COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

VIRGINIA LAW REVIEW

VOLUME 110 APRIL 2024 NUMBER 2

ARTICLES

RACE IN THE MACHINE: RACIAL DISPARITIES IN HEALTH AND MEDICAL AI

*Khiara M. Bridges**

What does racial justice—and racial injustice—look like with respect to artificial intelligence in medicine ("medical AI")? This Article offers that racial injustice might look like a country in which law and ethics have decided that it is unnecessary to inform people of color that their health is being managed by a technology that likely encodes the centuries of inequitable medical care that people of color have received. Racial justice might look like an informed consent process that is reformed in light of this reality. This Article makes this argument in four Parts. Part I canvases the deep and wide literature that documents that people of color suffer higher rates of illness than their white counterparts while also suffering poorer health outcomes than their white counterparts when treated for these illnesses. Part II then provides an introduction to AI and explains the uses that scholars and developers predict medical AI technologies will have in healthcare, focusing specifically on the management of pregnancy. Part III subsequently serves as a primer on algorithmic bias—that is,

* Professor of Law, UC Berkeley School of Law. I would like to thank W. Nicholson Price, Rebecca Wexler, and the participants of a faculty workshop at the UC Berkeley School of Law, who offered incisive feedback on earlier versions of this Article. I am grateful to Edna Lewis in the Berkeley Law Library for truly wonderful research support. Thanks are also owed to Arni Daroy, Natalie Giron, Taylor Fox, Simone Lieban Levine, Hayley MacMillen, Tiffanie Obilor, and Samira Seraji, who all provided superb research assistance related to this Article. As always, I am thankful for my perfect little Belgian endive, Gert Reynaert, who makes my life both sweet and savory.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

systematic errors in the operation of an algorithm that result in a group being unfairly advantaged or disadvantaged. This Part argues that we should expect algorithmic bias that results in people of color receiving inferior pregnancy-related healthcare, and healthcare generally, because medical AI technologies will be developed, trained, and deployed in a country with striking and unforgivable racial disparities in health.

Part IV forms the heart of the Article, making the claim that obstetricians, and healthcare providers generally, should disclose during the informed consent process their reliance on, or consultation with, medical AI technologies that likely encode inequities. To be precise, providers should have to tell their patients that an algorithm has informed the recommendation that the provider is making; moreover, providers should inform their patients how racial disparities in health may have impacted the algorithm's accuracy. It supports this argument by recounting the antiracist, anti-white supremacist—indeed radical—origins of informed consent in the Nuremberg Trials' rebuke of Nazi "medicine." This Part argues that the introduction into the clinical encounter of medical AI—and the likelihood that these technologies will perpetuate racially inequitable healthcare while masking the same—is an invitation to reform the informed consent process to make it more consistent with the commitments that spurred its origination. This Part proposes that, given the antiracist roots of the doctrine of informed consent, it would be incredibly ironic to allow the informed consent process to permit a patient—and, particularly, a patient of color—to remain ignorant of the fact that their medical care is being managed by a device or system that likely encodes racism. This Part argues that informing patients about the likelihood of race-based algorithmic bias—and the reasons that we might expect race-based algorithmic bias—may, in fact, be a prerequisite for actually transforming the inequitable social conditions that produce racial disparities in health and healthcare.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

INTRODUCTION

As artificial intelligence ("AI") technologies proliferate across sundry sectors of society—from mortgage lending and marketing to policing and public health—it has become apparent to many observers that these technologies will need to be regulated to ensure both that their social benefits outweigh their social costs and that these costs and benefits are distributed fairly across society. In October 2022, the Biden Administration announced its awareness of the dangers that "technology, data, and automated systems" pose to individual rights.[1] Through its

[1] See Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, The White House Off. of Sci. & Tech. Pol'y, https://www.whitehouse.gov/ostp/ai-bill-of-rights/ [https://perma.cc/E5GS-6ZP3] (last visited Jan. 5, 2024). Some states and cities have also initiated efforts to regulate AI. See, e.g., Laura Schneider, Debo Adegbile, Ariella Feingold & Makenzie Way, NYC Soon to Enforce AI Bias Law, Other Jurisdictions Likely to

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Office of Science and Technology Policy, the Administration declared the need for a coordinated approach to address the problems that AI technologies have generated—problems that include "[a]lgorithms used in hiring and credit decisions [that] have been found to reflect and reproduce existing unwanted inequities or embed new harmful bias and discrimination," "[u]nchecked social media data collection [that] has been used to threaten people's opportunities, undermine their privacy, or pervasively track their activity," and, most germane to the concerns of this Article, "systems [that are] supposed to help with patient care [but that] have proven unsafe, ineffective, or biased."[2]

As an initial measure in the effort to eliminate—or, at least, contain—the harms that automation poses, the Administration offers a *Blueprint for an AI Bill of Rights*, which consists of "five principles that should guide the design, use, and deployment of automated systems to protect the American public in the age of artificial intelligence."[3] Crucially, the *Blueprint* identifies "notice and explanation" as a central element in a

Follow, WilmerHale (Apr. 10, 2023), https://www.wilmerhale.com/insights/client-alerts/20230410-nyc-soon-to-enforce-ai-bias-law-other-jurisdictions-likely-to-follow [https://perma.cc/K47J-XZUQ] ("New York City's Department of Consumer and Worker Protection (DCWP) is expected to begin enforcing the City's novel artificial intelligence (AI) bias audit law on July 5, 2023. This law prohibits the use of automated decision tools in employment decisions within New York City unless certain bias audit, notice, and reporting requirements are met."); Jonathan Kestenbaum, NYC's New AI Bias Law Broadly Impacts Hiring and Requires Audits, Bloomberg Law (July 5, 2023, 5:00 AM), https://news.bloomberglaw.com/us-law-week/nycs-new-ai-bias-law-broadly-impacts-hiring-and-requires-audits [https://perma.cc/L94C-X3BN] (observing that the "New Jersey Assembly is considering a limit on use of AI tools in hiring unless employers can prove they conducted a bias audit," that "Maryland and Illinois have proposed laws that prohibit use of facial recognition and video analysis tools in job interviews without consent of the candidates," and that "the California Fair Employment and Housing Council is mulling new mandates that would outlaw use of AI tools and tests that could screen applicants based on race, gender, ethnicity, and other protected characteristics"); Attorney General Bonta Launches Inquiry into Racial and Ethnic Bias in Healthcare Algorithms, State of Cal. Dep't of Just. Off. of the Att'y Gen. (Aug. 31, 2022), https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-racial-and-ethnic-bias-healthcare [https://perma.cc/ERC4-GVJJ] ("California Attorney General Rob Bonta today sent letters to hospital CEOs across the state requesting information about how healthcare facilities and other providers are identifying and addressing racial and ethnic disparities in commercial decision-making tools. The request for information is the first step in a DOJ inquiry into whether commercial healthcare algorithms—types of software used by healthcare providers to make decisions that affect access to healthcare for California patients—have discriminatory impacts based on race and ethnicity.").

[2] See Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People, supra note 1.

[3] Id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

program that protects the rights of individuals in an increasingly automated society.[4] That is, the Biden Administration proposes that in order to ensure that AI does not threaten "civil rights or democratic values," individuals should be informed when "an automated system is being used," and they should "understand how and why it contributes to outcomes that impact" them.[5] To apply it to the context to which this Article is most attuned, if a hospital system or healthcare provider relies upon an AI technology when making decisions about a patient's care, then the patient whose health is being managed by the technology ought to know about the technology's usage.

Although the Biden Administration appears committed to the idea that an individual's rights are violated when they are unaware that an AI technology has had some impact on the healthcare that they have received, many actors on the ground, including physicians and other healthcare providers, do not share this commitment. As one journalist reports:

> [T]ens of thousands of patients hospitalized at one of Minnesota's largest health systems have had their discharge planning decisions informed with help from an artificial intelligence model. But few if any of those patients [have] any idea about the AI involved in their care. That's because frontline clinicians . . . generally don't mention the AI whirring behind the scenes in their conversations with patients.[6]

This health system is hardly unique in its practice of keeping this information from patients. "The decision not to mention these systems to patients is the product of an emerging consensus among doctors, hospital executives, developers, and system architects who see little value . . . in raising the subject."[7] Moreover, while these actors see few advantages associated with informing patients that AI has informed a healthcare decision or recommendation, they see lots of disadvantages, with the disclosure operating as a "distraction" and "undermin[ing] trust."[8]

We exist in a historical moment in which the norms around notice and consent in the context of AI in healthcare have not yet emerged—with

[4] Id.
[5] Id.
[6] Rebecca Robbins & Erin Brodwin, An Invisible Hand: Patients Aren't Being Told About the AI Systems Advising Their Care, STAT (July 15, 2020), https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/ [https://perma.cc/R3F5-NNX4].
[7] Id.
[8] Id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

some powerful actors in the federal government proposing that patients are harmed when they are *not* notified that AI has impacted their healthcare, and other influential actors on the ground proposing that patients are harmed when they *are* notified that AI has impacted their healthcare.[9] As we think about the shape that these norms ought to take, this Article implores us to keep in mind the fact of racial inequality and the likelihood that AI will have emerged from, and thereby reflect, that racial inequality. Indeed, this Article's central claim is that the well-documented racial disparities in health that have existed in the United States since the dawn of the nation demand that providers inform all patients—but especially patients of color—that they have relied on or consulted with an AI technology when providing healthcare to them.

Although much has been written about AI in healthcare,[10] or medical AI, very little has been written about the effects that medical AI can and should have on the informed consent process.[11] Moreover, no article to date has interrogated what the reality of racial disparities in health should mean with respect to obtaining a patient's informed consent to a medical intervention (or nonintervention) that an AI system has recommended. This Article offers itself as the beginning of that conversation. It makes

[9] See also Attorney General Bonta Launches Inquiry into Racial and Ethnic Bias in Healthcare Algorithms, supra note 1 (understanding as problematic the fact that some AI tools used in healthcare "are not fully transparent to healthcare consumers"); cf. Schneider et al., supra note 1 (noting that New York City's law regulating AI in employment requires an employer to provide "applicants and employees who reside in New York City notice of its use of AI in hiring and/or promotion decisions, either via website, job posting, mail or e-mail").

Interestingly, some investigations have shown that some patients do not want to know when physicians and hospital administrators rely on medical AI when managing their healthcare. See Robbins & Brodwin, supra note 6 (reporting that some patients who were interviewed stated that "they wouldn't expect or even want their doctor to mention" the use of medical AI and stating that these patients "likened it to not wanting to be privy to numbers around their prognosis, such as how much time they might expect to have left, or how many patients with their disease are still alive after five years"). However, other studies have shown that patients do desire this information. See Anjali Jain et al., Awareness of Racial and Ethnic Bias and Potential Solutions to Address Bias with Use of Health Care Algorithms, JAMA Health F., June 2, 2023, at 10, https://jamanetwork.com/journals/jama-health-forum/fullarticle/2805595 [https://perma.cc/9FMK-E4VV] (discussing a "recent, nationally representative survey" that showed that "patients . . . wanted to know when [AI] was involved in their care").

[10] Indeed, volumes have been written about algorithmic bias, what AI technologies mean with respect to data privacy, and how we ought to regulate AI inside the medical context. See generally The Oxford Handbook of Digital Ethics (Carissa Véliz ed., 2021).

[11] See I. Glenn Cohen, Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?, 108 Geo. L.J. 1425, 1428 (2020) (noting that his Article, which was published just three years ago, was "the first to examine in-depth how medical AI / [machine learning] intersects with our concept of informed consent").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

the case that we ought to reform the informed consent process to ensure that patients of color are aware that their health is being managed by a technology that likely encodes the centuries of inequitable medical care that people of color have received in this country and around the world.

The Article proceeds in four Parts. Part I canvases the deep and wide literature that documents that people of color suffer higher rates of illness than their white counterparts while also suffering poorer health outcomes than their white counterparts when treated for these illnesses. These racial disparities in health are also present in the context of pregnancy, a fact that is illustrated most spectacularly by the often-quoted statistic describing black women's three- to four-fold increased risk of dying from a pregnancy-related cause as compared to white women.[12] Part II then provides an introduction to AI and explains the uses that scholars and developers predict medical AI technologies will have in healthcare and, specifically, the management of pregnancy. Part III subsequently serves as a primer on algorithmic bias—that is, systematic errors in the operation of an algorithm that result in a group being unfairly advantaged or disadvantaged. This Part explains the many causes of algorithmic bias and gives examples of algorithmic bias in medicine and healthcare. This Part argues that *we should expect algorithmic bias from medical AI that results in people of color receiving inferior healthcare. This is because medical AI technologies will be developed, trained, and deployed in a country with striking and unforgivable racial disparities in health*.

Part IV forms the heart of the Article. It begins by asking a question: Will patients of color even *want* medical AI? There is reason to suspect that significant numbers of them do not. Media attention to the skepticism with which many black people initially viewed COVID-19 vaccines has made the public newly aware of the higher levels of mistrust that black people, as a racial group, have toward healthcare institutions and their agents. That is, the banality of racial injustice has made black people more suspicious of medical technologies. This fact suggests that ethics—and justice—require providers to inform their patients of the use of a medical technology that likely embeds racial injustice within it.

The Part continues by making the claim that healthcare providers should disclose during the informed consent process their reliance on medical AI. To be precise, providers should have to tell their patients that

[12] Elizabeth A. Howell, Reducing Disparities in Severe Maternal Morbidity and Mortality, 61 Clinical Obstetrics & Gynecology 387, 387 (2018).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

an algorithm has affected the providers' decision-making around the patients' healthcare; moreover, providers should inform their patients how racial disparities in health may have impacted the algorithm's predictive accuracy. This Part argues that requiring these disclosures as part of the informed consent process revives the antiracist, anti-white supremacist origins of the informed consent process. To be sure, the practice of informed consent originated in the Nuremberg Trials' rebuke of Nazi medicine. These defiant, revolutionary origins have been expunged from the perfunctory form that the informed consent process has taken at present. Resuscitating the rebelliousness that is latent within informed consent will not only help to protect patient autonomy in the context of medical AI but may also be the condition of possibility for transforming the social conditions that produce racial disparities in health and healthcare. That is, the instant proposal seeks to call upon the rebellious roots of the doctrine of informed consent and use it as a technique of political mobilization. A short conclusion follows.

Two notes before beginning: First, although this Article focuses on medical AI in pregnancy and prenatal care, its argument is applicable to informed consent in all contexts—from anesthesiology to x-rays—in which a provider might utilize a medical AI device. Concentrating on pregnancy and prenatal care allows the Article to offer concrete examples of the phenomena under discussion and, in so doing, make crystal clear the exceedingly high stakes of our societal and legal decisions in this area.

Second, the moment that a provider consults a medical AI device when delivering healthcare to a patient of color certainly is not the first occasion in that patient's life in which racial disenfranchisement may come to impact the healthcare that they receive. That is, we can locate racial bias and exclusion at myriad sites within healthcare, medicine, and the construction of medical knowledge well before a clinical encounter in which medical AI is used. For example: people of color are underrepresented within clinical trials that test the safety and efficacy of drugs—a fact that might impact our ability to know whether a drug actually is safe and effective for people of color.[13] For example: the

[13] See The Nat'l Acads. of Scis., Eng'g & Med., Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups 24 (Kirsten Bibbins-Domingo & Alex Helman eds., 2022), https://nap.nationalacademies.org/catalog/26479/improving-representation-in-clinical-trials-and-research-building-research-equity [https://perma.cc/FE2H-9YC5] (explaining that "research has demonstrated that many groups underrepresented and excluded in clinical research can have distinct disease

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

National Institute of Health ("NIH") and the National Science Foundation ("NSF") fund medical research conducted by investigators of color at lower rates than that conducted by white investigators[14]—a fact that might contribute to the underfunding of medical conditions that disproportionately impact people of color. For example: most medical schools still approach race as a genetic fact instead of a social construction, with the result being that most physicians in the United States have not been disabused of the notion that people of color—black people, specifically—possess genes and genetic variations that make them get sicker and die earlier than their white counterparts.[15] For

presentations or health circumstances that affect how they will respond to an investigational drug or therapy" and that "[s]uch differences contribute to variable therapeutic responses and necessitate targeted efficacy and safety evaluation"). An FDA report of clinical trials that took place between 2015 and 2019 revealed that while non-Hispanic white people constituted only 61% of the general population in the United States, they were 78% of trial participants. See id. at 35; see also id. at 44–45 ("Even recently completed trials have failed to include enrollment consistent with the distribution of disease across the population—a Phase 2 trial of crenezumab in Alzheimer's disease with 360 participants across 83 sites in 6 countries reported 97.5 percent of participants being white, and only 2.8 percent of all participants being Hispanic.").

Notably, clinical trials only rarely include pregnant and lactating people. See id. at 40. This means that when most medications are introduced into the market, their safety and efficacy vis-à-vis pregnant and lactating people are unknown—although it is quite common for people to take medications while pregnant or lactating. See id. ("During pregnancy and lactation, greater than 90 percent of these individuals take at least one medication, either to treat pregnancy-related complications or to treat ongoing medical issues.").

[14] See Christine Yifeng Chen et al., Meta-Research: Systemic Racial Disparities in Funding Rates at the National Science Foundation, eLife, Nov. 29, 2022, at 2, https://doi.org/10.7554/eLife.83071 [https://perma.cc/NFS8-T3LB] (showing that the National Science Foundation funded proposals by white principal investigators at +8.5% of the average funding rate while funding proposals by Asian, black, and Native Hawaiian/Pacific Islander principal investigators at –21.2%, –8.1%, and –11.3% of the average funding rate, respectively); Donna K. Ginther et al., Race, Ethnicity, and NIH Research Awards, 333 Science 1015, 1016 (2011), https://doi.org/10.1126/science.1196783 [https://perma.cc/NQA9-LYMG] (showing that the National Institute of Health funded proposals by black principal investigators at close to half the rate as white principal investigators).

[15] See Christina Amutah et al., Misrepresenting Race—The Role of Medical Schools in Propagating Physician Bias, 384 New Eng. J. Med. 872, 873–74 (2021). Funding for research into the imagined genetic causes of racial disparities in health outcomes vastly outstrips funding for research into social determinants of health or the physiological effects of stress and racism on people of color. Shawn Kneipp et al., Trends in Health Disparities, Health Inequity, and Social Determinants of Health Research, 67 Nursing Rsch. 231, 231 (2018). See also René Bowser, Racial Profiling in Health Care: An Institutional Analysis of Medical Treatment Disparities, 7 Mich. J. Race & L. 79, 114 (2001) (arguing that "physicians who focus on racism as opposed to cultural peculiarities or the genetic basis of disease are likely to be considered both as not 'real scientists' and as dangerous" and stating that producing

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

example: pulse oximeters, which use infrared light to measure an individual's blood saturation levels, are so common as to be called ubiquitous, even though it is well-known that the devices do not work as well on more pigmented skin.[16] For example: most clinical studies that are used to establish evidence-based practices are conducted in well-resourced facilities, making their generalizability to more contingently equipped and more unreliably funded facilities uncertain.[17] For example: many research studies do not report their findings by race, thereby impeding our ability to know whether the studies' results are equally true for all racial groups.[18] And so on. If providers ought to notify their patients (especially their patients of color) that the provider has relied upon medical AI when caring for the patient, then it is likely true that providers similarly ought to notify their patients about racial inequity in other contexts as well. That is, there is a compelling argument that when a provider prescribes a medication to a patient, they might need to notify the patient that preciously small numbers of people who were not white cisgender men participated in the clinical trial of the medication.[19] There is a compelling argument that when a provider tells a black patient that the results of her pulmonary function test were "normal," they might also need to inform that patient that if she were white, her results would be considered "abnormal," as the idea that the races are biologically distinct

research that explains racial disparities in health outcomes in terms of culture and genes, as opposed to structural racism and inherited disadvantage, "enhances the researcher's status"). This funding disparity undoubtedly contributes to the perpetuation of the myth of biological race.

[16] See Haley Bridger, Skin Tone and Pulse Oximetry: Racial Disparities in Care Tied to Differences in Pulse Oximeter Performance, Harv. Med. Sch. (July 14, 2022), https://hms.harvard.edu/news/skin-tone-pulse-oximetry [https://perma.cc/HZW8-YMAS].

[17] See The National Academies of Sciences, Engineering, and Medicine, supra note 13, at 25 (observing that "[c]linical research is often performed in well-resourced tertiary care sites in large urban centers, and may have limited applicability to community sites, less well-resourced safety net settings, and rural settings").

[18] See id. at 31 (stating that the "[l]ack of representative studies on screening for cancer or cardiometabolic disease may lead to recommendations that fail to consider earlier ages or lower biomarker thresholds to start screening that might be warranted in some populations" and observing that "due to [a] lack of studies that report findings by race," the guidelines for some screenings are universal, although there is some evidence that they should vary by race and age).

[19] See Barbara A. Noah, Racial Disparities in the Delivery of Health Care, 35 San Diego L. Rev. 135, 152 (1998) (noting that "[b]efore the National Institutes of Health (NIH) issued a directive in 1990, investigators almost uniformly tested new chemical entities only on white male subjects").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

has long informed notions of whether a set of lungs is healthy or not.[20] There is a compelling argument that when a provider affixes a pulse oximeter to the finger of a patient of color, they might also need to inform that patient that the oximeter's readings may be inaccurate—and the care that she receives based on those readings may be inferior[21]—given the widely known and undisputed fact that such devices do not work as well on darker skin. There is a compelling argument that when a physician tells a pregnant patient laboring in a safety net hospital that the evidence-based practice for patients presenting in the way that she presents is an artificial rupture of membranes ("AROM") to facilitate the progression of the labor, they might also need to inform the patient that the studies that established AROM as an evidence-based practice were conducted in well-funded research hospitals that were affiliated with universities.[22] There is a compelling argument that when a physician tells a forty-year-old black patient that he does not need to do a screening for colorectal cancer until age forty-five, they might also need to inform the patient that the studies that established forty-five as the age when such screenings should commence did not report their findings by race.[23] And so on.

It does not defeat this Article's claim to observe that racial bias and exclusion are pervasive throughout medicine and healthcare and that providers in many contexts outside of the use of medical AI ought to notify patients how this bias and exclusion may affect the healthcare that they are receiving. Indeed, it is seductive to claim in those other contexts that it is better to fix the inequities in the healthcare than to tell patients of color about them—a fact that is also true in the context of medical AI. However, fixing the inequities in healthcare in those other contexts and telling patients about them are not mutually exclusive—a fact that is also true in the context of medical AI. And as Part IV argues, *telling patients*

[20] See Lundy Braun, Breathing Race into the Machine: The Surprising Career of the Spirometer from Plantation to Genetics, at xv (2014).

[21] See Bridger, supra note 16 (describing a study that showed that pulse oximeters reported blood oxygen saturation levels for patients of color that were higher than what they actually were, leading these patients' providers to give them supplemental oxygen at lower rates).

[22] See, e.g., Alan F. Guttmacher & R. Gordon Douglas, Induction of Labor by Artificial Rupture of the Membranes, 21 Am. J. Obstetrics & Gynecology 485, 485 (1931) (establishing artificial rupture of the membranes as an evidence-based practice in obstetrics after studying the safety and efficacy of the procedure among patients cared for at a clinic affiliated with Johns Hopkins University).

[23] See Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement, 325 JAMA 1965, 1970 (2021), https://jamanetwork.com/journals/jama/fullarticle/2779985 [https://perma.cc/TV68-6W75].

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

about the inequities in those other contexts might be the condition of possibility of fixing the inequities—a fact that is also true in the context of medical AI.

Essentially, this Article's claim may be applied in a range of circumstances. In this way, this Article's investigation into how algorithmic bias in medical AI should affect the informed consent process is simply a case study of a broader phenomenon. This Article's insights vis-à-vis medical AI are generalizable to all medical interventions and noninterventions.

I. RACIAL DISPARITIES IN HEALTH AND HEALTHCARE

One of the many manifestations of structural racism in the U.S. is nonwhite people being sicker and dying earlier than their white counterparts.[24] This is the fact of racial disparities in health.

There is a key distinction that we ought to recognize—that between racial disparities in *health* and racial disparities in *healthcare*. The former refers to the verity that people of color suffer from higher rates of every major common illness, including hypertension, heart disease, diabetes, kidney disease, lung disease, and asthma.[25] With respect to obstetrics specifically, black people suffer higher rates of pregnancy-related complications, injuries, and deaths than white people;[26] further, black infants suffer higher rates of infant mortality and morbidity than white infants.[27] Racial disparities in health have contributed to racial disparities in life spans.[28] While racial disparities in life spans have been declining—owing, in large part, to the higher rates at which white people have been dying from opioid overdoses, suicide, and COVID[29]—it remains true that nonwhite people, on average, have shorter, sicker lives than white people.

[24] Khiara M. Bridges, Critical Race Theory: A Primer 320 (2018).

[25] Lesley Russell, Fact Sheet: Health Disparities by Race and Ethnicity, Ctr. for Am. Prog. (Dec. 16, 2010), https://www.americanprogress.org/issues/healthcare/news/2010/12/16/8762/fact-sheet-health-disparities-by-race-and-ethnicity/ [https://perma.cc/HQZ3-WRST].

[26] Jamila Taylor, Cristina Novoa, Katie Hamm & Shilpa Phadke, Eliminating Racial Disparities in Maternal and Infant Mortality, Ctr. for Am. Prog. (May 2, 2019), https://www.americanprogress.org/issues/women/reports/2019/05/02/469186/eliminating-racial-disparities-maternal-infant-mortality/ [https://perma.cc/LMU7-XSSU].

[27] Id.

[28] See Bridges, supra note 24.

[29] Anne Case & Angus Deaton, Deaths of Despair and the Future of Capitalism 64–65 (2020); Shannon Sabo & Sandra Johnson, COVID-19 Impacts on Mortality by Race/Ethnicity and Sex, U.S. Census Bureau (June 22, 2023), https://www.census.gov/library/stories/2023/

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

The causes of racial disparities in health are many. They include the fact that people of color disproportionately bear the burden of poverty, which is known to compromise the health of those forced to live in it;[30] the disproportionate rates at which people of color lack health insurance and, therefore, access to healthcare;[31] the polluted environments in which people of color are compelled to live;[32] the closure of hospitals in the neighborhoods that people of color call home and the consequent strain on the hospitals that remain;[33] the health-damaging, physiological effects of the racism-based stress that people of color experience;[34] and the inferior care that people of color receive—both from the segregated hospitals in which they obtain their care,[35] as well as from individual healthcare providers, who may have biases that decrease the quality of the care that they give their patients of color,[36] or who may not be as skilled as those that provide care to white patients.[37]

06/covid-19-impacts-on-mortality-by-race-ethnicity-and-sex.html [https://perma.cc/F2KR-N2B4].

[30] David R. Williams & Pamela Braboy Jackson, Social Sources of Racial Disparities in Health, 24 Health Affs. 325, 327 (2005).

[31] Id.

[32] Dorceta E. Taylor, Toxic Communities: Environmental Racism, Industrial Pollution, and Residential Mobility 3 (2014).

[33] Ruqaiijah Yearby, Sick and Tired of Being Sick and Tired: Putting an End to Separate and Unequal Health Care in the United States 50 Years After the Civil Rights Act of 1964, 25 Health Matrix 1, 12 (2015) (noting that "hospital closures in African American communities . . . leave[] predominately African American neighborhoods without access to quality health care services").

[34] See Arline Geronimus, Weathering: The Extraordinary Stress of Ordinary Life in an Unjust Society 14–15 (2023).

[35] Id. at 194; Yearby, supra note 33, at 15–16 (observing that "[i]n highly racially segregated areas, African Americans are more likely to undergo surgery in low-quality hospitals, whereas in areas with low degrees of racial segregation, African Americans and Caucasians are likely to undergo surgery at low quality hospitals at the same rate" and stating that "[t]his is significant because among Medicare patients, most of the racial disparities in risk-adjusted death rates for major surgery are a result of the site of care").

[36] See Dayna Bowen Matthew, Just Medicine: A Cure for Racial Inequality in American Health Care 2–3 (2015).

[37] See David R. Williams, Race, Health, and Health Care, 48 St. Louis U. L.J. 13, 32 (2003) ("Some evidence also suggests that non-white patients are more likely than their white counterparts to be treated by lower quality physicians."). It may be problematic to state that the physicians who treat patients of color are not as skilled as those who treat white patients, as the former are more likely to be people of color themselves. See Yearby, supra note 33, at 19 (explaining that black patients frequently turn to black physicians for healthcare). Moreover, black physicians may themselves be subjected to discrimination, which may impact their ability to access the goods and services that their patients need. See id. (observing that black physicians "report greater difficulties accessing high-quality specialists, . . . imaging,

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

In contrast to racial disparities in *health*, racial disparities in *healthcare* refer to the verity that providers sometimes offer their patients of color treatments and interventions that are different from, and in most cases inferior to, the treatments and interventions that providers offer their white patients. These racial disparities in healthcare typically persist even after one controls for all of the factors that would otherwise justify differences in care.

In 2003, the Institute of Medicine,[38] a nonprofit that conducts wide-ranging research related to health, medicine, and science, released a report that brought increased attention to the reality that racial disparities in *healthcare* undoubtedly contribute to racial disparities in *health*.[39] The report, *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*, documented the breathtaking extent of the poorer healthcare that people of color receive. The report extensively itemized the numerous ways in which people of color—black people, specifically—were less likely to receive health-conserving and life-saving treatments and interventions across a range of health conditions, including heart disease and cardiac emergencies, cancer, stroke, kidney disease, HIV/AIDS, asthma, diabetes, and mental illness.[40] The only interventions that black people were more likely to receive at higher rates than their white counterparts were those that are largely unwanted: antipsychotic medications in the event of mental health crises, amputations, and bilateral orchiectomies—that is, castrations.[41] The Institute of Medicine summarized its devastating findings:

and nonemergency admission of their patients . . . than physicians serving predominantly nonminority patients" (quoting Rene Bowser, Race and Rationing, 25 Health Matrix 87, 97–98 (2015))); see also W. Michael Byrd et al., African-American Physicians' Views on Health Reform: Results of a Survey, 86 J. Nat'l Med. Ass'n 191, 194–95 (1994) (reporting the results of a survey of black physicians in which the overwhelming majority of participants reported that they had experienced racial discrimination in various sites of their profession, including the peer review process, malpractice settlements, obtaining privileges to practice in hospitals, promotions, referrals from colleagues, and reimbursement by Medicare, Medicaid, and private insurance companies).

[38] The Institute of Medicine has been renamed the National Academy of Medicine. Institute of Medicine to Become National Academy of Medicine, Nat'l Acads. Sci., Eng'g & Med. (Apr. 28, 2015), https://www.nasonline.org/news-and-multimedia/news/april-28-2015-NAM.html [https://perma.cc/EMW4-D7LK].

[39] Inst. of Med., Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care 29 (Brian D. Smedley, Adrienne Y. Stith & Alan R. Nelson eds., 2003).

[40] See id. at 39–77.

[41] See id. at 70, 74.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

> Racial and ethnic minority patients are found to receive a lower quality and intensity of healthcare and diagnostic services across a wide range of procedures and disease areas. This finding is remarkably consistent and robust, as only a handful of the several hundred studies reviewed here and by others . . . find no racial and ethnic differences in care. In studies where patients' sociodemographic characteristics (e.g., education level, income), insurance status (e.g., public or privately funded insurance) and clinical factors (e.g., co-morbid illness, severity of disease) are controlled, these racial and ethnic differences are generally attenuated, but rarely disappear completely. Further, in a few well-designed, prospective studies, these disparities in care have been linked to poorer clinical outcomes and higher mortality among minorities.[42]

Not much has changed in the twenty years since the release of the Institute of Medicine's report. Research continues to reveal racial disparities in healthcare. Recent studies show that, even when one controls for all relevant characteristics, black patients are more likely to have their lower limbs amputated as treatment for diabetic foot ulcers,[43] are less likely to be treated with surgery for lung cancer,[44] and are less likely to be offered "treatment for depression, emergency department analgesia, influenza vaccine, and referral for cardiac catheterization."[45]

[42] Id. at 77.

[43] See Frank M. McClellan, James E. Wood & Sherin M. Fahmy, It Takes a Village: Reforming Law to Promote Health Literacy and Reduce Orthopedic Health Disparities, 8 J. Health & Biomedical L. 333, 338 (2013); Kartik K. Ganju, Hilal Atasoy, Jeffrey McCullough & Brad Greenwood, The Role of Decision Support Systems in Attenuating Racial Biases in Healthcare Delivery, 66 Mgmt. Sci. 5171, 5172 (Nov. 2020), https://doi.org/10.1287/mnsc.2020.3698 [https://perma.cc/H6HJ-JHSZ] ("In our focal context, peripheral arterial disease, black patients are more likely [to] have their limbs amputated, while white patients are more likely [to] have their limbs saved through revascularization. . . . Even more troubling, these differences persist after controlling for socio-demographic factors, comorbidities, and insurance provider.").

[44] See D.A. Paul, R. Locke, K. Zook, K.H. Leef, J.L. Stefano & G. Colmorgen, Racial Differences in Prenatal Care of Mothers Delivering Very Low Birth Weight Infants, 26 J. Perinatology 74, 74 (2006).

[45] Id. at 74; see also Yearby, supra note 33, at 22 (noting that "a patient's race can affect physicians' 'question-asking in clinical interview, diagnostic decision-making, referral to specialty care, symptom management, and treatment recommendations'" (quoting Michelle van Ryn, Disparities in Care and Unintended Biases in Clinical Decision-Making and Behavior, Address at the Case Western Reserve University School of Law, Law-Medicine Symposium 7 (Mar. 27, 2014) (on file with author))).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

With respect to obstetrics, the context to which this Article is most closely attuned, racial disparities in healthcare are also well-documented. It is well-established that obstetricians perform C-sections at higher rates on black birthing people.[46] It deserves underscoring that clinical factors or socioeconomic characteristics cannot explain these higher rates.[47] That is, a black person is more likely to have a C-section than a white person who is identical in terms of age, weight, parity, education, insurance status, and comorbid conditions.[48] Moreover, as legal scholar Colleen Campbell explains, "the disparity is visible even for low-risk pregnancies, i.e., those pregnancies for which there is no medical complication. In other words, perfectly healthy Black women who do not need a C-section are also receiving this major surgery . . . ."[49] Of course, C-sections are desirable when they are medically indicated—that is, when there is evidence that a birthing person or their fetus is suffering from a condition that makes a vaginal birth dangerous or life-threatening.[50] However, when a C-section is not medically indicated—and when a physician performs the surgery out of expedience or poor judgment, or as a matter of hospital policy[51]—then the physician unjustifiably subjects their patient to "a major surgery that poses greater fetal and maternal health risks than vaginal birth."[52]

There are many other examples of racial disparities in obstetrical healthcare: Providers are less likely to perform surgical interventions on black people who are suffering a hemorrhage after childbirth.[53] Providers are less likely to prescribe antenatal steroids, which can improve the health outcomes of premature infants, to black people who are at risk of preterm birth.[54] Providers are less likely to use tocolysis, medications that

[46] See Colleen Campbell, Medical Violence, Obstetric Racism, and the Limits of Informed Consent for Black Women, 26 Mich. J. Race & L. 47, 62 (2021).

[47] Id.

[48] Id.

[49] Id. at 62–63.

[50] Theresa Morris, Cut It Out: The C-Section Epidemic in America 5 (2013).

[51] The many reasons that a physician may perform a C-section despite a lack of a clinical indication include "the constant specter of malpractice liability; widespread bans on vaginal birth after C-section (VBAC); variation in hospital practices; and subjective, as opposed to objective, clinical indications." Campbell, supra note 46, at 61–62.

[52] Id. at 61.

[53] See Allison S. Bryant, Ayaba Worjoloh, Aaron B. Caughey & A. Eugene Washington, Racial / Ethnic Disparities in Obstetric Outcomes and Care: Prevalence and Determinants, 202 Am. J. Obstetrics & Gynecology 335, 338 (2010).

[54] See Paul et al., supra note 44, at 76–77.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

may delay the birth of an infant, on black people who are at risk of delivering prematurely.[55] Providers are less likely to give black pregnant people advice on smoking and drinking cessation during their prenatal care appointments.[56] Providers are also less likely to screen black people for lipid disorders,[57] which can cause gestational diabetes, preeclampsia, preterm birth, and other adverse pregnancy outcomes.[58]

Medical AI technologies will be developed, trained,[59] and deployed in a country with these dramatic and inexcusable racial disparities in health and healthcare. Now, many have taken the fact of racial disparities in health and healthcare as a reason to be excited about the prospect of medical AI. The idea is that medical AI will leave less discretion and decision-making responsibility to human physicians and healthcare providers, who so frequently labor under the weight of explicit and implicit biases.[60] With humans' flawed judgment supported, or supplanted, by a technology that contains no irrational racial aversions or

[55] Id. Paul and coauthors note that research has not definitively established that tocolytics produce better pregnancy outcomes for people at risk of a preterm birth. Thus, they frame their investigation as one into whether physicians offered "different," as opposed to "better," care to white patients. Id. at 74. They ultimately answer in the affirmative, explaining that "Cesarean delivery and receipt of tocolytic medications may reflect 'attentiveness to the patient' rather than quality of care." Id. at 77.

[56] See Michael D. Kogan, Milton Kotelchuck, Greg R. Alexander & Wayne E. Johnson, Racial Disparities in Reported Prenatal Care Advice from Health Care Providers, 84 Am. J. Pub. Health 82, 86 (1994).

[57] See Reed Mszar et al., Racial/Ethnic Disparities in Screening for and Awareness of High Cholesterol Among Pregnant Women Receiving Prenatal Care, 10 J. Am Heart Assoc., Jan. 2021, at 5.

[58] Lakshmi S. Tummala, Akanksha Agrawal & Gina Lundberg, Management Considerations for Lipid Disorders During Pregnancy, 23 Current Treatment Options Cardiovascular Med. 50, 51–52 (2021).

[59] For a definition of "training" in the context of medical AI, see infra notes 78–87 and accompanying text.

[60] See Sharona Hoffman & Andy Podgurski, Artificial Intelligence and Discrimination in Health Care, 19 Yale J. Health Pol'y, L. & Ethics no. 3, 2020, at 14 ("One reason for enthusiasm about AI is the hope that it will diminish human bias in health care."). Sociologist Ruha Benjamin's ruminations on the optimism with which people turn to high-tech fixes as a general matter is appropriate with respect to medical AI. See generally Ruha Benjamin, Race After Technology: Abolitionist Tools for the New Jim Code (2019) [hereinafter Benjamin, Race After Technology]. She writes:

> [A]s people become more attuned to racial biases in hiring, firing, loaning, policing, and a whole host of consequential decisions—an awareness we might take to be a sign of social progress—this very process also operates as a kind of opportunity for those who seek to manage social life more efficiently. The potential for bias creates a demand for more efficient and automated organizational practices . . . .

Id. at 30.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

associations, racial disparities in healthcare may be diminished substantially, or eliminated entirely. Or so the argument goes.

There is some reason to be optimistic that medical AI will operate to reduce or eliminate the opportunities for providers' biases to have a negative impact on the healthcare that they give their patients of color. One study showed that the use of clinical decision support systems ("CDSS")—computer programs that analyze electronic health records to help providers give evidence-based care[61]—worked to reduce racial disparities in physicians' decisions not to attempt revascularization of a diabetic patient's limb and to proceed with amputation.[62] This study showed that the CDSS worked to "attenuate bias under uncertainty" by standardizing the tests that physicians conduct on their patients and by prompting the primary care physician to seek an evaluation of the limb by a specialist, who would be better able to identify appropriate candidates for revascularization.[63] The authors of this study optimistically concluded:

> [A]lthough significant research has studied the effect of algorithms on racial biases, and the punitive effects they can yield . . . we demonstrate a means by which information systems can play a beneficial role. Specifically, the ability of CDSS to change organizational protocols, with the objective of adhering to best practice guidelines, can reduce the subjectivity in healthcare delivery.[64]

Another study showed that medical AI may be able to protect people of color from receiving inferior healthcare when that inferior healthcare is a product of the historical exclusion of people of color from the processes by which medical standards were generated.[65] The norm in

[61] Ctrs. for Disease Control & Prevention, Clinical Decision Support Systems (Feb. 28, 2022), https://hdsbpc.cdc.gov/s/topic/0TO3d0000000IpOGAU/clinical-decision-support-systems [https://perma.cc/F8NM-36XF].

[62] See Ganju et al., supra note 43, at 5172.

[63] Id.

[64] Id. (citations omitted). Interestingly, the CDSS that the authors studied does not contain a black-box algorithm. See id. at 5178. Indeed, the authors appear a bit more skeptical about the ability of black-box algorithms to eliminate the biases that contribute to racial disparities in healthcare. See id. ("This research indicates that the use of digital protocols and guidelines, *rather than a black box approach*, offers the opportunity to limit the presence of undesirable biases." (emphasis added)).

[65] See Emma Pierson, David M. Cutler, Jure Leskovec, Sendhil Mullainathan & Ziad Obermeyer, An Algorithmic Approach to Reducing Unexplained Pain Disparities in Underserved Populations, 27 Nature Med. 136, 136 (2021).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

medicine has been to base osteoarthritis severity on a measure—the KLG score—that was developed among a patient population that was entirely white (and British).[66] There tends to be a greater discrepancy between black patients' subjective reports of knee pain and the severity of osteoarthritis as determined by the KLG score, an "objective" measure.[67] Researchers designed an algorithm to identify osteoarthritis severity in patients based on x-rays of their knees.[68] Crucially, they trained the algorithm on a racially and socioeconomically heterogeneous dataset.[69] The result was an algorithm that more capably identified patients with severe knee pain—although these patients lacked severe radiographic disease as identified by traditional methods.[70] If introduced widely, the algorithm may have the effect of reducing the well-documented racial disparities in access to knee surgery and the overreliance on pain medications to manage knee pain in low-income patients and patients of color.[71]

It is vital that we soberly consider the introduction of medical AI into the clinical encounter. First, we have to recognize that while medical AI may reduce racial disparities in *healthcare*, it cannot entirely eliminate racial disparities in *health*. This is simply because implicit and explicit biases and "subjectivity in healthcare delivery"[72] do not entirely explain why people of color are sicker and die earlier than their white counterparts. Medical AI does nothing to improve the health-damaging conditions under which people of color live—health-damaging conditions that are doing most of the heavy lifting when it comes to making people of color unwell and shortening their lives. Essentially, we must not pretend that the reason why people of color are sicker and die earlier than white people is that our technology is not as advanced as it could be.

Second, we have to consider the possibility that medical AI is just as likely to perpetuate existing racial disparities in healthcare, and therefore

[66] Id. at 137.

[67] See id. at 139 ("[S]tandard radiographic measures of severity overlook objective but undiagnosed features that disproportionately affect diagnosis and management of underserved populations with knee pain.").

[68] Id. at 136.

[69] See id. at 137–38.

[70] See id. at 138 ("[M]odels trained on the diverse training sets achieved better predictive performance for pain and greater reductions in racial and socioeconomic pain disparities than models trained on the non-diverse training sets of the same size."). Fascinatingly, the researchers did not know "which features of the knee the algorithm is using." Id. at 139.

[71] See id. at 138.

[72] Ganju et al., supra note 43, at 5172.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

racial disparities in health, as it is to diminish them. This is simply because, as noted above, medical AI technologies will be developed, trained, and deployed in a country with dramatic and inexcusable racial disparities in health and healthcare. The next two Parts explain why we should expect health inequities to be built into the machine.

II. AI IN PRENATAL CARE

This Part provides a definition of AI and describes how the technology works. It then explains how AI has been applied in healthcare before exploring the particular usages that AI has had, and may have in the future, in prenatal care.

A. AI for the Uninitiated

"Artificial intelligence," or "AI," refers to a range of technological possibilities.[73] At the most general level, the term refers to programs that enable computers to perform tasks that ordinarily have been performed by humans,[74] e.g., driving cars, transcribing speech, or diagnosing a tumor as malignant. Technically speaking, "machine learning," or "ML," is a subset of AI, referring to computer programs that, through experience, become more proficient at the analytical task that they are designed to do.[75] Although there are important differences between the general

[73] See Kathleen Walch, Is There a Difference Between Assisted Intelligence vs. Augmented Intelligence?, Forbes (Jan. 12, 2020, 1:00 AM), https://www.forbes.com/sites/cognitiveworld/2020/01/12/is-there-a-difference-between-assisted-intelligence-vs-augmented-intelligence [https://perma.cc/QL6Q-RRNW] (explaining that AI consists of "a continuum of human-machine intelligence interaction ranging from situations where machines are basically repeating many of the tasks humans are already doing (assisted) to enabling humans to do more than they are currently capable of doing (augmented) to fully accomplishing tasks on their own without human intervention (autonomous)").

[74] See Lena Davidson & Mary Regina Boland, Enabling Pregnant Women and Their Physicians to Make Informed Medication Decisions Using Artificial Intelligence, 47 J. Pharmacokinetics & Pharmacodynamics 305, 306 (2020) ("AI is used to refer to the method by which computer systems can perform tasks that would typically require a human.").

[75] See Hoffman & Podgurski, supra note 60, at 8 (defining ML as a subfield of artificial intelligence involving computers that "automatically detect patterns in data, and then use the uncovered patterns to predict future data or to perform decision-making tasks under uncertainty" (quoting Kevin P. Murphy, Machine Learning: A Probabilistic Perspective 1 (2012) (internal quotation marks omitted))); Sandra L. J. Johnson, AI, Machine Learning, and Ethics in Health Care, 39 J. Legal Med. 427, 429 (2019) ("Machine learning is a branch of AI where computer systems learn from large bodies of data; then, through statistical methods and algorithms, they derive a rule to explain or categorize the data and predict future outcomes."); Davidson & Boland, supra note 74, at 306 (describing ML as using "statistical techniques to

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

category of AI and the specific category of ML, those differences are not relevant to this Article's main argument. For that reason, this Article uses the two terms interchangeably.

The basic building block of AI is the algorithm, which might be defined as a "set of instructions—a preset, rigid, coded recipe that gets executed when it encounters a trigger."[76] AI, in turn, "is a group of algorithms that can modify its algorithms and create new algorithms in response to learned inputs and data as opposed to relying solely on the inputs it was designed to recognize as triggers."[77]

Computer scientists instruct AI algorithms to engage in the analytical work that the programmers hope the programs eventually will do by giving them vast amounts of information in the form of "training data."[78] One cannot underscore enough that *large* quantities of information make AI possible; indeed, AI is unachievable without "big data."[79] AI programs sift through these enormous datasets to uncover otherwise unidentified patterns.

allow for the computer to 'learn,' or progressively improve performance on a given task, without being explicitly programmed").

[76] Kaya Ismail, AI vs. Algorithms: What's the Difference?, CMSWire (Oct. 26, 2018), https://www.cmswire.com/information-management/ai-vs-algorithms-whats-the-difference/ [https://perma.cc/8PCA-X6W9].

[77] Id.

[78] See Hoffman & Podgurski, supra note 60, at 9.

[79] As Glenn Cohen and Harry Graver explain, the meaning of "big data" is a bit slippery. See I. Glenn Cohen & Harry S. Graver, Cops, Docs, and Code: A Dialogue Between Big Data in Health Care and Predictive Policing, 51 U.C. Davis L. Rev. 437, 437 (2017) (describing the meaning of "big data" as "somewhat protean"). However, at least one of the meanings of "big data" refers to "information assets characterized by such a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value." Tokio Matsuzaki, Ethical Issues of Artificial Intelligence in Medicine, 55 Cal. W. L. Rev. 255, 257 (2018) (quoting Andrea De Mauro, Marco Greco & Michele Grimaldi, What Is Big Data? A Consensual Definition and a Review of Key Research Topics, 1644 AIP Conf. Proc. 97, 103 (2015)); see also Cohen & Graver, supra, at 437–38 (noting that from a legal perspective, big data is best defined as "the combination of a technology with a process—the technology 'is a configuration of information-processing hardware capable of sifting, sorting, and interrogating vast quantities of data in very short times' while the process 'involves mining the data for patterns, distilling the patterns into predictive analytics, and applying the analytics to new data'" (quoting Julie E. Cohen, What Privacy Is For, 126 Harv. L. Rev. 1904, 1920 (2013))). With respect to healthcare and medicine, the data in big data comes from "a myriad of sources, including patients, health-care providers, insurers, manufacturers, the government, and even mobile devices such as smartphones and wearables." Hoffman & Podgurski, supra note 60, at 9.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

In "supervised" ML, humans label the input data and, in so doing, identify the proper output.[80] For example, a computer scientist designing a program that can detect cancerous tumors in x-rays will feed the program tens of thousands of x-ray images of tumors; in this case, the x-ray images are the input data. Each x-ray will have been labeled as one in which the tumor is either cancerous or noncancerous; a tumor's status as either cancerous or noncancerous is the output. Whenever the algorithmic model improperly identifies a cancerous tumor as noncancerous, or vice versa, it adjusts the parameters of the model.[81] The process is repeated until the likelihood that the model will yield the correct answer cannot be increased.[82] The programmer then provides the program with a "test set" of data—that is, a different set of labeled data (previously unseen and, ideally, from an entirely different data source)—in order to confirm the accuracy of the program's predictions.[83] Once validated, the program can be unleashed on the world and exposed to input data with unknown labels, i.e., images of tumors that no human previously has identified as cancerous or noncancerous.[84]

"Unsupervised" ML differs from "supervised" ML inasmuch as the former involves training algorithms on unlabeled data—for example, a dataset composed of sundry x-ray images, CT scans, records of vital signs, and other data culled from electronic health records.[85] Without human intervention, the algorithmic model would then identify patterns in the data and come to predict outputs—be it the likelihood of an individual suffering a heart attack, developing preeclampsia during pregnancy, or responding well to a certain treatment regimen.[86] We have not yet arrived at an era of unsupervised ML. However, because labeling

[80] See W. Nicholson Price II & Arti K. Rai, Clearing Opacity Through Machine Learning, 106 Iowa L. Rev. 775, 777 (2021).

[81] See id.

[82] Id.

[83] Id. at 777–78; Johnson, supra note 75, at 430.

[84] See Hoffman & Podgurski, supra note 60, at 9.

[85] See Johnson, supra note 75, at 429 ("In unsupervised learning, the ML algorithm is provided with large volumes of data on every aspect of the object. The machine then separates images and learns details based on inherent characteristics of the object independent of human input and interaction."); Cohen, supra note 11, at 1430 (noting that in unsupervised ML, "the algorithms are presented with unlabeled data and tasked with discerning underlying patterns in the data").

[86] See Johnson, supra note 75, at 429.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

data for supervised ML is incredibly costly and time intensive, many observers long for the day when most ML will be unsupervised.[87]

It might be important to note that AI systems cannot establish causal relationships between inputs and outputs.[88] Thus, the outcomes that these systems generate are *predictive*.[89] Indeed, most of their outputs come with a prediction score, identifying the program's confidence that its result is, indeed, correct.[90]

Observers frequently refer to the "black box" nature of ML.[91] By "black box," these observers mean that the reasoning behind a tool's output is opaque and that it is impossible for humans to understand precisely how it reached its outcome.[92] This, of course, limits the ability of humans to verify that an algorithmic model's outcome is, in fact, correct.[93] As evocatively described by Ian Kerr and Vanessa Gruben, "[w]ith increasing frequency, the final dance between data and algorithm takes place without understanding, often without human intervention or oversight. Indeed, in many cases, humans have a hard time explaining how or why the machine got it right (or wrong)."[94] This inability to

[87] See Price & Rai, supra note 80, at 777 n.1 ("Because the costs of proper data curation by experts are often nontrivial, unsupervised learning has been described as the 'holy grail' of data-based machine learning.").

[88] See Davidson & Boland, supra note 74, at 306 (noting that ML is "not designed to demonstrate causality, and at best can provide likely candidates for causality").

[89] A. Michael Froomkin, Ian Kerr & Joelle Pineau, When AIs Outperform Doctors: Confronting the Challenges of a Tort-Induced Over-Reliance on Machine Learning, 61 Ariz. L. Rev. 33, 48 (2019) ("Neural networks, in particular, do not typically extract causal relationships between inputs and outputs; therefore, it is important to interpret any relationship between input and output as a predictive one, no matter how intuitive such relationships might look on the surface.").

[90] See Daniel Schönberger, Artificial Intelligence in Healthcare: A Critical Analysis of the Legal and Ethical Implications, 27 Int'l J.L. & Info. Tech. 171, 175 (2019) ("The output usually includes a probability score (e.g. 0.8) reflecting the confidence of the algorithm as regards this 'prediction.'").

[91] See id. at 177 (observing that AI "algorithms are hence often portrayed as 'black boxes'").

[92] See Froomkin et al., supra note 89, at 48 (observing that AI algorithms have "poor explainability" inasmuch as "they cannot extract a compact narrative explaining the logic behind their reasoning").

[93] W. Nicholson Price II, Medical Malpractice and Black-Box Medicine, *in* Big Data, Health Law, and Bioethics 295, 300 (I. Glenn Cohen, Holly Fernandez Lynch, Effy Vayena & Urs Gasser eds., 2018) (noting that that once an individual "has decided to use a particular black-box algorithm . . . he or she cannot understand and thus verify the algorithm's recommendation against his or her body of substantive expertise").

[94] Ian Kerr & Vanessa Gruben, AIs as Substitute Decision-Makers, 21 Yale J.L. & Tech. 78, 90 (2019).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

explain the *why* behind a prediction that an ML program generates is due to two primary reasons. First, the sheer complexity of the patterns that the program uncovers in the data—and, consequently, the rules that the model generates for understanding those patterns—might be beyond the comprehension of humans.[95] Second, once an ML program is designed, the program adjusts its own parameters—without programmers' involvement.[96] If those adjustments are unknown, then observers will find it impossible to decipher what led the program to the output at which it arrived.[97] At present, no method exists for enabling the programs themselves to offer reasons for their outputs.[98] Further, the approaches that computer scientists have developed to satisfy demands for AI that is explainable may not actually answer the question that they purport to answer.[99]

[95] See W. Nicholson Price II, Regulating Black-Box Medicine, 116 Mich. L. Rev. 421, 430 (2017) (stating that at times, "algorithms are nontransparent because, while they may rely on explicit rules, those rules are too complex for us to explicitly understand").

[96] See Robin C. Feldman, Ehrik Aldana & Kara Stein, Artificial Intelligence in the Health Care Space: How We Can Trust What We Cannot Know, 30 Stan. L. & Pol'y Rev. 399, 408 (2019) ("AI systems are opaque largely because they constantly modify their own parameters and rules.").

An additional cause of AI's opacity is a developer's choice to keep secret its algorithms and/or its training data. See Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations, 366 Science 447, 447 (2019) ("Algorithms deployed on large scales are typically proprietary, making it difficult for independent researchers to dissect them."). AI opacity that is due to this cause, however, is more easily remedied than opacity due to the other two factors identified above.

[97] See Price, supra note 95, at 430 (stating that, at times, "the relationships used in a black-box algorithm are literally unknowable because of the machine-learning techniques employed—that is, no one, not even those who programmed the machine-learning process, knows exactly what factors go into the ultimate decisions").

[98] See Price, supra note 93, at 296 (observing that ML algorithms "can find that sort of complex underlying pattern in the data" but that the algorithms "cannot explain or even state what those patterns are").

[99] See Boris Babic, Sara Gerke, Theodoros Evgeniou & I. Glenn Cohen, Beware Explanations from AI in Health Care: The Benefits of Explainable Artificial Intelligence Are Not What They Appear, 373 Science 284, 285 (2021) ("Explainable AI / ML . . . offers post hoc algorithmically generated rationales of black-box predictions, which are not necessarily the actual reasons behind those predictions or related causally to them. Accordingly, the apparent advantage of explainability is a 'fool's gold' . . . . [W]e are likely left with the false impression that we understand [the AI] better.").

In the face of being unable to explain an AI's prediction, some scholars have proposed that we should simply offer users counterfactual explanations, allowing them to see how the output would change upon a different input variable. See Sandra Wachter, Brent Mittelstadt & Chris Russell, Counterfactual Explanations Without Opening the Black Box: Automated Decisions

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Some have found ML's opacity disconcerting.[100] If we do not know how or why a program has arrived at its result, how will we know that the result is right? In response, many commentators seek to assuage these fears by reminding us that ML's opacity is not unique to it. We oftentimes do not know how or why *humans* reach the decisions that they make.[101] Nevertheless, that has not prevented us from relying on human decision-making since the dawn of time. Essentially, these observers insist that the brave new world that ML appears to be ushering in is not all that new; moreover, we do not have to be particularly brave to embrace it. Professor Nicolas Terry summarizes the choice that we face succinctly: "[A]t some stage we may have to simply trust AI's judgment or do without using it."[102] Many observers insist that we will lose so much if we decide against using AI; the gains, they say, are tremendous. Thus, they implore, the best way forward is to trust AI's judgment. However, in a country as stratified along racial lines as the United States, people's willingness to trust AI likely differs. Section IV.B explores this argument further.

and the GDPR, 31 Harv. J.L. & Tech. 841, 844 (2018) (making this argument and giving the example of the following counterfactual explanation: "You were denied a loan because your annual income was £30,000. If your income had been £45,000, you would have been offered a loan.").

[100] In fact, AI's black-box nature may put it in tension with the European Union's General Data Protection Regulation ("GDPR"), which provides for a "right to explanation." See Tae Wan Kim & Bryan R. Routledge, Why a Right to an Explanation of Algorithmic Decision-Making Should Exist: A Trust-Based Approach, 32 Bus. Ethics Q. 75, 76 (2022) (describing the debate over whether the GDPR grants a right to explanation, which would impose a legal duty on companies that use data generated from EU residents to provide "meaningful explanations about how their automated algorithmic decision-making and/or profiling systems reach final decisions"). But see Wachter et al., supra note 99, at 842 (arguing that the GDPR does not provide for a "legally binding right to explanation").

[101] Scott J. Schweikart, Who Will Be Liable for Medical Malpractice in the Future? How the Use of Artificial Intelligence in Medicine Will Shape Medical Tort Law, 22 Minn. J.L. Sci. & Tech. 1, 7 (2021) (noting the argument that "the 'black-box' is also endemic of human intelligence, in that '[h]uman intelligence can reason and make arguments for a given conclusion, but can't explain the complex, underlying basis for how we arrived at a particular conclusion'" (quoting Vijay Pande, Artificial Intelligence's 'Black Box' Is Nothing to Fear, N.Y. Times (Jan. 25, 2018), https://www.nytimes.com/2018/01/25/opinion/artificial-intelligence-black-box.html [https://perma.cc/3WNX-GVXG])).

[102] Nicolas P. Terry, Appification, AI, and Healthcare's New Iron Triangle, 20 J. Health Care L. & Pol'y 117, 160 (2018) [hereinafter Terry, Appification] (internal quotation marks omitted) (quoting Will Knight, The Dark Secret at the Heart of AI, MIT Tech. Rev. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ [https://perma.cc/8648-GEPB]).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

B. Medical AI

AI technologies have been used in a wide variety of contexts, including financing, policing, and hiring.[103] Importantly, all signs indicate that AI will have an enormous impact in healthcare as well. Indeed, the FDA expects that medical AI devices will play an "outsized role" in healthcare.[104] Some experts predict that expenditures on AI in healthcare systems will reach $36.1 billion by 2025.[105] As Professor Cohen notes, "in the real world many physicians are racing toward integrating AI/ML into diagnostics, prognostics, allocation of resources, and treatment itself," and "[m]any companies and healthcare providers are currently investing heavily in developing medical AI/ML systems."[106]

The uses of AI in healthcare can be schematized into the categories of "[p]rediction" (i.e., assessing an individual's risk of developing a condition based on past data), "[c]lassification" (i.e., diagnosis), "[a]ssociation" (i.e., research), and "[o]ptimisation" (i.e., administrative tasks involving scheduling and the like).[107] This Article primarily focuses on medical AI as it relates to prediction and classification.

The days of medical AI are not in some imagined future—a generation or two away. They are here now. The FDA currently has approved hundreds of medical AI devices,[108] and it expects the number of

[103] Tom C.W. Lin, Artificial Intelligence, Finance, and the Law, 88 Fordham L. Rev. 531, 551 (2019); see Stephanie Bornstein, Antidiscriminatory Algorithms, 70 Ala. L. Rev. 519, 521, 568 (2018).

[104] Ctr. for Devices & Radiological Health, U.S. Food & Drug Admin., Artificial Intelligence / Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan 1, 5 (2021).

[105] See Pulwasha Iftikhar, Marcela V. Kuijpers, Azadeh Khayyat, Aqsa Iftikhar & Maribel DeGouvia De Sa, Artificial Intelligence: A New Paradigm in Obstetrics and Gynecology Research and Clinical Practice, 12 Cureus 1, 2 (2020); see also Chiara Longoni, Andrea Bonezzi & Carey K. Morewedge, Resistance to Medical Artificial Intelligence, 46 J. Consumer Rsch. 629, 629 (2019) ("Medical AI is forecasted to become a $10 billion market in the United States by 2025, pervade 90% of hospitals, and replace as much as 80% of what doctors currently do." (internal citations omitted)).

[106] Cohen, supra note 11, at 1426–27.

[107] See Jessica Morley, Thinking Critically about AI in Healthcare 2 (May 2023), https://rogerswannell.com/wp-content/uploads/2023/12/wp-1703519698138.pdf [https://perma.cc/MFJ7-9MAW] (last visited Feb. 6, 2024).

[108] See U.S. Food & Drug Admin., Artificial Intelligence and Machine Learning (AI/ML)-Enabled Medical Devices, https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices [https://perma.cc/JFD6-TNWC] (last visited Jan. 5, 2024).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

applications to continue to increase.[109] The uses of the devices that already have been approved are sundry. For example, the FDA has approved an AI-based device that aids in the diagnosis of diabetic retinopathy (an eye disease that is a complication of diabetes)[110] and another that diagnoses skin cancer.[111] Even the cynic would find inspiring the possible uses of AI in medicine:[112]

> A doctor might feed the genetic sequence of a patient's tumor into a black-box algorithm, for instance, and receive a recommendation as to what drug is most likely to treat the tumor effectively. Alternatively, an opaque algorithm could continuously evaluate a trauma patient's electronic vital signs and sound an alarm at the earliest sign of trouble, perhaps even before trained providers could observe the need. Black-box medicine can also be used to allocate scarce healthcare resources by suggesting which patient might benefit most from an organ transplant, a hospital bed, or the attention of the first available healthcare provider.[113]

The use of technology in healthcare is nothing new, of course. Since the 1970s, physicians have employed CDSS to guide them in the

[109] See Ctr. for Devices & Radiological Health, U.S. Food & Drug Admin., supra note 104, at 2 ("FDA continues to receive a high volume of marketing submissions and pre-submissions for products leveraging artificial intelligence/machine learning technologies, and we expect this to increase over time.").

[110] See W. Nicholson Price II, Medical AI and Contextual Bias, 33 Harv. J.L. & Tech. 65, 66 (2019) [hereinafter Price, Contextual Bias]. Notably, the device does not require that an expert confirm its results. As such, the device is "usable by health care providers who may not normally be involved in eye care," like a primary care physician. Cohen, supra note 11, at 1428 (quoting FDA Permits Marketing of Artificial Intelligence-Based Device to Detect Certain Diabetes-Related Eye Problems, U.S. Food & Drug Admin. (Apr. 11, 2018), https://www.fda.gov/news-events/press-announcements/fda-permits-marketing-artificial-intelligence-based-device-detect-certain-diabetes-related-eye [https://perma.cc/P68T-35WB]).

[111] Andre Esteva et al., Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks, 542 Nature 115, 118 (2017); see U.S. Food & Drug Admin., FDA Executive Summary Reclassification Panel Meeting on Skin Lesion Analyzers 2, 4 (2022), https://www.fda.gov/media/160252/download [https://perma.cc/8GC4-P9MT].

[112] Of course, some of the uses of AI in medicine will be far from extraordinary and far more prosaic. As Professor Terry describes it, at present, "healthcare stakeholders are more likely to use these tools for more pedestrian, revenue-generating purposes such as reducing readmissions that otherwise would lead to Medicare readmission penalties . . . ." Nicolas Terry, Of Regulating Healthcare AI and Robots, 18 Yale J. Health Pol'y, L. & Ethics 133, 171 (2019) [hereinafter Terry, Regulating Healthcare AI]; see also Iftikhar et al., supra note 105, at 1 (stating that AI has the potential to reduce clinicians' workload and decrease healthcare costs).

[113] Price, supra note 93, at 297.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

construction of treatment plans for their patients.[114] However, existing CDSS are charmingly simple when compared to the AI tools that currently are under development and will be used in the future. Historically, CDSS have been rule-based, leading providers to decisions through a series of if-then rules.[115] (For example: If patient had a C-section with a previous pregnancy, then a vaginal birth is contraindicated and the patient should have a C-section with the present pregnancy.) Meanwhile, AI's capacities allow it to do much more than guide decision-making, with uses including diagnosing pathology, identifying effective treatments based on the patient's unique biology or genetics, and predicting health outcomes.[116]

It appears safe to forecast that AI tools will outperform human physicians.[117] Indeed, when it comes to diagnosis, there are already several examples of the machines besting their human counterparts:

> Google's neural net diagnoses skin cancer as effectively as do experienced dermatologists. Google has tested an AI-based system that successfully identified eye diseases in retinal fundus photographs and one that reaches or exceeds that of experts on a variety of sight-threatening retinal diseases. Other programs already beat humans: an AI beat humans at predicting heart attacks—without even considering the effects of diabetes or lifestyle. A different AI beat humans at diagnosing brain tumors and predicting hematoma expansion.[118]

[114] See Cohen, supra note 11, at 1430.

[115] See Davidson & Boland, supra note 74, at 306 (describing one of the first clinical decision support systems, MYCIN, which used a "series of if-then statements" to "guide clinicians to appropriate decision making").

[116] See Price, supra note 93, at 297.

[117] See Cohen, supra note 11, at 1430 (stating that "AI systems have achieved 'specialist-level performance in many diagnostic tasks [and] can better predict patient prognosis than clinicians' in many areas" (quoting Kun-Hsing Yu, Andrew L. Beam & Isaac S. Kohane, Artificial Intelligence in Healthcare, 2 Nature Biomedical Eng'g 719, 722 (2018))); Jane R. Bambauer, Dr. Robot, 51 U.C. Davis L. Rev. 383, 391–92 (2017) ("There is little reason to doubt that AI will eventually flourish and outperform physicians in many aspects of medical care. In fact, much of what the training and licensing process does for doctors is train them to act more like algorithms.").

[118] Froomkin et al., supra note 89, at 39–40. Cohen describes "image-based diagnosis . . . in areas such as radiology, cardiology, dermatology, ophthalmology, and pathology" as "low-hanging fruit." Cohen, supra note 11, at 1430. Nevertheless, he notes that AI has been impressive with respect to fruits that are harder to reach: "Data from health insurance claims can be used to predict mortality in elderly patients, patient attributes in the medical notes can be employed to classify cancer patients with different responses to chemotherapy, and clinical

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

We might have expected that medical AI would outperform humans, though.[119] This is due to the reality that human biology—the object of medicine—is a wondrously complex phenomenon.[120] AI is simply superior to its human counterparts when it comes to identifying the patterns within that complexity.[121] This may enable AI to produce better outcomes when it comes to diagnosis, prognosis, and treatment decisions.[122]

If one has read too much science fiction, one might believe that society is careening toward a future within which the machines have completely replaced the humans—one in which the only living, breathing people found in hospitals will be the patients being cared for by an army of insentient (albeit remarkably capable) robots. However, most commentators have concluded that the machines likely will do no more than *augment* humans.[123] That is, humans will be involved in the provision of healthcare for the foreseeable future.

C. AI in Prenatal Care

Healthcare providers have long made use of automated tools in the provision of prenatal care. As early as 1994, physicians were using a rule-based CDSS to help identify patients who were likely to experience

predictors for the prognosis of patients receiving thoracic organ transplantation can be identified." Id. at 1430–31 (quoting Yu et al., supra note 117, at 726).

[119] See Terry, supra note 102, at 174 ("Much of what physicians do (checkups, testing, diagnosis, prescription, behavior modification, etc.) can be done better by sensors, passive and active data collection, and analytics" (quoting Vinod Khosla, Technology Will Replace 80% of What Doctors Do, Fortune (Dec. 4, 2012, 9:26 AM), https://fortune.com/2012/12/04/technology-will-replace-80-of-what-doctors-do/ [https://perma.cc/KS3T-3Y7X])).

[120] See Price & Rai, supra note 80, at 778 (observing that "[b]iomedicine is an example where human experts often don't understand what is going on due to systemic complexity").

[121] See Claudia E. Haupt, Artificial Professional Advice, 21 Yale J.L. & Tech. (Special Issue) 55, 71 (2019) ("Unlike deep learning, expert human interpretation fails to capitalize on all the patterns, or 'regularities,' that can be extracted from very large data sets and used for interpretation of still and moving images." (quoting C. David Naylor, On the Prospects for a (Deep) Learning Health Care System, 320 JAMA 1099, 1099 (2018))).

[122] See id.

[123] See Terry, supra note 102, at 140 (noting that when AI is used in healthcare, "it will take a human 'to work with patients to understand and translate patients' symptoms, inform patients of treatment options, and guide patients through treatment plans'" (quoting Exec. Off. of the President, Artificial Intelligence, Automation, and the Economy 18 (2016), https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.PDF [https://perma.cc/B7Y6-Q8HV])).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

preterm births.[124] However, the advancement of technology has led to the proliferation of increasingly sophisticated tools to help manage pregnancy, childbirth, and the postpartum period.[125] As one might expect, the more cutting-edge tools on the horizon incorporate AI.

There are AI instruments under development that will identify perinatal asphyxia—a condition that is caused by a lack of blood flow to or from the fetus during childbirth and carries the risk of severe neurological damage to the newborn.[126] While these instruments would identify fetuses in distress, thus allowing providers to take necessary steps to prevent harm, they would also identify fetuses that are *not* in distress, thus avoiding unwarranted medical interventions on pregnant patients and their babies.[127] Observers also expect that AI will be incorporated into existing systems that monitor the fetal heart rate for signs of low blood oxygen levels.[128] A decrease in variation in the fetal heart rate is the surest sign that the health of the fetus is declining. However, "there is significant observer variation in interpreting this data."[129] Enhancing these systems with AI will regularize the interpretation of the data, possibly leading to better outcomes for both pregnant people and their babies.[130]

AI is also expected to impact one of the rites of passage of pregnancy—the glucose challenge test. This test, which is administered during the twenty-fourth week of pregnancy and is a screen for gestational diabetes, requires the pregnant person to drink fifty grams of glucose so that

[124] See Davidson & Boland, supra note 74, at 306.

[125] Interestingly, a lot of the pregnancy-related technology that observers expect will be introduced in the future does not constitute AI, but rather simple telecommunications tools. See Oliver Kim & Tamara Kramer, The Girl with the Cyber Tattoo: Applying a Gender Equity Lens to Emerging Health Technology, 12 Ne. U. L. Rev. 327, 353 (2020) ("[E]xperts believe there is an untapped potential for telemedicine to address maternal care, including prenatal and postnatal care. Virtual live visits can be used to replace . . . the common schedule of routine pregnancy care.").

[126] Brian S. Carter, Albert D. Haverkamp & Gerald B. Merenstein, The Definition of Acute Perinatal Asphyxia, 20 Clinics Perinatology 287, 287 (1993); see, e.g., Fatemeh Darsareh et al., Application of Machine Learning to Identify Risk Factors of Birth Asphyxia, 23 BMC Pregnancy & Childbirth 1, 2 (2023); Maria Ribeiro, Inês Nunes, Luísa Castro, Cristina Costa-Santos & Teresa S. Henriques, Machine Learning Models Based on Clinical Indices and Cardiotocographic Features for Discriminating Asphyxia Fetuses–Porto Retrospective Intrapartum Study, 11 Frontiers Pub. Health 1, 2, 5 (2023).

[127] See Iftikhar et al., supra note 105, at 2.

[128] See id.

[129] Id.

[130] See id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

providers can examine their ability to metabolize sugar.[131] The test is expensive and, because it is time-consuming, onerous for many pregnant people.[132] Researchers are developing an AI-powered, online calculator that can reduce some of the costs and burdens of the existing test.[133] While the calculator currently underperforms relative to the existing glucose challenge test, its developers expect that as it "continues to be exposed to more cases," it eventually will outdo the current approach to identifying patients with gestational diabetes.[134]

Other AI devices will address preterm birth, which affects approximately ten percent of all births[135] and can cause severe injuries in the child, ranging from dental problems to cerebral palsy and death.[136] Researchers are developing a tool that uses AI algorithms to analyze the electrical signals that uterine activity produces to more accurately diagnose preterm labor.[137] Other AI devices under development will evaluate a pregnant person's amniotic fluid for biomarkers that may indicate a high risk of preterm birth. "[P]hysicians can use this tool to guide their management, such as observation alone, or suggesting cervical cerclage and[/]or antenatal steroids if deemed necessary."[138]

Thus, we stand on the cusp of a future that witnesses the incorporation of AI into prenatal care. The hope, and expectation, is that these new technologies will not only conserve healthcare costs but will also better the experience of pregnancy and improve pregnancy outcomes. Yet, there is reason to be wary about the future that AI in prenatal care heralds. This is due to algorithmic bias—an undoubtedly real phenomenon that threatens to cheapen AI by making it into nothing more than a fancy, digital technique for reproducing the racially stratified present. The next Part defines algorithmic bias, explains its causes, and provides examples of it in medicine.

---

[131] Catherine A. Carr, Evidence-Based Diabetes Screening During Pregnancy, 46 J. Midwifery & Women's Health 152, 155 (2001).

[132] Id.

[133] See Iftikhar et al., supra note 105, at 3 (describing the tool).

[134] Id.

[135] Preterm Birth, Ctrs. for Disease Control & Prevention (Oct. 24, 2023), https://www.cdc.gov/reproductivehealth/maternalinfanthealth/pretermbirth.htm [https://perma.cc/5BFY-MVZ6].

[136] Premature Birth, Mayo Clinic (Feb. 25, 2023), https://www.mayoclinic.org/diseases-conditions/premature-birth/symptoms-causes/syc-20376730 [https://perma.cc/PCX7-GRGG].

[137] See Iftikhar et al., supra note 105, at 3–4.

[138] Id. at 3.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

III. ALGORITHMIC BIAS

Programmers are developing AI tools with an optimism that they will improve society. Indeed, hopes of increased ease and diminished discomfort might always accompany advances in technology. However, some observers of AI have claimed that AI's promises go beyond merely cutting costs and making tasks easier to do. Some have claimed that AI can address profound injustices in our society.[139] The argument is that AI might be an effective weapon against longstanding and deeply ingrained social wrongs—one of the most resilient of which is racial inequality.[140] Considering this possibility in the context of racial disparities in pregnancy outcomes is illuminating.

There are many factors that contribute to racial disparities in maternal and infant mortality and morbidity.[141] However, at least some portion of these disparities can be explained in terms of the inferior care that providers give to their black patients.[142] Most have proposed that if providers are doling out different, second-rate care to their black patients, then it is likely true that providers' implicit biases are to blame.[143] AI, then, is a mechanism that can neutralize these implicit biases. Providers will no longer need to rely on their fallible judgments and the unconscious racial associations and aversions on which that judgment so often is based. Instead, they will only need to rely on the AI tool.[144]

[139] See Irene Y. Chen, Shalmali Joshi & Marzyeh Ghassemi, Treating Health Disparities with Artificial Intelligence, 26 Nature Med. 16, 16 (2020) ("Used purposefully, AI can be leveraged as a tool to level the playing field, to help highlight and erase the numerous and well-documented inequities in health."). An organization using AI to right a social wrong is the Geena Davis Institute, which used a machine learning tool to document just how little women are seen and heard on film. See Schönberger, supra note 90, at 187.

[140] Chen et al., supra note 139, at 16.

[141] See Khiara M. Bridges, Racial Disparities in Maternal Mortality, 95 N.Y.U. L. Rev. 1229, 1257–62 (2020) (discussing the multiple factors that contribute to racial disparities in maternal mortality, including poverty and access as well as stress and weathering).

[142] See supra notes 35–37 and accompanying text.

[143] Bridges, supra note 141, at 1263.

[144] The issue of physicians' reliance on AI in order to circumvent reliance on their implicit biases raises interesting questions about the changes that medical AI may bring to the standard of care in a practice area. As a bit of background: a patient may bring a tort claim against her provider for medical malpractice if she believes that she was given inferior healthcare. B. Sonny Bal, An Introduction to Medical Malpractice in the United States, 467 Clinical Orthopaedics & Related Rsch. 339, 340 (2009). A provider will only be liable for malpractice if the care that she provided represents a departure from the standard of care that has been established in that practice area. See Schweikart, supra note 101, at 13 ("A physician's deviation from the standard of care will amount to a breach in their duty to the patient.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

The problem is that there is an abundance of evidence that AI tools and systems can be just as biased as the humans that the technology is intended to augment or supersede. "[C]redit-scoring algorithms predict outcomes related to income, thus incorporating disparities in employment and salary. Policing algorithms predict measured crime, which also reflects increased scrutiny of some groups. Hiring algorithms predict

Typically, a court evaluates this standard 'by reference to what a reasonable *physician* would have done—i.e., a person with the same kind of technical background, training, and expertise as the defendant.' " (quoting Michael D. Greenberg, Medical Malpractice and New Devices: Defining an Elusive Standard of Care, 19 Health Matrix 423, 428 (2009))). Medical AI may or may not shift that standard of care.

For example, imagine that the medical AI device that the provider utilizes counsels that a pregnant patient ought to be prescribed benzodiazepines to treat her seizures, although most providers would avoid prescribing the drug to a pregnant person. Mohammad Masud Iqbal, Tanveer Sobhan & Thad Ryals, Effects of Commonly Used Benzodiazepines on the Fetus, the Neonate, and the Nursing Infant, 53 Psychiatric Servs. 39, 47 (2002). If the provider follows the tool's advice and prescribes the medication, and the drug adversely affects the pregnancy, should the physician be liable for malpractice because the care that she provided departed from established standards? Suppose that, despite the tool's recommendation, the provider chooses not to provide the medication; the patient subsequently suffers a severe seizure, harming herself and the fetus that she is gestating. Should the physician be liable for malpractice because she chose *not* to follow the tool's recommendation and *not* to depart from established standards? Relatedly, will there arrive a day wherein using medical AI is the standard of care and a failure to use AI will constitute a departure from the standard, opening up a provider to liability? See Iria Giuffrida & Taylor Treece, Keeping AI Under Observation: Anticipated Impacts on Physicians' Standard of Care, 22 Tul. J. Tech. & Intell. Prop. 111, 113 (2020) ("Over time, a physician's standard of care may be heightened and may even come to require the use of AI."); Kerr & Gruben, supra note 94, at 88 (noting the argument that "existing medical malpractice law will come to require superior ML-generated medical diagnostics—rather than doctors—as the standard of care in clinical settings").

Essentially, these are all questions about whether it is proper to rely on medical AI—an issue that other scholars capably have addressed. See, e.g., Schweikart, supra note 101, at 21 ("[T]he most natural solution is to have common-law medical malpractice modify its standard of care over time for physicians to reflect the newfound reality of medical AI in clinical practice."); Price, supra note 93, at 305 (concluding that providers should meet the standard of care if they ensure that an AI tool that recommends unconventional treatment was properly validated and was created by a reputable developer). Future scholarship might put these questions in conversation with racial disparities in health: that is, if there is some hope that medical AI will improve outcomes for people of color because it will eliminate the ability of providers' explicit and implicit biases to affect the care that they give, will it ever be appropriate for a provider to disregard the recommendation that a medical AI device provides? See, e.g., David Leslie, Anjali Mazumder, Aidan Peppin, Maria K. Wolters & Alexa Hagerty, Does 'AI' Stand for Augmenting Inequality in the Era of COVID-19 Healthcare?, 372 Brit. Med. J. 1, 3 (2021), https://www.bmj.com/content/bmj/372/bmj.n304.full.pdf [https://perma.cc/86CU-4XP9] ("[C]linicians who over-rely on AI decision support systems might take their recommendations at face value, even when these models might be faulty. On the other hand, *clinicians who distrust AI decision support systems might discount their recommendations, even if they offer corrections to discrimination*." (emphasis added)).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

employment decisions or supervisory ratings, which are affected by race and gender biases."[145] One of the most well-known cases of algorithmic bias involved an algorithm that was designed to guide judges' determinations of bail amounts by identifying which defendants facing criminal charges were most likely to re-offend.[146] Quite famously, an investigation revealed that the algorithm was twice as likely to *erroneously* identify black defendants as posing a high risk of re-offending when compared to their white counterparts.[147] This is to say that the algorithm identified black defendants as likely to re-offend when, in fact, they ultimately did not re-offend; the algorithm did not tend to make this error when it came to white defendants. Additionally, the algorithm was twice as likely to *erroneously* identify white defendants as posing a low risk of re-offending when compared to their black counterparts.[148] This is to say that the algorithm identified white defendants as unlikely to re-offend, when, in fact, they ultimately did re-offend; the algorithm did not tend to make this error when it came to black defendants.

Algorithmic bias is a consequence of a couple of different factors, including the design of the algorithmic model, the underrepresentation of populations in the training data, and the training data as the sedimentation of past and present racially inequitable practices. Further, in the context of medical AI, the myth of biological race leads to algorithmic bias. Each of these factors is discussed in turn.

A. Choices in Model Design as a Source of Bias

Algorithmic bias can be a product of the choices that programmers make when designing the model.[149] Indeed, an unfortunate choice in model design resulted in a widely noted example of algorithmic bias in healthcare. Physician and researcher Ziad Obermeyer and coauthors investigated an algorithmic model that health systems use to identify

[145] Obermeyer et. al, supra note 96, at 453.

[146] Julia Angwin, Jeff Larson, Surya Mattu & Lauren Kirchner, Machine Bias, ProPublica (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/4BS2-32SA].

[147] Id.

[148] Id.

[149] See Benjamin, Race After Technology, supra note 60, at 11–12 ("[T]ech designers encode judgments into technical systems but claim that the racist results are entirely exterior to the encoding process.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

patients with complex health needs.[150] Identified patients were enrolled in "high-risk care management" programs.[151] These programs help coordinate care and ensure that patients receive the more thorough medical attention that they need.[152] Obermeyer and his coauthors discovered that, although the model specifically excluded race as a variable, black patients were much less likely to be identified as appropriate candidates for these programs.[153] That is, black patients had to be much sicker than their white counterparts before the model classified them as sick enough for inclusion in the programs.[154] Why was this? Well, the programmers had assumed that healthcare costs indicated healthcare needs—i.e., the more money that an individual spent on healthcare, the sicker she was.[155] Accordingly, they built the model to identify patients whom it predicted would have higher healthcare costs.[156] However, for a variety of reasons that have nothing to do with medical need and everything to do with structural racism and its sequelae, white people are likely to have higher healthcare costs than black people.[157] Accordingly, a black person has to be much sicker than her white counterpart to accrue the same healthcare costs as the latter. Thus, because of a choice in model design—the programmers' decision to use the healthcare cost label as a

[150] Obermeyer et al., supra note 96, at 447.

[151] Id.

[152] Id.

[153] Id. at 449.

[154] Id. at 448, 449 ("[A]t the same level of algorithm-predicted risk, Blacks have significantly more illness burden than Whites. . . . [L]ess-healthy Blacks scored at similar risk scores to more-healthy Whites . . . . Across all of these important markers of health needs—severity of diabetes, high blood pressure, renal failure, cholesterol, and anemia—we find that Blacks . . . are substantially less healthy than Whites at any level of algorithm predictions . . . .").

[155] Id. at 450–51.

[156] Id.

[157] Id. at 450. As the authors explain, people of color are more likely to be poor than white people, and "poor patients face substantial barriers to accessing health care, even when enrolled in insurance plans." Id. These barriers include "geography and differential access to transportation, competing demands from jobs or child care, or knowledge of reasons to seek care." Id. The authors go on to explain that research indicates that black people trust healthcare systems less than their white counterparts, and this reduced trust makes them less likely to use healthcare systems. See id. ("Thus, whether it is communication, trust, or bias, something about the interactions of Black patients with the health care system itself leads to reduced use of health care. The collective effect of these many channels is to lower health spending substantially for Black patients . . . ."). Part IV of this Article explores the issue of race and trust in greater depth.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

predictor of healthcare needs—the algorithm discriminated against black people.[158]

B. Unrepresentative Training Data as a Source of Bias

Algorithmic bias may be a function of the data on which the model was trained. Indeed, the prediction that a model makes is only as good as its training data. If people of color are not adequately represented in the population from which the training data is gathered,[159] or if the information that has been gathered about people of color is laden with errors and gaps,[160] then the AI tool will not perform as well vis-à-vis people of color.

When it comes to medical AI, we have every reason to expect that the structural impediments to accessing healthcare that people of color face will result in their being underrepresented in training data.[161] As Professor Price has explained, medical AI usually is trained in "high-resource settings: academic medical centers or state-of-the-art hospitals or hospital systems."[162] Vulnerable people frequently find these high-resource settings inaccessible—because these hospital systems do not accept Medicaid, or because they are located far from the communities that vulnerable folks call home.[163] Further, vulnerable people have a hard time

[158] Notably, the cost label is used in many algorithms in healthcare. Id. at 451 ("[T]he cost label reflects the industry-wide approach. For example, the Society of Actuaries's comprehensive evaluation of the 10 most widely used algorithms, including the particular algorithm we study, used cost prediction as its accuracy metric."). Professor Obermeyer and his coauthors note that "[s]imilar algorithms are developed and used by non-profit hospitals, academic groups, and governmental agencies, and are often described in academic literature on targeting population health interventions." Id.

[159] See Price, Contextual Bias, supra note 110, at 66–67, 92–93.

[160] See Hoffman & Podgurski, supra note 60, at 6 ("Data sources such as electronic health records (EHR) or insurance claims can be rife with errors, systemic biases, and data gaps that might be particularly pronounced for minorities who do not receive optimal care.").

[161] See Schönberger, supra note 90, at 181 ("[W]here minorities and even whole populations are excluded from health services, no health records of them exist."); Cohen & Graver, supra note 79, at 448 (identifying two factors that are responsible for vulnerable populations' exclusion from medical AI training data as "'lack of insurance and the inability to access healthcare'" (quoting Sarah E. Malanga, Jonathan D. Loe, Christopher T. Robertson & Kenneth S. Ramos, Who's Left Out of Big Data? How Big Data Collection, Analysis, and Use Neglect Populations Most in Need of Medical and Public Health Research and Interventions, *in* Big Data, Health Law & Bioethics 98, 99 (I. Glenn Cohen, Holly Fernandez Lynch, Effy Vayena & Urs Gasser eds., 2018))).

[162] Price, Contextual Bias, supra note 110, at 66–67.

[163] See Michael R. Daly & Jennifer M. Mellor, Racial and Ethnic Differences in Medicaid, 77 Med. Care Rsch. Rev. 85, 85–86 (2018).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

accessing *any* healthcare—not just that dispensed in high-resource settings. This is true for a panoply of reasons, including the inability to secure health insurance, transportation, or childcare.[164] Marginalized people "also often lack a primary care physician and visit multiple facilities when they do seek medical attention," resulting in medical records that "are fragmented and do not contain comprehensive information."[165] Even further, due to the budget-strained, personnel-taxed healthcare institutions on which people of color are forced to rely, the data that these institutions generate about the patients under their care more frequently will be inferior—inaccurate, incomplete—when compared to the data generated about patients with privileges of class and race.[166] Additionally, fitness wearables have been a significant source of data for medical AI.[167] It may be stating the obvious to observe that these expensive technologies are not the province of the class-unprivileged (and, therefore, the disproportionately race-unprivileged).[168]

These factors allow us to predict, with some degree of confidence, the likelihood that people of color's health data will not be adequately included in the training data on which medical AI tools rely. Thus, these tools will underperform with respect to people of color and other marginalized populations—a result that is doubly devastating inasmuch as these may be "the very people most in need of increased health research, intervention, and care."[169]

Disturbingly, a developer's attestations that an algorithmic tool is perfectly, or sufficiently, accurate is no guarantee that the tool will be accurate vis-à-vis the population that was underrepresented in the training

[164] See Hoffman & Podgurski, supra note 60, at 13.

[165] Id.

[166] See Leslie et al., supra note 144, at 2–3 ("[R]esources needed to ensure satisfactory dataset quality and integrity might be limited to digitally mature hospitals that disproportionately serve a privileged segment of a population to the exclusion of others."); Hoffman & Podgurski, supra note 60, at 13 ("[L]ow-income individuals may seek care at teaching clinics where practitioners are less meticulous about recordkeeping. Data gathered from these facilities may have more errors than data from facilities frequented by higher-income patients.").

[167] Jessilyn Dunn et al., Wearable Sensors Enable Personalized Predictions of Clinical Laboratory Measurements, 27 Nature Med. 1105, 1105 (2021).

[168] See Terry, Appification, supra note 102, at 166 ("[S]martphones and high-end wearables (and their data plans) or health-conscious connected automobiles are not the healthcare tools of the poor or the unhealthy."); Kim & Kramer, supra note 125, at 354 ("There are significant racial disparities related to who can access and engage with technology to track their health . . . .").

[169] Cohen & Graver, supra note 79, at 449 (quoting Malanga et al., supra note 161, at 99).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

data. This is best illustrated by an IBM-developed facial recognition system that had a 34.7% error rate when asked to identify the gender of women with darker skin tones.[170] However, the system's overall error rate was much lower—12.1%.[171] Indeed, with respect to men with light skin, the system approximated perfection, with an error rate of 0.3%.[172] The problem was that the data that was used to evaluate the algorithm's performance was desperately underinclusive of women with dark skin.[173] Thus, it was true both that the algorithm was almost 88% accurate as a general matter and 65% accurate when it came to women with dark skin. This might be terrifying in the context of medical AI.[174] We need only imagine the harm, and possible death, produced by an AI system that is only 65% accurate in predicting the likelihood that a black patient is suffering from preeclampsia, a pulmonary embolism, or an obstetric hemorrhage.

C. Racially Inequitable Healthcare as a Source of Bias

Algorithmic bias may simply be a function of the fact that inequity is built into the very fabric of the training data. This is different from the above point, which is about the underinclusiveness of a dataset. In order to address the problem of underinclusion, people of color simply need to be adequately represented in the dataset. In contrast, if inequity is built into the very fabric of the training data, then including data from more people of color solves nothing. This is the simple point that training data is nothing more and nothing less than a "record of human history"[175]; it is an assemblage of our past and present practices.[176] If people of color are

[170] Joy Buolamwini & Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, 81 Proc. Mach. Learning Rsch. 1, 9 (2018).

[171] Id.

[172] Id.

[173] Id.

[174] Tyler Dueno, Note, Racist Robots and the Lack of Legal Remedies in the Use of Artificial Intelligence in Healthcare, 27 Conn. Ins. L.J. 337, 344 (2020). ("This same problem of defective statistical analysis can exist within clinical AI. The clinical software could be guessing at diagnosis but still claim a high success rate. If the underlying data is underinclusive for subpopulations, then AI can produce skewed results.").

[175] Jonathan H. Chen & Abraham Verghese, Planning for the Known Unknown: Machine Learning for Human Healthcare Systems, 20 Am. J. Bioethics, Nov. 2020, at 1–2.

[176] See Leslie et al., supra note 144, at 2 ("The datasets which are the basis of data driven AI and machine learning models thus reflect complex and historically situated practices, norms, and attitudes. . . . [They] might incorporate the biases of previous inequitable practices . . . .").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

treated inequitably, then the training data simply will be an accumulation of those inequities.[177]

Consider the example of a vaginal birth after a C-section ("VBAC"). If providers, for no clinically relevant reason, disallow their black patients from attempting a VBAC more frequently than they disallow their white counterparts from making this attempt,[178] then black people will have lower rates of successful VBACs. The training data will reflect the lower rates of successful VBAC among black people, and this fact will be built into the model. The result is that the algorithmic tool upon which a provider relies will recommend against allowing their black patients to attempt a VBAC. The past produces the future. Inequity is reproduced.[179]

Consider also that our society is one in which racial segregation exists in innumerable aspects of life—including in hospitals.[180] Dr. Elizabeth Howell and her coauthors conducted a study that showed that there are hospitals in the United States that see disproportionate numbers of black

[177] See Schönberger, supra note 90, at 176 ("[I]f the underlying datasets reflect existing biases against minorities or other vulnerable groups prevalent in society, the algorithms will inadvertently adopt and reproduce them in their outputs." (internal quotation marks omitted) (citing Andrea Romei & Salvatore Ruggieri, A Multidisciplinary Survey on Discrimination Analysis, 29 Knowledge Eng'g Rev. 582, 582–638 (2014))); Benjamin, Race After Technology, supra note 60, at 59 ("To the extent that machine learning relies on large, 'naturally occurring' data sets that are rife with racial (and economic and gendered) biases, the raw data that robots are using to learn and make decisions about the world reflect deeply ingrained cultural prejudices and structural hierarchies.").

[178] See supra notes 46–52 and accompanying text.

[179] This realization of this hypothetical is not outside the realm of possibility. Hoffman and Podgurski describe an algorithm that tends to recommend less aggressive measures for women when it comes to treating hypertension and heart failure—even through women are more likely to suffer from those conditions and warrant additional interventions. See Hoffman & Podrugski, supra note 60, at 16. The reason the algorithm was biased against women in this way was that women have always been undertreated with respect to cardiac conditions. See id. This undertreatment simply was reflected in the algorithm. See id. Hoffman and Podgurski also describe the potential for an AI system to "recommend lower doses of pain drugs to African American patients regardless of their need for relief," reflecting the empirically documented fact that "African American patients receive, on average, less pain treatment than Caucasians." Id. at 18. They note the possibility that AI systems will recommend cardiac catheterization for black women presenting with chest pain less frequently than for their white male counterparts, a consequence of the fact that physicians currently are more likely to deny this intervention to black women. See id. at 18–19.

[180] See David A. Asch & Rachel M. Werner, Segregated Hospitals Are Killing Black People. Data from the Pandemic Proves It, Wash. Post (June 18, 2021, 11:15 AM), https://www.washingtonpost.com/opinions/2021/06/18/segregated-hospitals-are-killing-black-people-data-pandemic-proves-it/ [https://perma.cc/V4ES-AYA4] ("Decades of racial residential segregation have concentrated Black people in some areas and White people in others . . . . People tend to seek health care near home.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

patients.[181] These "high black-serving" and "medium black-serving" hospitals,[182] which only account for twenty-five percent of the nation's hospitals, are the sites for seventy-five percent of births by black people.[183] This statistic is worth repeating: "[O]ne-quarter of hospitals provided care for three-quarters of all black deliveries in the United States."[184] In essence, our hospitals are segregated. As anyone paying attention might have guessed, "high black-serving" and "medium black-serving" hospitals have higher rates of adverse pregnancy outcomes.[185] After parsing the data, Howell and coauthors conclude that the higher rates of poor pregnancy outcomes found at these hospitals are, in part, a function of the inferior care that these hospitals are providing their patients.[186] This is to say that the data being generated—the data on which medical AI systems will be trained—show that black birthing people hemorrhage, go into shock, need to be ventilated, require hysterectomies, experience kidney failure, suffer from sepsis, almost die, etc., more frequently than their white counterparts. This inequity will be encoded into medical AI systems.

D. The Myth of Biological Race as a Source of Bias

In addition to the choices and processes described above that might *unintentionally* result in algorithms discriminating against people of color and compounding the disadvantage that they experience, harmful evaluations of people of color are sometimes *intentionally* baked into the algorithms—with the technology, unsurprisingly, working to disadvantage the people of color that it encounters. That is, there are numerous examples of programmers knowingly and deliberately encoding into technology the idea that there is something simply *different* about black people, as a race, that warrants the provision of different

[181] See Elizabeth A. Howell, Natalia Egorova, Amy Balbierz, Jennifer Zeitlin & Paul L. Hebert, Black-White Differences in Severe Maternal Morbidity and Site of Care, 214 Am. J. Obstetrics & Gynecology 122.e1, 122.e2 (2016).

[182] As Dr. Howell and her coauthors explain, "Black-serving hospitals were more likely to be located in an urban area, to be located in the South, to be a teaching hospital, to have a higher delivery volume, to have larger bed size, and to have a higher proportion of Medicaid deliveries." Id. at 122.e3.

[183] See id. at 122.e2, 122.e5.

[184] Id. at 122.e5.

[185] Id.

[186] See id. ("Hospitals that disproportionately cared for black deliveries had higher severe maternal morbidity rates after adjustment for patient and hospital characteristics.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

healthcare. It is worthwhile to itemize these occurrences of deliberate algorithmic bias exhaustively[187]:

- An algorithm that the National Football League ("NFL") used to evaluate neurocognitive function in former players who suffered concussions over the course of their careers assumes that black people have a lower baseline function.[188] The result is that "Black players have to show steeper cognitive declines" in order to qualify for payouts from a settlement fund that the NFL established for players with brain injuries.[189] The NFL discontinued the use of the "race-norming" or "race-correcting" algorithm after public outcry.[190]
- An algorithm that is intended to evaluate how well a person's kidney is functioning estimates that black people have higher glomerular filtration rates ("eGFR"), an estimation that proposes that they have better kidney function than comparable nonblack people.[191] The algorithm developers defended the algorithm with the claim that black people are "more muscular" than their nonblack counterparts, a muscularity that would affect the serum creatinine levels on which the eGFR measurement is based.[192] Despite scores of researchers calling this claim into question, "the 'race-corrected' eGFR remains the standard."[193] If the idea of racial body type on which this algorithm is based is indeed a myth, it would mask kidney failure in black patients and, in so doing,

[187] Benjamin helpfully describes examples of deliberate algorithmic bias like these as "health interventions that embrace a racialized conception of 'problem people,' and thereby influence the direction of treatment by improper means." Benjamin, supra note 60, at 155.

[188] See Ken Belson, Black Former N.F.L. Players Say Racial Bias Skews Concussion Payouts, N.Y. Times (Oct. 20, 2021), https://www.nytimes.com/2020/08/25/sports/football/nfl-concussion-racial-bias.html [https://perma.cc/6YTJ-2VN6].

[189] Id.

[190] Lucia Trimbur & Lundy Braun, The NFL's Reversal on 'Race Norming' Reveals How Pervasive Medical Racism Remains, NBC News (June 8, 2021, 4:09 PM), https://www.nbcnews.com/think/opinion/nfl-s-reversal-race-norming-reveals-how-pervasive-medical-racism-ncna1269992 [https://perma.cc/B5B2-XMG6].

[191] See Darshali A. Vyas, Leo G. Eisenstein & David S. Jones, Hidden in Plain Sight—Reconsidering the Use of Race Correction in Clinical Algorithms, 383 New Eng. J. Med. 874, 875 (Debra Malina, ed.) (2020).

[192] See id.

[193] Id. at 875.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

delay their commencement of dialysis and placement on kidney transplant lists.[194]

- Algorithms that are designed to evaluate lung function assume that black patients have a lower lung volume.[195] Thus, lung volumes that are identified as "normal" among black patients are categorized as "abnormal" among white people.[196] The consequence is that black patients with lung diseases have a harder time obtaining a diagnosis and consequent treatment and disability support.[197]
- The American Heart Association's Heart Failure Risk Score predicts whether a patient who is admitted to the hospital is likely to die from heart failure.[198] The algorithm, which clinicians use to inform referrals to cardiologists and other decisions about care, assigns nonblack patients three additional points, suggesting that black patients are at a lower risk of death.[199] The American Heart Association "does not provide a rationale for this adjustment."[200]
- An algorithm that evaluates the risks posed to potential kidney donors assesses a higher risk for black patients, marking them as less suitable donors than comparable nonblack patients.[201]
- An algorithm that predicts whether a person who has had a C-section during a previous pregnancy will have a successful VBAC gives lower scores to black patients, increasing the likelihood that they will have surgical interventions during their childbirths.[202] "The study used to produce the algorithm found that other variables, such as marital status and insurance type, also

[194] See Jessica P. Cerdeña, Marie V. Plaisime & Jennifer Tsai, From Race-Based to Race-Conscious Medicine: How Anti-Racist Uprisings Call Us to Act, 396 Lancet 1125, 1125–26 (2020).

[195] See Adam W. Gaffney, Steffie Woolhandler & David U. Himmelstein, Are Lung Function Algorithms Perpetuating Health Disparities Experienced by Black People?, Stat News (Sept. 15, 2020), https://www.statnews.com/2020/09/15/lung-function-algorithms-health-disparities-black-people/ [https://perma.cc/T8T4-TLCT].

[196] See id.

[197] See Cerdeña et al., supra note 194, at 1125–26.

[198] See Vyas et al., supra note 191, at 874.

[199] See id.

[200] Id.

[201] See id. at 875.

[202] See id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

correlated with VBAC success. Those variables, however, were not incorporated into the algorithm."[203]

- The STONE algorithm, which assesses the likelihood that a patient complaining of severe pain on the side of their body is suffering from a kidney stone, may give lower scores to black patients, decreasing the likelihood that a physician will discover a kidney stone in a black patient if they do, in fact, have one.[204] "The developers of the algorithm did not suggest why black patients would be less likely to have a kidney stone. An effort to externally validate the STONE score determined that the origin/race variable was not actually predictive of the risk of kidney stones."[205]

The idea of race-based difference reflected in these examples of "race correction" or "race norming" in medicine is based, in part, on the myth of biological race, i.e., the belief that the groups that we understand to be races are genetically homogeneous groups.[206] As Dr. Darshali Vyas and coauthors explain, when AI developers offer rationales for encoding the idea of black racial difference into their models, the reasons they articulate oftentimes stem from "outdated, suspect racial science"—a fact that leads Vyas and coauthors to conclude that "[m]ost race corrections implicitly, if not explicitly, operate on the assumption that genetic difference tracks reliably with race."[207] Scholars have extensively analyzed the claim that the groups that we think of as races are genetically distinct from one another. It is worth repeating, time and again, that this analysis has revealed the idea of "biological race" to be unsupported by

[203] Id.

[204] See id. at 877, 879.

[205] Id. at 879.

[206] The myth of biological race lives on in other sites in medicine. Dr. Cerdeña and her coauthors note that medicine, for the most part, continues to assume that "Asian patients have higher visceral body fat than do people of other races" and, because of this assumption, are "at risk for diabetes at lower body-mass indices." Cerdeña et al., supra note 194, at 1125. Further, "[a]ngiotensin-converting enzyme (ACE) inhibitors are considered less effective in Black patients than in White patients, and they might not be prescribed to Black patients with hypertension." Id. They also reference a study that found that significant numbers of medical students "endorsed the false beliefs that Black patients had longer nerve endings and thicker skin than White patients," a finding that might correlate with the studies that document that physicians are less likely to prescribe analgesics to their black patients. Id.; Joseph Friedman et al., Assessment of Racial/Ethnic and Income Disparities in the Prescription of Opioids and Other Controlled Medications in California, 179 JAMA Internal Med. 469, 473 (2019).

[207] See Vyas et al., supra note 191, at 879.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

scientific evidence.[208] Although the weight of good science establishes that races are not genetically coherent entities, the myth of biological race lives on.[209] "Race correction" algorithms, in many cases, are simply a site of the persistence of the myth of biological race.

This, of course, is not to propose that black people do not have different health outcomes from their nonblack counterparts—it is undeniable that they do.[210] As discussed in Part I above, black people, on the whole, are sicker and die earlier than their nonblack counterparts.[211] But, when "black race" is used as a variable in medical AI, we have to know exactly to what "black race" refers. If "black race" refers to a group that is genetically distinct from other racial groups, then the algorithmic model will be wrong, as the weight of good science does not support the notion that black people are genetically distinct from nonblack people. If, however, "black race" refers to a group that has lived, and continues to live, under conditions of antiblack racism, then the algorithmic model will be more precise if it uses variables that more accurately capture that disadvantage, i.e., zip code, education level, income, wealth, health insurance status, comorbidities, allostatic load, and history with anxiety and depression.[212] Vyas and coauthors describe this tension when they

[208] See id. ("Studies of the genetic structure of human populations continue to find more variation within racial groups than between them."); Richard S. Cooper, Genetic Factors in Ethnic Disparities in Health, *in* Nat'l Rsch. Council, Critical Perspectives on Racial and Ethnic Differences in Health in Late Life 269, 270–71 (Norman B. Anderson, Rodolfo A. Bulatao & Barney Cohen eds., 2004), https://www.ncbi.nlm.nih.gov/books/NBK25517/ [https://perma.cc/7M3N-LYDT]; Osagie K. Obasogie, The Return of Biological Race? Regulating Innovations in Race and Genetics Through Administrative Agency Race Impact Assessments, 22 S. Cal. Interdisc. L.J. 1, 8 (2012).

[209] See, e.g., David Reich, How Genetics Is Changing Our Understanding of 'Race,' N.Y. Times (Mar. 23, 2018), https://www.nytimes.com/2018/03/23/opinion/sunday/genetics-race.html [https://perma.cc/8JQS-33GE].

[210] Alan Nelson, Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care, 94 J. Nat'l Med. Ass'n 666, 666 (2002).

[211] See supra notes 24–29 and accompanying text.

[212] See Rachel R. Hardeman, J'Mag Karbeah & Katy B. Kozhimannil, Applying a Critical Race Lens to Relationship-Centered Care in Pregnancy and Childbirth: An Antidote to Structural Racism, 47 Birth 3, 5 (2020), https://pubmed.ncbi.nlm.nih.gov/31630454/ [https://perma.cc/GGV8-4F4Q] (arguing that clinicians should start attributing poor health outcomes to "racism" as opposed to "race," and observing the "empirically documented impact that racial discrimination has on cortisol levels and biological weathering"); Black Women Scholars & Rsch. Working Grp. of Black Mamas Matter All., Black Maternal Health Research Re-Envisioned: Best Practices for the Conduct of Research with, for, and by Black Mamas, 14 Harv. L. & Pol'y Rev. 393, 397 (2020) ("[Black r]ace is not the risk factor—racism is . . . . There is nothing inherent about Black skin that is physiologically different from any other type of skin except its capacity to overexpose those who have it to racism.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

write that "the racial differences found in large data sets most likely often reflect effects of racism—that is, the experience of being black in America rather than being black itself—such as toxic stress and its physiological consequences."[213] In this way, "race correction" in medicine risks homogenizing people whose racial identity or ascription is black—an incredibly heterogeneous group. The encounters with antiblack racism that the individuals in this group will have had will vary wildly. However, "race corrections" in medical algorithms do not take that variability into account. Moreover, "if adjustments deter clinicians from offering clinical services to certain patients"—as they would if, for example, a physician relies on an algorithm that predicts the likelihood of a successful VBAC and denies their black patient the opportunity of a trial of labor because the algorithm predicts lower rates of success for black people—"they risk baking inequity into the system."[214] This is the difference between using "race in descriptive statistics, where it plays a vital role in epidemiologic analyses, and in prescriptive clinical guidelines, where it can exacerbate inequities."[215]

It is worth acknowledging that we are operating under conditions of uncertainty. It is possible that the science disproving the idea of genetic race is wrong and that there *are*, in fact, as-yet undiscovered genetic variations that are distinct to racial groups—race-based genetic variations that make it make sense to include race as a variable in algorithmic models. It is also possible that the science disproving the idea of genetic race is right. (As Professor Dorothy Roberts has succinctly explained, "It is implausible that one race of people evolved to have a genetic predisposition to heart failure, hypertension, infant mortality, diabetes, and asthma. There is no evolutionary theory that can explain why African ancestry would be genetically prone to practically every major common illness."[216]) We do not know for sure. Moreover, *we will not know definitively until we can talk about racism in the past tense*. As a set of researchers explain in the context of algorithms that assume that black people have inferior lung function, "[u]ntil we eliminate redlining, unequal access to housing and medical care, racially biased hiring and job assignment, environmental racism, and the psychosocial toll that racism

[213] Vyas et al., supra note 191, at 879.
[214] Id.
[215] Id. at 880.
[216] Dorothy E. Roberts, What's Wrong with Race-Based Medicine?: Genes, Drugs, and Health Disparities, 12 Minn. J.L. Sci. & Tech. 1, 15 (2011).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

imposes, we cannot know whether Black people's lower-than-average lung function is normal, or merely the norm."[217] The question for medicine, and the laws that may regulate it, becomes: in the face of this uncertainty, which error causes more harm? Does the harm caused by the erroneous inclusion of the "black race" variable outweigh the harm caused by its erroneous exclusion? Many scholars of race writing in this area urge us to risk the effects of the erroneous exclusion of the "black race" variable, as they are willing to bet their lives that when we have dismantled structural racism and eliminated all instances of individual racism, we will see that what has been "the norm" is not at all "normal."[218]

Before moving on, it is important to observe that the examples of race-norming or race-correcting algorithms discussed here are all transparent algorithms—that is, we know what went into the algorithms. Because we know what went into the algorithm, we can correct the algorithm by removing the race variable.[219] However, as explained above, AI is frequently a "black box," denying us the ability to know whether the algorithm is using race (or a proxy for race) as a variable and, if so, how that variable is being used. Thus, we will not know that the AI algorithm needs fixing and, even if we acquire this knowledge, we will not know how to fix the algorithm.

[217] See Gaffney et al., supra note 195.

[218] See generally Dorothy Roberts, Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-First Century (2011) (analyzing how the myth of the biological concept of race exacerbates inequality).

[219] However, as law professors Anya Prince and Daniel Schwarcz explain, even if programmers remove the race variable from an AI algorithm, we should expect the algorithm to simply find proxies for race, thus producing the same race-based outcomes that the original algorithm generated. See Anya E.R. Prince & Daniel Schwarcz, Proxy Discrimination in the Age of Artificial Intelligence and Big Data, 105 Iowa L. Rev. 1257, 1275 (2020) (stating that "unintentional proxy discrimination by AIs cannot be avoided merely by depriving the AI of information on individuals' membership" because "AIs can and will use training data to derive less intuitive proxies for directly predictive characteristics when they are deprived of direct data on these characteristics"). This is simply because the nature of AI is to find correlations between characteristics in the data on which it is trained and a target output. See id. at 1263. If the AI is explicitly denied access to a characteristic in the data—that is, if we remove the race variable—it will simply find characteristics that correlate with, or correspond to, the excised characteristic that allowed it to predict the target output. See id. at 1303–04 ("AIs that are deprived of direct information about suspect characteristics . . . will inevitably identify other proxy variables for directly predictive data. . . . [A]n AI that did not have access to information about race or hairstyles would inevitably tend to construct alternative proxies, such as Netflix shows watched, or even hair products purchased.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

E. Algorithmic Bias and Racial Disparities in Health and Healthcare

The concern that animates this Article is the possibility—indeed, the likelihood—that the increasingly sophisticated automated tools that will be introduced to manage prenatal care will embed within them the discriminatory treatment and inequitable care that black birthing people currently experience, and have always experienced, during the course of their pregnancies, childbirths, and postpartum periods. The fear is that over-interventions and under-interventions will be built into the machine. The worry is that the algorithm will counsel providers that the black pregnant person sitting in front of them, complaining of blurry vision, severe headaches, and excruciating abdominal pain is not, in fact, suffering from preeclampsia.[220] The apprehension is that the model will recommend that the physician perform a C-section immediately on another perfectly healthy black patient, thereby perpetuating the higher rates of unnecessary C-sections among black people.[221] The concern is that the software will advise the care team that the black postpartum patient complaining of difficulty breathing is not, in fact, experiencing a medical emergency—although, if a proper evaluation were conducted, it would reveal that she had a pulmonary embolism.[222]

Essentially, the anxiety is that the use of AI in prenatal care will perpetuate the status quo. When the status quo is as unjust as it currently is, this is a terrifying possibility.

But, the critique here goes further. The critique is not *solely* that medical AI has been promised to ameliorate some of the inequitable treatment and outcomes that exist in healthcare, but it is unlikely to achieve that noble aim. The critique is also that medical AI will perpetuate the inequitable status quo *while giving the outcomes that it produces the veneer of objectivity, unimpeachability, and scientific overdetermination*. That is, we should be disquieted by the possibility that medical AI will actually make black birthing people worse off—although their pregnancy outcomes will remain the same.[223] Black birthing people will be worse off

[220] See Pre-eclampsia, Nat'l Health Serv. (Sept. 28, 2021), https://www.nhs.uk/conditions/pre-eclampsia/ [https://perma.cc/U796-G5C2].

[221] See supra notes 46–52 and accompanying text.

[222] See, e.g., Rob Haskell, Serena Williams on Motherhood, Marriage, and Making Her Comeback, Vogue (Jan. 10, 2018), https://www.vogue.com/article/serena-williams-vogue-cover-interview-february-2018 [https://perma.cc/CNJ6-MUPK].

[223] Some scholars have cautioned us that medical AI does not need to be perfect; it just needs to be better than what we are currently doing. See Cohen & Graver, supra note 79, at 449

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

because unfair and unjust outcomes will be protected by a discourse proclaiming these outcomes' correctness and innocence, as an "unerring" technology produced them.

Sociologist Ruha Benjamin writes that "technical fixes too often reinforce and even deepen the status quo."[224] She calls this "'the New Jim Code': the employment of new technologies that reflect and reproduce existing inequities but that are promoted and perceived as more objective or progressive than the discriminatory systems of a previous era."[225] Medical AI is a home for the New Jim Code. New technologies for managing pregnancy threaten to perpetuate racial inequality while, simultaneously, shielding the processes that produce that inequality from investigation and interrogation. After all, how can a robot be racist? How can an algorithm have animus? The unease that courses beneath this Article is the possibility that black people will continue to be subjected to all of the inequitable care to which they currently are subjected in doctor's offices and obstetric wards across the nation—inequitable care that contributes to the poorer pregnancy outcomes that are so well-documented. However, because "the machine" will have dictated their treatment regimens, the observing society will sit back, shrug its shoulders, and conclude that there is no problem at all.

One final observation before continuing: some have noted that AI's black-box nature makes it difficult to determine whether an AI tool is biased and, if so, the causes of the bias. First, as noted above, because the technology oftentimes cannot provide comprehensible explanations for its reasons, it will be impossible in many circumstances for a healthcare provider relying on a medical AI device to know whether the device's recommendation is right or wrong. Second, if clinicians or auditors establish that an AI technology is, in fact, unacceptably inaccurate in relation to a specific population, they will find it exceedingly challenging

(stating that when it comes to medical AI (as well as the use of AI in the criminal legal system), "we must never forget to ask the 'as against what' question, and consider whether even imperfect use of predictive analytics here may be better than the status quo way minorities are treated currently"); Schönberger, supra note 90, at 187 ("The ultimate test, though, should be, if AI on balance puts protected groups in a better position compared with the *status quo*, that is being subjected to conventional forms of bias that have been persisting within society for many decades."). This Article makes the point that if the imperfect use of predictive analytics is equivalent to—or even slightly better than—the way people of color are treated currently, people of color may *still* be worse off, as the outcomes would be sheltered from critique by a discourse that declares them to be right, true, and inevitable.

[224] Benjamin, supra note 60, at 5.

[225] Id. at 5–6.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

in most cases to identify just why the tool is biased. As one set of scholars explains, investigation of the algorithm's bias will be "complicated . . . by the myriad potential causes of unfairness (prejudice, structural bias, choice of training data, complex interactions of human behavior with machine learning models, unforeseen supply and demand effects of online bidding processes, etc.)."[226]

The opacity of algorithmic bias has not staunched many observers' enthusiasm for medical AI—and AI in general. Many commentators are optimistic that algorithmic bias can be adequately policed and AI tools can be prevented from harming vulnerable populations. Their sense is that we have at our disposal a range of regulatory approaches—from pre-market FDA regulation of medical devices[227] to licensing[228] and tort liability[229]—that could function to contain algorithmic bias and, in so doing, prevent algorithms from harming marginalized people. Essentially, they propose that we can reap the benefits of AI while managing its harms. Two points deserve mentioning in response.

First, although we know that algorithmic bias in medical AI is likely—and although we know of strategies that could at least limit it—we currently lack the will or the interest to do anything meaningful about it. That is, although we know that algorithmic bias in medical AI can be dangerous—indeed, deadly—to people of color, we seem to have conceptualized those injuries and deaths as an acceptable price to pay for the benefits that medical AI can bring to the "general population," i.e., those who are white. Consider that a set of researchers examined publicly available data that the FDA released concerning the 130 medical AI devices that received FDA approval between January 2015 and December 2020.[230] The results of the examination were disturbing. It showed that close to half of approved devices did not report the number of patients enrolled in the evaluation study.[231] Moreover, "[o]f the 71 device studies

[226] Allison Woodruff, Sarah E. Fox, Steven Rousso-Schindler & Jeff Warshaw, A Qualitative Exploration of Perceptions of Algorithmic Fairness 2 (Proc. 2018 CHI Conf. on Hum. Factors in Computing Sys., 2018).

[227] See Price & Rai, supra note 80, at 804.

[228] See Bambauer, supra note 117, at 391 (asking "what would happen if medical AI were regulated like their closest substitutes—doctors—instead of like devices?" and suggesting that medical AI might be regulated through licensing).

[229] Hannah R. Sullivan & Scott J. Schweikart, Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI?, 21 AMA J. Ethics 160, 162–63 (2019).

[230] Eric Wu et al., How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals, 27 Nature Med. 582, 582 (2021).

[231] Id. at 583.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

that had this information, the median evaluation sample size was 300. Only 17 device studies reported that demographic subgroup performance was considered in their evaluations."[232] Further, the overwhelming majority of the approved devices did not confirm that they had evaluated the device at multiple sites.[233] Indeed, "[o]f the 41 devices with the number of evaluation sites reported, 4 devices were evaluated in only one site, and 8 devices were evaluated in only two sites"—leading the authors to conclude that "a substantial proportion of approved devices might have been evaluated only at a small number of sites."[234] The failure to evaluate a medical AI device at multiple sites impedes our ability to ascertain whether the apparent predictive accuracy of the algorithm is due to features that were present at the evaluation site but not present in all of the sites where the device will the used.[235] In short, despite everything that we already know about algorithmic bias and its potential to harm people of color, the FDA has approved medical AI devices that have not been extensively tested for the presence of algorithmic bias that might harm people of color. What is the import of a regulatory tool if we choose not to use it?

Consider as well that the FDA will not review many medical AI devices, as its authority only extends to devices that are so opaque as to deny the provider the ability to evaluate the recommendation and independently decide whether to follow or reject it.[236] Further, most of the medical AI devices currently on the market that the FDA has approved went through the 510(k) pathway.[237] This means that the FDA did not evaluate the new device's safety, but rather its similarity to a device that

[232] Id.

[233] See id.

[234] Id.

[235] See id. ("Multi-site evaluations are important for the understanding of algorithmic bias and reliability, and can help in accounting for variations in the equipment used, technician standards, image-storage formats, demographic makeup, and disease prevalence.").

[236] See W. Nicholson Price II, Sara Gerke & I. Glenn Cohen, Liability for Use of Artificial Intelligence in Medicine, *in* Research Handbook on Health, AI and the Law (Barry Solaiman & I. Glenn Cohen eds.) (forthcoming 2024) (noting that the FDA will not regulate devices "that make the bases for a recommendation available for a physician to question and to independently decide whether to accept the recommendations").

[237] See id. at 18 ("Almost all AI-based medical devices on the U.S. market received 510(k) clearances . . . . [O]nly one AI-based medical device has so far received [approval through the more rigorous premarket approval process].").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

was already on the market.[238] Again, what is the import of a regulatory tool if we choose not to use it?

Second, we might doubt that the manifold regulatory tools that are available to us, should we actually elect to deploy them, are capable of taming algorithmic bias and the racial stratification that it likely will reproduce and protect. This is simply because taming algorithmic bias may require us to know, with some exacting degree of precision, how racism works. Which is to say: to correct a biased algorithm, we may have to know the specific mechanisms by which racial inequality was built into the machine and in what measures. If, for example, a tool that is designed to support physicians' recommendations around VBAC is biased against black patients, we would first need to realize that the tool is, in fact, biased. That is, we would have to realize that the tool is recommending that physicians perform C-sections on black patients who could have had successful vaginal births if they were permitted to labor without surgical intervention. It may be difficult to arrive at this realization. Once we become aware of the algorithm's bias, we would need to divine the mechanisms behind the bias if we hope to correct the algorithm. We would need to divine whether black people were underrepresented in the training data, maybe because the data was culled from a hospital system that, like many across the nation, is "low black-serving."[239] Alternatively, were black people adequately represented in the training data, but did the data do nothing more than *accurately* reflect physicians' practice of disallowing black people to attempt VBACs while allowing their medically identical white patients to do the same? Or did programmers inappropriately label the data, perhaps predicting the likelihood of a successful VBAC by a factor that correlates with white race—like a person's enjoyment of private insurance, or their status as married, or the

[238] See Medtronic, Inc. v. Lohr, 518 U.S. 470, 493 (1996) (stating that the "510(k) process is focused on *equivalence*, not safety" and that "substantial equivalence determinations provide little protection to the public" (internal quotation marks omitted)). The upshot to this is that because the FDA has not confirmed the safety of devices approved through the 510(k) pathway, individuals whom these devices injure can bring a claim under state tort law against the manufacturer. See id. at 493–94 (holding that the federal Food, Drug, and Cosmetic Act does not preempt state law tort claims when a medical device received a 510(k) clearance); see also Riegel v. Medtronic, Inc., 552 U.S. 312, 322–23 (2008) (holding that the federal Food, Drug, and Cosmetic Act preempts state law when a device was approved through the premarket approval process, during which the FDA evaluates a device's safety and efficacy).

[239] See Howell et al., supra note 181, at 122.e1 (noting the existence of racial segregation in hospitals in the United States and the fact that hospitals could be understood as "high black-serving" and "low black-serving").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

presence of few or no comorbidities? Or did programmers fail to include in the model clinically relevant demographic data—for example, the fact that a pregnant patient will have a decreased likelihood of needing or undergoing a C-section if a doula cares for them during their labor and childbirth?[240] Or are all of the above true? Did all of these possibilities contribute to the technology's practice of informing physicians that vaginal births are contraindicated for their black patients? Do we have the knowledge necessary to perform this racial autopsy on a piece of software?[241] *Do we know with the requisite degree of exactitude how racism works?*

The next Part begins an exploration of an approach to addressing algorithmic bias (and medical AI, generally) that has not received a lot of attention: the informed consent process. That is, perhaps the informed consent process—the pre-intervention dialogue between provider and patient that is designed to enable the patient to be an active agent in the healthcare that they receive—presents an occasion for confronting the promises and perils of medical AI. Essentially, the next Part weds two observations. The first observation is the claim that law and ethics may require providers to disclose during the informed consent process that the intervention that they are recommending (or not recommending) to the patient is based, in part, on the prediction of an algorithm.[242] The second observation is that social justice concerns have not featured prominently in our discussions of how AI ought to be regulated—and whether AI ought to be introduced in the first instance.[243] As legal scholar Calvin

[240] Katy B. Kozhimannil et al., Potential Benefits of Increased Access to Doula Support During Childbirth, 20 Am. J. Managing Care e340, e346–47 (2014).

[241] It may be worth noting that in order to perform this racial autopsy, developers would need to grant access to the algorithms used in the model and the data on which the model was trained. See Obermeyer et al., supra note 96, at 447 (explaining that "[a]lgorithms deployed on large scales are typically proprietary, making it difficult for independent researchers to dissect them" and "[w]ithout an algorithm's training data, objective function, and prediction methodology, we can only guess as to the actual mechanisms for the important algorithmic disparities that arise").

[242] See Price, supra note 95, at 436 n.76 (noting that "deliberately or inevitably opaque algorithms that direct medical care may have implications for informed consent" because providers might have to "disclose that they are using a complex algorithm they do not understand").

[243] Some scholars have posed the question, though. See, e.g., Terry, Appification, supra note 102, at 166 ("Automation must address a fundamental question—will healthcare data technologies increase or decrease healthcare disparities?"); Terry, Regulating Healthcare AI, supra note 112, at 187 ("[A] fundamental inquiry must be whether healthcare AI will increase or decrease healthcare disparities.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Wai-Loon Ho describes, "[t]here is clearly overwhelming regulatory focus on values relating to safety and effectiveness. It also renders more apparent considerations of equity or social justice as a possible gap in regulatory attention."[244] The next Part investigates whether the informed consent process might be a space wherein social justice values can manifest in our discussions of medical AI.

IV. REFORMING THE INFORMED CONSENT PROCESS: RACE AND MEDICAL AI

This Part begins with an exploration of the fact of algorithmic aversion, whereby individuals do not trust algorithms to make decisions in some contexts. It discusses research showing that black people with higher levels of medical mistrust specifically want information about the accuracy of healthcare technologies with respect to black people. That is, black people who do not trust healthcare institutions and medical knowledge tend to want to receive particularized information about a medical intervention's safety and efficacy vis-à-vis black people. This research counsels in favor of disclosing the use of medical AI—and, particularly, information about the near inescapability of algorithmic bias given the country's history and present of racially inequitable healthcare—in the informed consent process.

The Part continues by recounting the traditional story of informed consent—a story that describes informed consent as a practice that emerged in response to medical paternalism. In this rendering, the doctrine of informed consent arose as an effort to transform the patient from an object upon which the physician acts to a subject who, in consultation with the physician, decides upon the healthcare that they will receive. While this traditional story of informed consent is certainly true, it is not the entire truth. That is, the traditional story obscures the antiracist, anti-white supremacist, and indeed *radical* origins of informed consent. This Part argues that the introduction into the clinical encounter of medical AI—and the likelihood that these technologies will perpetuate racially inequitable healthcare while masking the same—is an invitation to reform the informed consent process to make it more consistent with the commitments that spurred its origination. This Part proposes that,

[244] Calvin Wai-Loon Ho, Deepening the Normative Evaluation of Machine Learning Healthcare Application by Complementing Ethical Considerations with Regulatory Governance, 20 Am. J. Bioethics 43, 44 (2020).

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

given the antiracist roots of the doctrine of informed consent, it would be incredibly ironic to allow the informed consent process to permit a patient—and, particularly, a patient of color—to remain ignorant of the fact that their medical care is being managed by a device or system that likely encodes racism. This Part argues that informing patients about the likelihood of race-based algorithmic bias—and the reasons that we might *expect* race-based algorithmic bias—may, in fact, be a prerequisite for actually transforming the social conditions that produce racial disparities in health and healthcare.

A. Race, Medicine, and Trust: Will People of Color Want Medical AI?

When we are considering the question of what the informed consent process should look like in the advent of medical AI—querying whether we should require disclosures around the use of medical AI as well as relevant facts regarding racial disparities in health and healthcare—we should bear in mind an important fact: some patients may not want their providers to rely on or consult with medical AI when providing care to them. To be precise, *black patients* may not want their physicians to rely on or consult with medical AI when providing care to them. Black patients may not trust the technology.

Some scholars have proposed that physicians may serve as conduits through which patients' trust in medical AI could be established.[245] They have argued that patients initially may be wary of the use of a new, unestablished technology in their healthcare; however, their physicians could provide assurances that the technology is safe and effective and will operate to their benefit.[246] As a result, these proponents say, patients will come to trust medical AI and welcome its management of their healthcare.

There is a problem with this narrative, however: it is well-documented that black people are more likely than their nonblack counterparts to distrust physicians and healthcare institutions[247]—a fact that received a

[245] See Feldman et. al., supra note 96, at 411–12 (writing that patients trust "drugs whose mechanism of action is unknown" in part because of their doctors, "whose focus on patient care and patient bonding help bridge the understanding gap and establish trust").

[246] See id. at 413 (questioning whether "physicians themselves [will] be able to sprinkle the fairy dust necessary to establish a patient's comfort with an AI system").

[247] See Lillie D. Williamson, Beyond Personal Experiences: Examining Mediated Vicarious Experiences as an Antecedent of Medical Mistrust, 37 Health Commc'n 1061, 1061 (2022) ("African Americans consistently report higher levels of medical mistrust than their White counterparts."); Marcella Alsan, Owen Garrick & Grant Graziani, Does Diversity Matter for Health? Experimental Evidence from Oakland, 109 Am. Econ. Rev. 4071, 4072–73 (2019)

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

tremendous amount of attention when initial studies revealed that black people, as a racial group, reported the most hesitancy toward receiving a vaccine against COVID-19.[248] If black people do not trust the physicians

("Recent studies in public health demonstrate that African American men continue to score higher on medical mistrust measures than other groups."); Sirry Alang, Donna D. McAlpine & Rachel Hardeman, Police Brutality and Mistrust in Medical Institutions, 7 J. Racial & Ethnic Health Disparities 760, 762 (2020) (reporting the results of their study revealing that "[m]istrust was higher for all minority racialized groups compared to Whites"); Adolfo G. Cuevas & Kerth O'Brien, Race Centrality May be Linked to Mistrust in Healthcare Institutions for African Americans, 24 J. Health Psych. 2022, 2022 (2019) ("African Americans report significantly lower trust in healthcare providers and the healthcare system than do European Americans."). But see Marla B. Hall et al., Cervical Cancer Screening Behaviors and Perceptions of Medical Mistrust Among Rural Black and White Women, 29 J. Health Care for Poor & Underserved 1368, 1374 (2018) (finding that the poor, rural white participants in the study were less likely to trust the healthcare system than their black counterparts).

This mistrust has sundry negative consequences. See Jessica Jaiswal & Perry N. Halkitis, Towards a More Inclusive and Dynamic Understanding of Medical Mistrust Informed by Science, 45 Behav. Med. 79, 79 (2019) ("There are myriad negative health consequences associated with medical mistrust, including lower utilization of health care, and poorer management of health conditions including HIV, cancers, and diabetes."); Williamson, supra, at 1061 (stating that medical mistrust is associated with "decreased likelihood of engagement in several health behaviors including routine health checkups, cancer screenings, treatment adherence, and registering as an organ donor" (citations omitted)).

[248] See Florence Momplaisir et al., Understanding Drivers of Coronavirus Disease 2019 Vaccine Hesitancy Among Blacks, 73 Clinical Infectious Diseases 1784, 1784 (2021) (noting that "[l]arge surveys demonstrate that vaccine acceptance among blacks varies between 39% and 55%, the lowest reported acceptance rates among racial/ethnic groups"); Justin Stoler, Adam M. Enders, Casey A. Klofstad & Joseph E. Uscinski, The Limits of Medical Trust in Mitigating COVID-19 Vaccine Hesitancy among Black Americans, 36 J. Gen. Internal Med. 3629, 3629 (2021) (noting that there was significant racial variation among study participants regarding willingness to take a COVID-19 vaccine, with 70.4% of white participants, 61.5% of Latinx participants, and 44.3% of black participants expressing willingness to be vaccinated, and stating that "even at the highest levels of [trust in health institutions], Black respondents were significantly less willing to take a COVID-19 vaccine . . . than White respondents"); Simar Singh Bajaj & Fatima Cody Stanford, Beyond Tuskegee—Vaccine Distrust and Everyday Racism, 384 New Eng. J. Med. e12(1), e12(1) (2021) (reporting that in November 2020, a study showed that "only 14% of Black survey respondents trusted the vaccines' safety and only 18% said they would definitely get vaccinated"). Some scholars have observed the irony: the communities most harmed by COVID-19 would be least likely to be protected against it. See Natasha Crooks, Geri Donenberg & Alicia Matthews, Ethics of Research at the Intersection of COVID-19 and Black Lives Matter: A Call to Action, 47 J. Med. Ethics 205, 205 (2021) ("[U]nethical research practices over many years, combined with persistent health disparities and lack of access to effective treatments for Black people, discourage the very groups most in need of new innovations from receiving them.").

Over time, the Republican Party's participation in vaccine misinformation eventually led white people, as a racial group, to be the most hesitant to take a COVID-19 vaccine. See Andrea Michelson, White Republicans Are More Likely to Reject the COVID-19 Vaccine

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

who serve as conduits for trust in medical AI, they will not trust the medical AI.

One of the most frequently offered explanations for black mistrust of medicine and healthcare institutions is the Tuskegee Experiment.[249] The idea is that this event has traumatized generations of black people, teaching black people that institutions that purport to offer healthcare are duplicitous, deceitful, and dangerous. The thought is that this study, which ended half a century ago, has instructed black people that the people in white coats are just as likely to be acting against their patients' interests as they are to be acting in them. The Tuskegee explanation of

than Any Other Group in America, Bus. Insider (Mar. 15, 2021, 4:04 PM), https://www.businessinsider.com/white-republicans-more-likely-to-reject-covid-19-vaccine-2021-3 [https://perma.cc/9W35-QDAZ]; David Leonhardt, COVID and Race, N.Y. Times (June 9, 2022), https://www.nytimes.com/2022/06/09/briefing/covid-race-deaths-america.html [https://perma.cc/7SPF-TVZF]. But see MGH News & Pub. Affs., Willing but Unable to Get COVID Shot, Harv. Gazette (Feb. 1, 2022), https://news.harvard.edu/gazette/story/2022/02/study-examines-racial-ethnic-disparities-in-covid-vaccine-rate/ [https://perma.cc/Q5BQ-6PEW] (showing that people of color in the United States and United Kingdom were up to three times likelier than white people to report vaccine hesitancy); Juana Summers, Little Difference in Vaccine Hesitancy Among White and Black Americans, Poll Finds, NPR (Mar. 12, 2021, 5:00 AM), https://www.npr.org/sections/coronavirus-live-updates/2021/03/12/976172586/little-difference-in-vaccine-hesitancy-among-white-and-black-americans-poll-find [https://perma.cc/WJ2D-W3K5] (reporting that a national poll from 2021 showed similar levels in vaccine hesitancy among black and white respondents).

[249] Vernellia R. Randall, Slavery, Segregation and Racism: Trusting the Health Care System Ain't Always Easy! An African American Perspective on Bioethics, 15 St. Louis U. Pub. L. Rev. 191, 197–98 (1996); see also Jaiswal & Halkitis, supra note 247, at 80 (noting that "much of the public health and medical literature cites the infamous Tuskegee study as a main catalyst for this persistent health-related mistrust among racial and ethnic minority people").

The Tuskegee Experiment, originally called the Tuskegee Study of Untreated Syphilis in the Negro Male, was a study that that the United States Public Health Service conducted from 1932 to 1972. The Untreated Syphilis Study at Tuskegee Timeline, Ctrs. for Disease Control & Prevention (Dec. 5, 2022), https://www.cdc.gov/tuskegee/timeline.htm [https://perma.cc/U7J2-7NLM]. Researchers enrolled 600 black men—399 of whom had syphilis—in order to study how untreated syphilis affected black men's bodies. See id. The men were not informed that they were being studied. See id. Neither were they given treatment for syphilis when penicillin was found to be an effective treatment for the disease. See id. The researchers permitted the participants to infect their sexual partners and expose the children that they conceived with those partners to congenital syphilis. See Deleso Alford Washington, Examining the "Stick" of Accreditation for Medical Schools Through Reproductive Justice Lens: A Transformative Remedy for Teaching the Tuskegee Syphilis Study, 26 St. John's J.C.R. & Econ. Dev. 153, 193 (2011); Tuskegee Syphilis Experiment, Equal Just. Initiative (Oct. 31, 2020), https://eji.org/news/history-racial-injustice-tuskegee-syphilis-experiment/ [https://perma.cc/X85F-R2DG]. The study was terminated only after a whistleblower exposed its existence to the public—some four decades after it began. See Centers for Disease Control and Prevention, supra.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

black mistrust of medicine proposes that over the course of the fifty years that have passed since the study came to light, black people have developed a "culture" of healthcare mistrust and, consequently, avoidance.[250]

While the higher levels of medical mistrust among black people are empirically established, explaining this fact in terms of a "culture" kindled by the Tuskegee experiment is problematic. First, cultural explanations of health behaviors are reductive. They tend to imagine that subjects engage in health behaviors baselessly and unthinkingly—because that is just the way people in their "culture" behave. In truth, however, there often are good reasons for the relationships that individuals have with healthcare institutions and providers. We can understand more incisively and productively the decisions of those whose health behaviors are dismissively explained in terms of "culture" if we look to the structures that constrain (or expand) their choices[251]—or even if we look to the particular medical condition that they confront.[252]

[250] Peter A. Clark, A Legacy of Mistrust: African-Americans, the Medical Profession, and AIDS, 65 Linacre Q. 66, 66–67 (1998); see also Darcell P. Scharff et al., More than Tuskegee: Understanding Mistrust About Research Participation, 21 J. Health Care for Poor & Underserved 879, 880 (2010) (noting that "the Tuskegee syphilis study is widely recognized as a reason for mistrust").

[251] See Jonathan M. Metzl & Helena Hansen, Structural Competency: Theorizing a New Medical Engagement with Stigma and Inequality, 103 Soc. Sci. & Med. 126, 128, 130 (2014) (defining "structural competency" as "the trained ability to discern how a host of issues defined clinically as symptoms, *attitudes*, or diseases . . . also represent the downstream implications of a number of upstream decisions about such matters as health care and food delivery systems, zoning laws, urban and rural infrastructures, medicalization, or even about the very definitions of illness and health" and writing that the framework of structural competency might shift the "diagnostic focus from the 'culture' of individual patients" (emphasis added)); Jaiswal & Halkitis, supra note 247, at 81 (arguing that understanding medical mistrust in cultural terms "is problematic and likely racist in that it situates the onus to overcome medical mistrust on the population experiencing structural, social, political, and economic exclusion and marginalization, rather than the institutions and entities that have created environments that engender mistrust and sustain institutionalized inequalities").

[252] Benjamin conducted research at an organization that banked blood for families with children who had either sickle cell anemia or beta-thalassemia. See Ruha Benjamin, Cultura Obscura: Race, Power, and "Culture Talk" in the Health Sciences, 43 Am. J.L. & Med. 225, 232 (2017) [hereinafter Benjamin, Cultura Obscura]. The families negotiating sickle cell anemia tended to be black and were less likely to participate in the organization's transplant program. See id. at 232. Meanwhile, the families negotiating beta-thalassemia tended to be Asian American and were much more likely to participate in the transplant program. See id. Benjamin states that workers at the organization leaned on cultural explanations to understand the differences in participation. She quotes a caseworker, who offered that "sickle cell patients act like they don't have any control over what happens—fatalistic—and it may be that they don't trust medicine and science. But then thalassemia patients are so controlling. They have

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Further, the Tuskegee explanation of black medical mistrust proposes that the genesis of this mistrust is in a faraway past: a lone experiment that ended several generations ago. In this way, the Tuskegee explanation suggests that black mistrust is *irrational*.[253] The experiment, singular and exceptional, came to an end fifty years ago, after all. And if black mistrust is irrational, then physicians and healthcare institutions do black people a service when they attempt to disabuse them of their irrationality.

However, other scholars have argued that the genesis of black medical mistrust is not in a lone experiment that ended half a century ago, but rather in events that happen today.[254] As one set of researchers explains,

a completely different perspective on medicine and science. They absolutely trust it." Id. at 232–33 (quoting Ruha Benjamin, People's Science: Bodies and Rights on the Stem Cell Frontier 115 (2013)). Benjamin writes that the caseworker was "invoking popular, racialized notions about science-philia among Asian Americans contrasted with science-phobia among African Americans, suggesting possible cultural differences among these patient populations to explain the disparity in transplantation rates." Id. at 233. But, as Benjamin incisively explains:

> [A] beta-thalassemia diagnosis entails much more certainty as to the progression of the illness and the available forms of care, whereas for sickle cell anemia there is much more variability in symptoms and treatments. For sickle cell patient families, therefore, the uncertainty of the transplant outcome is set against the uncertainty of the illness itself, compounding what cannot be known in advance. Rather than reverting to static notions of culturally-based distrust towards vague ideas of "science" and "medicine," parents' *ambivalence-in-action* is informed by and produced through ongoing interaction with a host of different practitioners, diagnoses, treatment options, and institutions.

Id. In essence, "culture" does little work to explain differences in the willingness to engage with medicine and healthcare institutions among black and Asian American families.

[253] See Camisha A. Russell, Questions of Race in Bioethics: Deceit, Disregard, Disparity, and the Work of Decentering, 10 Phil. Compass 43, 45 ("While the Tuskegee Syphilis Study and its effects on the black community were undoubtedly traumatic, repeated mentions of the story *in isolation* can make it appear as an aberration, which risks casting African American wariness of the medical establishment as 'an overreaction to a single event . . . .'").

[254] See Jaiswal & Halkitis, supra note 247, at 80 (arguing that we should conceptualize "mistrust as a phenomenon created by and existing within a system that creates, sustains and reinforces racism, classism, homophobia and transphobia, and stigma"); Lillie D. Williamson, Marisa A. Smith & Cabral A. Bigman, Does Discrimination Breed Mistrust? Examining the Role of Mediated and Non-Mediated Discrimination Experiences in Medical Mistrust, 24 J. Health Commc'n 791, 795 (2019) (reporting the results of a study that showed participants' trust in medical institutions declined after being exposed to a news story about police brutality); Alang et al., supra note 247, at 763, 765 (stating that previous studies establish that medical mistrust "is significant among persons who frequently experience unfair treatment in the broader social environment as well as neighborhood stressors" and that their own research "emphasizes that conditions outside the medical system impact the quality of medical encounters"); Sirry Alang, Donna McAlpine, Malcolm McClain & Rachel Hardeman, Police Brutality, Medical Mistrust and Unmet Need for Medical Care, 22 Preventive Med. Reps. 1,

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

"[H]istorical traumas certainly provide critical context for interpreting present-day occurrences. But attributing distrust primarily to these instances ignores the everyday racism that Black communities face. Every day, Black Americans have their pain denied, their conditions misdiagnosed, and necessary treatment withheld by physicians."[255] These researchers propose that when black people cite the Tuskegee Experiment as the reason for their refusal to trust physicians and healthcare institutions, "those patients are probably not historicizing their frustration by recalling Tuskegee, but rather contemplating how an institution sworn to do no harm has failed them."[256]

In other words, we have organized society in such a way that black people disproportionately bear the burdens of poverty,[257] are disproportionately incarcerated,[258] disproportionately die violent deaths,[259] and disproportionately suffer from every major common illness—from heart disease to asthma.[260] *What is trustworthy about the institutions that such a country would generate? Indeed, what is trustworthy about the institutions that would generate such a country?*

If black medical mistrust is a product not of the Tuskegee experiment, but rather of injustices perpetrated today, then we can understand this mistrust as perfectly rational.[261] If framed in this way, then the solution to

1 (2021) (reporting the results of a study that revealed that "people who have experienced police brutality are more likely to mistrust medical institutions compared to their peers who have not experienced police brutality" and explaining the results as a function of people "bring[ing] the social context of their lives with them to the medical encounter").

[255] Bajaj & Stanford, supra note 248, at 1.

[256] Id.

[257] See Bettina M. Beech, Chandra Ford, Roland J. Thorpe Jr., Marino A. Bruce & Keith C. Norris, Poverty, Racism, and the Public Health Crisis in America, 9 Frontiers Pub. Health 1, 1–3 (2021); see also Gilbert C. Gee & Chandra L. Ford, Structural Racism and Health Inequities: Old Issues, New Directions, 8 Du Bois Rev. 115, 115–17 (2011) (explaining that racism has contributed to health inequities).

[258] Wendy Sawyer, Visualizing the Racial Disparities in Mass Incarceration, Prison Pol'y Initiative (July 27, 2020), https://www.prisonpolicy.org/blog/2020/07/27/disparities/ [https://perma.cc/Y48F-RZFV].

[259] See, e.g., Nada Hassanein, Young Black Men and Teens Are Killed by Guns 20 Times More Than Their White Counterparts, CDC Data Shows, USA Today (Feb. 25, 2021, 7:40 AM), https://www.usatoday.com/story/news/health/2021/02/23/young-black-men-teens-made-up-more-than-third-2019-gun-homicides/4559929001/ [https://perma.cc/5QHJ-B4W8].

[260] See discussion supra notes 25, 39, 40 and accompanying text.

[261] Benjamin, Cultura Obscura, supra note 252, at 234 (asking "[w]hy, then, are trust issues conceived of as a cultural trait and anomaly to be 'overcome'—rather than a perfectly rational, even incisive, disposition toward biomedicine in a socially stratified society?"); Jaiswal &

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

black mistrust is not to beneficently disabuse black people of their irrationality, but rather to address the banal injustices that make it make sense to withhold trust from these institutions. Indeed, the solution is to make institutions trustworthy.[262]

Reproductive health, unsurprisingly, is no exception to the phenomenon of patients of color mistrusting the physicians who are supposed to provide care to them. In qualitative studies investigating people of color's experiences with their reproductive healthcare providers, physicians' failure to provide adequate information appears as a common complaint.[263] Moreover, providers' failure to give full

Halkitis, supra note 247, at 81 (arguing that we should understand "medical mistrust as a protective response against the pervasive, interlocking structural inequalities that result in restricted access to resources, including housing, educational opportunities, employment, and healthcare, in addition to daily experiences of racism, stigma and discrimination"); Cuevas & O'Brien, supra note 247, at 2027 ("[M]istrust should not be viewed as a problematic characteristic of the patient. Mistrust may be a justifiable, adaptive response to experiences of discrimination, both contemporary and historical experiences.").

[262] See Benjamin, Cultura Obscura, supra note 252, at 234 (writing that we should "reorient[] ourselves—away from a fixation with distrust and towards the problem of institutional trustworthiness"); Williamson, supra note 247, at 1068 ("Instead of locating the solution within individuals, the solution should be properly placed on systems.").

[263] See Anu Manchikanti Gomez & Mikaela Wapman, Under (Implicit) Pressure: Young Black and Latina Women's Perceptions of Contraceptive Care, 96 Contraception 221, 223 (2017) (stating that study participants felt that "their questions were not fully answered" and that "they received incomplete descriptions of [birth control] methods and side effects"); Rachel D. Godsil & L. Song Richardson, Racial Anxiety, 102 Iowa L. Rev. 2235, 2255–56 (2017) (referencing a study showing that racial anxiety reduces the quality of information passed between doctor and patient and sharing the story of a black women being treated for breast cancer who reported that she had the "constant sense that her health care providers were not sharing as much information with her about treatment options because of her race—despite her status as a professional with high-quality health care insurance"); Brittany N. Edwards et al., What About the Men? Perinatal Experiences of Men of Color Whose Partners Were at Risk for Preterm Birth, a Qualitative Study, 20 BMC Pregnancy & Childbirth 1, 6 (2020) (noting that participants in a study of black men's experiences with preterm birth shared "[m]ultiple examples" of instances "where providers did not disclose or withheld desired information"). In fact, these studies show that not only are patients of color dissatisfied with the quality and quantity of the information that their providers give them, but they are also dissatisfied with their providers' receipt of the information that their patients share. That is, patients of color frequently complain that their providers do not *listen* to them. See, e.g., Rachel G. Logan, Ellen M. Daley, Cheryl A. Vamos, Adetola Louis-Jacques & Stephanie L. Marhefka, "When Is Health Care Actually Going to Be Care?" The Lived Experience of Family Planning Care Among Young Black Women, 31 Qualitative Health Rsch. 1169, 1174 (2021) (noting that not being listened to was a complaint among study participants); Molly R. Altman, Monica R. McLemore, Talita Oseguera, Audrey Lyndon & Linda S. Franck, Listening to Women: Recommendations from Women of Color to Improve Experiences in Pregnancy and Birth Care, 65 J. Midwifery & Women's Health 466, 466 (2020) ("In the

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

information led patients of color to suspect that their providers were experimenting on them or doing interventions for the providers' financial benefit.[264] One response to this phenomenon is to provide more information during the clinical encounter to patients seeking healthcare. Informing them about the use of medical AI and how that technology interacts with racial disparities in health and healthcare is as good a place as any to begin. It is an appropriate place to begin repairing fractured trust.

1. The Fact of Algorithmic Aversion

Studies have shown that, as a general matter, people do not trust AI systems to do the tasks that humans traditionally have done[265]—a phenomenon that the literature has come to call "algorithm aversion."[266] The existing research demonstrates that on the whole, people have higher levels of trust in the recommendations that human actors, as opposed to machine actors, make in myriad contexts, ranging from employment to dating.[267] The reasons for algorithmic aversion are many, including people's belief in:

context of pregnancy and birth, women of color have repeatedly shared that they are not listened to and/or that their concerns are ignored."). The consequence is that, as a general matter, the quality of physician-patient communication that black people report experiencing is poor.

[264] See Edwards et al., supra note 263, at 8 (stating that participants in a study of men whose partners had preterm births offered that "certain tests and procedures seemed experimental and without reason" and that "[s]ome felt as if they were being used for research purposes").

[265] See Min Kyung Lee, Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management, 5 Big Data & Soc'y 1, 14 (2018) (finding reinforcement for the claim that "the general public does not fully trust algorithms"); Longoni et al., supra note 105, at 630 ("Given the superior accuracy of statistical models over human intuition, people should prefer to follow the advice of statistical models over human intuition. Yet, in most cases, people do not."). But see Jennifer M. Logg, Julia A. Minson & Don A. Moore, Algorithm Appreciation: People Prefer Algorithmic to Human Judgment, 151 Organizational Behav. & Hum. Decision Processes 90, 92 (2019) (conducting experiments that revealed that "people consistently give more weight to equivalent advice when it is labeled as coming from an algorithmic versus human source"—a phenomenon that the researchers call "algorithmic appreciation").

[266] See Berkeley J. Dietvorst, Joseph P. Simmons & Cade Massey, Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err, 144 J. Experimental Psych. 114, 114 (2015).

[267] See Noah Castelo, Maarten W. Bos & Donald R. Lehmann, Task-Dependent Algorithm Aversion, 56 J. Mktg. Res. 809, 810 (2019); Lee Rainie, Monica Anderson, Colleen McClain, Emily A. Vogels & Risa Gelles-Watnick, AI in Hiring and Evaluating Workers: What Americans Think, Pew Rsch. Ctr. (2023), https://www.pewresearch.org/internet/2023/04/20/ai-in-hiring-and-evaluating-workers-what-americans-think/ [https://perma.cc/ZB3E-T6BU]; Dalia L. Diab, Shuang-Yueh Pui, Maya Yankelevich & Scott Highhouse, Lay Perceptions of

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

> the inability of algorithms to learn, the presumed ability of human forecasters to improve through experience, the notion that algorithms are dehumanizing, the notion that algorithms cannot properly consider individual targets, concerns about the ethicality of relying on algorithms to make important decisions, and the presumed inability of algorithms to incorporate qualitative data.[268]

Of course, there is nuance. Studies have shown that algorithmic aversion varies in light of the stakes of the situation in which an AI system is used.[269] That is, algorithmic aversion declines when individuals perceive the stakes of the situation to be low—as when an AI system is used to help job candidates hone their interview skills.[270] However, when individuals perceive the stakes of the situation to be high—as when an AI system is used to evaluate job candidates and to make hiring decisions—algorithmic aversion grows significantly.[271]

Studies also show that people's willingness to rely on an AI system's recommendation depends on the task that the system is performing.[272] People are more inclined to trust an AI system and to rely on its prediction when the system performs "objective tasks" (or tasks that require "cognitive abilities")—that is, an undertaking that "involves facts that are quantifiable and measurable."[273] However, algorithmic aversion rises when the AI system performs "subjective tasks" (or tasks that require "emotional abilities")—that is, an undertaking that is "open to interpretation and based on personal opinions or intuition."[274] Interestingly, researchers have shown that individuals' aversion to algorithms is manipulable.[275] A person's willingness to use an AI system

Selection Decision Aids in US and Non-US Samples, 19 Int'l J. Selection & Assessment 209, 214–15 (2011).

[268] Dietvorst et al., supra note 266, at 115 (citations omitted).

[269] See Markus Langer, Cornelius J. König & Maria Papathanasiou, Highly Automated Job Interviews: Acceptance Under the Influence of Stakes, 27 Int'l J. Selection & Assessment 217, 229 (2019).

[270] See id. at 229–30.

[271] See id.

[272] See Castelo et al., supra note 267, at 811.

[273] Id. at 811–12; see also Lee, supra note 265, at 4 (hypothesizing that people will distinguish between human skills and mechanical skills when analyzing AI).

[274] Castelo et al., supra note 267, at 811–12; see also Lee, supra note 265, at 11–12 (reporting that with tasks that seem to require more "human" skills, "participants judged algorithmic decisions as less fair, trusted algorithmic decisions less, and felt less positive toward algorithmic decisions than human decisions").

[275] See Castelo et al., supra note 267, at 811.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

will increase if an actor reframes the task that the technology is being asked to perform as one that is more objective than subjective; further, algorithmic aversion decreases even for subjective tasks if the user is informed that AI systems have robust emotional abilities.[276] Research also shows that algorithmic aversion increases when users witness the machine make an error;[277] notably, individuals are much more forgiving of the human decision-makers whose errors they witness.[278]

The general aversion that people have to algorithms carries over to the health context. Patients tend not to trust medical AI, and they tend to much prefer a recommendation that comes from a human than a machine.[279]

[276] See id. at 811. The average reader may be surprised to learn that AI systems do, in fact, have emotion-like capacities. Professor Castelo and coauthors explain that "[a]lgorithms can already create paintings that sell for hundreds of thousands of dollars, write compelling poetry and music, predict which songs will be hits, and even accurately identify human emotion from facial expressions and tone of voice and respond accordingly." Id. at 812 (citations omitted).

[277] See Dietvorst et al., supra note 266, at 124. Disturbingly, Professor Dietvorst and coauthors appear to take this as a reason to conceal the use of algorithms, which will eliminate the possibility that an individual will witness an AI system err. See id. Even more disturbingly, they identify the healthcare context as a site wherein the use of AI should be concealed. See id. ("[O]ur findings do suggest that people will be much more willing to use algorithms when they do not see algorithms err, as will be the case when . . . the algorithm is unseen (as it often is for patients in doctors' offices) . . . ."). What may seem to these researchers to be a reasonable effort to prevent people's irrationality from stopping them from relying on an effective and accurate tool may seem to others as an unfortunate effort to deny people information that they deem material.

[278] See id.

[279] See Romain Cadario, Chiara Longoni & Carey K. Morewedge, Understanding, Explaining, and Utilizing Medical Artificial Intelligence, 5 Nature Hum. Behav. 1636, 1636 (2021) (noting that "patients are reluctant to utilize medical AI," viewing such technology "as unable to meet their unique needs, . . . as performing more poorly than comparable human providers," and lacking an accountability that human providers have (citation omitted)); Castelo et al., supra note 267, at 809, 810–811 (citing research that shows that people "trust medical recommendations less when they come from an algorithm than from a human doctor" and referencing a study that "found that participants rated physicians who made an unaided diagnosis significantly more positively than a physician who used an algorithm to assist with the diagnosis, but no differently than a physician who consulted a colleague to assist with the diagnosis" (citations omitted)); Longoni et al., supra note 105, at 636 (reporting the results of an experiment that showed strong algorithmic aversion to medical AI and noting that "consumer resistance to medical AI is not driven merely by the belief that the performance of an automated provider is objectively inferior to that of a human provider," but that "[p]articipants were resistant to medical AI even when the performance of AI providers was explicitly specified to be superior to that of human providers").

Some scholars have proposed that patients will be more willing to trust medical AI if they understand, at some superficial level, how the AI system works. See Price & Rai, supra note 80, at 796 ("For medical algorithms, patients might be more willing to follow algorithmic recommendations if at least some facets of those algorithms are disclosed and potentially

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Psychological experiments have shown that part of an individual's antipathy to medical AI is a function of the belief that, as an individual, they are wildly unique and that algorithmic models are unable to recognize and appreciate uniqueness in the individuals with whom it engages.[280] Further, there is evidence that individuals' trust in their doctors is weakened when they know that their doctors are relying on AI to assist in healthcare provision.[281] The use of medical AI in the clinical encounter threatens to reframe physicians as "mere button-pushers," as opposed to wise, knowledgeable caretakers of their patients' lives and health.[282]

Of course, as always, there is nuance. Studies have shown that algorithmic aversion to medical AI is reduced when patients believe that a physician is using an AI system for *support* as opposed to supplantation.[283]

The copious amounts of research that show that people, as a general matter, are not enthusiastic when it comes to the prospect that AI will be used in their healthcare raises the question of whether they will be able to control whether or not AI is, in fact, used in their healthcare.[284] This supports the argument that the informed consent process—the assigned vehicle for protecting the patient's autonomy—should include the disclosure that a healthcare provider relied on or consulted with an AI system or device to manage a patient's care.[285]

understood—and, eventually, if we come to learn more about the underlying biological systems on which their recommendations are based."). Price holds out the possibility that patients will blindly trust medical AI as long as they believe that it works. See W. Nicholson Price, Big Data and Black-Box Medical Algorithms, 10 Sci. Translational Med. 1, 3 (2018).

[280] See Longoni et al., supra note 105, at 639.

[281] See Emily LaRosa & David Danks, Impacts on Trust of Healthcare AI 214 (Proc. 2018 AAAI/ACM Conf. on AI, Ethics, & Soc., 2018) (noting that "that patient-doctor trust will likely be damaged if doctors are perceived (socially) to abdicate their current roles to AI systems").

[282] Id.

[283] See Longoni et al., supra note 105, at 645.

[284] Some scholars have assumed that patients will determine whether or not medical AI will be used. See Longoni et al., supra note 105, at 630 ("In the context of medical AI, "consumers will *directly* drive medical AI's adoption. In conventional clinical settings, where patients' interactions with AI may still be mediated by physicians, consumers will *indirectly* determine medical AI's adoption."). These scholars may be assuming that patients have power that they, in reality, do not have. That is, it is not at all guaranteed that patients will control whether or not providers and healthcare institutions will adopt medical AI.

[285] See LaRosa & Danks, supra note 281, at 214 (proposing that until "healthcare AI is accepted as 'standard of care,' the doctor must provide the alternative of a human performing the assigned task or function").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

2. Race and Algorithmic Aversion

It would be surprising if race was *not* a fault line marking differences in perceptions of medical AI. Interestingly, the empirical research on this question is scant.[286] The few existing studies demonstrate an abiding skepticism among black people about AI. In one qualitative research study, black participants expressed their sense that algorithmic bias is a problem.[287] Many of the participants connected the racially discriminatory results that AI systems might generate with racial discrimination in society more broadly.[288] Essentially, these participants were skeptical that AI technologies would be immune from the antiblackness that the country, in a variety of ways, demonstrates to them.

Another study investigated the specific question of black people's perceptions of medical AI.[289] Recognizing that black people, as a group, exhibit more mistrust of doctors and healthcare institutions than their nonblack counterparts, this study asked: Given black people's mistrust of *human* decision-makers in the medical context, are they more willing to trust *machine* decision-makers?[290] The results were illuminating. They revealed that people with lower levels of medical mistrust were more likely to prefer human decision-makers over AI.[291] However, people with higher levels of medical mistrust did not necessarily prefer machine decision-makers over human decision-makers; they did not believe that AI was *more* accurate and trustworthy. Instead, people with higher levels of medical mistrust believed that machine decision-makers were *as*

[286] See Woodruff et al., supra note 226, at 3 ("[V]ery little research has explored understandings of algorithmic (un)fairness, and there is currently little insight into how the general public and in particular people affected by algorithmic unfairness might perceive it."). Woodruff and her coauthors report the results of a study that stands as a "rare exception," which examines perceptions of algorithmic bias among "a broad population in the US, including a near-census representative panel . . . ." Id. This study sought to investigate participants' reactions to algorithms that direct targeted ads to different racial groups. See id. It found that "almost half of the respondents viewed the scenarios as a moderate or severe problem, with Black respondents finding them to be of higher severity." Id.

[287] See id. at 5.

[288] See id. at 7 (stating that some participants considered "algorithmic unfairness [to be] a modern incarnation of familiar forms of discrimination, an unwelcome extension of offline discrimination into the online arena").

[289] See generally Min Kyung Lee & Kate Rich, Who Is Included in Human Perceptions of AI?: Trust and Perceived Fairness Around Healthcare AI and Cultural Mistrust (Proc. 2021 CHI Conf. on Hum. Factors in Computing Systems, May 2021), http://minlee.net/materials/Publication/2021-CHI-AIInclusion.pdf [https://perma.cc/A3BG-CY3B].

[290] See id. at 1.

[291] See id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

untrustworthy as their human counterparts.[292] Because the black participants in the study were more likely to have higher levels of medical mistrust,[293] the study might stand for the proposition that black people are more likely to be agnostic about the use of AI in their healthcare. Some may take the study as support for the claim that disclosing the use of medical AI is likely unnecessary for everyone, but *definitely* unnecessary for black people—who, because of their mistrust of medicine and healthcare institutions as a general matter, consider medical AI just as likely to be biased as their human physicians.

However, this conclusion would be erroneous. The study goes on to show that patients with higher levels of medical mistrust *wanted* more information about the technology.[294] Relevantly, black participants with higher levels of mistrust specifically wanted information about the accuracy of the technology with respect to black people. As one black participant pointedly queried, "'[W]hat was the demographic of the people tested on because for a variety of reasons, generally speaking, any kind of AI or computer systems tend to be biased [against] Black people.'"[295] Indeed, the researchers were led to propose that "building trust with Black communities with high [medical] mistrust requires . . . explicitly demonstrating how equitable the process of building their product is. This reveals the need for a scaffolding interface for healthcare AI information . . . ."[296] That is, the researchers take their study's results as a reason to give *more* information to patients with higher levels of medical mistrust. This Article proposes that an appropriate site for the dissemination of information about a medical AI system is the informed consent process. It proposes that not only should physicians disclose their use of medical AI to the patient, but that they should also inform patients about how and why algorithmic bias might manifest in the prediction that the medical AI has generated.

[292] See id. ("Participants with low mistrust trusted human decisions more than algorithmic decisions and regarded them as fairer. However, participants with high mistrust in human systems perceived algorithmic and human decisions to be equally trustworthy and fair.").

[293] See id. at 1, 4.

[294] See id. at 7 ("[P]articipants with higher mistrust sought out more details for each information type, particularly privacy, length of use, and accuracy specific to Black people or different skin colors."). The medical AI about which the researchers sought the participants' perceptions was a dermatological tool used to diagnose skin cancer. See id. at 5.

[295] Id. at 6.

[296] Id. at 9. The researchers also propose the need for "research on healthcare AI-specific guidelines for the regulation of training data sets that can gain people's trust." Id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Cohen and Professor Graver have observed that "distributional effects and the need to consider them in designing an analytics system are inevitable."[297] They continue: "both the status quo and the post-implementation worlds involve winners and losers and the redistributions of gains and losses is a moral matter that needs to be thought about and decisions on redistribution need to be defended."[298] The thing is: historical and current events have taught people of color that harms to them often are tolerated or dismissed. Indeed, there is ample evidence that society is perfectly content when the losers happen to be people of color—when losses are distributed along familiar racial lines. It seems a bit cruel for us, as a society, to make a conscious decision *not* to inform people of color that new technologies are being introduced that might enhance health and healthcare but are just as likely to reiterate the health inequities that run rampant in society. It seems like the *least* that we can do is to ensure that people of color receive this information before they endeavor to consent to healthcare.

B. The Traditional Story of Informed Consent

The doctrine of informed consent regulates the doctor-patient relationship. It is intended to protect the patient's autonomy by ensuring that the medical interventions that are done to them only take place after they have explicitly authorized them.

The traditional story of informed consent offers that the requirement that providers acquire the informed consent of their patients before performing any medical interventions on them began to develop as society's ideas around the physician's role and responsibility vis-à-vis the patient shifted. For much of the nation's history, society largely believed that the physician's duty of beneficence towards the patient required the physician to direct the patient's care—without the patient's input or involvement.[299] The idea was that the physician knows much more about medicine and the patient's condition than does the patient.[300] Because of

[297] Cohen & Graver, supra note 79, at 463.

[298] Id. at 464.

[299] See Benjamin Moulton & Jaime S. King, Aligning Ethics with Medical Decision-Making: The Quest for Informed Patient Choice, 38 J.L. Med. & Ethics 85, 86 (2010) (stating that physicians' duty of beneficence was understood as requiring them to act to the patient's medical benefit).

[300] See id. at 86.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

the physician's superior knowledge, he was in charge.[301] The physician's fiduciary duty towards the patient required the physician to do what was best for the patient—with "best" being understood as the most advisable course of action according to the physician's judgment.[302] The patient's responsibility was simply to do what his doctor told him to do.

This paternalistic model of the doctor-patient relationship transformed over time into the currently dominant one—one that foregrounds the self-determination of the patient.[303] In this transformed model, the patient possesses the right to decide what third parties will or will not do to their body.[304] In order to exercise that right of self-determination, the physician has to supply the patient with the knowledge that they need to direct their medical care.[305] This is to say that the autonomy model recognizes that the physician-patient relationship is one in which there is a disparity in knowledge between the parties.[306] However, instead of simply ceding to the doctor's authority over medical decision-making in light of that disparity, the autonomy model requires the doctor to reduce that disparity

[301] McClellan et al., supra note 43, at 365 (describing a physician "culture" that embraced the "belief in a hierarchy that placed the power of deciding what to do about medical care rightfully in the hands of licensed physicians").

[302] See Haupt, supra note 121, at 64 (explaining that providers have fiduciary duties to their patients that require them "to act in the interests" of their patients); Moulton & King, supra note 299, at 86 (describing the belief that "physicians bore the responsibility of acting as agents for their patients, determining the best treatment options to fulfill the singular goal of improved health").

[303] See McClellan et al., supra note 43, at 365 ("Law changed the power positions, and eventually the culture of medicine transformed itself, albeit begrudgingly, to accept the concept of patient autonomy that gave the patient the ultimate authority to decide what medical care, if any, he or she wants."); Moulton & King, supra note 299, at 85, 87 (stating that "medical ethicists have shifted from guiding physicians to focus on beneficence and improving patient health as emphasized in the Hippocratic oath toward a more subjective and 'patient-centered' practice, which also prioritizes patient autonomy in medical decision-making" and describing this autonomy-protecting model as "the dominant and controlling principle in both informed consent law and medical ethics").

[304] See Moulton & King, supra note 299, at 86–87 ("In the late 20th century, respect for autonomy began to supersede beneficence and paternalism as the leading medical decision-making paradigm due to strong arguments . . . that patients should determine what happened to their bodies, as they ultimately had to experience the consequences of the chosen treatment decision.").

[305] See McClellan et al., supra note 43, at 366–67 (noting healthcare professionals' duty to "educate" the patient so that "he or she will be in a position to make appropriate health care decisions").

[306] See Barbara A. Noah, The Invisible Patient, 2002 U. Ill. L. Rev. 121, 144 (reviewing Sally Satel, How Political Correctness Is Corrupting Medicine (2000)) ("Physicians and patients nearly always have unequal levels of medical knowledge.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

by providing the patient with the information that they need to understand their options when it comes to their healthcare.[307] The doctrine of informed consent manages the requisite information sharing.[308]

The law initially understood a doctor's failure to acquire informed consent before a medical intervention as a species of battery—an unauthorized touching.[309] Providers could be liable for a battery if they did not obtain the patient's informed consent, if they obtained consent for a procedure that was significantly different from the one that they ultimately performed, or if they did not sufficiently inform the patient about all of the possible consequences of the procedure.[310] Patients did not need to prove that the unauthorized touching—that is, the treatment to which the patient did not consent or fully understand—resulted in a physical harm.[311] The law understood the unsanctioned contact as harming the patient's dignity, and it allowed patients to recover damages in light of that dignitary injury.[312]

Over time, many jurisdictions reconceptualized the theory of liability underlying an action in tort for a failure to obtain informed consent. These jurisdictions, which now constitute the majority, propose that the intrusion is not a battery, but rather a species of negligence.[313] That is, the physician's obligation to exercise due care with respect to the patient requires them to obtain the patient's informed consent.[314] The failure to

[307] See id. at 144–45 ("The informed consent process is designed to reduce this disparity of medical knowledge by mandating the provision of sufficient information to the patient to allow participation in making choices about health care."); Nadia N. Sawicki, Modernizing Informed Consent: Expanding the Boundaries of Materiality, 2016 U. Ill. L. Rev. 821, 870 ("Informed consent is a common law doctrine grounded in the ethical obligations of medical professionals to correct for the information imbalance between patients and physicians.").

[308] The law also carves out a couple of exceptions whereby the physician does not need to obtain the patient's informed consent before a medical intervention. See Peter H. Schuck, Rethinking Informed Consent, 103 Yale L.J. 899, 919 (1994) (explaining that the physician does not have to obtain informed consent when (1) it would harm the patient to provide them with information about their condition, (2) the patient is incompetent, (3) there is an emergency, (4) the risks of the treatment are common knowledge, (5) the risk is remote, or (6) the physician is reasonably unaware of a material risk).

[309] See Ruth R. Faden & Tom L. Beauchamp, A History and Theory of Informed Consent 26 (1986).

[310] See id. at 28.

[311] See id. at 27.

[312] See id.

[313] Alina-Emilia Ciortea, What Medical Risks Should Physicians Disclose to Their Patients? Towards a Better Standard in American and French Medical Malpractice Law, 10 J. Civ. L. Stud. 173, 181 (2017).

[314] See Faden et al., supra note 309, at 28–29.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

obtain informed consent is a failure to exercise due care, thereby making the physician negligent.[315] While causes of action for failure to obtain informed consent that sound in battery do not require patients to suffer a physical harm, those that sound in negligence do.[316]

At its simplest, the law requires physicians to disclose the risks of a proposed treatment, the benefits of that treatment, and alternatives to that treatment—including non-treatment and the consequences thereof.[317] However, it is clear that the physician does not need to disclose every single fact relevant to the risks, benefits, and alternatives to an identified treatment. Instead, the law requires that the physician disclose only those facts that are *material*.[318] Notably, jurisdictions have embraced different standards of materiality. Some jurisdictions propose that materiality should be judged from the physician's perspective, asking whether a reasonable physician would deem the fact material.[319] Other jurisdictions propose that materiality should be judged from the patient's perspective, asking whether a reasonable patient would deem the fact material.[320] According to the reasonable patient standard, a "risk is thus material when

[315] See id. at 29.

[316] See Noah, supra note 306, at 146 (noting the requirement that "the patient must suffer physical harm in order to receive compensation for a negligent failure to provide informed consent"). Of course, requiring patients to show physical harm in order to recover for a negligent failure to provide informed consent may be incompatible with the tort's purpose of protecting an individual's dignity. See Aart Hendricks, Personal Autonomy, Good Care, Informed Consent and Human Dignity—Some Reflections from a European Perspective, 28 Med. & L. 469, 475 (2009). That is, the informed consent process is supposed to guard against the harm to dignity that a person suffers when their physician performs a procedure to which they did not consent. See Noah, supra note 306, at 146 (noting that "commentators have suggested that nondisclosure of risks or treatment alternatives represents a similar type of dignitary harm that deserves compensation, even in the absence of a negative health outcome"); James L. Bernat & Lynn M. Peterson, Patient-Centered Informed Consent in Surgical Practice, 141 Archives Surgery 86, 87 (2006). That dignitary harm may occur even in the absence of a physical harm. See Noah, supra note 306, at 146. Moreover, the "law compensates other types of intangible harms, such as emotional distress and breach of privacy." Id.

[317] See Schuck, supra note 308, at 921.

[318] See Sawicki, supra note 307, at 827.

[319] See Jill Wieber Lens, Medical Paternalism, Stillbirth & Blindsided Mothers, 106 Iowa L. Rev. 665, 678–79 (2021).

[320] See Faden et al., supra note 309, at 32. The reasonable patient standard was first articulated in *Canterbury v. Spence*, which involved a physician who failed to inform a patient that the operation that he was to undergo carried a risk of paralysis. 464 F.2d 772, 776–77 (D.C. Cir. 1972). After the patient was, indeed, paralyzed after the operation, he sued. See id. at 778. The court found in his favor, establishing the reasonable patient standard along the way. See id. at 779, 787.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

a reasonable person . . . would be likely to attach significance to the risk or cluster or risks in deciding whether or not to forego the proposed therapy."[321]

Most observers of informed consent report that the doctrine in action is wildly different from what medical ethicists and lawyers had hoped it would be. In practice, informed consent does not look like a dialogue between provider and patient, with the provider sharing information that reduces the knowledge disparity between the two parties and the patient becoming an empowered and active agent in their healthcare.[322] Instead, most attest that the informed consent process typically is stunted and formulaic, with a patient signing an informed consent form containing information that they have not read or do not fully understand.[323] As law professor Peter Schuck explains, when one sees "informed consent in action," one sees that "[m]any physicians discuss risk in a more or less perfunctory manner and without much regard to how well the patient comprehends the information. Many patients appear to understand little of the risk information and, shortly after the discussion, to recall even less."[324]

There are scores of observers who would like to reform the informed consent process and the law around it. These commentators believe that the process should be a true conversation between the provider and their patient—a discussion during which information is exchanged, mutual

[321] *Spence*, 464 F.2d at 787 (citation omitted).

[322] See Schuck, supra note 308, at 933 ("Studies indicate that patients rarely initiate a dialogue by asking questions, that physicians often discourage genuinely dialogic, open-ended discourse about treatment, and that patients seldom ask follow-up questions. Physicians discourage active, give-and-take dialogue.").

[323] See Moulton & King, supra note 299, at 90 ("[L]egal informed consent requirements have reduced the medical disclosures in many hospital settings to obtaining a patient's signature on a written form, which is rarely read and even less frequently understood."); Schuck, supra note 308, at 934 (stating that "anecdotal and social science evidence alike demonstrate that informed consent law in action is often ritualistic, formalistic, and hollow"); Campbell, supra note 46, at 68–69 (describing informed consent as "illusory" and "vacuous" on account of physician paternalism and patients' desire to be "good" patients "by placing deferential trust in physicians").

[324] Schuck, supra note 308, at 948.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

understanding is achieved, and the two (or more)[325] parties jointly decide the course of treatment that they will pursue.[326]

The introduction of medical AI presents an opportunity to reconsider our practices around informed consent. In truth, given the well-documented limitations of our existing rituals around informed consent, we ought to have reconsidered them long before this socio-technical moment. But, the launching of technologies into the clinical encounter that *likely encode the health inequities that are pervasive in our society* presents us with a marvelous occasion for interrogating—and revamping—our informed consent practices.

This Article proposes that as part of the informed consent process, providers should inform patients that they have relied on or consulted with a medical AI device or system that, given a social context rife with racial disparities in health and health outcomes, likely has inferior predictive accuracy vis-à-vis patients of color. The most assured consequence of this reform is that patients of color will ask questions about the technology, prompting physicians to, hopefully, answer those questions.[327] This

[325] See Dayna Bowen Matthew, Race, Religion, and Informed Consent—Lessons from Social Science, 36 J.L. Med. & Ethics 150, 156 (2008) (stating that "[m]any minority patients want to involve different decision makers than most majority patients wish to include, and more than medical providers typically allow").

[326] See Lisa C. Ikemoto, Racial Disparities in Health Care and Cultural Competency, 48 St. Louis U. L.J. 75, 128 (2003) (lamenting that the law focuses on the "content of the physician's risk disclosures," which "fails to provide a place for substantive patient participation," thereby reducing the patient's role to simply saying "yes or no"); Campbell, supra note 46, at 65–66 ("When operationalized, [informed consent] is a *process* of communication between provider and patient that produces either affirmative consent or refusal. . . . [I]nformed consent is more than simply signing a consent form—it is the 'mutual sharing of information' between the clinician and patient to facilitate the patient's active engagement in their treatment."). Professor Schuck has called those who would reform the informed consent process to make it more conversational "idealists." Schuck, supra note 308, at 903. He writes that these

> idealists emphasize the *qualitative* dimension of physician-patient interactions concerning treatment decisions. They insist that these interactions be dialogic rather than authoritative, tailored to the individual patient's emotional needs and cognitive capacities rather than formulaic, aimed at maximizing patient autonomy and comprehension rather than mere information flow, and sensitive to the distortions that can be created by power differentials between physician and patient.

Id.

[327] See Rebecca Robbins & Erin Brodwin, An Invisible Hand: Patients Aren't Being Told About the AI Systems Advising Their Care, Stat News (July 15, 2020), https://www.statnews.com/2020/07/15/artificial-intelligence-patient-consent-hospitals/ [https://perma.cc/6654-2TTV] (stating that many of the physicians interviewed for the story reported not wanting to inform patients about their use of medical AI because of concerns that "mentioning AI would derail a conversation" and quoting a patient who, when told about this reason for

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

creates the conditions for a patient-provider *dialogue*—the structure that informed consent, in its best form, takes. If the informed consent process assumed the shape that it would assume in its most perfected iteration, it would be much more likely to accomplish what theorists and ethicists initially had hoped that it would accomplish, i.e., reduce the knowledge disparity between provider and patient, and protect patient autonomy.

Now, undoubtedly, some will oppose transforming the informed consent process from a *pro forma* routine into a meaningful dialogue because a dialogue will require more time from physicians—and thereby cost money.[328] But, there is a compelling argument that part of the reason the U.S. healthcare system is as dysfunctional as it is—the United States has the highest healthcare spending among industrialized nations, but poorer health outcomes[329]—is that the United States has allowed profit maximization to guide the decisions that we make around the structure of our healthcare system and the practice of medicine. If so, perhaps we should not permit the impulse toward profit maximization to lead us into staying committed to a debased informed consent process. Perhaps we should permit ourselves to allow the informed consent process to assume the form that it ought to take if it is to achieve its valuable intended aims.[330]

Importantly, the traditional story of informed consent, described in this section, gives only a partial, somewhat sterile history of the development of the practice of informed consent. In truth, we can locate informed consent's origins not in a generalized dissatisfaction with physician paternalism that ripened in the 1970s, but rather several decades earlier—

physicians' objection to notifying patients, said, "'Woe to the professional that as you introduce a topic, a patient might actually ask questions and you have to answer them'").

[328] See Nan D. Hunter, Rights Talk and Patient Subjectivity: The Role of Autonomy, Equality, and Participation Norms, 45 Wake Forest L. Rev. 1525, 1536 (2010) (noting that "the economics of medical practice seem increasingly likely to impede the expansion of the informed-consent process as an ingredient in doctor-patient interaction").

[329] Munira Z. Gunja, Evan D. Gumas & Reginald D. Williams II, U.S. Health Care from a Global Perspective, 2022: Accelerating Spending, Worsening Outcomes, Commonwealth Fund (Jan. 31, 2023), https://www.commonwealthfund.org/publications/issue-briefs/2023/jan/us-health-care-global-perspective-2022 [https://perma.cc/SCB8-SUTT].

[330] That said, it is important to keep in mind that part of hospitals' and healthcare systems' interest in medical AI is that it may cut costs. See Narendra N. Khanna et al., Economics of Artificial Intelligence in Healthcare: Diagnosis vs. Treatment, 10 Healthcare, Dec. 9, 2022, at 2–3, https://www.mdpi.com/2227-9032/10/12/2493 [https://perma.cc/6XRP-NF95]. If so, then the costs attendant to extending the informed consent process in light of medical AI-inspired dialogues might be offset by the money saved from introducing medical AI in the first instance.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

in 1947, to be precise.[331] That is, the bioethical commitment to informed consent emerged at the end of World War II during the Nuremberg Trials, at which the United States, the United Kingdom, and the Soviet Union prosecuted several political, military, and economic leaders who were responsible for the crimes and atrocities committed in Nazi Germany.[332] This fuller history reveals that informed consent emerged as a rebuke of Nazi "medicine," eugenics, antisemitism, racism, and white supremacy. As such, this more complete history unearths a certain *radicalness* embedded in the practice of informed consent—a radicalness that might still remain. The proposal that this Article advances—healthcare providers should notify patients during the informed consent process that they have relied on or consulted with medical AI when making decisions and recommendations around patients' healthcare—seeks to excavate this intrinsic radicalness and bring it to the fore. There is a possibility that when patients of color are told that inequity likely has been baked into the machinery of their healthcare, it will breed discontent among these patients with the way things are. Discontent might inspire the development of social movements. And social movements have always been the condition precedent for social transformation. To put it simply, the instant proposal endeavors to revive the rebellious roots of the doctrine of informed consent and use them as a technique of political mobilization.

C. The Radical History of Informed Consent

While many scholars providing a history of informed consent would begin the story in the 1970s, other scholars have observed that a more comprehensive history would begin three decades earlier, during the Nuremberg Doctors' Trials.[333] These trials involved the prosecution of

[331] George J. Annas, The Legacy of the Nuremberg Doctors' Trial to American Bioethics and Human Rights, 10 Minn. J.L. Sci. & Tech. 19, 26 (2009).

[332] David P. Stewart, A Brief History of International Criminal Law, *in* International Criminal Law in a Nutshell 31–34 (2014).

[333] Annas, supra note 331, at 23 ("Although the World War II origin of American bioethics is easier to see at the beginning of the twenty-first century, mainstream bioethics historians, while acknowledging the Nuremberg Doctors' Trial and the Nuremberg Code as important historical events, continue to prefer seeing American bioethics as a 1960s and 1970s response to medical paternalism."); Michael J. Malinowski, Choosing the Genetic Makeup of Children: Our Eugenics Past—Present, and Future?, 36 Conn. L. Rev. 125, 149–50 (2003) (arguing that it was during the Doctors' Trial at Nuremberg that "informed consent evolved from a principle into law, eventually becoming the cornerstone ethos for contemporary medical research and

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

twenty-three physicians and researchers for the inhumane, torturous, and ultimately fatal experiments that they conducted on victims in Nazi concentration camps.[334] The Doctors' Trials resulted in the Nuremberg Code—"ten principles developed to judge the actions of the Nazi doctors in the absence of preexisting codified guidance."[335] Importantly, the first precept of the Code was that "there must be no research on human subjects without their voluntary, informed consent."[336]

The reason that most scholars adhere to the traditional story of informed consent—and, consequently, ignore the role of the Nuremberg Trials in the development of the modern doctrine of informed consent—is likely that they draw a distinction between the regulation of physicians as they provide healthcare to their patients (to which the traditional story speaks more directly) and the regulation of researchers as they conduct experiments on research subjects (to which the Nuremberg Trials most directly relate).[337] However, law professor and bioethicist George Annas rejects the tendency to conceptualize these two contexts as fundamentally disconnected.[338] He argues that while the traditional story proposes that informed consent emerged as a response to physician "paternalism" in the 1960s and 1970s—and while the language of "paternalism" fails to capture the *evil* that Nazi doctors inflicted on Jewish and other prisoners during the Holocaust—a thread nevertheless connects the two: as paternalistic physicians in the 1960s and 1970s enjoyed an

the practice of medicine"); Hunter, supra note 328, at 1530 ("The initial cultural groundwork for informed consent was laid by public revulsion against experiments conducted by Nazi physicians, and its legal codification derived from the Nuremberg Code's prohibition against using human beings as research subjects without their consent.").

[334] See Annas, supra note 331, at 21 (noting that some of the "most infamous" experiments that Nazi physicians performed "resulted in the planned death of the research subjects"); Malinowski, supra note 333, at 155–56 (describing some of the most gruesome of the experiments and noting that "death was the target endpoint of these studies, and in fact, surviving research subjects were sent to gas chambers as soon as studies were deemed complete").

[335] Malinowski, supra note 333, at 161.

[336] Id.; Annas, supra note 331, at 26 ("The first response of the American judges to the horror of the Nazi doctors was to articulate, in the first precept of the Nuremberg Code, the doctrine of informed consent.").

[337] It is not unreasonable for scholars and observers to separate healthcare provision and research conceptually, of course. Indeed, the Belmont Report, which was issued after the country learned about the Tuskegee Experiment and articulates ethical guidelines for the conduct of research on human subjects, establishes a "guideline for distinguishing biomedical research and the practice of medicine." See Malinowski, supra note 333, at 165.

[338] Annas, supra note 331, at 26.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

"unaccountable power," Nazi doctors were also "unaccountable in the exercise of their power over their subjects."[339] Thus, he argues that it is "unreasonable" to separate informed consent as a response to the horrors of the Holocaust in the 1940s from informed consent as a response to physician paternalism in the 1970s.[340] He concludes: "The modern doctrine of informed consent was not born either of U.S. health law in 1972, or of American bioethics shortly thereafter, but at Nuremberg in 1947."[341]

It is important to understand that the problem that the Doctors' Trials at Nuremberg identified was not simply that the Nazi researchers failed to inform their research subjects about the risks and benefits of the experiments and failed to get the participants' consent before torturing them to death. Instead, one might understand that the Nuremberg Trials rebuked Nazi medicine and Nazi "science" in their entirety. The trials rejected eugenics, a pseudoscience that proposes that genes determine all manner of social outcomes[342]—a pseudoscience whose logic may inexorably lead to genocidal attempts to cleanse populations of "problematic" genetic strains. The trials rejected the racism that defined Jewish people as a race of people that was entirely separate from, and woefully inferior to, the "Aryan race."[343] The trials rejected the white supremacy in the proposition that a small segment of the white race was superior to all those who did not belong to that narrow slice of humanity.[344] If we are correct in concluding that Nuremberg is the birthplace of the modern practice of informed consent, then it is no stretch to understand the modern practice of informed consent as, similarly, a rejection of antisemitism, racism, and white supremacy.

When informed consent is framed as a denunciation of antisemitism, racism, and white supremacy, the answer to the question of whether patients—particularly patients of color—should be notified that their provider is managing their healthcare with a tool that likely encodes racism and racial inequality seems apparent: *of course* patients should be notified. Denying patients information about the possibility—indeed, the

[339] Id.
[340] Id.
[341] Id.
[342] Khiara M. Bridges, White Privilege and White Disadvantage, 105 Va. L. Rev 449, 462 (2019).
[343] See Richard D. Heideman, Legalizing Hate: The Significance of the Nuremberg Laws and the Post-War Nuremberg Trials, 39 Loy. L.A. Int'l & Comp. L. Rev. 5, 22 (2017).
[344] See id.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

likelihood—that their health is being managed by a technology that has racial disenfranchisement embedded in its cogs and wheels obscures the racial inequality from which the technology emerged. Once obscured, it is easy for the racial inequality to be normalized—and then perpetuated.

To be clear, the instant proposal is *not* motivated by the hope that if the informed consent process were reformed to require physicians to inform patients that they are using a technology that likely encodes racial inequality, it would make it easier for patients to sue in tort if the technology ultimately causes an injury to them.[345] Neither is the instant proposal primarily motivated by the expectation that including this information in the informed consent process will lead to patients of color refusing to follow the AI-influenced recommendations of their providers.[346] Neither is the instant proposal principally motivated by the desire for patients of color to demand that their providers answer counterfactuals: what kind of care would the provider recommend for the patient if the provider did not know what the medical AI predicted? While all of these possible consequences might result from the adoption of the instant proposal—and while they might all be welcomed—it is not the proposal's immediate aim to bring them about. Instead, the instant proposal is motivated by a recognition of the relationship between the visibility of injustice and social movements. When injustice becomes visible, social movements, at times, form to demand an end to the injustice.[347] If patients are forced to confront the facts of racial injustice *as it affects them* during their doctors' visits, it creates the condition of possibility for the formation of a social movement that will demand an

[345] Plaintiffs who sue providers for failing to obtain informed consent are infrequently successful. See Hunter, supra note 328, at 1536 (observing that informed consent's "utility as a remedy is questionable because of the difficulty of proving causation").

[346] Law professor Nan Hunter similarly recognizes that the value of informed consent is not that it provides patients with an opportunity to refuse treatment. Hunter, supra note 328, at 1538–39 (observing that the "ostensible purpose of the informed-consent requirement is [that it] provide[s] the patient with a right to exit, i.e., a right to determine whether to proceed with treatment that the doctor recommends," but recognizing that the "utility in securing the right to exit may be more aspirational than real, because patients often rely on physicians for the expertise that determines their decisions"). In her view, informed consent's value is that it encourages physicians to communicate with their patients. Id. at 1539 (arguing that the "primary function [of informed consent] is less about exit than about voice"). Informed consent may also derive its value from its ability to make patients aware about facts about their society that may prompt them to demand a reorganization of that society.

[347] See Sean Illing, How Black Lives Matter Fits into the Long History of American Radicalism, Vox (July 2, 2020, 8:50 AM), https://www.vox.com/policy-and-politics/2130677 1/black-lives-matter-george-floyd-protest-michael-kazin [https://perma.cc/HQU7-CSQW].

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

end to the myriad practices that have made it so that we can expect racial inequality to be built into medical AI. Informed consent, then, can be a tool of political mobilization.

It is not naïve to identify a connection between social movements and informed consent. As law professor Nan Hunter has explained, although "[t]he legal literature on informed consent and the literature on rights-oriented social movements are seldom in conversation," the two contexts are not distinct worlds.[348] Indeed, some social movements have identified informed consent as the change that they want to see in the world. She observes that while physician paternalism in the 1960s and 1970s prompted interest in the doctrine of informed consent, patients were not equally positioned vis-à-vis this paternalism. The form that physician paternalism took often differed along lines of patient race and gender.[349] She writes, "[p]hysician disrespect of patients had long been exacerbated by race and gender, and equality movements of the mid-twentieth century included these issues as part of their agendas."[350] Indeed, "[t]he women's and racial justice movements were especially significant in the move toward recognition of patient-autonomy rights."[351]

This Article offers that the relationship between social movements and informed consent practices can be bidirectional. While Hunter observes

[348] Hunter, supra note 328, at 1529; see also Faden et al., supra note 309, at 86–87 (noting that informed consent "first appeared as an issue in American medicine in the late 1950s and early 1960s" and proposing that "it seems likely that increased legal interest in the right of self-determination and increased philosophical interest in the principle of respect for autonomy and individualism were but instances of the new rights orientation that various social movements . . . introduced into society").

[349] We can safely assume that physician paternalism also uniquely harmed patients unprivileged along lines of class and gender identity, among others.

[350] Hunter, supra note 328, at 1531.

[351] Id. at 1530–31. Hunter notes that feminists took an interest in informed consent in light of the sexism that had become normalized in the practice of medicine. Indeed, performing medical interventions on women without telling them about the intervention was the norm. She states that included in "[t]he catalog of deprivations of patient autonomy that motivated" the feminist investment in informed consent were "Cesarean sections for physician convenience and sterilization and mastectomy without specific consent." Id. at 1532. Indeed, *Schloendorff v. Society of New York Hospital*, a landmark case that helped to establish the principle that physicians would be liable in tort if they failed to obtain their patients' consent prior to a medical intervention—a case that provided Justice Benjamin Cardozo with the opportunity to write his famous dictum, "[e]very human being of adult years and sound mind has a right to determine what shall be done with his own body"—involved medical misogyny. 105 N.E. 92, 93 (N.Y. 1914). Hunter observes that the case arose after a physician performed a hysterectomy on his patient without her consent after deciding that she "was 'too nervous' for a vaginal examination." Hunter, supra note 328, at 1531.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

that social movements have prompted the development of robust practices of informed consent, this Article proposes that robust practices of informed consent can prompt the development of social movements. If patients of color are forced to encounter facts during the informed consent process about the ubiquity of racial disenfranchisement and how that disenfranchisement affects their health and healthcare, it creates the possibility that they will organize to demand an end to it.

D. Informed Consent as One of Many Efforts to Eliminate Algorithmic Bias in Medical AI

The instant proposal has a radical spirit. It is radical insofar as it recognizes that the phenomena that make it reasonable to expect that medical AI will provide inferior healthcare to patients of color—i.e., choices in model design, unrepresentative training data, racially inequitable healthcare, the myth of biological race—are complex. Most of these phenomena are structural in nature, historically rooted, and wildly durable. The instant proposal proceeds from the hope that if patients of color are notified about how these structural, systemic, huge, large-scale, abstracted issues impact the immediate, material healthcare that they are receiving in their doctor's office or hospital bed, they might be inspired to demand better from this country. This is the radical spirit behind the instant proposal.

At the same time, the instant proposal is meek. Revising the informed consent process in the manner proposed will not immunize programmers from making choices in algorithmic model design that perpetuate existing inequities in healthcare. The instant proposal will not guarantee that the data on which medical AI is trained is representative in ways that ensure the accuracy of the technology's prediction. It will not prevent providers from exercising their discretion in ways that lead them to provide healthcare to their patients of color that differs from that provided to their white patients. It will not cause the hegemonic reign of the myth of biological race to finally come to an end—centuries after the myth's creation.

This is to say that the instant proposal, in and of itself, will not eliminate the problem of algorithmic bias in medical AI. Instead, it may create the conditions under which the problem will be eliminated. It may inspire the political mobilization that will transform society in ways that will finally end the deep-seated racial inequality that medical AI simply replicates and perpetuates.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Thus, adopting the instant proposal cannot be the only thing that we do to protect patients of color from racially biased medical AI. In fact, we ought to pursue multiple strategies at once. Other efforts that we ought to pursue simultaneously to a reform of the informed consent process are numerous:

- During medical school, physicians should not only be disabused of the myth of biological race, but they should also be educated about the structures and processes that make it so that people of color in the United States receive healthcare that is inferior to their white counterparts (i.e., the two-tiered healthcare system and residential segregation that often leads to racially segregated hospitals). They should learn how medical AI may replicate these inequities and fail to provide quality care to their future patients of color. Further, they should learn how to evaluate the quality of the medical AI that they may choose to deploy in their practices—a task that many observers have argued physicians should undertake whenever they rely on or consult with a medical AI device for patient care.[352]
- Medical societies, like the American College of Obstetrics and Gynecology, should evaluate the quality of medical AI devices that physicians might use to manage pregnancy, childbirth, and the postpartum period, and they should publish assessments of these devices' predictive accuracy vis-à-vis patients of color.[353]

[352] See W. Nicholson Price II, Sara Gerke & I. Glenn Cohen, Potential Liability for Physicians Using Artificial Intelligence, 322 JAMA 1765, 1766 (2019) (arguing that "physicians should learn how to better use and interpret AI algorithms, including in what situations an available medical AI should be applied and how much confidence should be placed in an algorithmic recommendation").

Of course, in order to be able to evaluate the quality of a device, providers will need information about the studies that purport to establish the device's safety and efficacy—information that, currently, may not be readily available. See Wu et al., supra note 230, at 583 (noting that information that will assist a provider in evaluating a device, such as the number of sites in which a device was tested, must be "consistently reported in the public summary document" that the FDA publishes "in order for clinicians, researchers, and patients to make informed judgments about the reliability of the algorithm").

[353] See Price et al., supra note 352, at 1766 (arguing that medical societies should "take active steps to evaluate practice-specific algorithms" as "societies will be well placed to provide additional guidelines to evaluate AI products at implementation and to evaluate AI recommendations for individual patients").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

- The possibility of algorithmic bias should be addressed during the FDA approval process. Indeed, the FDA should not approve medical AI devices that have failed to demonstrate some degree of predictive accuracy vis-à-vis patients of color.[354] Notably, this is a regulatory effort to ensure the safety and efficacy of medical AI that the FDA is not adequately undertaking at present.[355]
- Medicare authorities should consider the possibility of algorithmic bias when determining whether to grant a certificate of medical necessity.[356] Indeed, Medicare should not cover the cost of any durable medical equipment that employs AI technology that has not demonstrated some degree of predictive accuracy vis-à-vis patients of color.
- Patients of color harmed by medical AI should be able to recover for their injuries in tort. Scholars should develop legal theories as to how injured patients might successfully sue developers of flawed medical AI, and courts should adopt these theories as law.[357] Further, legislators should pass laws that prevent

[354] Sama Kahook, Left to Their Own Devices: Addressing Racial Bias in the FDA Approval Process for Medical Devices, 30 Annals Health L. Advance Directive 153, 164 (2021) (arguing that if an AI medical device "fails to take into account a diverse data set, then it should not reach the market" and arguing that we should take the approach of preventing medical AI devices "from even passing the FDA approval process unless they adhere to incorporating diverse data sets"). Notably, Kahook argues that if the FDA nonetheless approves the device, "then prospective users should be informed of the device's potential limitations." Id.

[355] See discussion supra notes 230–35 and accompanying text.

[356] 42 U.S.C. § 1395m(j)(2)(B) ("[A] form or other document containing information required by the carrier to be submitted to show that an item is reasonable and necessary for the diagnosis or treatment of illness or injury or to improve the functioning of a malformed body member.").

[357] See, e.g., Price et al., supra note 236, at 14 (arguing that patients of color who are injured by a biased algorithm in a medical AI device might be able to bring a successful product liability suit, as "a claim for a marketing defect may be given if the labeling of the AI did not include a warning that the model may likely not give reliable/correct recommendations when used in non-Caucasian patients").

Law professors Barbara Evans and Frank Pasquale have proposed that patients of color injured by medical AI that employs a racially biased algorithm might be able to bring a design-defect product liability suit against developers. See Barbara J. Evans & Frank Pasquale, Product Liability Suits for FDA-Regulated AI / ML Software, *in* The Future of Medical Device Regulation: Innovation and Protection 22 (I. Glenn Cohen, Timo Minssen, W. Nicholson Price II, Christopher Robertson & Carmel Shachar eds., 2022); see also Price et al., supra note 236, at 14 (arguing that "a model that has not been trained on a diverse patient population should not be placed on the market in the first place and may thus also trigger a

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

> developers from shielding themselves from liability for racially biased algorithms.[358]

This is all to say that there is nothing about the instant proposal that demands that we pursue it to the exclusion of other efforts to ensure that patients of color receive quality healthcare. Given the stakes of this issue—indeed, given the fact that algorithmic bias in medical AI may be a matter of life or death—we ought to take an "all-hands-on-deck" approach. Medical AI should be regulated at all relevant levels—including at the level of the clinical encounter.

It is important to wed regulation of the informed consent process to regulation of medical AI at other levels. This is because there are dangers attendant to reforming the informed consent process to the exclusion of other efforts to address algorithmic bias. Further, there are dangers attendant to pursuing other efforts to address algorithmic bias to the exclusion of reforming the informed consent process. The next Subsections discuss these two sets of dangers in turn.

---

design defect suit"). Evans and Pasquale observe that the FDA regulates medical products, not healthcare services. Evans & Pasquale, supra, at 24. Further, they explain that Congress has authorized the FDA to regulate black-box medical AI devices that generate nontransparent recommendations that the provider cannot challenge. Id. at 27 ("CDS software that makes recommendations falls under the FDA's regulatory jurisdiction if these recommendations are not intended to be transparent to the health care professionals using the software."). Evans and Pasquale argue that these medical AI devices, as FDA-regulated products, are subject to strict liability. Id. at 35. Moreover, strict product liability requires plaintiffs to demonstrate that although a "reasonable alternative design" existed at the time of the development of the product, the developer failed to use that alternative design, instead employing the defective design that injured the plaintiff. See id. at 33. Evans and Pasquale assert that patients of color whom a medical AI device harms could show that "[a]n alternative design seemingly always exists, i.e., train the software on a larger, more appropriate, more accurate, less biased dataset that better reflects the intended patient population." Id. They conclude that "it would be left for the trier of fact to decide whether it would have been reasonable for the software developer to have used that alternative, better dataset, in view of the cost, delay, availability, and accessibility of additional data." Id.

[358] Evans and Pasquale observe that developers frequently shield themselves from lawsuits for their products by including terms in licensing agreements that transfer liability to the providers who use their products. See Evans & Pasquale, supra note 357, at 30. They observe that "[t]he result is to channel negligence claims toward providers while the software developer goes unscathed." Id. If we are interested in holding developers and manufacturers accountable for the potentially racially biased algorithms that they employ in their devices, then we might consider passing laws or pursuing judicial outcomes that would render such licensing terms unenforceable.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

1. The Perils of Informed Consent Reform in Isolation

It is perilous to pursue reform of the informed consent process in isolation from other efforts because of the legal consequences that flow from an individual having given informed consent to a medical treatment. As explained above, although informed consent was born of lofty interests in patient autonomy, self-determination, and dignity—and although informed consent emerged as a product of a commitment to antiracism and a rejection of white supremacy—the practice of informed consent today, as it is implemented in doctor's offices, hospitals, and other healthcare institutions across the country, appears to be completely divorced from its noble origins. It has become prosaic, formulaic, and perfunctory.[359] In many cases, a provider's endeavor to obtain a patient's informed consent is less about reducing the information disparity between the two parties and protecting the patient's autonomy and more about obtaining the approval that will release the provider from liability. If this is what informed consent has become, then including information about the racial inequity that likely is encoded in a medical AI device may similarly function to release providers—and developers—from liability. In releasing providers from liability, it may remove incentives for providers to do their diligence to ensure that the medical AI that they use has some degree of predictive accuracy vis-à-vis their patients of color. In releasing developers from liability, it may remove any incentives that they have to avoid encoding racial inequity in the first instance, or to test for and correct algorithmic bias before selling the product in the second instance.

This is why it is important for the instant proposal to be joined with regulation of medical AI at other levels. If reforming the informed consent process was the sole effort that we took to address algorithmic bias, it may function simply to offload responsibility onto patients. Patients would be charged with the task of protecting themselves from potentially harmful devices, although a slew of other entities—ranging from device developers to the FDA and providers—are much better positioned to identify whether there is algorithmic bias in a device. Moreover, if augmenting the information shared during the informed consent process were the only effort that society made to address the harms that medical AI might pose to people of color, it would be consistent with "a broader trend in neoliberal regulatory reform in which information regulation

[359] See supra notes 314–24 and accompanying text.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

alone is considered sufficient and regulation of substantive issues . . . is deemed invasive and unnecessary."[360] Indeed, if the instant proposal were the sole effort that we took to address algorithmic bias, it may be read as acquiescence to FDA nonaction, provider complacency, developer negligence, and the structural processes that generate algorithmic racial bias. For these reasons, reform of the informed consent process must be tethered to interventions at every possible location.

2. The Perils of Forgoing Informed Consent Reform

Although informed consent reform is not the *only* strategy that we should pursue to address racially biased algorithms in medical AI, it is an *essential* element in a multi-scale attack on algorithmic bias and the social contexts that produce them. It is perfectly reasonable to believe that social movements are the condition precedent to the societal transformation that will eradicate the structural and interpersonal inequalities that lead to algorithmic bias. The instant proposal seeks to provoke that political mobilization.

It is important to be clear about why we must insist on efforts that will eliminate structural racism *alongside* other regulatory interventions to eliminate algorithmic bias: we should be *terrified* about the possibility that algorithmic bias might be rooted out while maintaining the status quo with respect to everything else.

To explain: one fear that motivates the insistence upon informed consent reform is the possibility that medical AI will *not* work for people of color. But, there might be another, more dire fear that we ought to have: What if medical AI *works* for people of color*?* That is, what if these new technologies manage to eliminate racial disparities in maternal and infant mortality and morbidity? What if they make pregnancy safer for the vulnerable people who are so frequently felled—or who come treacherously close to being felled—on the path to motherhood? What if marginalized people's babies are born healthier? What if fewer of these babies are born preterm and with low birth weight? What if smaller numbers of them die before their first birthdays?

Benjamin has worried about the "allure of tech fixes."[361] She has written of her concern that technologies like medical AI "offer pragmatic

[360] Margot J. Pollans, Eaters, Powerless by Design, 120 Mich. L. Rev. 643, 673 (2022).
[361] Benjamin, Race After Technology, supra note 60, at 156.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

inclusion in place of political and social transformation."[362] She explains that, in many cases, these technologies "are a permanent placeholder for bolder change. Sure, some people may live longer, but they are still living in neighborhoods and schools that those in power are less motivated to spot and fix. Medical inclusion, in short, can be a lucrative stand-in for social and political justice."[363]

In truth, pregnancy is a dicey proposition for so many black people in the United States because they live in segregated neighborhoods with inferior housing stock, a cornucopia of environmental toxins, a dearth of affordable and fresh foods and vegetables, and higher rates of violent crime. Pregnancy is hazardous for scores of black people because prenatal care often is the first time in their lives that they have had regular contact with a healthcare institution and a provider with a duty to conserve their health; indeed, childbearing is a perilous attempt for black people because many of them enter pregnancy with health conditions that are undetected and unmanaged. Pregnancy is dangerous for so many black people because the physiological impacts of stress have weathered their bodies and aged their organ systems at faster rates than those of their nonblack counterparts. Pregnancy is precarious for black people because the racial segregation that is a banal feature of American life characterizes hospitals as well; moreover, the high and medium "black-serving hospitals" to which most black people turn are known to offer inferior care. Pregnancy is risky for black people because the United States has "maternity care deserts"—counties "with no hospital offering obstetric care and no OB/GYN or certified nurse midwife providers."[364] The fact of maternity care deserts make it difficult for hundreds of thousands of black people to access the prenatal care that will increase their odds of successfully navigating pregnancy and bringing home a healthy, living baby.

That is, structural racism has made pregnancy a dicey proposition for so many black people. So: What if the introduction of medical AI into prenatal care works? What if technology manages to reduce or eliminate the effects of structural racism? More people will live. But structural racism will remain intact.

[362] Id.

[363] Id. at 156–57.

[364] See Maeve Wallace et al., Maternity Care Deserts and Pregnancy-Associated Mortality in Louisiana, 31-2 Women's Health Issues 122, 122–24 (2021), https://pubmed.ncbi.nlm.nih.gov/33069560/ [https://perma.cc/25ER-UDFX].

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

The public health literature calls all of reasons listed above that explain why pregnancy is less safe for black people the "social determinants of health." These social determinants of health are the "conditions in which people are born, grow, work, live and age, and the wider set of forces and systems shaping the conditions of daily life."[365] Medical AI has the potential to improve and regularize the care provided to the bodies living under damaging and deadly social determinants of health. But, medical AI will do nothing to improve their social determinants of health in the first instance.[366]

However, improving the unhealthy environments in which marginalized people live and removing the toxic stress that marginalized people endure is precisely what justice requires. And so, medical AI threatens to improve health outcomes without achieving justice. In this way, the state, which might be the entity that bears the primary obligation to organize society so that *all* of its citizens can thrive, gets relieved of the responsibility of guaranteeing justice for all. In this way, medical AI accomplishes "a desirable neoliberal result: better living through technology."[367]

To date, the dramatic inequalities that scar our social landscape have not moved us to remedy them—even though they are causing the most vulnerable to be sicker and die earlier than others. If sickness and premature death have not compelled us to do something about structural racism, then we should have little hope that we will feel compelled to act if technology has managed to eliminate sickness and premature death. For those of us who believe in a full-throated racial justice, this is a reason to remain ever wary of technological "fixes."

And so: There is a danger involved in eradicating algorithmic racial bias in medical AI without eradicating structural racism. To the extent that reform of the informed consent process might be a lever by which we can inspire the political energies that lead to social transformation, we

[365] See Social Determinants of Health at CDC, Ctrs. for Disease Control & Prevention (Dec. 8, 2022), https://www.cdc.gov/socialdeterminants/index.htm [https://perma.cc/XUM9-7NMK].

[366] See Terry, Regulating Healthcare AI, supra note 112, at 187 (stating that the health equality question that medical AI poses is whether the improvements that these technologies will make "will accrue to all or only a section of the population," but stating that "[t]he health equity question is broader, asking whether we can reduce not just health disparities but also their determinants").

[367] Froomkin et al., supra note 89, at 36.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

have to pull that lever alongside all the other levers within our reach to address algorithmic bias in medical AI.

E. Informed Consent, Redux: Disclosing the Use of Medical AI and the Fact of Racial Disparities in Health and Healthcare

Commentators have long lamented the fact that we take a one-size-fits-all approach to informed consent.[368] The informed consent process for a wealthy, highly educated, white, thirty-three-year-old woman who is having her labor induced will be identical in all relevant respects to the informed consent process for an indigent, poorly educated, black, sixteen-year-old undergoing the same procedure. However, we have every reason to suppose that those two individuals will want and need different things from the informed consent process.[369] For one, a number of studies show that an individual's race might affect their requirements and expectations

[368] See Noah, supra note 306, at 144 (noting that "[l]ower levels of 'health literacy' may make the informed consent process" less beneficial to populations that are "less able to comprehend complex medical concepts and thus cannot participate meaningfully in making decisions about their medical care"); Schuck, supra note 308, at 957 ("[A] doctrine that treats all patients and physician-patient relationships as essentially homogenous when in fact they are not exacts a price. Specifically, the law requires a level of informed consent that is different from the level that many consumers or groups of consumers want."); McClellan et al., supra note 43, at 362 ("[P]atients enter the physician-patient relationship with widely varying values and knowledge, and, consequently, the law should not embrace a legal mandate of communication on the false premise that one size fits all.").

[369] Interestingly, studies show that context matters tremendously. While research has shown that, as a general matter, people tend to have poor recall of the information a doctor gives them during the informed consent process, this is not true when it comes to obstetrics. See Schuck, supra note 308, at 954 (describing a study that "indicates that patients retain, for long periods and with stunningly accurate recall, information that their obstetricians provide about risks, tests, and alternatives" and explaining that "[t]his finding contrasts sharply with many other studies, which reveal that patients in other contexts recall little, even just after the medical intervention"). With respect to information shared during informed consent processes concerning pregnancy, pregnant people tend to have much higher levels of information retention and understanding than others. See id. In light of the fact that context matters, Schuck proposes a "contextualized informed consent doctrine" which

> might . . . distinguish between elective and nonelective treatments, between the informed consent duties owed by a patient's family physician and those owed by a hospital-based anesthesiologist or tertiary care subspecialist who encounters the patient only fleetingly, and between treatment through a single medical event such as surgery and treatment occurring over a long period of time.

Id. at 955.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

when it comes to informed consent.[370] Several studies have shown that black patients tend to want their doctors to give them lots of information about their treatment options—even while preferring that their physicians ultimately decide which treatment option is best for them.[371]

Critiques of the one-size-fits-all nature of our current approach to informed consent are based on the recognition that patients are different from one another—and, therefore, might need different things from their encounters with healthcare providers. These critiques emerge from the acknowledgment that patients are heterogeneous when it comes to the amount of information they want from their providers, their ability to understand the potentially technical information that is given to them, their willingness to share non-medical facts about their lives with their physicians, and their desire to make autonomous decisions about their medical care.[372] However, there is another sense in which patients are different from one another—a sense that warrants an explicit discussion of race in the informed consent process. *Patients are differently situated with respect to the medical facts that physicians share with them*.

An example is illustrative: a physician in the United States discussing with a patient the medical risks involved with carrying pregnancy to term may explain that the risks are relatively low, as only 32.9 out of 100,000 live births in the United States end in the pregnant person's death each year.[373] But, patients are differently situated with respect to this statistic. This statistic will exaggerate the risk of death for a non-Hispanic white

[370] See, e.g., Alexia M. Torke, Giselle M. Corbie-Smith & William T. Branch, African American Patients' Perspectives on Medical Decision Making, 164 Archives Internal Med. 525, 526 (2004).

[371] See id. at 529 (noting that in a study of older African American patients, the participants indicated that they wanted their healthcare providers "to provide them with as much information as possible" even though they "regard[ed] the recommendation of the physician as one of the most important factors in their decision making"); Matthew, supra note 325, at 161 (describing a study that showed that "Hispanic and African-American respondents were more likely than white patients to rely on their physicians, both for gathering medical knowledge, and for treatment decisions" and noting that "[t]his ethnic disparity remained even after controlling for socio-economic and health insurance status").

[372] See, e.g., Ikemoto, supra note 326, at 127 (noting "language or other cultural differences" that may "interfere with provider-patient communication"); McClellan et al., supra note 43, at 367 (noting the "different values and personalities" that patients have, which may "produce differing needs for deferring to, digesting, relying on, or rebelling against suggestions and advice received from health care professionals").

[373] Donna L. Hoyert, Maternal Mortality Rates in the United States, 2021, Ctrs. for Disease Control & Prevention (Mar. 16, 2023), https://www.cdc.gov/nchs/data/hestat/maternal-mortality/2021/maternal-mortality-rates-2021.htm [https://perma.cc/2K6Q-LKVH].

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

patient, as only 26.2 out of 100,000 non-Hispanic white people attempting a live birth suffer a pregnancy-related death each year.[374] Meanwhile, the statistic will understate the risk of death for a non-Hispanic black patient, as 69.9 out of 100,000 non-Hispanic black people attempting a live birth suffer a pregnancy-related death each year.[375] Informed consent standards currently do not require physicians to explain medical risks apropos of one of the most health-impactful characteristics that the patient possesses—their race.[376] This is true even though studies show that some patients—namely, black patients—desire this information.[377]

In fact, it is odd that race is not *already* part of the informed consent dialogue. This is simply because physicians are very much aware of a patient's race, as medicine has long embraced the notion that race is a clinically relevant characteristic.[378] Indeed, among the very first facts about the patient that are shared with providers in a patient's care team is the patient's race (as well as their gender and age).[379] It may represent the height of dissonance for physicians to believe that race is relevant when making diagnoses and prognoses and constructing treatment plans, yet fail to mention during the informed consent process how a patient's race may have been relevant when the physician made the diagnosis and prognosis and constructed the treatment plan.[380]

---

[374] Id.

[375] Id.

[376] See generally Medical Treatment and Informed Consent State Law Survey, Lexis+ (Mar. 27, 2023), https://plus.lexis.com/api/permalink/c6ea4792-134a-4d3d-a03c-aa78a7bacbd8/?context=1530671 [https://perma.cc/B5HJ-H2JU] (failing to identify patient race as a relevant characteristic when obtaining informed consent); Schuck, supra note 308, at 917 ("Courts have not clarified whether the probabilities that the physician provides should refer to the average risk to large populations or must instead refer to the risk to the smallest group of which the patient is a member and for which the risk information exists.").

[377] See Matthew, supra note 325, at 162 (describing a study that investigated the type of information that individuals want before consenting to a prostate cancer screening and found that black participants in the study "were aware that blacks are generally at higher risk for prostate cancer and wanted risk information that expressly compared the mortality rates for black patients compared to whites").

[378] Linda M. Hunt, Nicole D. Truesdell & Meta J. Kreiner, Genes, Race, and Culture in Clinical Care: Racial Profiling in the Management of Chronic Illness, 27 Med. Anthropology Q. 253, 254 (2013).

[379] See Thomas E. Finucane, Mention of a Patient's "Race" in Clinical Presentations, 16 AMA J. Ethics 423, 423–24 (2014).

[380] I am reminded here of Professor Neil Gotanda's description of what legally mandated "color blindness" in hiring requires. See Neil Gotanda, Failure of the Color-Blind Vision: Race, Ethnicity, and the California Civil Rights Initiative, 23 Hastings Const. L.Q. 1135, 1140 (1996). He observes that although "[c]olor blindness is often described as a race-neutral

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

The introduction of medical AI into the clinical encounter presents us with an opportunity to reform the informed consent process to require providers to disclose race-related disparities as part of this process and to explain why algorithmic racial bias might have affected the predictive accuracy of the AI technology on which the provider relied.[381] Requiring disclosures of this information seems appropriate, as it may be overly optimistic to hope that the history and present of racially inequitable healthcare in the United States (and, to the extent that training data comes from countries outside of the United States, around the world) will *not* be encoded into medical AI.[382]

process," it is not neutral inasmuch as "certain characteristics were recognized, calculated, and then discounted." Id. For a person who is not medically colorblind to be colorblind as required by laws and policies that forbid the consideration of a candidate's race in hiring, a person "would first 'see' the color, then pretend that the colors could not be seen. It is the process of taking something that one knows to exist—colors—and then consciously discounting their existence." Id. Physicians' consideration of the patient's race as a clinically relevant factor, but refusal to talk about race *to* the patient during the informed consent process or otherwise, may be a quintessential example of the "color blindness" that Gotanda describes. Id.

[381] In Schuck's analysis of the doctrine of informed consent, he arrives at the conclusion that it might be appropriate for physicians to disclose disparities in the rates at which physicians in the United States and physicians in other nations perform certain interventions. See Schuck, supra note 308, at 956. He writes:

> There may be some situations in which physicians should be required to disclose not merely the existence of a reasonable nonsurgical alternative (a disclosure required under existing doctrine), but also the fact, bearing on the necessity for the procedure, that the surgery rate for it is much lower in other countries with comparable standards of health care. Notable examples of this disparity are the far higher rates of elective hysterectomies, caesarean deliveries, and tonsillectomies in the United States than in other such countries despite the failure of these higher rates to produce significantly better medical outcomes.

Id. This Article builds on this claim by observing that race-related disparities are at least as important as disparities across countries.

[382] Professor Larry J. Pittman has made a similar, but distinct, argument about informed consent. See Larry J. Pittman, A Thirteenth Amendment Challenge to Both Racial Disparities in Medical Treatments and Improper Physicians' Informed Consent Disclosures, 48 St. Louis U. L.J. 131, 186–89 (2003). He proposes that:

> [I]nformed consent law should be refined to specifically impose the obligation that physicians do three additional things before providing medical treatment: (1) inform their minority patients that there is a racial disparity between them and white patients regarding the types of treatments that the physician disproportionately provides to white patients; (2) explain the specific reasons why there is a disparity in the treatments among the different races of patients, and (3) explain the specific risks and benefits of the treatments, including, but not limited to, the risks and benefits of the specific treatments that the physician is recommending to minority patients and the specific risks and benefits of the treatments that the physician disproportionately recommends and gives to his or her white patients.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

Consider the example of a medical AI device that predicts an adverse outcome if a black patient attempts a VBAC, leading the physician to recommend that the patient undergo a C-section. Informed consent to the C-section should require not only that the physician disclose to the patient that medical AI has impacted the recommendation—a proposition to which many, but not all,[383] scholars are coming around[384]—but also that

Id. at 187. Pittman's proposal differs from the present one inasmuch as it would require a physician to notify patients of racial disparities in the rates at which the physician, personally, offers patients specific medical treatments. This Article's proposal does not require physicians to take an accounting of their own practices to determine if there are racial disparities in the rates at which they perform interventions. This is because the biases encoded in medical AI are not a product of the practices of an individual physician. Rather, they are an aggregation of the practices of thousands of physicians (alongside choices made during the algorithmic model's design and biases in the training data).

[383] See, e.g., Cohen, supra note 11, at 1446 n.51 (discussing the Obermeyer study, described above, and stating that "[o]ne can believe that this *is* a real problem, and yet also believe the law of informed consent is an unlikely place to solve it, preferring, for example, FDA or other forms of premarket review as more plausible"); Price, supra note 93, at 299 n.15 (noting that when considering the question of informed consent with respect to black-box medical AI, "it is hard to imagine precisely what 'informed' means in the context of a recommendation where no one knows exactly how it works" and concluding that "[i]t is entirely possible that in most circumstances neither a reasonable provider nor a reasonable patient would find information about black-box medicine's development or opacity material to disclose, just as patients need not be informed about the strength of clinical trial evidence for most interventions recommended today").

[384] See LaRosa & Danks, supra note 281, at 214 ("AI should not be used for patient care without the *educated* consent of the patient or caregiver. Educated consent is more stringent that informed consent . . . [inasmuch as the former] involves patients in a conversation about these protocols and procedures, and requires more active forms of consent."); Jessica S. Allain, From Jeopardy! To Jaundice: The Medical Liability Implications of Dr. Watson and Other Artificial Intelligence Systems, 73 La. L. Rev. 1049, 1063–64 (2013) (arguing that patients should be "fully informed" when a physician relies on a piece of medical AI to assist in diagnosis and the construction of a treatment plan and arguing that "informed consent may require that a patient be fully informed of [the medical AI's] results, including the options the physician chose not to pursue"); Matsuzaki, supra note 79, at 272 (stating that "adequate information and education about AI decisions should be provided to the patient"); Johnson, supra note 75, at 436 (arguing that "[i]t is imperative that patients are informed when AI is to be used in their treatment and that clinicians understand the level of accuracy of the AI system"); Hoffman & Podgurski, supra note 60, at 46 (concluding that "[e]ven if there is no danger of liability, discussing AI use might be the right thing to do in order to be candid with patients and keep them fully informed about their care"); see also Thomas P. Keenan, Trying for the Trifecta: Telehealth Meets AI Meets Cybersecurity, 17 SciTech Law. 10, 14 (2020) ("[P]atients may be asked to trust an AI algorithm that they (or even the creators) cannot fully understand. So the concept of 'informed consent' becomes problematic."); Terry, Regulating Healthcare AI, supra note 112, at 189 ("[H]ow can a patient make an informed decision about proffered healthcare without understanding, even in very general terms, how the decision about his or her health is being made? The 'transparency' answer to these questions is that we should be able to interrogate decision-making algorithms.").

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

black people have higher rates of C-sections, many of which seem to be unnecessary.[385] Moreover, informed consent should require the physician to notify the patient that it may be safe to assume that the medical AI on which the recommendation is based reflects the higher rates of unnecessary C-sections performed on black people. If known to the physician, they might also share information with the patient about the representativeness of the data on which the AI was trained that makes it safer to trust—or question—the technology's predictive accuracy vis-à-vis black people.

Another example may be instructive: consider a pregnant black patient who complains of blurry vision, severe headaches, and excruciating abdominal pain, but is sent home without intervention because the AI technology with which the physician consults predicts that she is not suffering from preeclampsia. Informed consent should require the physician to inform the patient that it may be reasonable to assume that the medical AI on which the physician's decision is based reflects the lower rates at which physicians adequately treat preeclampsia and eclampsia among black women—a phenomenon that has led to deaths from preeclampsia and eclampsia being much more common among black people than their white counterparts.[386]

Indeed, one of the benefits of reforming the informed consent process in the manner that this Article proposes is that not only will it cause the patient to be aware of social facts that may impact the quality of the healthcare that they are receiving, but it also disciplines the *provider*. The instant proposal will require the provider to repeat undisputed facts about the structural inequities and clinical decisions that, in many cases, are literally killing people of color in the country today. What if the provider has to repeat that the technology that they are using might replicate inequities? What if the provider has to repeat that the country is failing

[385] Elise G. Valdes, Examining Cesarean Delivery Rates by Race: A Population-Based Analysis Using the Robson Ten-Group System, 8 J. Racial & Ethnic Health Disparities 844, 848–49 (2021).

[386] See Building the U.S. Capacity to Review and Prevent Maternal Deaths: Report From Nine Maternal Mortality Review Committees 16–17 (2018), https://www.cdcfoundation.org/sites/default/files/files/ReportfromNineMMRCs.pdf [https://perma.cc/2C25-8ZZB]. It bears noting that the high numbers of deaths that black women suffer from preeclampsia and eclampsia appear to be, on the whole, avoidable. "Over a three-year period, the United Kingdom had only two deaths from preeclampsia and eclampsia, suggesting deaths from these hypertensive disorders of pregnancy are highly preventable." Id. at 6.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

people of color in so many ways? What will this do to the provider?[387] There is a possibility that in repeating these facts to patients, the provider will do even more than they would have done to ensure that the medical AI that they use to help manage the healthcare of their patients of color has some degree of predictive accuracy vis-à-vis those patients—a fact that is especially important when considering that many patients will trust the medical AI because their provider trusts the medical AI.[388]

Cohen has conducted the most thorough analysis to date of the impact that the introduction of medical AI might have on informed consent.[389] On the whole, he is skeptical that these new technologies would, or should, require any changes to present informed consent practices—denying that current doctrine requires physicians to inform their patients about their reliance on medical AI and rejecting the idea that physicians should disclose such information, even if the law does not require it.[390]

However, Cohen is most sympathetic to disclosures about the use of medical AI in the context of race-based algorithmic bias. He gives the example of an AI system that analyzes mammograms but was trained with data within which black people were underrepresented.[391] Because of the bias in the training data, the predictions that the system gives black patients are likely to be less accurate than those that it gives nonblack

[387] There is a substantial literature that establishes that when providers consciously consider the possibility or likelihood that they have implicit biases, it mutes the effect that their implicit biases will have on their decisions. Alexander R. Green et al., Implicit Bias Among Physicians and Its Prediction of Thrombolysis Decisions for Black and White Patients, 22 J. Gen. Internal Med. 1231, 1235 (2007). If doctors must recite facts during the informed consent process about the structural inequities and interpersonal interactions that have had deleterious effects on the health of people of color, this may be just the type of recitation that will make providers aware that their unconscious aversions and associations may impact the healthcare that they deliver. It may be just the type of recitation that will lead them away from acting on any implicit biases that they may have.

[388] See Babic et al., supra note 99, at 286 (arguing that we should think of medical AI as a "credence good," in which an individual trusts the device not because they understand it, but rather because someone else has vouched for the device). Interestingly, the example that Professor Babic and coauthors give involves a *physician* who trusts a drug because someone else—the authorities at the FDA—have warranted the drug's safety and efficacy. Id. It is a bit unsettling to think that physicians will trust medical AI because the FDA has warranted the device's safety and efficacy in light of the fact almost all of the medical AI devices that the FDA has approved have been cleared through the 510(k) process, in which the FDA does not evaluate the device's safety. See discussion supra notes 236–38.

[389] See generally Cohen, supra note 11 (examining the potential impacts of medical AI and ML on informed consent).

[390] Id. at 1448–49.

[391] See id. at 1464.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

patients.[392] Cohen asks, "Should informed consent look different in such a case?"[393] He concludes that it might, writing:

> One possibility would be to just say that in such cases, the AI should not be used. We might achieve that result through regulatory premarket approval regimes or on the back end through malpractice law . . . . But one might worry that in some instances this will make the Perfect the enemy of the Good, that even if the [AI] performs slightly worse in this population it produces better results than the alternative . . . . This might carve out some role for informed consent—physicians should disclose to these populations when they use [AI] involving training-set data that might make medical recommendations for these populations less accurate then [sic] they might otherwise be.[394]

This Article agrees with Cohen. However, the instant proposal for reforming the informed consent process departs from Cohen's appraisal inasmuch as Cohen suggests that disclosures of the use of AI should be required only in those cases where there is evidence that the training data is underinclusive of a population. The immediate proposal insists that disclosures are appropriate in every case, as the causes of race-based algorithmic bias are many, and the fact of race-based algorithmic bias will often go undetected for long periods of time, if ever.

Indeed, when Cohen discusses the Obermeyer study, described in detail above,[395] he admits that the "bias and mechanism" that led to the system requiring black patients to be much sicker than their white counterparts before identifying them as eligible for a program for very sick patients "were largely unknown before the leaders in the field investigated, and not the kind of thing that the average physician or even hospital system could know ahead of time and disclose to a patient."[396] This is all the more reason to disclose the use of medical AI to patients and to explain how a recommendation interacts with documented racial disparities in health and healthcare. Essentially, while Cohen takes the inability to know in advance the specifics of race-based algorithmic bias and the mechanisms by which it occurs as a reason *not* to inform patients about the use of a

[392] See id.
[393] Id.
[394] Id.
[395] See discussion supra notes 149–58, 383 and accompanying text.
[396] Cohen, supra note 11, at 1446 n.51.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

potentially biased tool, this Article takes those same facts as a reason *to* inform patients about the use of a potentially biased tool.

The claim that a properly conceived and implemented informed consent process requires a physician to disclose to the patient not only that a recommendation is based on a medical AI device or system but also that there is a good chance that the recommendation reflects racial disparities in overinterventions and underinterventions, touches on the robust discussion that scholars have been having about the explainability of AI technologies. As discussed above, the black box nature of AI has made many observers wary inasmuch as this opacity precludes the affected public from knowing whether the results that the AI generates are, in fact, accurate.[397] In response, many have proposed that the outputs of any given AI technology must be explainable; that is, an expert must be able to determine how an algorithmic model produced any given result and, further, communicate that explanation to end users in terms that they can understand.[398] This will be unachievable in many cases on account of the complexity of the rules that AI systems generate and the fact that these systems adjust the parameters of their models without human intervention.[399] If explainability is a prerequisite of the use of AI technologies, the impossibility of explaining outputs leads to the conclusion that AI technologies should not be introduced into society—a conclusion that many find deeply unsatisfying.[400]

[397] See discussion supra notes 96–102 and accompanying text.

[398] See Johnson, supra note 75, at 437.

[399] See discussion supra notes 81–87, 91–97 and accompanying text.

[400] See Schönberger, supra note 90, at 188, 195 ("An explanation of the inner workings of the respective algorithms would not empower patients to make an informed choice about a given treatment" and "making AI more explainable will mean to limit its complexity, which will almost certainly adversely impact performance."); Babic et al., supra note 99, at 286 (arguing that if explainability were a requirement for the use of AI in healthcare, then we would not use AI in healthcare, and contending that instead of making explainability a prerequisite for medical AI's use, we "should place more emphasis on well-designed clinical trials," thus ensuring that the device performs as it should "in the hands of its intended users").

The most frequently made riposte to the proposition that AI should not be used because it is not explainable is the reminder that we cannot explain lots of medicine as well as many human decisions; however, that has not stopped us from relying on these unexplainable phenomena. See Cadario et al., supra note 279, at 1636 ("[D]ecisions made by human providers appear more transparent, but this perception is an illusion . . . . Human decision making is often as much a black box as decisions made by algorithms."); Haupt, supra note 121, at 75 (noting the argument that "[t]he black-box problem in clinical practice . . . has existed well before the introduction of AI" and "[w]hy certain medical treatments work is much less explainable than we would think"); Feldman et al., supra note 96, at 411 ("[T]here are numerous drugs today whose mechanisms of action are unknown, including the muscle relaxant metaxalone, the

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

This Article takes the position that the inability to explain the results of medical AI makes it all the more necessary that physicians not only disclose their use of the technology, but also put the technology's recommendations in conversation with documented racial disparities in health and healthcare. In essence, the lack of explainability ought to create a presumption that the technology has encoded biases—be they the result of problematic labels that the programmers chose, or nonrepresentative training data, or perfectly representative training data that *accurately* reflects the inequitable healthcare that black people receive. This presumption would only be rebutted if an expert can explain the results and adduce evidence that racism is not, in fact, embedded in the machine, i.e., the labels used in the model are appropriate, the training data was representative and did not reflect race-based medicine and medical practices that are scientifically unsupported.

This reveals how the instant proposal diverges from Cohen's assessment of the necessity of informed consent in the context of race-based algorithmic bias, discussed above.[401] While Cohen appears to presume that informed consent is *inappropriate* unless there is evidence that the tool might be *biased*, i.e., the system's training data was underinclusive of a population, the instant proposal presumes that informed consent is *appropriate* unless there is evidence that the tool is *unbiased*. The instant proposal presumes the latter simply because of an awareness of the banality of racism, as evidenced by the fact of persisting racial disparities in health and healthcare.

CONCLUSION

Philosopher Camisha Russell explains that the discipline of bioethics—the field that has theorized the import of informed consent most extensively—historically concerned itself with concepts like autonomy, freedom, and self-determination. These are all concepts that center the individual as divorced from social context. However, she writes about the

diabetes-related drug metformin, and the cough suppressant [guaifenesin] . . . . [R]egulatory bodies determine *whether* many drugs and treatments are safe and effective, but the answer of *how* the drug works is not a necessary condition."); Chen & Verghese, supra note 175, at 2 ("[M]any of the tools and modalities we use in medicine are just as invisible or opaque. Do we see or know how acetaminophen or anesthesia works, or how the machine generating the number for the serum sodium or creatinine works?").

[401] See supra notes 393–96 and accompanying text.

COPYRIGHT © 2024 VIRGINIA LAW REVIEW ASSOCIATION

possibility of the irruption of "the political" into bioethics, an irruption that would widen the scope of the field's interests.[402] She offers:

> The ethical, where it is centered on autonomy conceived in terms of personal freedom, comes to be concerned only with what is or is not permissible in biomedical practice in terms of individually conceived ethical rights, duties, obligations, or prohibitions. With ethical rules in place, much of patient and physician decision-making is taken to be a private matter, with little relevance to politics or social justice. By contrast, the view from the margins suggests that bioethics ought to be at least as concerned with what we might label the *political*—that is, social responsibility, collective life, the power dynamics and inequalities of social orders, and the role that concepts like race have played in creating and maintaining such inequalities.[403]

This Article observes the introduction into healthcare of AI technologies that threaten to encode the indefensible racial disparities in health and healthcare in the United States that are so well-documented. It concludes that this is an appropriate occasion by which to invite the political to irrupt into the informed consent process—a space that, despite its origins in the Nuremberg Trials' rebuke of racism, antisemitism, and white supremacy, many imagine to be unconcerned with social life.

[402] See Russell, supra note 253, at 49.
[403] Id. at 49–50.