



BIOETHICS FORUM ESSAY

## For Ethical Use of AI in Medicine, Don't Overlook Maintenance and Repair

---

by [Kellie Owens](#)



Published May 9, 2023

Posted in [Artificial Intelligence](#), [Hastings Bioethics Forum](#)

Last month, Microsoft and the electronic health records vendor [Epic announced that they would partner](#) to use generative artificial intelligence in online portals to help doctors answer patient's questions. This project added to the growing examples of AI being used in medicine, raising ethical concerns about how to equitably and effectively implement these tools.

A number of federal institutions have provided frameworks for thinking about the ethical integration of AI in society, most notably the White House Office of Science and Technology Policy's "[Blueprint for an AI Bill of Rights](#)" and the National Institute of Standards and

Technology's "Artificial Intelligence Risk Management Framework." Both documents outline the persistent and pervasive challenges of automated systems in our society and propose protections for individuals under the jurisdiction of these systems.

While these documents are a welcome step forward, their real power will be in their execution across a range of domains. How should our health care system apply the "AI Risk Management Framework" and the "AI Bill of Rights" to assure the accurate and ethical use of AI tools? How can these two frameworks help us manage the firehose of new automated systems at our disposal? I urge attention to an often-overlooked piece of the puzzle: the maintenance and repair of these powerful tools over time to improve them in practice, including to detect and eliminate bias.

Concerns about bias, fairness, equity, justice, and privacy with AI and machine learning models that identify key patterns from clinical datasets are well documented. Many of them encode and reinscribe longstanding health inequities. In a particularly egregious case, one widely-used algorithm for distributing health resources prioritized white patients over Black patients, "reducing the number of Black patients identified for extra care by more than half." Some creators of AI/ML models recognize that overlooking the performance of their tools across a variety of subgroups over time has led to significant harm and are seeking to build fairness and equity into the design of new models. Yet they often center this assessment solely on the moment of initial deployment—focusing more on innovation than sustainability over time. While this work is critical to preventing harm, it is insufficient to expect that all potential harms of AI/ML models can be anticipated in the design phase, even with robust ethical and technical reviews.

Health care is a setting where change is expected—model performance and impact evolve over time or in new contexts. Researchers have found, for example, that even small shifts in data can lead to debilitating declines in the performance of AI/ML models used to predict risk of health complications or length of hospital stay. Most notably, a sepsis prediction model became no better than a coin flip, suggesting that patients may often be subjected to the "ineffective systems" that the AI Bill of Rights seeks to eliminate. Problems related to dataset shift can occur for many reasons, including changes in technology, population, or behavior (prompted, for example, by new reimbursement incentives). Current regulatory frameworks from the Food and Drug Administration do not sufficiently grapple with the potential ethical and social implications of AI/ML model deterioration over time and in new settings, often requiring models to be "locked" or unmodifiable without additional review. While the FDA is developing a more sophisticated regulatory framework, its focus on safety and efficacy is not the same as a full analysis of the ethical and social implications of AI/ML models over their lifecycles. Because shifts in model performance may generate new inequities and harms, ongoing social and technical review of these models over time is essential, alongside mechanisms to redress harms caused by AI/ML models.

What might these maintenance systems look like? First, they will require material and human resources—health systems must recognize the importance of devoting time and money to maintaining AI/ML tools. Second, they must include substantive and sustained participation from a range of stakeholders, including developers, clinicians, ethicists, and patients. Next, they require clear delineations of responsibility for the utility and effects of AI/ML in health care. This could include, for example, establishing separate centers within health care systems that are devoted to ensuring good performance of AI/ML tools over time. Such centers could minimize the role of strategic ignorance, where institutions deny liability for harms caused by the tools that they use by claiming they had no way to know about them. Finally, these maintenance systems will require mechanisms for repair—both technical repair of poor model performance and societal repair of any harms done to communities and individuals under the jurisdiction of those models. While researchers have begun to identify specific harms of some AI/ML tools, little effort has been invested to repair the damage.

AI/ML tools have tremendous potential to streamline care, identify high-risk patients, and augment processes of diagnosis and treatment, but this potential will not be realized without systems of maintenance and repair. Our current approach of ad-hoc audits to assess continued performance of clinical decision support tools is missing critical inequities that patients should be protected from under “AI Risk Management Framework” and the “AI Bill of Rights.” Once a tool is broken enough to notice, it’s too late. Trust has been lost and harms are accruing.

The importance of evaluating, maintaining, and repairing health technology is not unique to AI/ML tools, but may be particularly crucial when models are opaque and the pace of innovation is rapid. Recent U.S. government guidance provides an opportunity to innovate new infrastructural solutions to support the equitable and fair use of AI/ML tools in health care. As the push for innovation increases, our interest in creating new models and applications should not blind us to the essential urgency of building new ways to support the systems already in place—improving our existing systems in service of their eventual successors. Without this infrastructure, we will never unlock the potential of AI/ML in health care, unwittingly contributing to destructive social inequities and denying patients the rights they deserve.

*Kellie Owens, PhD, is an assistant professor in the Division of Medical Ethics of NYU Grossman School of Medicine. [@\\_kellie\\_owens\\_](#)*



Tagged [Medical AI](#)

[Read More Like This](#)

1 Comment on

“For Ethical Use of AI in Medicine, Don’t Overlook Maintenance and Repair.”