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

# Five Things Every Clinician Should Know about AI Ethics in Intensive Care

Contact Information

James A Shaw, PhD
Research Director of Artificial Intelligence, Ethics & Health
Joint Centre for Bioethics
University of Toronto
Scientist
Institute for Health System Solutions and Virtual Care
Women's College Hospital
76 Grenville Street
Toronto, Ontario
Canada
M5S1B2
Jay.shaw@wchospital.ca
1-416-323-6400, ex 4224

Nayha Sethi, PhD
Centre for Biomedicine, Self and Society,
Usher Institute, University of Edinburgh
Edinburgh, United Kingdom
nayha.sethi@ed.ac.uk

Brian L Block, MD
Assistant Professor
Department of Pulmonary, Allergy, Critical Care and Sleep Medicine
University of California, San Francisco
brian.block@ucsf.edu

Word count: 1,016

**Five Things Every Clinician Should Know about AI Ethics in Intensive Care**

You've just admitted a patient to your ICU with COVID-19 requiring invasive mechanical ventilation. You know that using appropriate ventilator settings is important, but each trip into the room to make adjustments requires donning and doffing personal protective equipment, which is in short supply. You wish the ventilator could autotitrate in real-time; monitoring pressures and volumes, incorporating blood gas results from the medical record, and adjusting settings in an evidence-based way. You have heard about a newly marketed artificial intelligence (AI) system that can support autotitration in the ICU, and are considering whether it would be safe, effective, and worth advocating for its procurement at your hospital.

As applications of AI become a routine part of clinical practice, intensive care clinicians will need to develop comfort working with AI to deliver patient care. They must also 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.

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.[1] 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 Intensive Care Unit (ICU) to patient-facing applications such as automated chat functions (a chat bot) of which families can ask questions.[1] Understanding the basic principles underlying

healthcare AI will allow practicing clinicians to recognize what algorithms can and cannot do, and promote appropriate use of healthcare AI.

Second, clinicians should understand that patients and the public do not necessarily trust or embrace the technology underlying AI applications in clinical care. 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.[2] 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.[3] 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 about the role of AI in medical care.[4]

Understanding the provenance of training data is a third facet of healthcare AI that clinicians must grasp. Algorithms derive their power and accuracy from the data they are trained on, and in most cases, this means basing statistical predictions on individual patient-level health data. While using patient-level data is not inherently objectionable, the use of such data without consent is morally problematic.[5] And yet, such unannounced data mining is occurring worldwide.

In Denmark, for example, health authorities made population-wide health record data available to digital health innovators.[6] 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 United States, where health systems shared large numbers of health records with large technology companies.[5] Although such sharing is 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. Understanding the sources of training data and mode by which it was acquired is therefore an important ethical consideration, and one that patients may ask about.

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.[7] 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.[8] 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 the evidence underlying each use of healthcare AI. A statistical prediction from AI is just that and must not be weighed too heavily; especially if the training data for such an algorithm is not reflective of the patient in front of you. Just as an evidence base is required before utilizing novel chemotherapy, AI algorithms must also be tested to ensure they are delivering intended results. Using algorithms without appropriate skepticism or testing—as some seem to be advocating for amid the current COVID-19 pandemic—risks causing patient harm.[9] Clinicians adopting new

technologies such as healthcare AI must understand that algorithms are not infallible, and have access to information about impact that such algorithms have on patient outcomes.[10]

Returning to the patient mentioned at the outset, a clinician hoping to leverage AI for ventilator titration should understand what information the algorithm is using to make suggestions about ventilator settings, understand how such information relates to the individual patient in front of them, and demand outcomes data to ensure the algorithm is safe and effective. Moreover, clinicians must understand that working with AI does not alter their moral or legal obligations to deliver appropriate, compassionate care. Finally, clinicians must see working with AI as a partnership, rather than an offloading of tasks. Maintaining human touch is critical for preserving trust and addressing the human experience of illness.

**References**


1. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. *J Med Internet Res*. 2019;21(7):e13659.

2. Canadian Medical Association. The Future of Connected Health Care: Reporting Canadians' Perspectives on the Health Care System. Published online August 2019. Accessed July 7, 2020. https://www.cma.ca/sites/default/files/pdf/Media-Releases/The-Future-of-Connected-Healthcare-e.pdf

3. Longoni C, Bonezzi A, Morewedge CK. Resistance to medical artificial intelligence. *J Consum Res*. 2019;46(4):629–650.

4. Nundy S, Montgomery T, Wachter RM. Promoting trust between patients and physicians in the era of artificial intelligence. *Jama*. 2019;322(6):497–498.

5. Wachter RM, Cassel CK. Sharing health care data with digital giants: overcoming obstacles and reaping benefits while protecting patients. *Jama*. 2020;323(6):507–508.

6. Wadmann S, Hoeyer K. Dangers of the digital fit: Rethinking seamlessness and social sustainability in data-intensive healthcare. *Big Data Soc*. 2018;5(1):2053951717752964.

7. Suresh H, Guttag JV. A framework for understanding unintended consequences of machine learning. *ArXiv Prepr ArXiv190110002*. Published online 2019.

8. 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.

9. Wynants L, Van Calster B, Bonten MM, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *bmj*. 2020;369.

10. Sethi N. Regulating for uncertainty: bridging blurred boundaries in medical innovation, research and treatment. *Law Innov Technol*. 2019;11(1):112–133.