Healthcare and Medical Devices, Vol. 79, 2023, 1–6
https://doi.org/10.54941/ahfe1003466

Artificial Intelligence in Healthcare: The Explainability Ethical Paradox

Patrick Seitzinger[1,2] and Jawahar (Jay) Kalra[3,4]

[1]Department of Pediatrics, College of Medicine, University of Saskatchewan, 103 Hospital Drive, Saskatoon, S7N 0W8, Canada
[2]Jim Pattison Children's Hospital, Saskatchewan Health Authority, 103 Hospital Drive, Saskatoon, S7N 0W8, Canada
[3]Department of Pathology and Laboratory Medicine, College of Medicine, University of Saskatchewan, Saskatoon, Canada
[4]Royal University Hospital, Saskatchewan Health Authority, 103 Hospital Drive, Saskatoon, Saskatchewan, S7N 0W8, Canada

ABSTRACT

Explainability is among the most debated and pivotal discussions in the advancement of Artificial Intelligence (AI) technologies across the globe. The development of AI in medicine has reached a tipping point in medicine with implications across all sectors. How we proceed with the issue of explainability will shape the direction and manner in which healthcare evolves. We require new tools that bring us beyond our current levels of medical understanding and capabilities. However, we limit ourselves to tools that we can fully understand and explain. Implementing a tool that cannot be fully understandable by clinicians or patients violates medical ethics of informed consent and autonomy. Yet, denying patients and the population attainable benefits of a new resource violates medical ethics of justice, health equity and autonomy. Fear of the unknown is not by itself a reason to halt the progression of medicine. Many of our current advancements were implemented prior to fully understanding its intricacies. To convey competence, some subfields of AI research have emphasized validity testing over explainability as a way to verify accuracy and build trust in AI systems. As a tool AI has shown immense potential in idea generation, data analysis, and pattern identification. AI will never be an independent system and will always require human oversight to ensure healthcare quality and ethical implementation. By using AI to augment, rather than replace clinical judgement, the caliber of patient care that we provide can be enhanced in a safe and sustainable manner. Addressing the explainability paradox in AI requires a multidisciplinary approach to address the technical, legal, medical, and ethical aspects of this challenge.

Keywords: Artificial intelligence, Healthcare quality, Explainability, Explainability paradox, Transparency, Ethics, Black-box

INTRODUCTION

Technological advancements have provided the ability to perform tasks that would otherwise be unattainable. These technological advances often pose novel ethical dilemmas that require careful consideration balancing benefits

and risks to ensure the safety and sound progression of healthcare quality. Healthcare quality improvement aims to develop tools and processes to expand our capabilities and understanding. We require new tools that bring us beyond our current levels of medical understanding and capabilities. However, we limit ourselves to tools that we can fully understand and explain. This juxtaposition of healthcare innovation and understandability has created an ethical paradox when it comes to providing the highest possible calibre of medical care in a timely manner.

AI is an example of such a tool with tremendous potential to facilitate desired outcomes and is being implemented without a complete grasp of its underlying processes. Artificial intelligence refers to technologies that can perform tasks without specifically being programmed to do so. The inherent nature of AI makes it susceptible to multiple factors relevant to explainability. AI systems are only as accurate as the data with which it is provided (Seitzinger et al. 2021). Due to various sources of error, heterogeneous data, and biases in the underlying assumptions and input data, AI technology is unlikely to ever reach complete accuracy (Amann et al. 2020, Reddy, 2022). The challenges of explainability raise various ethical, societal, and legal challenges (Amann et al. 2020). The development of AI in medicine has reached a tipping point in medicine with implications across all sectors.

Explainability is among the most debated and pivotal discussions in the advancement of artificial intelligence technologies (Amann et al. 2020). Explainability refers to the ability to have systems understood by the user, including inputs, the process implemented, and the rationale for conclusions that were drawn (Rudin, 2019). It has been described as the intersection between usefulness, understandability and usability (Combi et al. 2022). Explainability reduces development time and expenses (Amann et al. 2020). How healthcare systems navigate the issue of explainability will shape the direction and manner in which these systems evolve and progress. The circumstances of modern medicine represent a timely and necessary opportunity to address the explainability paradox in healthcare improvement.

NECESSITIES OF ETHICAL MEDICAL CARE

Transparency is a cornerstone of clinical care and therapeutic relationships (Amann et al. 2020, Kundu, 2021, Yoon et al. 2021, Reddy, 2022). When it comes to providing medical care for patients, high standards the transparency, trustworthiness, and standards of care are not only a privilege, but a necessity (Olsen et al. 2019, Amann et al. 2020) In order to gain trust and acceptance of new processes, clinicians and patients must be able to explain and understand these processes (Kundu, 2021, Reddy, 2022). Implementing a tool that cannot be fully understandable by clinicians or patients violates medical ethics of informed consent, non-maleficence, and autonomy (Kundu, 2021, Yoon et al. 2021, Reddy, 2022). To date the advancement of Western Medicine whether it be pharmaceuticals or medical devices has relied on validation processes such as randomized control trials, to ensure patient safety (Reddy, 2022). The United States Food and Drug Administration demands an 'appropriate amount' of transparency and clarity in the design and outputs of

AI systems used in healthcare (Amann et al. 2020). Inadequate transparency is a pivotal reason for the limited adoption of AI technologies, especially those considered black box algorithms (Amann et al. 2020, Reddy, 2022). Without carefully examining the role of explainability in medical AI, these technologies might violate fundamental moral and professional standards, neglect legal and regulatory requirements, and result in significant harm (Obermeyer et al. 2019, Amann et al. 2020).

MEDICAL ETHICS IN THE AI-ASSISTED ERA OF MEDICAL CARE

In order to be implemented in healthcare settings, medical AI systems are required to withstand rigorous testing to ensure adequate clinical validation (Higgins and Madai 2020, Amann et al. 2020). As the complexity of tools healthcare tools evolves, so must the implementation of concepts of medical ethics. Modern deep learning models have achieved levels of performance and necessary complexity with billions of parameters (Marcus et al. 2018, Reddy, 2022). The roles of each parameter can only be appropriately understood in the context of relationships to other similarly contingent and complex parameters within the models (Marcus et al. 2018, Reddy, 2022). Due to the complexity and fluid nature of these models, traditional definitions of explainability cannot be easily applied to AI systems (Marcus et al. 2018, Reddy, 2022). Given the challenges of applying traditional methods of validity testing to AI systems, different techniques have been implemented to approximate explainability. Emerging fields of AI have moved toward validity measures of AI programs rather than explainability to convey the competence and trustworthiness of these systems (Ghassemi et al. 2021, Cutillo et al. 2020, Reddy, 2022). However, currently, only exploratory beginning attempts to quantitatively rank explainability methodologies exist (Islam et al. 2019, Amann et al. 2020).

DUTY TO PROVIDE THE HIGHEST POSSIBLE CALIBRE OF PATIENT CARE

Medical AI systems have the potential to make considerable advancements in the calibre of healthcare delivery that can be provided to patients. Denying patients and populations attainable benefits of a new tool or resource violates medical ethics of justice, health equity and autonomy. As a tool AI has shown immense potential in idea generation, data analysis, and pattern identification. The inherent benefit of AI is also its ability to recognize novel patterns and discover new biomarkers without the need for pre-selection of traits (Amann et al. 2020). The potential for these systems to augment idea generation, pattern recognition, and synthesize the newest available evidence has tremendous potential to alleviate stressors in the healthcare system. It has been demonstrated that AI-powered systems produce overall lower error rates than conventional techniques (Weng et al. 2017, Kakadiaris et al. 2018, Liu et al. 2019, Amann et al. 2020).

ACCEPTABLE LEVELS OF RISK AND EXPLAINABILITY

Balancing innovation and advancement of medical care with quality assurance requires clearly defined permissibility parameters when it comes to lack of explainability. The advancement of medicine and healthcare has included countless tools and treatments that have provided benefits without a complete understanding of precise underlying pathophysiological processes. Public health practitioners and researchers have started employing AI in a variety of tasks, including scanning for emerging outbreaks around the globe (Dion et al. 2015). In typical healthcare interactions, clinicians do not explain their sources of information, and data when making day-to-day clinical decisions. Similar to getting consent for Magnetic Resonance Imaging, the patient typically does not need to understand the details of the physics and mechanics of the technique but must be able to understand the fundamental concept of the test and the risks associated with it (Amann et al. 2020).

The necessity for specialized training and professional development in the field of medical AI is clearly highlighted by the fact that this also raises significant questions regarding the role and responsibilities of physicians (Amann et al. 2020). Doctors and patients need to be informed about the underlying processes of data input and assumptions on which conclusions derived from AI systems are based (Amann et al. 2020). An understanding of these components is necessary to implement input from AI systems in the right circumstances, for the appropriate cases in which the data and assumptions used by AI systems are likely to be accurate. The extent to which the patient must be informed that treatment choices, such as those made by a clinical decision support system, may be influenced by AI, as well as the legal and litigation implications of whether the doctor followed the machine's recommendation or disregarded it (Amann et al. 2020).

NEXT STEPS

Innovation and regulation are often at odds, and this balance requires careful deliberation and anticipation of potential implications (Amann et al. 2020). A certain degree of skepticism is not only warranted but is necessary for the implementation of AI. It is time to define clearly what are permissible levels of explainability in the context of medical AI systems in healthcare at a practical and regulatory level. It is a timely opportunity to define the bounds of what the technology can be used for with our current understanding and move ahead with implementation. At a regulatory level, a framework for determining the appropriate level of explainability is required (Beaudouin et al. 2020, Amann et al. 2020). Perfect accuracy is unattainable due to inherently flawed medical datasets and as such should not be a requirement or reason to delay progress in legislation related to AI systems in healthcare (Amann et al. 2020). Clarification is required on the liability of AI-assisted medical decisions, both in the context of incorporating AI outputs as well as choosing to ignore information provided by AI systems. On a clinical level, explainability can help physicians assess a system's suggestions in light of their clinical expertise and experience. Clinicians are not to follow blindly the recommendations of AI. Similar to other decision-making tools such as scoring tools

and decision trees AI systems enable users to decide for themselves whether or not to believe the system's recommendations as one of many data points used to inform clinical judgement A learning opportunity for clinicians to hone skills, and be aware of newest evidence.

CONCLUSION

Explainability is likely to play a pivotal role in the adoption of new AI technologies and the progress of advancements of medical processes (Amann et al. 2020). The goal of such tools is to allow us to do that which we otherwise could not. Fear of the unknown is not by itself a reason to halt the progression of medicine. Many of the tools and techniques we currently use remain beyond our understanding. Artificial intelligence has demonstrated the potential they have transformative effects on many aspects of medical care. However, despite its clear potential, AI is not a panacea. Some aspects of the block box model are unavoidable in the context of complex AI systems that have the ability to learn and modify their underlying algorithms. Implementing a tool that cannot be fully understandable by clinicians or patients violates medical ethics of informed consent and autonomy. Yet, denying patients and the population attainable benefits of a new resource violates medical ethics of justice, health equity and autonomy. The need for a multidisciplinary approach is necessitated by the technological aspects of AI in certain cases and the legal, medical, and patient viewpoints in others (Amann et al. 2020). Implementation must not be considered an all-or-nothing approach. The contribution and potential AI should be gauged on its capability as a data point in clinical decision-making processes, guided by human oversight and clinical judgement.

REFERENCES

Amann, J., Blasimme, A., Vayena, E., et al. (2020) Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Medical Informatics and Decision Making [online], 20(1): 1–9.

Beaudouin, V., Bloch, I., Bounie, D., et al. (2020) Identifying the 'Right' Level of Explanation in a Given Situation. CEUR Workshop Proceedings [online], 2659: 63–66.

Combi, C., Amico, B., Bellazzi, R., et al. (2022) A manifesto on explainability for artificial intelligence in medicine. Artificial Intelligence in Medicine, 133.

Cutillo, C. M., Sharma, K. R., Foschini, L., et al. (2020) Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency. npj Digital Medicine [online], 3(1): 1–5.

Dion, M., AbdelMalik, P. and Mawudeku, A. (2015). Big Data: Big Data and the Global Public Health Intelligence Network (GPHIN). Canada Communicable Disease Report [online], 41(9): 209.

Ghassemi, M., Oakden-Rayner, L. and Beam, A. L. (2021) The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11): e745–e750.

Higgins, D. and Madai, V. I. (2020) From Bit to Bedside: A Practical Framework for Artificial Intelligence Product Development in Healthcare. Advanced Intelligent Systems [online], 2(10): 2000052.

Islam, S. R., Eberle, W. and Ghafoor, S. K. (2019) Towards Quantification of Explainability in Explainable Artificial Intelligence Methods. Proceedings of the 33rd International Florida Artificial Intelligence Research Society Conference, FLAIRS 2020 [online], 22 November 2019: 75–81.

Kakadiaris, I. A., Vrigkas, M., Yen, A. A., et al. (2018) Machine Learning Outperforms ACC / AHA CVD Risk Calculator in MESA. Journal of the American Heart Association [online], 7(22).

Kundu, S. (2021). AI in medicine must be explainable. Nature Medicine, 27(8): 1328.

Liu, T., Fan, W. and Wu, C. (2019) A hybrid machine learning approach to cerebral stroke prediction based on imbalanced medical dataset. Artificial intelligence in medicine [online], 101.

Marcus, G., Christina, I., Chollet, F., et al. (2018) Deep Learning: A Critical Appraisal. arXiv preprint, 1801: 631.

Obermeyer, Z., Powers, B., Vogeli, C., et al. (2019) Dissecting racial bias in an algorithm used to manage the health of populations. Science (New York, N. Y.) [online], 366(6464): 447–453.

Olsen, H. P., Slosser, J. L., Hildebrandt, T. T. and Wiesener, C. (2019) What's in the Box? The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration. SSRN Electronic Journal, 162: 2019–2084.

Reddy, S. (2022) Explainability and artificial intelligence in medicine. The Lancet Digital Health [online], 4(4): e214–e215.

Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. Nature machine intelligence [online], 1(5): 206–215.

Seitzinger, P., Rafid-Hamed, Z. and Kalra, J. (2021) Healthcare Delivery: Leveraging Artificial Intelligence to Strengthen Healthcare Quality. In: Lecture Notes in Networks and Systems. Springer, Cham, 263: 16–21.

Weng, S. F., Reps, J., Kai, J., et al. (2017) Can machine-learning improve cardiovascular risk prediction using routine clinical data?. PLOS ONE [online], 12(4): e0174944.

Yoon, C. H., Torrance, R. and Scheinerman, N. (2021). Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned?. Journal of medical ethics 48(9): 581–585.