

ACLU WHITE PAPER: AI IN HEALTH CARE MAY WORSEN MEDICAL RACISM

Crystal Grant, Technology Fellow with the ACLU's
Speech, Privacy, and Technology Project

INTRODUCTION

“Of all the forms of inequality, injustice in health is the most shocking and inhuman”

— Martin Luther King Jr.

Artificial intelligence (“AI”) and algorithmic decision-making systems are in increasingly common use in the criminal legal system, education, hiring, housing, and other contexts. Each day seems to bring new headlines about how an AI tool is perpetuating and reinforcing racism and historical biases: a faulty facial recognition tool has led to yet another [false arrest of a Black person](#); speech-recognition systems commonly used in commercial products are worse [at recognizing the speech](#) of non-White customers; a social media platform’s image recognition algorithms have [once again](#) labeled Black people as primates. But not nearly enough attention has been paid to questions about algorithmic bias in health care and medicine, where these tools have also proliferated. Decisions in health care can have serious ramifications, including in some instances life-or-death consequences for patients. It is therefore especially important to attend to how automated bias in the health care space can be identified, ameliorated, and properly regulated.

While the promise of medical AI is that it will improve health care for everyone, there is also the potential for it to worsen care for those already affected by medical racism. There are currently gaps in the federal regulatory structure governing medical devices and algorithms used in health care which lead to the widespread adoption of medical AI and algorithmic tools that have not been properly vetted to ensure they don’t cause harm.

I. Current medical AI, medical devices, and algorithmic decision-making tools display harmful biases

Racial bias directing the provision of worse care to communities of color has already been found in a number of health and medical tools, including tools that use AI or algorithms as well as more traditional hardware medical devices. For example:

- An AI tool meant to decide how to best distribute the limited resource of extra care to new mothers at risk of postpartum depression [was found to show racial bias](#)—directing care away from Black mothers and favoring White mothers.
- A widely used clinical algorithm¹ indicating kidney health is adjusted based on whether a patient is Black, and systematically indicates Black patients are healthier than they may actually be; in fact, an [October 2020 study](#) found that without this explicit race-based adjustment, nearly a third of Black patients would be reclassified as having more severe kidney disease. (Only in [September 2021](#), after increased pressure from lawmakers and advocates, was the algorithm updated to remove the use of race. Still, [recent](#)

¹ The term “clinical algorithm” or “algorithm” will be used to refer to tools as simple as calculators to more complex algorithms used in the clinical space to adjust patient tests based on their self-reported race or to determine the cost of treating a patient or assess their risk is for an adverse outcome. This differs from more complex tools referred in the text as “AI” or “medical AI,” which refer to a medical device eligible for FDA clearance or approval.

[reports suggest](#) the old algorithm is still being used by federal courts to make determinations about health-based early prison release despite litigation indicating that it functions in a clearly biased way.)

- A recent [meta-analysis](#) found the vast majority of machine learning (ML) studies in dermatology did not include information on different skin tones as part of algorithm development. As a result, the validity of model results varied based on skin tone, with some models performing worse on darker skin.
- [A 2020 study on pulse oximeters](#), a medical device used especially in the COVID-19 pandemic to monitor patients' oxygen levels, detailed that the devices are less accurate among patients with darker skin and could even increase risk of adverse health outcomes for those patients. In fact, a [2022 retrospective study](#) confirmed that patients of color, likely due to this known bias, received less supplemental oxygen than White patients, contributing to their morbidity. While this is a hardware issue, it shows an existing bias associated with patient's skin color in medical devices, instances like this are alarming considering that this issue was arguably more predictable than issues that may arise from the use of AI as a medical device.

Even when race is purposely left out of a model, the model can still reflect societal biases and lead to worse outcomes for people of color. This is exactly what happened in a commercial tool widely used by hospitals to decide which patients should get extra medical care. [Though race was excluded, the model still led to a biased outcome](#) in which Black patients had to be much sicker than White patients to be recommended the same level of care. The model worked by predicting what a patient's total medical expenditures were expected to be, then used that amount as an indicator of the patient's health care needs. However, the data used to train the model reflects the reality that socioeconomic and historical factors, including lack of access to health care, have led Black and Brown patients to spend less on medical expenditures than their White counterparts, even when they are just as sick. In using this model, hospitals made racially biased allocation decisions—a result of the inappropriate assumption by the health systems that they could implement a tool that measured money spent to infer care needed.

If an algorithm that purposely excludes race as a factor can exhibit racial bias and encourage decisions that hurt patients of color, it is clear that regulatory protections are needed. Removing race as an explicit variable in AI models isn't enough to cure the biased effects of some tools—for example those trained on historical data known to reflect societal biases. Instead, health care tools must be subject to explicit tests to identify racial and ethnic bias in their outcomes.

II. AI may identify race even when humans cannot

Last year it was [reported](#) that an AI model trained on patient data from X-rays and other medical images had learned to guess a patient's self-reported race of 'White,' 'Asian,' or 'Black' with stunning accuracy—varying between 80% and 90%. The [study, published in June, details that regardless of the aim of a number of deep-learning models—in other words, whether or not they were trained with the explicit goal of identifying race or trained only with the clinical aim of diagnosing the images as normal or abnormal—the](#) models had still somehow learned to tell patients' race from their medical images. This was true even in external validation and does not appear to be due to proxies or other known confounders for race in medicine. This means AI is capable of determining race even though human specialists can't tell anything about a patient's race from their X-ray or CT image.

The study authors are unsure *how* the AI model learned to discern self-reported race. The authors found the parts of the image used to discern race were not localized to any specific anatomical regions, thus, they are unsure how to adjust the model to *not* learn a patient's race. Additionally, several AI devices trained on the same medical images used by these researchers have already been cleared by the FDA.² That means there may be medical devices already in use that identify or make decisions based on patients' race unbeknownst to human specialists, with unknown and possibly harmful effects.

While this finding may not sound dangerous in the context of X-ray analysis, the prospect that AI tools might be able to infer race where humans cannot perceive it has staggering implications. The promise of AI in medicine has always been that it could act objectively in a field that continues to be poisoned by racial discrimination and that has a history of dehumanizing people with disabilities. But the revelation that algorithms can detect race when even trained clinicians cannot means that AI is capable of discrimination *even when it's impossible for humans*. In other words, if a medical AI tool can detect self-reported race, it may also use race as a factor in other outputs or decisions it renders, in ways that are not relevant to the medical context in which the tool is being used. This turns the promise of AI in medicine on its head, suggesting that its use could become even more dangerous than human decision-making if not properly audited.

It is not just race-based discrimination that patients need to worry about when medical AI tools are used without proper assessment. These tools must be tested to ensure they are doing what they set out to do and that they perform fairly across different populations. Without such robust audits, we're likely to see outcomes that are discriminatory not only on the basis of race, but also on the basis of other characteristics:

- Discrimination based on disability status. An algorithm trained to find the best person to receive a transplant organ may learn from [biased historical data](#) to discriminate against patients with disabilities who would benefit from such a transplant. Historical data may reflect bias because some people with disabilities are commonly denied transplants in violation of the Americans with Disabilities Act.
- Discrimination based on immigration status or income. An algorithm that uses electronic health records (EHR) to maximize health outcomes could be more error-prone in people with more sparse health histories such as undocumented immigrants and people who are unhoused.
- Discrimination based on sex. An algorithm that aims to dole out treatments with the goal of extending patients' lifespans could decide to recommend treatments only to women since they, on the whole, tend to live longer than men.
- Discrimination based on gender identity. An algorithm that uses genetic data from a patient to infer their gender identity may incorrectly classify patients based on this data and/or may recommend inappropriate treatments.

² The website of the American College of Radiology's Data Science Institute lists all FDA cleared and approved AI tools on the market currently. See Am. Coll. of Radiology, *AI Central*, <https://aicentral.acrdsi.org/> (last visited July 22, 2022). Many of the devices listed under the subspecialty "Chest Imaging" were trained on the exact same datasets as the tools in a recent study demonstrating AI tools' ability to detect patients' self-reported race. See Judy Wawira Gichoya et al., *AI Recognition of Patient Race in Medical Imaging: A Modelling Study*, 4 *The Lancet* E406 (June 1, 2022), [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(22\)00063-2/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00063-2/fulltext).

- Discrimination based on geographic location. An algorithm disproportionately trained on more populated states or higher income areas may underperform among patients from a different geographic area. In fact, a [2020 study](#) found that 71% of data used in deep learning of U.S. medical data came from California, Massachusetts, or New York, while the other 47 states had little to no representation.

These outcomes could happen without anyone intending them, because many AI tools' outcomes are black boxes (i.e. their outputs cannot be explained based on their inputs), or because AI tools can infer and consider characteristics about patients that human decision-makers cannot detect.

III. Race-based medicine: From Present to Past

The consideration of race in medicine has been widespread, but we need to interrogate its use. While race exists as a social category that may help explain a patient's lived experience and cultural context, and in turn inform treatment, there is no relationship between racial categories and biological factors.³ Medical conditions can be caused by environmental factors, genetic and other biological factors, or some combination of the two. Though the use of genetics [has the potential to aid in treatment and diagnosis](#) for some conditions that are caused by discrete genetic mutations or events, genomics is not yet widely used in medical treatment. [This is because the diseases and disorders most common in humans are multifactorial and polygenic](#), meaning they result from complex environmental interactions involving multiple genes contributing to a disease state. This is the case with conditions like high blood pressure, diabetes, cancer, and mental health disabilities.

Environmental causes can also be difficult to tease out and can appear to correlate with race—a social construct with only nebulous ties to biology. Environmental factors correlated with health include behaviors like smoking, poor diet, or living in more polluted neighborhoods. But these environmental factors can be difficult to separate from societal inequities—especially since these and other factors like stress, poverty, access to medical care, and racism can themselves be considered environmental risk factors. Because of common lived experience and shared environment, racial and ethnic minorities may be more likely to suffer from poor health outcomes which are then misattributed to their genes.⁴

A person's geographic origin or that of their ancestors can sometimes be linked to specific health risks through genetic factors. Examples include the increased risk of sickle cell disease among those whose ancestors came from areas where malaria is common, or the [finding](#) that many people with red hair, often with ancestors from northern Europe, may require different doses of

³ Because the CDC has identified racism as a fundamental driver of racial and ethnic health disparities, *see* Ctrs. For Disease Control & Prevention, *CDC's Efforts to Address Racism as a Fundamental Driver of Health Disparities* <https://www.cdc.gov/healthequity/racism-disparities/cdc-efforts.html>, race can be used to predict the sociological effects of racism on a patient. In these instances, the accuracy of certain predictive measures may depend on the inclusion of race. This inclusion represents race-conscious health care. But unlike race-based medicine, which attempts to use race as a proxy for biology, race-conscious health care attempts to use self-reported race as a sociological tool in health care to address and attempt to correct past and present injustices against racial and ethnic minorities.

⁴ Nat'l Acad. of Scis., Eng'g, & Med., *Communities in Action: Pathways to Health Equity* 3 (James N. Weinstein et al. eds., National Academies Press 2017), https://www.ncbi.nlm.nih.gov/books/NBK425848/pdf/Bookshelf_NBK425848.pdf.

pain medication. But these somewhat relevant links between disease risk and geographic ancestry have often been simplistically reduced to an assumed link between disease risk and race. This is an oversimplification that can incorrectly inform care. Furthermore, race is sometimes just eyeballed by medical practitioners instead of being self-reported, leading to inaccurate application of an already imprecise medical shortcut. There are no universal, international standards for racial categories in medicine, nor standards on how to classify multi-racial individuals.

This substitution of race for actual biological information constitutes the current practice of [race-based medicine](#) and makes little sense scientifically. Take, for example, the term “African American,” which can be applied to essentially all Americans whose ancestors originated from the continent of Africa (this term is also applied to many multiracial individuals). Africa is known to be the most genetically diverse continent in the world, yet many African Americans, regardless of their ancestors’ African ethnic group of origin, are lumped into a single group by the medical field: “Black.” This flattening of an incredibly biologically diverse population was also common in eugenics texts, as was the racist insistence that anyone who could be described as Black was both [inferior and fundamentally different](#) from those of European descent. This way of using race is a direct intellectual descendant of the eugenics movement and it continues to plague medicine today. In short, race is too subjective and simplistic a construct to replace actual biological information in directing medical care, and its use encourages racial profiling.

IV. The role of the federal government in regulating medical AI

There is clear evidence that some medical AI and clinical algorithms exhibit harmful racial biases, intensifying the longstanding problems that our medical system has in misusing the concept of race or, conversely, failing to account for how socioeconomic factors have led to differential health outcomes by race or ethnicity. Given how high the stakes are when these tools are used to make decisions about health care, regulatory protections are needed for AI-based medical devices and tools. How should those protections be crafted, and what should they look like?

There is no one agency regulating the hundreds of AI tools and clinical algorithms in use in the health and medical field today. Instead, a patchwork of regulatory powers has led to gaps and permitted the continued use of potentially harmful technologies without sufficient oversight. Even where AI-enabled medical devices are subject to regulation and face testing before they can be marketed, these processes may not be rigorous enough given the tools’ potential for automated bias. It is particularly important for future AI devices that a compliance and auditing framework is established that allows regulators to detect if and when a device which performed one way when it was allowed to market is beginning to stray in its performance.

1. The role of the FDA in the regulation of medical AI devices

The U.S. Food and Drug Administration (FDA) is tasked with regulating medications, medical treatments, and medical devices. In order to be legally marketed in the United States, many medical devices and products must undergo a review by the FDA to ensure they are safe. The process of regulating medical devices is complex, in part because a wide variety of items are considered medical devices, “[from a simple tongue depressor to a life-sustaining heart valve.](#)” Exactly which regulatory process a medical device must go through is a function of 1) its potential for harm, and 2) its novelty when compared to existing medical devices or technology.

Devices are grouped into three classes based on their increasing potential for harm. Class III devices have the highest risk and/or employ novel methods, and are required to undergo the

most intensive forms of FDA review in order to obtain Premarket Approval (PMA). Class II devices that are substantially similar to existing devices undergo a less rigorous process (known as “510(k)”) to obtain clearance. Most Class I devices are considered so low risk that they can be legally marketed in the U.S. without FDA review. Class I or II devices that are novel but pose low enough risk in their potential for harm go through the de novo process to obtain clearance.

<i>Classification</i>	<i>Premarket requirement</i>
Class I	Most Class I devices are exempt from premarket review. For some devices, a 510(k) requirement can be specified. De novo technologies that are not similar to existing devices but are not high risk enough to require a PMA.
Class II	Class II devices require a 510(k) review to obtain clearance (unless specified otherwise). De novo technologies that are not similar to existing devices but are not high risk enough to require a PMA.
Class III	Class III devices require a PMA (or Humanitarian Device Exemption, “HDE”) to obtain approval.

Table 1: FDA-regulated medical devices fall into three classifications based on the devices’ novelty and risk, each with its own pre-market review requirements.

Thus, “approval” represents the most rigorous path to market for devices considered to have the most potential for harm, while “clearance” allows devices thought to have lower risk to go to market. In order for a device manufacturer to obtain approval, a rigorous clinical trial is often required as well as preclinical data—typically this kind of approval is for Class III devices. (These clinical studies however themselves often suffer from an underrepresentation of minority populations and women.⁵ Recently, the FDA released [draft guidance](#) suggesting manufacturers submit a Race and Ethnicity Diversity Plan to enroll more participants from underrepresented racial and ethnic populations, which represents a positive step if manufacturers choose to comply.) The FDA can also impose post market requirements in which manufacturers report real-world data on their device in a “post-approval study.” This post-approval monitoring would be ideal for AI tools. However, some device manufacturers [fail to conduct](#) such a study even when the FDA orders

⁵ Nat’l Acads. of Scis., Eng’g, & Med., *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups* (National Academies Presss 2022), <https://nap.nationalacademies.org/catalog/26479/improving-representation-in-clinical-trials-and-research-building-research-equity>.

it, because the FDA does not use its enforcement authority to rescind approval for failure to conduct such a study.

For other devices, often Class II devices, all that is required is a less rigorous study showing similarity to an approved device as well as some clinical data. In that situation, there are often no postmarket requirements nor a required post-approval study. Medical devices that go through this less rigorous process thus have the potential to be trained on data that is not diverse or representative of the real world. They could then be used in real world settings, and could cause disproportionate harm to some groups without detection. Many AI-enabled medical devices are only subject to the less rigorous clearance process; this may be of concern for future AI-enabled devices that are dynamic⁶ and may have little to no human input in their ongoing development—in other words, for AI-enabled devices where machine learning processes may render the tool’s performance very different from the one that was studied at the premarket stage. Even among currently cleared AI-enabled devices, all of which are static, the populations on which they were trained may differ from the current makeup and needs of the treatment population.

In addition to medical hardware devices, the FDA is also tasked with the regulation of AI-enabled medical devices including “software as a medical device” (SaMD) and “software in a medical device” (SiMD). In 2017, the agency [published recommendations](#) that medical devices be tested for algorithmic bias based on patients’ age, sex, and race/ethnicity. But few manufacturers have made the results of such evaluations public, including vital data on the demographics of the patients used to train the software. An [evaluation](#) found that just seven of 161 AI products cleared in recent years include any public information about the racial and ethnic composition of their datasets. Unless the FDA makes its approval or clearance of medical devices contingent on evaluations for algorithmic bias, manufacturers have little incentive to perform such tests.

Databases that are used in the development of AI tools must also be evaluated for harmful bias. Worryingly, in recent years FDA’s oversight of medical products was repeatedly included on the Government Accountability Office’s list of high-risk areas⁷ needing reform. To respond to these concerns, the FDA in 2016 conducted internal reviews and created the new National Evaluation System for health Technology (NEST) which was aimed at recommending practices to improve the regulation of medical devices. A 2020 publication from NEST’s Coordinating Center includes frameworks that [emphasize the use of more real-world evidence and data](#) as a means to reduce bias in medical devices. While this step would be an improvement, it is not a panacea, as real-world data will still capture the current systemic biases and inequities plaguing our health system.

The FDA appears poised to increase engagement and transparency around its approach to the regulation of medical AI and algorithms; the agency recently made public an initial [list](#) of AI-enabled medical devices marketed in the U.S. The FDA has also been soliciting and incorporating

⁶ The term “dynamic” here refers to models that can analyze and evaluate the new data it encounters, using this data to alter their decision rules, thus they are continuously learning and can update after the approval/clearance process with the FDA. While “static” models are trained once on training data and have essentially fixed rules governing their decision making. Currently, no dynamic models have gone through the FDA approval/clearance process.

⁷ According to the Government Accountability Office’s website, “[t]he High Risk List is a list of federal programs and operations that are vulnerable to fraud, waste, abuse, and mismanagement, or need transformation. The list is issued every 2 years at the start of each new session of Congress[.]” See U.S. Gov’t Accountability Office, *High Risk List*, <https://www.gao.gov/high-risk-list> (last visited July 22, 2022).

feedback from concerned stakeholders on how to best regulate this technology through the publication of a 2019 [Regulatory Framework](#), a 2021 [Action Plan](#), and most recently [Draft Guidance](#). These documents address that the regulation of medical AI will feature novel considerations and challenges. For example, AI-enabled devices—especially dynamic devices that learn and iterate autonomously based on their use—will need to undergo ongoing evaluation to identify and regulate emergent biases. And regulators will need to decide whether device testing should focus on real-world populations or on artificial datasets adjusted for real-world biases. Across this new regulatory frontier, it is vital that equity be given consideration alongside safety and effectiveness. While the Regulatory Framework mentions transparency as an issue and explicitly states the FDA’s concern about racial and ethnic bias creeping into AI-enabled medical devices and algorithms, it is unclear whether the agency will go so far as to tie its concerns over the potential for bias to a device’s ability to be approved or cleared.

2. *The role of the FDA in the regulation of AI clinical and public health tools*

In medicine and health care, AI and algorithmic decision making are being integrated into an ever-increasing variety of processes without being considered “devices” that trigger the FDA’s Class II or III review. AI and algorithmic decision making are used in deciding who is able [to receive medical benefits and aid](#), who is [considered to be at higher risk of certain conditions](#), who will be examined for possible [prescription abuse and fraud](#). There are even [autonomous tools](#) that work without any physician opinion or input. While AI and algorithmic tools’ involvement in patient wellbeing decisions is widespread and increasing across many medical disciplines, it is unclear whether these tools are being adequately evaluated for bias before entering into use. These kinds of tools are used in clinical care and public health but were not considered medical devices to be regulated by the FDA.

Such tools include those that rely on electronic health record (EHR) data to allow medical administrators to make predictions about patients, including their need for care. While the FDA does regulate some “Clinical Decision Support” (CDS) tools used by medical practitioners to aid in their diagnoses and treatment of patients, requiring them to go through the usual premarket clearance or approval process, in its [2019 guidance](#), it excluded some tools from premarket review based on a lower perceived risk. These more administrative algorithmic tools and CDS software which, for example, predict risk of mortality, readmission, or developing sepsis but that do not diagnose directly, would instead be regulated by the Federal Trade Commission (FTC). However, in September 2022, the FDA released [new guidance](#) redefining some of these tools as medical devices subject to their review and regulation.

Even with the existing fragmentation of regulatory authority, through collaborative practices, both the FDA and FTC, as well as other federal collaborators, could work to establish best practices to detect and mitigate harmful bias in AI tools used in the practice of medicine, regardless of how the tools are defined or whether they are used in clinical care, health care administration, or public health settings.

To create policies that center equity in human–technology interactions, the regulation of AI in medicine and health will require [a systems approach](#), involving stakeholders from different government agencies, health system administrators, academics, industry professionals in AI model manufacturing and evaluation, medical educators and students, and patient advocates.

V. Recommendations

- The FDA has engaged the public and other stakeholders in discussions about how to best regulate AI devices with the issue of safety at the forefront. But there must be an equal focus on and requirement of equity in the operation of these devices. While safety certainly has an equity dimension, regulators should specifically seek to ensure that medical AI tools do not contribute to differential outcomes across racial groups. To ensure these tools are truly safe, the agency must incorporate into its safety audits questions specifically designed to stem AI bias against marginalized patient populations. Already, there has been harm resulting from the use of biased clinical algorithms; similarly, medical devices must be evaluated for their use or detection of data on race or ethnicity, bias against people with disabilities, and disparate accuracy and performance. Such evaluation must become a requirement for device clearance or approval, and the FDA should use its enforcement authority to deny or rescind clearance or approval when such evaluations are not conducted. Increased transparency around the demographics of training and validation data must also be part of the regulatory approval or clearance process as well as data on subgroup performance.
- The Department of Health and Human Services (HHS) and its subagencies including but not limited to the FDA, Office of the National Coordinator for Health Information Technology (ONC), National Institutes of Health (NIH) and others must study the potential for harmful biases in AI-enabled medical devices, while the FTC must study the same issues in more simple algorithmic tools used in health and medicine. The federal government must take a [systems approach](#) to regulating all medical AI tools, facilitating cross-agency partnerships. The FDA and FTC must work collaboratively to identify common sources of harmful bias occurring in both medical devices and clinical tools. In a positive step, earlier this year, HHS acknowledged the potential for clinical algorithms to unintentionally discriminate and, [in a proposed rule, prohibited this discrimination “in covered health programs and activities.”](#)
- Hospitals and other medical systems must take care to ensure AI devices: are not used for clinical applications without FDA approval or clearance, are not used on patient populations they were not intended for, and that cleared tools are not used outside of their intended use cases—as the [FDA cautioned in one example in April](#). If the medical system is using a commercial clinical tool not regulated by the FDA, they should, at minimum, ensure the tool has been evaluated in a peer-reviewed study on a population representative of their patient population and has not been shown to reflect racial and ethnic bias. For non-commercial tools like those developed in-house, at minimum assessments around the effects of the tool’s use in different patient populations and the potential sources of bias in the training data should be performed. This requirement can be made part of the hospital accreditation process and quality monitoring metrics in order to ensure compliance.
- Medical students and staff should continue to push for evidence-based research into the use of devices and tools that recommend adjusting patients’ treatment or medication based on broad racial categories in the absence of information on genetics or socio-cultural risk factors. Race is not an appropriate proxy for genetics nor does it represent any kind of [“objective biological reality.”](#)
- Medical educators, medical professional societies, and medical systems must establish some proficiency in AI and ML topics in order to standardize the use of these devices. Embedded in these teachings must be mention of sources of bias already in medicine and

ways this bias will be encoded into AI tools and clinical algorithms; this could help providers exercise caution before using a tool on a population or in a context for which it wasn't originally designed. Medical practitioners, their education around technology, and the use of this technology in health care all form an interconnected system and should be regulated as one. Making an increased understanding of the inner workings of AI tools and their limitations a part of medical education may also help lessen the [human tendency](#) to attribute too much authority to decisions made by algorithms.

- Developers of AI devices, even in the absence of formal requirements, must ensure their tools aren't leading to disparate impacts that have no legitimate correlation to the question being asked of the tool. These companies must actively ensure that historically disadvantaged populations are well represented in their training data. Additionally, in order to encourage public trust, demographic data of the population used to train a model and used in the testing and clinical trial datasets must be made public. Any differences in its performance metrics between groups should also be made public. Additionally, the specific question asked (i.e., the label the algorithm is predicting) needs to be very clear in marketing materials, because algorithms are not flexible tools and cannot be used for purposes other than those for which they were specifically designed without substantial risk of bias. Algorithm developers must work with medical systems to ensure their tool will be applied appropriately. Developers should work to include, in the user interface of the tool, prompts or other methods that encourage use of the device only with appropriate data inputs and on appropriate populations.
- Device manufacturers should work alongside researchers in the nonprofit and academic space as well as entities collecting the real-world data from health systems used to train and test models to research and publish best practices on how to detect and address biases in data as well as to understand technical changes needed for hardware devices found to show racial bias, like pulse oximeters. Many in the private sector may have access to datasets that researchers do not, so working collaboratively should be encouraged as it can lead to the development of better tools and methods to address bias.
- Legislation should be drafted authorizing the FDA to make and maintain a public list of software as a medical device (SaMD) products and provide demographic information about the subjects in which the devices were calibrated or trained. Additionally, as part of the approval or clearance of these devices, an impact assessment of the effectiveness and accuracy of the device in different demographic groups must be performed and evaluated, with devices showing disparate impact having this finding reflected either in the patient- and/or physician-facing labeling on the product. Such assessments should be performed repeatedly as [models' performance can degrade over time once deployed in real world populations](#).