Challenges and Ethical Considerations to Successfully Implement Artificial Intelligence in Clinical Medicine and Neuroscience: a Narrative Review

Authors
Scott Monteith[1], Tasha Glenn[2], John R. Geddes[3], Eric D. Achtyes[4], Peter C. Whybrow[5], Michael Bauer[6]

Affiliations
1 Department of Psychiatry, Michigan State University College of Human Medicine, Traverse City Campus, Traverse City, MI, USA
2 ChronoRecord Association, Fullerton, CA, USA
3 Department of Psychiatry, University of Oxford, Warneford Hospital, Oxford, UK
4 Department of Psychiatry, Western Michigan University Homer Stryker M.D. School of Medicine, Kalamazoo, MI, USA
5 Department of Psychiatry and Biobehavioral Sciences, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles (UCLA), Los Angeles, CA, USA
6 Department of Psychiatry and Psychotherapy, University Hospital Carl Gustav Carus Faculty of Medicine, Technische Universität Dresden, Dresden, Germany

Key words
artificial intelligence, implementation, machine learning, safety

received 25.04.2023
revised 13.06.2023
accepted 16.06.2023
published online 29.08.2023

Bibliography
Pharmacopsychiatry 2023; 56: 209–213
DOI 10.1055/a-2142-9325
ISSN 0176-3679

Georg Thieme Verlag, Rüdigerstraße 14,
70469 Stuttgart, Germany

Correspondence
Scott Monteith, MD
Michigan State University College of Human Medicine
Traverse City Campus
1400 Medical Campus Drive
Traverse City, MI 49684
USA
monteit2@msu.edu

ABSTRACT

This narrative review discusses how the safe and effective use of clinical artificial intelligence (AI) prediction tools requires recognition of the importance of human intelligence. Human intelligence, creativity, situational awareness, and professional knowledge, are required for successful implementation. The implementation of clinical AI prediction tools may change the workflow in medical practice resulting in new challenges and safety implications. Human understanding of how a clinical AI prediction tool performs in routine and exceptional situations is fundamental to successful implementation. Physicians must be involved in all aspects of the selection, implementation, and ongoing product monitoring of clinical AI prediction tools.

Introduction

The use of artificial intelligence (AI) to augment human intelligence in medicine is expected to reshape healthcare [1, 2]. The safe and effective use of clinical AI prediction tools requires recognition of the importance of human involvement and the technical limitations of AI. Successful use of clinical AI prediction tools needs human intelligence, creativity, situational awareness, and professional knowledge to interpret and integrate results, and determine and handle exceptions. The implementation of clinical AI prediction tools may change the workflow in medical practice, resulting in multiple new challenges and safety implications. For example, research is focused on developing AI tools to predict treatment response to specific drugs, such as antidepressants [3–5]. In the future, this will change the workflow, requiring the physician to agree or disagree with the recommendation. To achieve the potential benefits, this narrative review will discuss some of the important

and diverse challenges involved in the successful implementation of clinical AI prediction tools in medicine.

Fundamental differences between human intelligence and artificial intelligence

The successful implementation of AI tools in medicine requires recognition of the unique importance of human involvement. Although human brains and computers are often compared, there are many fundamental differences. A single human brain stores roughly the same amount of information as the entire Internet [6, 7]. Typically, a human brain has about 200 billion nerve cells, connected via trillions of synapses, which is more than all the computers and routers and Internet connections on Earth [8, 9]. Human brain cells, primarily neurons and synapses, perform both data storage and data processing [10]. In contrast, computers separate data storage from data processing, and must spend considerable energy moving data [10]. The adult brain is extraordinarily energy efficient requiring only about 20 watts of power [10, 11]. A brain can perform an exaflop (billion-billion) mathematical operations per second with 20 watts of power, as compared to an advanced supercomputer that required 20 megawatts for the same computations, a million times more power [11].

Human intelligence is very different from AI. Human intelligence deals with uncertainty, and responds to very small amounts of data [12]. Humans evaluate the trustworthiness of new information, and integrate it with accumulated wisdom [13]. Humans frame decision making using mental models that allow us to understand and make abstractions [13, 14]. Causality is a fundamental aspect of human decision making, and causal knowledge underlies much of what humans do even if we don't understand the underlying mechanisms [15]. Human reasoning about cause and effect allows humans to ask why [16]. Human evaluation of information includes the creation of constraints, abstractions, and counterfactuals, such that different answers will be given to people having the same data. In contrast to human intelligence, AI systems do not understand causality [17]. AI does not capture human understanding that if x causes y, it does not mean that y causes x [17]. AI cannot ask why [16]. AI cannot create constraints or counterfactuals, or generate abstractions [14]. AI assumes that the same inputs will always result in the same prediction [18]. Judea Pearl noted that although the achievements of modern deep learning AI are impressive, they can be described today as "curve fitting" [17].

Artificial intelligence technical challenges

The physician should expect technical challenges during the implementation of clinical AI prediction tools. Currently most AI, including in medicine, is based on data-intensive machine learning (ML) methods [19, 20]. ML uses very large training datasets to determine the best model (data variables and equations) for predicting an outcome, with the model remaining an opaque black box. The accuracy of the clinical AI prediction tools is tied to the training data where better quality data produces better quality predictions. The electronic medical records (EMR) and claims data that are routinely used as training data in medicine have quality problems related to inaccuracy, missing data, biases, coding errors, lack of diversity, unrepresentative samples, and lack of vendor software interoperability [21]. There are additional data quality concerns for psychiatry due to the high frequency of missing behavioral health data in the EMR [21, 22]. Compared to nonmedical domains, the size of training data available for clinical AI prediction tools in psychiatry is much smaller [23], and a small training data size will decrease the accuracy of predictions [24, 25]. It is harder to test ML applications than conventionally coded applications [26], and there is no standard for communicating the amount of uncertainty in a ML prediction [27].

Another area of concern with the data used to train ML clinical AI prediction tools is dataset shift, where the data collected from the population used to train the model is different from the population where the model is deployed. When a clinical AI prediction tool is implemented in a setting where the patient population characteristics differs from the training data, AI often does not perform well [28–30]. Many diverse factors contribute to dataset shift in medicine including changes in patient demographics, standards of care, treatment practices, disease prevalence, and technology use [28]. Additionally, there is a reproducibility crises in all scientific fields that use ML [31], including healthcare [32], and a need to address the reproducibility challenge for clinical AI prediction tools [33–35]. The problem of reproducibility of ML models emphasizes the need for validated clinical AI prediction tools that have received approval from appropriate regulatory bodies.

Formal implementation process

A carefully considered, clearly defined process for implementing clinical AI technology in medical settings will improve the quality of the results. Despite the high expectations, there is a well documented productivity paradox, a delay in years between the adoption of a new technology and productivity increases, including in medicine [21]. The introduction of any new clinical technology, including AI prediction tools, will change the workflow, often in unexpected ways, and may result in new types of human errors and failure paths [36]. Physicians, clinical support staff, management and technical staff should all be involved in selection, implementation, and ongoing product monitoring and maintenance of clinical AI prediction tools [37]. This includes physician training, workflow changes and clinical impacts, product testing, integration into current systems, and ongoing monitoring and reporting of product performance and accuracy [37]. In medicine, many clinical AI prediction tools are developed internally. This is very expensive in the long-term, as the ongoing costs to maintain reliable systems are much higher for ML than for traditional software [38]. Vendor contracts for clinical AI prediction tools should explicitly define responsibilities related to ML maintenance, administration, and enhancements.

Expect artificial intelligence errors

A fundamental assumption by physicians in an AI implementation should be that clinical AI prediction tools will make errors. There will be errors from any predictive model, whether based on AI or traditional statistics. The consequences of false positives, false negatives, and other errors will vary with the situation [24]. Much of the commercial use of AI is for low risk situations, such as a product recommendation on Amazon, where the costs of errors are financial [24]. In contrast, the potential impacts of errors from clinical AI prediction tools in medicine emphasize the need for human

oversight and error tracking. The clinical AI prediction tools must perform in exceptional situations and boundary cases, as well as during routine activities. AI tools are not good at predicting rare events [24]. The results of AI prediction tools may conflict with current practice guidelines [39]. The results of an AI prediction could be plausible but incorrect, and potentially dangerous for an individual patient [40]. Additionally, when the result from an opaque black box algorithm is incorrect, it is not clear what went wrong or what should be fixed [41]. Physicians who learned their skills based on interpretation of raw data values may not function as well when only a prediction is presented [36]. AI failures may lead to unexpected safety hazards not seen previously, emphasizing the importance of physician training on potential ML dysfunction [42].

A comprehensive implementation plan for AI will include error tracking and implementing enhancements as an integral part of using any ML tool. There must be clearly defined and documented methods for physicians who use clinical AI prediction tools to identify, record and track errors. Physicians must understand what is expected of them in relation to error reporting and tracking. A framework for continuous tracking and reporting of errors in clinical AI prediction tools is especially important with ML, due to lack of model transparency, including for some Food and Drug Administration- approved ML tools [43].

Central importance of humans

An AI implementation plan must recognize that humans are central to the successful use of clinical AI prediction tools. The diverse types of errors from clinical AI prediction tools, and potential for serious consequences, highlights the importance and need for human intelligence in the process [24]. Increasing complexity in an automated system increases the need for human judgement and situational awareness when unexpected errors occur [36, 44]. The implementation of clinical AI prediction tools must ensure adequate physician review such that predictions can be overridden if necessary. Unlike ML, human clinical decision making is tied to context [36, 45]. The black-box nature of ML can make it difficult to detect biases, understand, or trust the results [43, 46]. Although there are ongoing efforts to provide explainability to ML models, there are limitations and drawbacks to explainability techniques [47, 48]. There are also complex problems in medicine where humans disagree on what is the best solution. Concern about the quality of evidence available for many clinical AI prediction tools is widespread, along with recognition of the need to improve and expand government regulation [49, 50].

Unintended consequences of artificial intelligence

The implementation of any new technology, including AI, results in unintended consequences [36, 51, 52]. The introduction of AI into routine clinical practice can lead to overreliance on the technology, automation bias and complacency. Automation bias occurs when the user gives greater authority to automated advice than to other sources of advice [53]. The risk for automation bias increases when it is hard to verify if automation is performing correctly, as found in clinical medicine [54]. Automation complacency occurs when a user in a multitasking environment focuses on the manual tasks, not noticing errors in the automated tasks [53]. One concern is that most psychiatrists have no formal training in technology and may be unaware of the risks and drawbacks of AI [55]. Another possible consequence of using AI is that overreliance on AI will lead to deskilling, reducing the clinical knowledge and the patient communications and examination skills of a physician [45, 56].

Limitations

There are many limitations to this discussion. With a focus on implementation, the important topic of validation standards was not discussed. Unsolved technical issues with ML that may negatively impact safety, including biases, the presentation of uncertainty in results [57] and cybersecurity [58, 59], were not discussed. The use of interpretable rather than ML models was not discussed [60]. The unique challenges for ongoing governing and regulating of ML in healthcare were not reviewed [49, 61]. Specific measures to mitigate the risks of implementation of AI were not discussed. Legal issues including physician responsibility and liability for related errors made by AI products were not included [62, 63].

Conclusion

The use of clinical AI prediction tools should emphasize the importance of human intelligence. The fundamental determinant of implementation success will be human understanding of how AI technology performs in routine and exceptional situations. Human intelligence, creativity, situational awareness, and professional knowledge is required to interpret, integrate, handle results, and recognize exceptions. Implementation of a clinical AI prediction tool may assist a physician, after the results are evaluated in the appropriate context for the individual patient. Physicians must be involved in all aspects of the selection, implementation, and ongoing product monitoring of clinical AI prediction tools.

Author Contributions

SM and TG wrote the initial draft. All authors reviewed and approved the final manuscript.

Conflict of Interest

The authors have no conflicts of interest to declare.

References

[1] Haug CJ, Drazen JM. Artificial intelligence and machine learning in clinical medicine, 2023. N Engl J Med 2023; 388: 1201–1208

[2] Yu KH, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018; 2: 719–731

[3] Chekroud AM, Bondar J, Delgadillo J et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 2021; 20: 154–170

[4] Kambeitz-Ilankovic L, Koutsouleris N, Upthegrove R. The potential of precision psychiatry: What is in reach? Br J Psychiatry 2022; 220: 175–178

[5] Lin E, Lin CH, Lane HY. Precision psychiatry applications with pharmacogenomics: Artificial intelligence and machine learning approaches. Int J Mol Sci 2020; 21: 969

[6] Bartol TM Jr, Bromer C, Kinney J et al. Nanoconnectomic upper bound on the variability of synaptic plasticity. eLife 2015; 4: e10778

[7] Ghose T. The human brain's memory could store the entire internet. Live Science 2016 https://www.livescience.com/53751-brain-could-store-internet.html

[8] Micheva KD, Busse B, Weiler NC et al. Single-synapse analysis of a diverse synapse population: proteomic imaging methods and markers. Neuron 2010; 68: 639–653

[9] Moore EA. Human brain has more switches than all computers on Earth. CNET 2010 https://www.cnet.com/tech/computing/human-brain-has-more-switches-than-all-computers-on-earth/

[10] Mehonic A, Kenyon AJ. Brain-inspired computing needs a master plan. Nature 2022; 604: 255–260

[11] Madhavan A. Brain-inspired computing can help us create faster, more energy-efficient devices – if we win the race. Taking Measure 2023 https://www.nist.gov/blogs/taking-measure/brain-inspired-computing-can-help-us-create-faster-more-energy-efficient

[12] Gigerenzer G. One data point can beat big data. Behavioral Scientist 2022 https://behavioralscientist.org/gigerenzer-one-data-point-can-beat-big-data/

[13] Gopnik A. AIs versus four-year olds. In: Brockman J, ed. Possible minds: Twenty-five ways of looking at AI. New York: Penguin; 2020: p219–p230

[14] Cukier K, Mayer-Schönberger V, de Véricourt F. Framers: Human advantage in an age of technology and turmoil. Penguin; 2022

[15] Marcus G, Davis E. Insights for AI from the human mind. Communications of the ACM 2020; 64: 38–41

[16] Pearl J, Mackenzie D. The book of why: The new science of cause and effect. New York, NY: Basic books; 2018

[17] Hartnett K. Quanta Magazine. How a pioneer of machine learning became one of its sharpest critics. The Atlantic 2018 https://www.theatlantic.com/technology/archive/2018/05/machine-learning-is-stuck-on-asking-why/560675/

[18] Gilder G, Gaming AI. Discovery Institute Press 2020

[19] Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015; 349: 255–260

[20] Rajkomar A, Dean J, Kohane I. Machine learning in medicine. N Engl J Med 2019; 380: 1347–1358

[21] Monteith S, Glenn T, Geddes J et al. Expectations for artificial intelligence (AI) in psychiatry. Curr Psychiatry Rep 2022; 24: 709–721

[22] Madden JM, Lakoma MD, Rusinak D et al. Missing clinical and behavioral health data in a large electronic health record (EHR) system. J Am Med Inform Assoc 2016; 23: 1143–1149

[23] Bzdok D, Meyer-Lindenberg A. Machine learning for precision psychiatry: Opportunities and challenges. Biol Psychiatry Cogn Neurosci Neuroimaging 2018; 3: 223–230

[24] Agrawal A, Gans J, Goldfarb A. Prediction machines: The simple economics of artificial intelligence. Boston, MA: Harvard Business Review Press; 2022

[25] Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. Neuroimage 2018; 180: 68–77

[26] Shwartz S. Evil robots, killer computers, and other myths: The truth about AI and the future of humanity. Greenleaf Book Group 2021

[27] Kompa B, Snoek J, Beam AL. Second opinion needed: communicating uncertainty in medical machine learning. NPJ Digit Med 2021; 4: 4

[28] Finlayson SG, Subbaswamy A, Singh K et al. The clinician and dataset shift in artificial intelligence. N Engl J Med 2021; 385: 283–286

[29] Subbaswamy A, Saria S. From development to deployment: Dataset shift, causality, and shift-stable models in health AI. Biostatistics 2020; 21: 345–352

[30] Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: Insights from multi-site COVID-19 screening. NPJ Digit Med 2022; 5: 69

[31] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in ML-based science. arXiv preprint arXiv:2207.07048 2022

[32] McDermott MB, Wang S, Marinsek N et al. Reproducibility in machine learning for health research: Still a ways to go. Sci Transl Med 2021; 13: eabb1655

[33] Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. JAMA 2020; 323: 305–306

[34] Rajpurkar P, Chen E, Banerjee O et al. AI in health and medicine. Nat Med 2022; 28: 31–38

[35] Sohn E. The reproducibility issues that haunt health-care AI. Nature 2023; 613: 402–403

[36] Bauer M, Monteith S, Geddes J et al. Automation to optimise physician treatment of individual patients: examples in psychiatry. The Lancet Psychiatry 2019; 6: 338–349

[37] Daye D, Wiggins WF, Lungren MP et al. Implementation of clinical artificial intelligence in radiology: Who decides and how? Radiology 2022; 305: 555–563

[38] Sculley D, Holt G, Golovin D et al. Hidden technical debt in machine learning systems. Adv Neural Inf Process Syst 2015; 28:

[39] Vellido A. Societal issues concerning the application of artificial intelligence in medicine. Kidney Diseases 2019; 5: 11–17

[40] Whitby B. Automating medicine the ethical way. In: Machine medical ethics, 2015. Springer, Cham. Van Rysewyk SP and Pontier M, eds.. 2015: 223–232

[41] Pearl J. The limitations of opaque learning machines. Possible minds. In: Brockman J, editor. Possible minds: Twenty-five ways of looking at AI. Penguin; 2020: 13–19

[42] Mongan J, Kohli M. Artificial intelligence and human life: Five lessons for radiology from the 737 MAX disasters. Radiol Artif Intell 2020; 2: e190111

[43] Bizzo BC, Dasegowda G, Bridge C et al. Addressing the challenges of implementing artificial intelligence tools in clinical practice: Principles from experience. J Am Coll Radiol 2023; 20: 352–360

[44] Strauch B. Ironies of automation: Still unresolved after all these years. IEEE Transactions on Human-Machine Systems 2017; 48: 419–433

[45] Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. JAMA 2017; 318: 517–518

[46] Quinn TP, Jacobs S, Senadeera M et al. The three ghosts of medical AI: Can the black-box present deliver? Artif Intell Med 2022; 124: 102158

[47] Babic B, Gerke S, Evgeniou T et al. Beware explanations from AI in health care. Science 2021; 373: 284–286

[48] Petch J, Di S, Nelson W. Opening the black box: The promise and limitations of explainable machine learning in cardiology. Can J Cardiol 2022; 38: 204–213

[49] Mashar M, Chawla S, Chen F et al. Artificial intelligence algorithms in health care: Is the current Food and Drug Administration regulation sufficient? JMIR AI 2023; 2: e42940

[50] Niemiec E. Will the EU Medical Device Regulation help to improve the safety and performance of medical AI devices? Digital Health 2022; 8: 20552076221089079

[51] Challen R, Denny J, Pitt M et al. Artificial intelligence, bias and clinical safety. BMJ Qual Saf 2019; 28: 231–237

[52] Sujan M, Furniss D, Grundy K et al. Human factors challenges for the safe use of artificial intelligence in patient care. BMJ Health Care Inform 2019; 26: e100081

[53] Parasuraman R, Manzey DH. Complacency and bias in human use of automation: An attentional integration. Human Factors 2010; 52: 381–410

[54] Lyell D, Coiera E. Automation bias and verification complexity: A systematic review. J Am Med Inform Assoc 2017; 24: 423–431

[55] Bauer R, Glenn T, Monteith S et al. Survey of psychiatrist use of digital technology in clinical practice. Int J Bipolar Disord 2020; 8: 1–9

[56] Hoff T. Deskilling and adaptation among primary care physicians using two work innovations. Health Care Manage Rev 2011; 36: 338–348

[57] Tomsett R, Preece A, Braines D et al. Rapid trust calibration through interpretable and uncertainty-aware AI. Patterns 2020; 1: 100049

[58] DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. J Am Med Inform Assoc 2020; 27: 2020–2023

[59] Hendrycks D, Carlini N, Schulman J et al. Unsolved problems in ML safety. arXiv preprint arXiv:2109.13916 2021

[60] Rudin C, Radin J. Why are we using black box models in AI when we don't need to? A lesson from an explainable AI competition. Harvard Data Science Review 2019; 1: https://hdsr.mitpress.mit.edu/pub/f9kuryi8/release/8

[61] Zanca F, Brusasco C, Pesapane F et al. Regulatory aspects of the use of artificial intelligence medical software. Semin Radiat Oncol 2022; 32: 432–441

[62] Price WN, Gerke S, Cohen IG. Potential liability for physicians using artificial intelligence. JAMA 2019; 322: 1765–1766

[63] Stöger K, Schneeberger D, Holzinger A. Medical artificial intelligence: The European legal perspective. Communications of the ACM 2021; 64: 34–36