pakistan

ARTICLE INFO

REVIEW ARTICLE

Corresponding author:
Nauman Ismat Butt
Email:
nauman_ib@yahoo.com
Assistant Professor, Department of Medicine & Allied, Azra Naheed Medial College, Superior University Lahore Pakistan

Vol: 4 | Issue: 2
ISSN Print: 2960-2580
ISSN Online: 2960-2599

Publisher:
Medical Research and Statistical Consultancy Training Centre (SMC-PRIVATE) Limited

CONTRIBUTION
Nauman Ismat Butt:
Conception and design, literature research, manuscript writing, review and corrections

Keywords: Artificial intelligence (AI), AI governance, Regulatory frameworks, Clinical medicine, Patient safety

ABSTRACT

Artificial intelligence (AI) is transforming the field of clinical medicine at a fast rate. AI is being used in the field of clinical medicine in areas such as diagnostic imaging, prediction, documentation, and large language model-based clinical reasoning tools such as ChatGPT. Algorithm-based triage tools are also being used in clinical settings such as hospitals. Radiology departments are also using AI-based detection tools. However, the rate at which AI is transforming the field of clinical medicine is much higher than the rate at which the regulatory bodies are keeping pace with the developments in the field. Even though the Food and Drug Administration (FDA) has developed a pathway to regulate AI-based Software as a Medical Device (SaMD) and machine learning-based technologies, the current system that is in place to ensure the safety and efficacy of these technologies was developed to ensure the safety and efficacy of static technologies and not those that have the capability to learn. The rate at which AI is being incorporated in the field of clinical medicine is much higher than the rate at which these technologies are being regulated. The issue of clinical AI governance is not about the regulation of AI; it is about patient safety. This article outlines the need for a structured framework of five principles of safety, accountability, fairness, explainability, and adaptive oversight.

https://doi.org/10.61171/pioneerjbiostat.4.2.2026.129

Acknowledgements: The manuscript was written, reviewed and revised by the author. However, the author acknowledges the use of ChatGPT (OpenAI) in assisting with the language refinement of this manuscript. All work related to the study design and intellectual content was conducted solely by the author.

INTRODUCTION

Artificial intelligence (AI) is no longer considered a futuristic concept or idea; rather, it is a reality in the field of medicine. It has been integrated into the workflow of medicine, and many such systems are having a significant impact on the outcome of the treatment of a patient or a particular disease.[1,2] It has also started attracting the attention of many regulatory agencies, and many such agencies have started formulating rules and regulations for the management of AI in the field of medicine. The U.S. Food and Drug Administration (FDA) has come up with

a pathway for the regulation of AI/ML-Based Software as a Medical Device, which includes tools such as a predetermined change control plan for making limited changes to the algorithm itself.[3,4] At the international level, a new form of legislation is the European Union's AI Act, which has classified many AI systems used in the field of healthcare as 'high-risk products'.[5,6]

Nevertheless, this progress still faces the fact that the development of regulations is a slow process in comparison to the fast rate at which AI technology is being adopted. This process of regulation may be fragmented and may not always be effective. The retrainable and performance-drifted adaptive algorithms pose a challenge to the traditional paradigm for static devices. The burden of safety for AI lies on the shoulders of healthcare organizations. The concept of clinical AI governance extends beyond just obtaining regulatory approvals to include the full lifecycle of an AI system, from procurement to validation, and eventually monitoring and recalibration.

The Expanding Role of AI in Clinical Care

AI technologies used in healthcare can be broadly classified into three domains: diagnostic augmentation, workflow optimization, and predictive analytics.[1,2] In the field of diagnostic medicine, AI technologies are used for radiologic interpretation, dermatologic image classification, digital pathology analysis, and electrocardiographic interpretation. In the field of workflow optimization, natural language processing is used for clinical documentation generation. In predictive analytics, risk prediction of sepsis, hospital readmission, and mortality is performed, often influencing triage and resource allocation decisions.

In contrast to conventional rule-based decision support systems, many current models of AI are data-based, probabilistic, and opaque in terms of their internal logic.[7,8] Their data bases may not be representative in terms of demographics, and their results may be affected by changes in the mix of patients or changes in practice patterns. These are new risks, including algorithmic bias, which can increase health disparities; automation bias, in which clinicians may rely too heavily on the results of AI; model drift, in response to a changing clinical environment; and accountability, in terms of the influence of AI on decision-making.[7,8,9] The transformative potential of AI is substantial. However, the magnitude of its influence demands governance structures proportionate to its clinical impact.

Limitations of the Current Regulatory Landscape

Although some regulatory bodies have initiated the process of developing guidelines for the use of AI in healthcare, many issues remain to be addressed. First, the traditional approach to the approval of medical devices implies that the device will behave predictably. However, the device, being an AI device that can be retrained, operates in a changing environment. Second, the institutional oversight capability varies. Some academic institutions have the capability for data science infrastructure, while others do not have the capability for critical

evaluation. Thirdly, the concept of clinical accountability still remains unclear. In this case, it is imperative to understand that even though the role of AI exists, the main accountability of the health outcome remains on the practitioner's shoulders. The issue of ambiguity represents a limitation of the role of AI in health governance. Therefore, this article outlines the need for a structured framework of five principles of safety, accountability, fairness, explainability, and adaptive oversight.

1. Safety and Clinical Validation:

Any AI system meant for clinical use has to be validated, which is not limited to performance metrics. It is not true that the accuracy of the AI model on a given dataset translates into real-world accuracy. A thorough pre-deployment validation also includes real-world testing, which has to be done on a wide range of patients and also include subgroup analyses for performance variability and also be compared with clinical standards.[10,11] Significantly, the validation is not a one-time task; rather, it is a continuous process. Surveillance is also required after the deployment of the AI model to ensure that the performance is not compromised over time.[10,11] Such changes need to be monitored to ensure continued safety, not simply assumed. AI models that impact diagnosis, triage, or treatment should be subject to scrutiny similar to other significant clinical interventions.

2. Accountability and Liability Clarity:

The involvement of AI in decision-making has blurred the lines of responsibility, making it a complicated issue. It has been argued that, in the context of decision-making, there are several entities to be held responsible, including software developers, institutions, and clinicians.[12,13] Therefore, it has been suggested that a clear definition of the role of AI in decision-making, considering it to be assistive rather than autonomous, must be included in the governance structure, keeping in view the fact that the final responsibility must lie with a licensed professional.[13,14] It has been argued that a structure of liability, considering the entire system to be liable, must be put in place, instead of leaving it to individual clinicians, as it may undermine the purpose of implementing the technology in a manner that is appropriate.[12,14]

3. Fairness and Bias Monitoring:

AI systems trained on past data may learn existing inequities within healthcare systems. The risk of AI systems learning from historical data and perpetuating inequities present within the structure of healthcare systems is high.[15,16] It is important to note that the performance of AI systems must not perpetuate inequities based on race, gender, age, disability, and socioeconomic status. Therefore, a lack of proper governance may cause such AI systems to reinforce or increase inequities based on race, gender, age, disability, or socioeconomic status. For a healthcare AI system to be effectively governed, it is essential to ensure that the composition of the data used to train the AI is transparent and that the performance of the AI is evaluated across various groups of people.[17,18] It is imperative to note that fairness is not equivalent to accuracy. An ethical AI must exhibit reliability or accuracy

within various groups of people within a healthcare setup. Equity monitoring is therefore an ongoing institutional responsibility rather than a one-time regulatory requirement.[16,19]

4. Explainability and Transparency:

A significant number of AI models, especially those that are deep learning-based, are "black boxes," in the sense that there are no clear ways to interpret how the model arrived at a particular output.[20,21] Though it may not be possible to make AI models fully transparent, there are ethical and practical issues to consider. It is important for clinicians to be familiar with the purpose, restrictions, and confidence level of AI models being used to make clinical decisions.[20,21] It is suggested that there are clear standards set by governance policies to ensure that there is a clear explanation of the development of the AI model, the data set on which the model was trained, and any restrictions on the model's performance. In addition, the output of the AI model should include a range of probabilities to assist in understanding the results of the model. Transparency can aid in informed clinical judgment, enhance patient communication, and reduce blind reliance on algorithm results.[22,23] Trust in AI-assisted care is not only dependent on performance but also on transparency about how and when to use these tools.

5. Adaptive Oversight and Continuous Monitoring:

This is due to the fact that AI is used in a dynamic clinical environment.[24,25] For example, a change in the population, rate of disease, or way of diagnosis might affect the performance of AI. This is called a model drift. In this case, the governance system must also be dynamic. For this reason, healthcare institutions must develop performance surveillance dashboards and thresholds for review and suspension of approval, as well as a periodic review of the approved AI. Multidisciplinary oversight committees, which consist of clinicians, data scientists, ethicists, and legal experts, are a good way of governing AI tools in different domains of evaluation.[24,25]

Dynamic oversight of AI tools is a delicate balance between innovation and caution.[12,26] It is imperative to understand that while enacting laws, they should not act as a barrier to the development of AI. Similarly, they should not remain static while AI is becoming more and more dynamic.

Implementation Recommendations

In order to operationalize the principles of governance, organizational commitment is also required, as depicted in Table 1. This could be done in the following ways:

1. Formation of AI oversight committees in the health systems
2. Development of standardized procurement and evaluation checklists based on safety, accountability, fairness, transparency, and monitoring principles
3. Development of registries that include information on AI tools used, their validation status, and results from monitoring

4. Integrating AI literacy into undergraduate, graduate, and continuing education for physicians
5. Designing feedback systems that enable front-line practitioners to raise concerns regarding AI performance

Embedding governance into institutional processes helps to promote a more proactive oversight strategy rather than a reactive approach to risk management.[3,7] AI implementation should be considered a system-level clinical intervention requiring coordinated evaluation and monitoring.

Table 1: Core Principles for Clinical AI Governance

Principles	Key Actions	Purpose / Outcome
Safety and Clinical Validation	Prospective validation Real-world performance testing Subgroup bias analysis Benchmarking against clinician performance	Ensures AI tools are reliable, accurate, and safe across patient populations
Accountability and Liability	Define assistive vs autonomous roles Standardize documentation of AI-influenced decisions Establish escalation protocols Clarify institutional liability	Maintains clear responsibility and protects clinicians and patients
Fairness and Bias Monitoring	Report training dataset demographics Evaluate performance across demographics Conduct periodic bias audits Transparently disclose limitations	Prevents inequities and promotes equitable healthcare delivery
Explainability and Transparency	Summarize model purpose and limitations Provide understandable methodology and confidence metrics Educate clinicians on interpretation	Builds trust, facilitates informed clinical decisions, and reduces "black-box" risks
Adaptive Oversight and Continuous Monitoring	Continuous performance surveillance Trigger thresholds for review Periodic recertification Multidisciplinary oversight committees	Detects model drift, adapts to changing populations, and ensures ongoing safety and compliance

Ethical Imperatives and Policy Considerations

The integration of AI into the field of medicine raises some fundamental issues. For example, how should the concept of informed consent change to accommodate the impact of AI on medical decisions? Can the allocation of resources based on the predictions of analytics potentially harm certain groups of people? Does the increasing reliance on the recommendations of algorithms affect the autonomy of the practitioner?

Governance structures need to be congruent with the ethical principles of beneficence, non-maleficence, autonomy, and justice.[12,15] Trusting relationships with the patient require transparency about the use of AI in patient care and the avoidance of unintended patient harm.[21,25] As the complexity of AI systems increases, including the potential use of multimodal data and semi-autonomous decision-making processes, the maturity of the oversight processes must also increase. The healthcare industry is at an important juncture. It can choose to use AI reactively and deal with the consequences of harm afterwards, or it can use structured approaches to the governance of AI. The need for AI governance in healthcare is not optional. It is an essential component of ensuring patient well-being and the integrity of the patient-physician relationship.

CONCLUSION

Artificial intelligence has transitioned from being a promise to being a reality in the field of clinical medicine. The tools that impact diagnosis, documentation, prediction, and allocation are now being incorporated into practice. However, the governance structure is not well established or aligned to the evolving nature of these tools. A structure that is based on safety, accountability, equity, fairness, and oversight is a vital structure that should be used to ensure the safe implementation of these AI technologies. By incorporating a governance structure in the health organization, the maximum potential of artificial intelligence can be realized. Not only will the future of artificial intelligence in medicine be based on the potential of the technology itself, but also on the governance structure that is put in place to ensure the safe implementation of these tools.

REFERENCES

1. Sanchez M, Alford K, Krishna V, Huynh TM, Nguyen CDT, Lungren MP, et al. AI-clinician collaboration via disagreement prediction: A decision pipeline and retrospective analysis of real-world radiologist-AI interactions. Cell Rep Med. 2023;4(10):101207. doi: 10.1016/j.xcrm.2023.101207.
2. Almanaa M. Trends and Public Perception of Artificial Intelligence in Medical Imaging: A Social Media Analysis. Cureus. 2024;16(9):e70008. doi: 10.7759/cureus.70008.
3. Joshi G, Jain A, Araveeti SR, Adhikari S, Garg H, Bhandari M. FDA-Approved Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices: An Updated Landscape. Electronics. 2024;13(3):498. doi: 10.3390/electronics13030498.

4. Almarie B, Gonzalez-Gonzalez LF, Dos Santos Barbosa LA, Lutz A, Grosse U, Fregni F. Machine Learning-Enabled Medical Devices Authorized by the US Food and Drug Administration in 2024: Regulatory Characteristics, Predicate Lineage, and Transparency Reporting. Biomedicines. 2025;13(12):3005. doi: 10.3390/biomedicines13123005.
5. van Kolfschooten H, van Oirschot J. The EU Artificial Intelligence Act (2024): Implications for healthcare. Health Policy. 2024;149:105152. doi: 10.1016/j.healthpol.2024.105152.
6. Gilbert S. The EU passes the AI Act and its implications for digital medicine are unclear. NPJ Digit Med. 2024;7(1):135. doi: 10.1038/s41746-024-01116-6.
7. van de Sande D, van Genderen ME, Huiskens J, Gommers D, van Bommel J. Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. Intensive Care Med. 2021;47(7):750-760. doi: 10.1007/s00134-021-06446-7.
8. Cerezo-Martínez P, Nicolás-Sánchez A, Castro-Toledo FJ. Analyzing the European institutional response to ethical and regulatory challenges of artificial intelligence in addressing discriminatory bias. Front Artif Intell. 2024;7:1393259. doi: 10.3389/frai.2024.1393259.
9. Kubba A, Khunda A, Al-Ameri L, Hameed EK. Artificial Intelligence in Medical Education: Balancing Emerging Opportunities with Real-World Limitations. Al-Kindy Col Med J. 2025;21(3):165-7. doi: 10.47723/tsa6a061.
10. Shulha M, Hovdebo J, D'Souza V, Thibault F, Harmouche R. Integrating Explainable Machine Learning in Clinical Decision Support Systems: Study Involving a Modified Design Thinking Approach. JMIR Form Res. 2024;8:e50475. doi: 10.2196/50475.
11. Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. NPJ Digit Med. 2019;2:1–7. doi: 10.48550/arXiv.1905.05134.
12. Wang AQ, Karaman BK, Kim H, Rosenthal J, Saluja R, Young SI, et al. A Framework for Interpretability in Machine Learning for Medical Imaging. IEEE Access. 2024;12:53277-53292. doi: 10.1109/access.2024.3387702.
13. Vigdorovits A, Köteles MM, Olteanu GE, Pop O. Breaking Barriers: AI's Influence on Pathology and Oncology in Resource-Scarce Medical Systems. Cancers (Basel). 2023;15(23):5692. doi: 10.3390/cancers15235692.
14. Muter S, Al-Ani N. A Cautionary Perspective on Artificial Intelligence and Novel Imaging Technologies in Patient Selection for Retrograde Intrarenal Surgery. Al-Kindy Col Med J. 2025;21(2):161-4. doi: 10.47723/e4j95114.

15. Beauchamp T, Childress J. Principles of Biomedical Ethics: Marking Its Fortieth Anniversary. Am J Bioeth. 2019;19(11):9-12. doi: 10.1080/15265161.2019.1665402.
16. Eriksen J, Ebbesen M, Eriksen KT, Hjermitslev C, Knudsen C, Bertelsen P, et al. Equity in digital healthcare - the case of Denmark. Front Public Health. 2023;11:1225222. doi: 10.3389/fpubh.2023.1225222.
17. Chang T, Nuppnau M, He Y, Kocher KE, Valley TS, Sjoding MW, et al. Racial differences in laboratory testing as a potential mechanism for bias in AI: A matched cohort analysis in emergency department visits. PLOS Glob Public Health. 2024;4(10):e0003555. doi: 10.1371/journal.pgph.0003555.
18. Rajkomar A, Hardt M, Howell MD, Corrado G, Chin MH. Ensuring Fairness in Machine Learning to Advance Health Equity. Ann Intern Med. 2018;169(12):866-872. doi: 10.7326/M18-1990.
19. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. 2019;366(6464):447-453. doi: 10.1126/science.aax2342.
20. Sendak MP, D'Arcy J, Kashyap S, Gao M, Nichols M, Corey K, et al. A path for translation of machine learning products into healthcare delivery. EMJ Innov. 2020;4(1):1–8. doi: 10.33590/emjinnov/19-00172.
21. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med. 2019;25(1):44-56. doi: 10.1038/s41591-018-0300-7.
22. 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(3):283-286. doi: 10.1056/NEJMc2104626.
23. Rudin C. Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nat Mach Intell. 2019;1(5):206-215. doi: 10.1038/s42256-019-0048-x.
24. Goddard K, Roudsari A, Wyatt JC. Automation bias: a systematic review of frequency, effect mediators, and mitigators. J Am Med Inform Assoc. 2012;19(1):121-7. doi: 10.1136/amiajnl-2011-000089.
25. Kellogg KC, Sadeh-Sharvit S. Pragmatic AI-augmentation in mental healthcare: Key technologies, potential benefits, and real-world challenges and solutions for frontline clinicians. Front Psychiatry. 2022;13:990370. doi: 10.3389/fpsyt.2022.990370.
26. Javaid A, Zghyer F, Kim C, Spaulding EM, Isakadze N, Ding J, et al. Medicine 2032: The future of cardiovascular disease prevention with machine learning and digital health technology. Am J Prev Cardiol. 2022;12:100379. doi: 10.1016/j.ajpc.2022.100379.