


EDITORIAL



Five things every clinician should know about AI ethics in intensive care

James A. Shaw^{1,2*} , Nayha Sethi³ and Brian L. Block⁴

© 2020 Springer-Verlag GmbH Germany, part of Springer Nature

You have just admitted two patients to your intensive care unit (ICU) with coronavirus disease 2019 (COVID-19), both needing intubation. You only have the resources to offer mechanical ventilation to one of them. In your view, both are equally ill and warrant a trial of mechanical ventilation. Your hospital uses artificial intelligence (AI) to make recommendations for the allocation of scarce resources, to reduce subjectivity and remove treating clinicians from triage decisions. Without showing the data or reason behind its decision, the algorithm recommends offering mechanical ventilation to one of the patients, who is White, rather than the other, who is Black. You wonder why the algorithm made this recommendation and whether it is morally “right”.

As applications of AI become a routine part of clinical practice, intensive care clinicians will need to develop an understanding of the ethics and responsibilities that come with healthcare AI. In this brief paper, we outline five things every clinician should know to inform the ethical use of AI technologies in intensive care (see Fig. 1 for a summary). We highlight issues that clinicians must understand to engage in ethical deliberation about the uses of AI more generally. Readers seeking additional information and a principlist approach to issues of AI in healthcare would do well to read other articles in this special series on AI, or consult other authoritative publications [1, 2].

First, clinicians should have a basic fluency with the technology underlying AI because they will ultimately remain ethically and legally responsible for treatment decisions. As a general-purpose technology, AI refers to

computer algorithms that run complex computations on data using advanced statistical analyses [3]. These algorithms are generally trained on large datasets, which permit more accurate predictions than can be made with other methodologies. Healthcare applications of AI range from clinician-facing tools to predict clinical deterioration in the ICU to patient-facing applications such as automated chat functions (a chatbot) of which families can ask questions [3]. The purpose of becoming familiar with the technology underlying AI is not to become an expert in developing such technologies. Rather, practicing clinicians must understand what algorithms can and cannot do, promote the appropriate use of healthcare AI, and recognize when the technology is not performing as desired or expected.

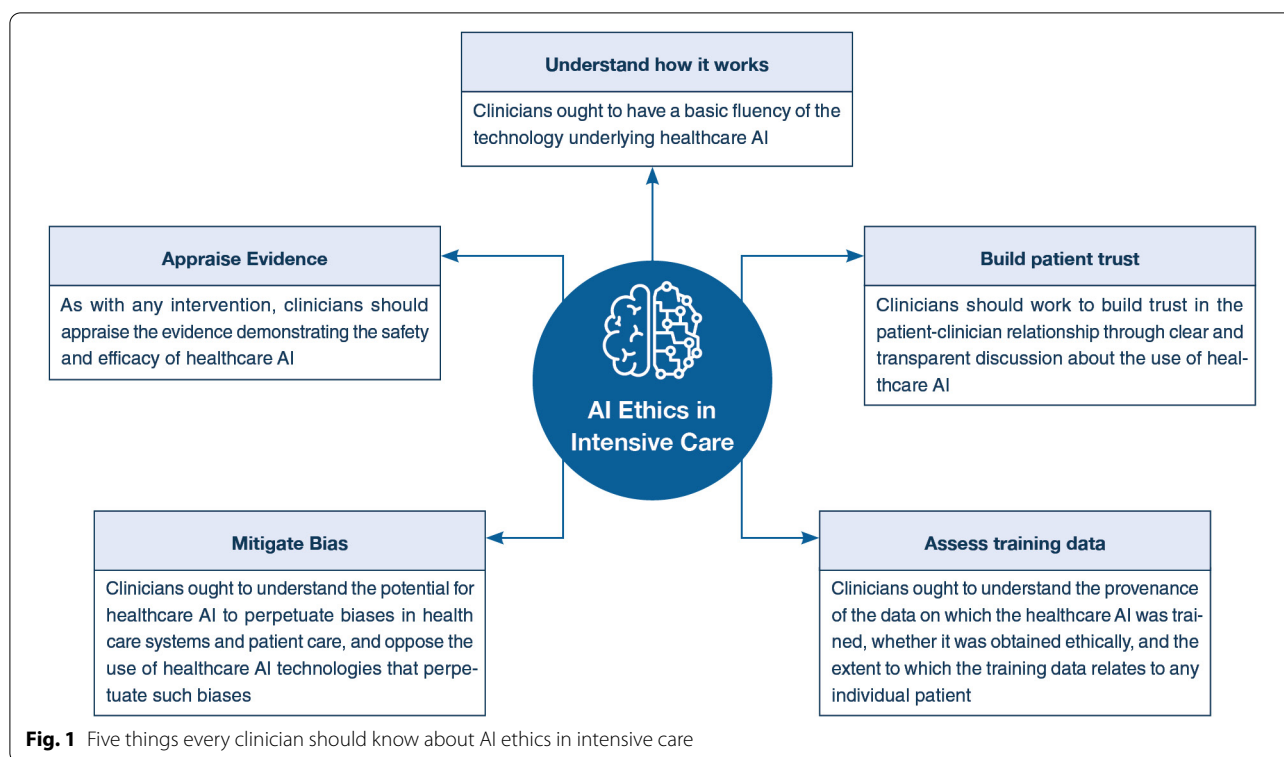
Second, clinicians should understand that patients and the public will not necessarily trust or embrace healthcare AI. A 2019 survey of members of the Canadian public found that 58% of respondents believed it was very or somewhat likely that AI technologies would be delivering health services in the next 10 years [4]. Two-thirds of respondents believed that such advances in the role of AI in medicine would have a positive impact on their lives. And yet, experimental evidence in the United States suggests that people are less likely to use medical services when such services are known to use AI [5]. Trust—it turns out—is still dependent on having a clinician remain in charge of decision-making. Thus, clinicians planning to utilize healthcare AI must develop strategies for communicating clearly that the clinician will only employ AI in ways that are safe and effective [6].

Third, clinicians must understand the provenance of training data used in healthcare AI. Algorithms derive their power and accuracy from the data they are trained on. In most cases, training data comes from individual patients. While using patient-level data is not inherently objectionable, the use of such data without consent

*Correspondence: Jay.shaw@wchospital.ca

¹ Research Director of Artificial Intelligence, Ethics & Health, Joint Centre for Bioethics, University of Toronto, Toronto, Canada

Full author information is available at the end of the article



is morally problematic [7]. In Denmark, for example, health authorities made population-wide health record data available to digital health innovators [8]. When a group of physicians discovered this was taking place, they raised public awareness and advocated for ending such data sharing. Similar events have taken place in the United Kingdom and the United States, where health systems shared large numbers of health records with large technology companies [7]. Although morally problematic, sharing de-identified data to generate AI is legal in most of the world. European countries tend to demand stronger justifications for such initiatives, but even in Europe, patients need not expressly grant permission for their health record to be shared. Just as with other technologies, frameworks to inform ethical data use lag behind the deployment of AI, which is occurring across sectors at a blistering pace [9].

Fourth—in addition to considering whether training data was acquired with patient consent, it is also important to understand that algorithms may perpetuate bias. Suresh and Guttag (2019) outline six ways in which bias can be incorporated into the process of AI development and deployment [10]. Although a detailed exploration of each form of bias is beyond the scope of this paper, we provide one example here. Obermeyer et al. (2019) identified the case where a health system in the United States

deployed a model to identify patients with complex needs that was allocating systematically fewer resources to Black patients than White patients based on past expenditures to those patients [11]. The algorithm was perpetuating historical inequities in access by providing less care to Black patients. It was not the case that such patients needed less care.

The fifth and final consideration we emphasize is that clinicians should understand whether algorithms are achieving appropriate and desired results. Just as an evidence base is required before utilizing novel chemotherapy, AI algorithms must also be tested to ensure they are delivering intended results. Healthcare AI is not infallible. Clinicians must have access to information about the impact that AI has on patient outcomes to avoid causing patient harm [12]. Recent work has proposed a multi-phased approach to generating evidence on healthcare AI, and clinicians should become familiar with the interpretation of such evidence for this novel collection of technologies [13].

Returning to the patients mentioned at the outset, the clinician is apt to question the reasoning behind the algorithm. They must have access to information about how the algorithm was trained, understand how the training data relates to the patients in front of them, and review patient outcomes associated with the algorithm. In this

case, there is concern that the algorithm may perpetuate inequities by calculating survival probabilities on the basis of historical data showing that Black patients—who often receive worse medical care and face other social inequities—are less likely to survive.

Author details

¹ Research Director of Artificial Intelligence, Ethics & Health, Joint Centre for Bioethics, University of Toronto, Toronto, Canada. ² Institute for Health System Solutions and Virtual Care, Women's College Hospital, 76 Grenville Street, Toronto, ON M5S1B2, Canada. ³ Centre for Biomedicine, Self and Society, Usher Institute, University of Edinburgh, Edinburgh, UK. ⁴ Department of Pulmonary, Allergy, Critical Care and Sleep Medicine, University of California, San Francisco, USA.

Compliance with ethical standards

Conflicts of interest

The authors declare no conflicts of interest.

Publisher's Note

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

Received: 24 August 2020 Accepted: 3 October 2020

Published online: 19 October 2020

References

1. Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. *Nat Mach Intell* 1(9):389–399
2. Morley J et al. (2020) The ethics of AI in health care: A mapping review. *Soc Sci Med*, 113172
3. Shaw J, Rudzicz F, Jamieson T, Goldfarb A (2019) Artificial intelligence and the implementation challenge. *J Med Internet Res* 21(7):e13659
4. Canadian Medical Association, "The Future of Connected Health Care: Reporting Canadians' Perspectives on the Health Care System." Aug. 2019, Accessed: Jul. 07, 2020. [Online]. <https://www.cma.ca/sites/default/files/pdf/Media-Releases/The-Future-of-Connected-Healthcare-e.pdf>.
5. Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical artificial intelligence. *J Consum Res* 46(4):629–650
6. Nundy S, Montgomery T, Wachter RM (2019) Promoting trust between patients and physicians in the era of artificial intelligence. *JAMA* 322(6):497–498
7. Wachter RM, Cassel CK (2020) Sharing health care data with digital giants: overcoming obstacles and reaping benefits while protecting patients. *JAMA* 323(6):507–508
8. Wadmann S, Hoeyer K (2018) Dangers of the digital fit: rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data Soc* 5(1):2053951717752964
9. Einav S, Ranzani OT (2020) Focus on better care and ethics: Are medical ethics lagging behind the development of new medical technologies? *Intensive Care Med* 46(8):1611–1613. <https://doi.org/10.1007/s00134-020-06112-4>
10. Suresh H, Gutttag JV (2019) A framework for understanding unintended consequences of machine learning. *ArXiv Prepr. ArXiv190110002*
11. Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453
12. Wynants L et al (2020) Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 369:m1328. <https://doi.org/10.1136/bmj.m1328>
13. McCradden MD, Stephenson EA, Anderson JA Clinical research underlies ethical integration of healthcare artificial intelligence. *Nat Med* 26(9), Art. no. 9, Sep. 2020. <https://doi.org/10.1038/s41591-020-1035-9>