Artificial intelligence for health care equity

March 24 2021

(Left to right:) Regina Barzilay is the School of Engineering Distinguished Professor for AI and Health and faculty co-lead of AI for the MIT Jameel Clinic. Collin Stultz is a professor of electrical engineering and computer science and a cardiologist at Massachusetts General Hospital. Fotini Christia is a professor of political science and director of the Sociotechnical Systems Research Center. Credit: Massachusetts Institute of Technology

The potential of artificial intelligence to bring equity in health care has spurred significant research efforts. Racial, gender, and socioeconomic disparities have traditionally afflicted health care systems in ways that are difficult to detect and quantify. New AI technologies, however, are providing a platform for change.

Regina Barzilay, the School of Engineering Distinguished Professor of AI and Health and faculty co-lead of AI for the MIT Jameel Clinic; Fotini Christia, professor of political science and director of the MIT Sociotechnical Systems Research Center; and Collin Stultz, professor of electrical engineering and computer science and a cardiologist at Massachusetts General Hospital—discuss here the role of AI in equitable health care, current solutions, and policy implications. The three are co-chairs of the AI for Healthcare Equity Conference, taking place April 12.

Q: How can AI help address racial, gender, and socioeconomic disparities in health-care systems?

Stultz: Many factors contribute to economic disparities in health care systems. For one, there is little doubt that inherent human bias contributes to disparate health outcomes in marginalized populations. Although bias is an inescapable part of the human psyche, it is insidious, pervasive, and hard to detect. Individuals, in fact, are notoriously poor at detecting preexisting bias in their own perception of the world—a fact that has driven the development of implicit association tests that allow one to understand how underlying bias can affect decision-making.

AI provides a platform for the development of methods that can make personalized medicine a reality, thereby ensuring that clinical decisions are made objectively with the goal of minimizing adverse outcomes across different populations. Machine learning, in particular, describes a set of methods that help computers learn from data. In principle, these

methods can offer unbiased predictions that are based only on objective analyses of the underlying data.

Unfortunately, however, bias not only affects how individuals perceive the world around them, it also influences the datasets we use to build models. Observational datasets that store patient features and outcomes often reflect the underlying bias of health care providers; e.g., certain treatments may be preferentially offered to those who have high socioeconomic status. In short, algorithms can inherit our own biases. Making personalized medicine a reality is therefore predicated on our ability to develop and deploy unbiased tools that learn the patient-specific decisions from observational clinical data. Central to the success of this endeavor is the development of methods that can identify algorithmic bias and suggest mitigation strategies when bias is identified.

Informed, objective, and patient-specific clinical decisions are the future of modern clinical care. Machine learning will go a long way to making this a reality—achieving data-driven clinical insights devoid of implicit prejudice that can influence health-care decisions.

Q: What are some current AI solutions being developed in this space?

Barzilay: In most cases, biased predictions can be attributed to distributional properties of the training data. For instance, when some population is underrepresented in the training data, the resulting classifier is likely to underperform on this group. By default, models are optimized for the overall performance, thus inadvertently preferring to fit the majority class, at the expense of the rest. If we are aware of such minority groups in the data, we have multiple means to steer our learning algorithm towards fair behavior. For example, we can modify the learning objective where we enforce consistent accuracy across different

groups, or reweigh the significance of training examples, amplifying "the voice" of the minority group.

Another common source of bias relates to "nuisance variations" where classification labels exhibit idiosyncratic correlations with some input features which are dataset-specific and are unlikely to generalize. In one infamous dataset with such property, health status of patients with the same medical history depended on their race. This bias was an unfortunate artifact of the way training data was constructed, but it resulted in systematic discrimination of Black patients. If such biases are known beforehand, we can mitigate their effect by forcing the model to reduce the effect of such attributes. In many cases though, biases of our training data are unknown. It is safe to assume that the environment in which the model will be applied is likely to exhibit some distributional divergence from the training data. To improve a model's tolerance to such shifts, a number of approaches (like invariant risk minimization) explicitly train the model to robustly generalize to new environments.

However, we should be aware that algorithms are not magic wands that can correct all wrongs in messy, real-world training data. This is especially true when we are not aware of the peculiarity of a specific dataset. The latter scenario is unfortunately common in the health care domain where data curation and machine learning are often performed by different teams. These "hidden" biases have already resulted in deployed AI tools that systematically err on certain populations (like the model described above). In such cases, it is essential to provide physicians with tools that enable them to understand the rationale behind model predictions and detect biased predictions as soon as possible. A large body of work in machine learning is dedicated today to developing transparent models that can communicate their internal reasoning to users. At this point, our understanding of what types of rationales are particularly useful for doctors is limited, since AI tools are not yet part of routine medical practice. Therefore, one of the key goals of MIT's

Jameel Clinic is to deploy clinical AI algorithms in hospitals around the world and empirically study their performance in different populations and clinical settings. This data will inform the development of the next generation of self-explainable and fair AI tools.

Q: What are the policy implications for government agencies and the industry of more equitable AI for health care?

Christia: The use of AI in health care is now a reality and for government agencies and the industry to reap the benefits of a more equitable AI for health care, they need to create an AI ecosystem. They have to work together closely and engage with clinicians and patients to prioritize the quality of the AI tools that get employed in this space, making sure they are safe and ready for prime-time. This means that AI tools that get deployed have to be well-tested and to lead to improvements in both clinician capacity and patient experience.

To that effect, government and industry players need to think about educational campaigns that inform health practitioners of the importance of specific AI interventions in complementing and augmenting their work to address equity. Beyond clinicians, there also has to be a focus on building confidence with minority patients that the introduction of these AI tools will result in overall better and more equitable care. It is particularly important to also be transparent about what the use of AI in health means for the individual patient, as well as assuage data privacy concerns of patients from minority populations who often lack trust in a "well-intentioned" health care system, given historical transgressions against them.

In the regulatory realm, government agencies would need to put together a framework that would allow them to have clarity over AI funding and

liability with the industry and health care professionals so the highest-quality AI tools get deployed while also minimizing the associated risks for clinicians and patients using them. Regulations would need to make clear that the clinicians are not fully outsourcing their responsibility to the machine and outline the levels of professional accountability for their patients' health. Working closely with the industry, clinicians and patients, government agencies would also have to monitor through data and patient experience the actual effectiveness of AI tools in addressing health care disparities on the ground, and be attuned to improving them.

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology