Racial Equity Implications of Artificial Intelligence in Health Care

PATRICK ROSS*

ABSTRACT

Decision-making using artificial intelligence (AI) aims to reduce human error and bias in the clinical decision process. The advanced pattern-finding capabilities of AI and machine learning serve as both promise and pitfall, as pattern recognition can unintentionally cause algorithms to incorporate human biases, such as racial, gender, or socioeconomic biases. Considering the health care applications of these tools, ensuring they are safe to deploy is critical. This requires a clear oversight structure from development to deployment. However, the unique ability to learn and adapt that fuels AI's promise also presents a challenge to current regulatory frameworks.

I. INTRODUCTION

Artificial intelligence (AI) and machine learning is transitioning from a mainstay of science fiction to a workday tool. Advances in computational power, joined with a deep well of data, have resulted in a flourishing of uses for AI. However, regulation and oversight of these algorithmic tools have not kept up with the rate of technological innovation. What sets these tools apart is their capacity for continuous learning—given additional data, the mathematical model can be refined and adapt, giving the appearance of "intelligence." This unique ability of AI tools presents a challenge to current regulation and oversight structures. Without proper guidance, these futuristic tools pose a risk of further embedding biases and health disparities captured in today's data.

When thinking of the future of AI in medicine, many point to the possibilities to reduce human error and improve the quality of care. AI and machine learning tools have been used to reduce adverse events such as pressure ulcers, surgery complications, and diagnostic errors.[1] However, from image detection tools to algorithms used to guide population health decisions, these powerful technologies are susceptible to replicating or entrenching human racial biases due to the same pattern recognition that gives them such potential.

[*] Patrick Ross is the Associate Director of Federal Affairs at The Joint Commission. Patrick leads The Joint Commission's Washington office efforts regarding new and developing health care delivery models, including artificial intelligence. Prior to joining The Joint Commission, Patrick was a research assistant for the National Academies of Sciences, Engineering, and Medicine, focusing on health care services and cancer care. Patrick holds a Master of Public Health degree from the Harvard T.H. Chan School of Public Health. The views in this Article are the author's own and do not necessarily reflect the official policy or position of The Joint Commission.

[1] David W. Bates, Gretchen Purcell Jackson & Kyu Rhee, *The Potential of Artificial Intelligence to Improve Patient Safety: A Scoping Review*, 4 NPJ DIGITAL MEDICINE 54, 55–56 (2021).

Across a wide variety of disciplines and use cases, evidence is already showing that AI tools can produce results that discriminate by race, sex, or socioeconomic status, either by incorporating human biases or being used in the wrong setting or with the wrong population.[2] While clinical care and public health are not the only sectors to grapple with these challenges, the risks posed to users are greater, as the use of AI solutions without careful development and review may further encode biased care into the health system.

II. INTRODUCING BIAS INTO AI MODELS THROUGH DATA

Preventing racial bias in AI requires an understanding of the importance of training data in AI and machine learning development. When developing new AI tools, the accuracy and relevancy of the data used to "train" an algorithm is crucial. No algorithm can exceed the level of performance reflected in the training data. If an AI is producing incorrect output in training, no amount of "learning" will lead it to correct outcomes.

Many models rely on the "ground truth" contained in the training data, which is intended to represent the best-case scenario where researchers know the outcome in question. However, establishing this truth is not always straightforward. Even when drawing from data elements such as medical records, insurance claims, or device readings, the data is the result of human decisions.[3] There might be data missing, no data that are not relevant to a particular disease state, or demographic information might be removed in order to protect patient privacy.[4]

Even training data that are accurate may reflect current or historical disparities in clinical care or outcomes, and thus produce biased results. In one such case, a commercial algorithm that was widely used by health systems and insurers was intended to flag patients with complex care needs who might need additional resources or health services. However, in an analysis, the tool was shown to under-select Black patients for additional follow-up care. Instead of clinical risk factors, the algorithm predicted needs based on the projected dollar amount spent on care.[5] While the health spending data was accurate, white patients have historically received a disproportionate amount of the total spending on care, resulting in the algorithm's biased outcome for Black patients.[6]

This example also illustrates the problem of an algorithm's *ideal* target versus the *actual* target. The ideal target is the researcher's original intent. The actual target may be dramatically different, often as a result of mismatch between the desired outcome

[2] James Zou & Lona Schiebinger, *AI Can Be Sexist and Racist–It's Time to Make It Fair*, 559 NATURE 324, 324–25 (2018).

[3] Ravi B. Parikh, Stephanie Teeple & Amol S. Navathe, *Addressing Bias in Artificial Intelligence in Health Care*, 322 JAMA 2377, 2377 (2019).

[4] Robert Challen, Joshua Denny, Martin Pitt, Luke Gompels, Tom Edwards & Krasimira Tsaneva-Atanasova, *Artificial Intelligence, Bias and Clinical Safety*, 28 BMJ QUALITY & SAFETY 231, 234 (2019).

[5] Ziad Obermeyer, Brian Powers, Christine Vogeli & Sendhil Mullainathan, *Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations*, 366 SCIENCE 447, 449–50 (2019).

[6] Joseph L. Dieleman, Jackie Cao, Abby Chapin, Carina Chen, Zhiyin Li, Angela Liu, Cody Horst, Alexander Kaldjian, Taylor Matyasz, Kirstin Woody Scott, Anthony L. Bui, Madeline Campbell, Herbert C. Duber, Abe C. Dunn, Abraham D. Flaxman, Christina Fitzmaurice, Mohsen Naghavi, Nafis Sadat, Peter Shieh, Ellen Squires, Kai Yeung & Christopher J. L. Murray, *U.S. Health Care Spending by Race and Ethnicity, 2002–2016*, 326 JAMA 649, 652–53 (2021).

and proxy variables chosen by the developer or underrepresentation in training data.[7] Underrepresentation in the training data can lead to significantly lower performance for the minority population—a result seen in use cases from computer-aided diagnosis for thoracic X-rays to automated speech recognition tools that produce nearly twice as many errors in transcribing Black speakers compared to white speakers.[8]

Such a mismatch in proxy variables is especially likely when AI systems are built to model complex concepts. In cases where concepts cannot be measured directly or easily captured by data, developers use other data elements as a proxy for the concept they attempt to model. Developers may choose data sets that are readily available or accessible rather than the most appropriate, which can lead to harmful or discriminatory outcomes.[9] A frequently cited example beyond the health care space is the use of predominantly white subjects in existing training sets for facial recognition software, leading to inaccuracies in recognizing non-white faces.[10]

In many cases, developers only have access to limited datasets. Collecting large data sets can be time-consuming and expensive. Favoring easily accessible data over a thoughtful data collection process presents a risk of introducing racial bias into AI tools. For example, training on the available data may have led an Alzheimer's detection software trained on native English speakers to identify more non-native English speakers as having early signs of Alzheimer's. Increased pauses or different pronunciations were associated with signs of the disease instead of the possibility of non-native speakers.[11]

Choices about how algorithms are trained—including where and how the data is collected and variables are used in assembling the model—can lead to widespread consequences. Algorithms are commonly viewed as a "fair" or objective means of dividing resources, but the human choices involved in creating algorithms can reinforce existing disparities. During the COVID-19 pandemic, the algorithm used to allocate federal funds from the Provider Relief Fund aligned awards more closely with historic hospital revenue than COVID-19 burden, morbidities, or hospital financial health. Despite lawmakers' intent to allocate emergency funds to health systems with the highest health or financial needs, funds were distributed disproportionately to hospitals that previously brought in more revenue per patient, potentially diverting funds away from areas with more acute need.[12]

Deploying AI tools in new contexts or populations can also present the risk of a mismatch between the AI tool's intent and a new population. This is known as

[7] ZIAD OBERMEYER, REBECCA NISSAN, MICHAEL STERN, STEPHANIE EANEFF, EMILY JOY BEMBENECK & SENDHIL MULLAINATHAN, ALGORITHMIC BIAS PLAYBOOK 7–10 (2021).

[8] Allison Koenecke, Andrew Nam, Emily Lake, Joe Nudell, Minnie Quartey, Zion Mengesha, Connor Toups, John R. Rickford, Dan Jurafsky & Sharad Goel, *Racial Disparities in Automated Speech Recognition*, 117 PNAS 7684, 7685 (2020).

[9] REVA SCHWARTZ, LEANN DOWN, ADAM JONAS & ELHAM TABASSI, A PROPOSAL FOR IDENTIFYING AND MANAGING BIAS IN ARTIFICIAL INTELLIGENCE, 1270 NIST SPECIAL PUBLICATION 3 (2021).

[10] Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROC. MACH. LEARNING RSCH. 3 (2018).

[11] Dave Gershgorn, *If AI is Going to be the World's Doctor, it Needs Better Textbooks*, QUARTZ (Sept. 6, 2018), https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks.

[12] Pragya Kakani, Amitabh Chandra, Sendhil Mullainathan & Ziad Obermeyer, *Allocation of COVID-19 Relief Funding to Disproportionately Black Counties*, 324 JAMA 1000, 1001 (2020).

distributional shift, when an AI tool fails to recognize or incorporate a change in context or data and results in the algorithm continuing to make incorrect predictions based on out-of-sample inputs.[13] As a result, training an AI tool on a single population or geographic region can limit its generalizability to new areas or populations.[14] Algorithms developed based on an urban setting may struggle to provide accurate predictions in rural areas. Similar context constraints may be seen across different provider settings; an algorithm developed using data from a long-term care facility may be less effective in an acute care setting.

This poses a significant challenge for health systems intending to use AI tools across varying regions, especially considering that relatively few AI tools are developed using multisite cohorts, and most rely on data sets from California, Massachusetts, and New York.[15] The same factors that may contribute to the concentration of data sets originating from these states, such as large urban areas with significant investments in biomedical research, may also represent economic, social, and cultural factors that limit their generalizability to other states and regions. The specificity of some algorithms can be extremely narrow—some researchers have even found that algorithms need to be adjusted between individual imaging machines.[16] With health care organizations serving a wide variety of patient populations, it's very likely that third party AI tools are not ready to go out-of-the-box. It is critical that before they are deployed, AI tools are reviewed to ensure that they are appropriate for the setting and population where they will be used.

III. AI USE AND HUMAN BIAS

Regulators and health systems using AI tools must also carefully consider human–computer interactions as they look to prevent patient harm. Implicit racial bias in clinicians is well-documented, and these biases are applied most often when clinicians are rushed, distracted, or under time pressure.[17] In turn, implicit bias may lead to clinicians prioritizing AI-based decision support over a patient's input. This could result in scenarios where incorrect AI output is overlooked or a particular diagnosis is

[13] Challen et al., *supra* note 4, at 234.

[14] Junaid Nabi, *How Bioethics Can Shape Artificial Intelligence and Machine Learning,* 48 HASTINGS CTR. REP. 10, 12 (2018).

[15] Amit Kaushal, Russ Altman & Curt Langlotz, *Geographic Distributions of U.S. Cohorts Used to Train Deep Learning Algorithms*, 324 JAMA 1212, 1213 (2020).

[16] Jeffrey De Fauw, Joseph R. Ledsam, Bernardino Romera-Paredes, Stanislav Nikolov, Nenad Tomasev, Sam Blackwell, Harry Askham, Xavier Glorot, Brendan O'Donoghue, Daniel Visentin, George van den Driessche, Balaji Lakshminarayanan, Clemens Meyer, Faith Mackinder, Simon Bouton, Kareem Ayoub, Reena Chopra, Dominic King, Alan Karthikesalingam, Cían O. Hughes, Rosalind Raine, Julian Hughes, Dawn A. Sim, Catherine Egan, Adnan Tufail, Hugh Montgomery, Demis Hassabis, Geraint Rees, Trevor Back, Peng T. Khaw, Mustafa Suleyman, Julien Cornebise, Pearse A. Keane & Olaf Ronneberger, *Clinically Applicable Deep Learning for Diagnosis and Referral in Retinal Disease*, 24 NATURE MED. 1342, 1348 (2018).

[17] William J. Hall, Mimi V. Chapman, Kent M. Lee, Yesenia M. Merino, Tainayah W. Thomas, B. Keith Payne, Eugenia Eng, Steven H. Day & Tamera Coyne-Beasley, *Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review*, 105 AM. J. PUB. HEALTH e60, e61 (2015).

prematurely incorporated into a patient's care plan based on unsubstantiated AI output.[18]

Beyond implicit bias threats, relying on machine output can actually impede human decision-making. Automation bias is the result of humans trusting software uncritically and favoring computer-generated outcomes over human reasoning. Most software gains trust from human users by being relatively predictable. AI, which can generate new outcomes based on new data, does not follow this model. When deploying AI, the optimal level of trust still has some human skepticism of computer-generated results.[19] Automation bias results from maximized trust in an AI tool and causes decision-makers to stop looking for evidence after being provided with machine-generated output.[20]

Automation bias is closely linked to automation complacency. In automation complacency, human users or interpreters of AI output become overly reliant on AI support.[21] Complex tasks increase the likelihood of automation complacency, and studies have demonstrated that human performance declines as a result of repeated use of computer-aided decision support, such as that sometimes used in radiology.[22] In addition to more readily relying on machine-generated outcomes to make decisions, human users are also less likely to override incorrect computer output.[23] While not unique to AI, these biases may play an outsized role in safety given the importance of decisions AI may make and the typical role of humans as the final safety check.

How AI tools are integrated into a clinician's workflow is an essential consideration to promote patient safety. Patient examination skills may weaken as clinicians devote more time to data entry and lose the detail and nuance of patient histories.[24] The risk of automation bias, like implicit racial bias, is increased when the human user is distracted or busy—a common scenario in modern medicine where reviews suggest that nearly 40% of a physician's workday is spent interacting with electronic health records.[25] What prompts might a clinician click through or review only briefly in order

[18] Danton S. Char, Nigam H. Shah & David Magnus, *Implementing Machine Learning in Health Care–Addressing Ethical Challenges*, 378 NEW ENG. J. MED. 981, 982 (2018).

[19] Onur Asan, Alparslan Emrah Bayrak & Avishek Choudhury, *Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians*, 22 J. MED. INTERNET RES. e15154, e15158 (2020).

[20] David Lyell & Enrico Coiera, *Automation Bias and Verification Complexity: A Systemic Review*, 24 J. AM. MED. INFORMATICS ASSOC. 423, 423–24 (2017).

[21] Carl Macrae, *Governing the Safety of Artificial Intelligence in Healthcare*, 28 BMJ QUALITY & SAFETY 495, 495–96 (2019).

[22] Eugenio Alberdi, Andrey Povyakalo, Lorenzo Strigini & Peter Ayton, *Effects of Incorrect Computer-Aided Detection (CAD) Output on Human Decision-Making in Mammography*, 11 ACAD. RADIOLOGY 909, 918 (2004); Constance D. Lehman, Robert D. Wellman, Diana Buist, Karla Kerlikowske, Anna Tosteson, Diana L. Miglioretti & Breast Cancer Surveillance Consortium, *Diagnostic Accuracy of Digital Screening Mammography With and Without Computer-Aided Detection*, 175 JAMA INTERNAL MED. 1828, 1832 (2015).

[23] Mark Sujan, Dominic Furniss, Kath Grundy, Howard Grundy, David Nelson, Matthew Elliott, Sean White, Ibrahim Habli & Nick Reynolds, *Human Factors Challenges for the Safe Use of Artificial Intelligence in Patient Care*, 26 BMJ HEALTH & CARE INFORMATICS e100081, e100082–83 (2019).

[24] Jingyan Lu, *Will Medical Technology Deskill Doctors?*, 9 INT'L EDUC. STUD. 130, 130–31 (2016).

[25] Yuliya Pinevich, Kathryn J. Clark, Andrew M. Harrison, Brian W. Pickering & Vitaly Herasevich, *Interaction Time with Electronic Health Records: A Systematic Review*, 12 APPLIED CLINICAL INF ORMATICS 788, 796 (2021).

to move on to the next task? What signals from the patient might a clinician miss in favor of AI-assisted decision-making?

Regular training is necessary to prevent AI users from "deskilling," or losing human skills after a task is automated.[26] AI developers should also be charged with creating transparent, explainable AI models. Promoting transparency in AI tools makes understanding computer-generated outcomes easier, but having an explainable decision path improves error detection and allows AI users to learn from incorrect predictions or near-misses.[27] For patients belonging to systematically marginalized and excluded groups, the interaction between racial bias and automation biases could turn near misses into preventable patient harm.

To promote AI trustworthiness, tools should be explainable to patients, who should be informed if AI tools are involved in making key care decisions. While being too trusting of AI-generated output can lead to overlooked errors, a focus of developers and AI users must be to foster understanding of AI and machine learning if they are to be accepted as functional tools in the clinic, especially by patients.

IV. DEVELOPMENT AND REGULATION TO PREVENT RACIAL BIAS IN AI

Without thoughtful design and careful development, AI tools risk further entrenching bias as health care enters a new digital era. Guidelines for the safe development of AI are critical to ensuring patient safety, as is thoughtful regulation of the development and deployment of AI tools, including ongoing real-world data collection and review. Due to AI's capacity for continuous "learning" or adaptation while in use, updated frameworks for regulation are necessary.

AI regulation in the United States remains in its infancy. As it currently exists, federal oversight of AI in health care has been located within the U.S. Food and Drug Administration (FDA), which has jurisdiction over AI tools marketed as "Software as a Medical Device (SaMD)." In 2019, FDA released a proposed regulatory framework for AI and machine learning devices.[28] Under the proposed framework, FDA has begun to review and approve "locked" algorithms—those that do not engage in self learning as real-world or new data become available.[29]

In 2021, FDA released its AI and machine learning action plan, which lays out the agency's path forward in AI regulation, highlighting several focus areas where more specific guidance will be issued.[30] In the action plan, FDA responded to stakeholder concerns about the risk of racial bias in AI tools, acknowledging that AI devices are susceptible to mirroring biases present in training data.[31] As a first step to reduce the

[26] Federico Cabitza, Raffaele Rasoini & Gian Franco Gensini, *Unintended Consequences of Machine Learning in Medicine*, 318 JAMA 517, 517 (2017).

[27] Trevor Jamieson & Avi Goldfarb, *Clinical Considerations When Applying Machine Learning to Decision-Support Tasks Versus Automation*, 28 BMJ QUALITY & SAFETY 778, 779 (2019).

[28] U.S. FOOD & DRUG ADMIN., PROPOSED REGULATORY FRAMEWORK FOR MODIFICATIONS TO ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) (2019).

[29] *Id.* at 3.

[30] U.S. FOOD & DRUG ADMIN., ARTIFICIAL INTELLIGENCE/MACHINE LEARNING (AI/ML)-BASED SOFTWARE AS A MEDICAL DEVICE (SAMD) ACTION PLAN (2021).

[31] *Id.* at 5–6.

risk of algorithmic biases, FDA supports research efforts to develop methods that would be used to evaluate AI software to identify and eliminate bias during development and the product review stage. However, much of the regulatory framework developed by FDA remains undefined or in the form of guidelines and basic principles.

For the many AI tools that fall outside of FDA's jurisdiction, there is little in the way of a formal oversight process. To promote the development of safe and effective AI tools, industry stakeholders have launched several efforts to define standards to guide the creation and use of machine learning software. Many of these guidelines call for explicitly testing new AI tools for potential racial bias, as well as other biases that could harm marginalized or vulnerable populations. The American Medical Association released its first AI policy recommendations in 2018, advocating for transparent, reproducible software and a development process that takes proactive steps to identify and mitigate bias and avoid exacerbating health care disparities, calling out the testing of new AI tools on vulnerable populations.[32]

In a consensus-standards document on promoting AI trustworthiness in health care, the Consumer Technology Association (CTA) recommends that developers list an AI's potential use cases prior to development, to ensure the algorithm is properly scoped. Additionally, to improve AI accuracy, CTA recommends that developers determine whether existing data set is composed of raw or pre-processed data and determine what kind of pre-processing has been performed. Developers should be aware of how the original data was collected, which promotes an understanding of potential biases in the data.[33]

The American Health Law Association has even suggested introducing contractual requirements to meet certain diversity and representation benchmarks in training data to mitigate potential bias.[34] Some organizations have already begun to release inventories that can be used during development to review, screen, retrain, and prevent bias in the final AI product.[35] These principles should be considered as FDA gathers input on Good Machine Learning Practices, a key component of the agency's plan to address AI safety during development.[36] In addition to practice guidelines, regulators should provide readily accessible resources to developers to screen AI tools for bias before they are tested or used in clinical populations.

An additional focus for industry standards is ongoing real-world testing following device approval. In addition to testing during development, the American Medical Information Association calls for continuous review of algorithm results during use.[37]

[32] Press Release, American Medical Association, AMA Passes First Policy Recommendations on Augmented Intelligence (June 14, 2018), https://www.ama-assn.org/press-center/press-releases/ama-passes-first-policy-recommendations-augmented-intelligence.

[33] CONSUMER TECH. ASS'N, THE USE OF ARTIFICIAL INTELLIGENCE IN HEALTH CARE: TRUSTWORTHINESS (ANSI/CTA-2090) 15 *(*2021).

[34] AM. HEALTH LAW ASS'N (AHLA), DESIGNING A TRUSTED FRAMEWORK FOR THE APPLICATION OF AI IN HEALTH CARE 15 (2021).

[35] OBERMEYER ET AL., *supra* note 7, at 5–7.

[36] U.S. FOOD & DRUG ADMIN., *supra* note 28, at 9.

[37] Carolyn Petersen, Jeffery Smith, Robert R. Freimuth, Kenneth W. Goodman, Gretchen Purcell Jackson, Joseph Kannry, Hongfang Liu, Subha Madhavan, Dean F. Sittig & Adam Wright, *Recommendatio ns for the Safe, Effective Use of Adaptive CDS in the U.S. Healthcare System: An AMIA Position Paper*, 28 J. AM. MED. INFORMATICS ASS'N 677, 681–82 (2021).

The American Health Law Association has echoed this call, recommending a format for ongoing assessment of an AI device's compliance with legal and regulatory requirements throughout its life cycle, particularly including ongoing testing and re-validation of safety, efficacy, and privacy protections.[38]

Given the risk of distribution shifts as AI tools are used in new regions or for new patient populations, post-market surveillance is critical to proper oversight. The FDA framework provides an example of how to conduct oversight once AI tools are in use. Currently, FDA relies on pre-determined change control plans submitted by developers to set boundaries for how software can be updated once it has been deployed. Change plans should include requirements for regular testing and data submission to oversight bodies. The level of risk to patients and the level of AI software autonomy in decision-making should be factored into how frequently these programs are monitored.[39] How often data is submitted to oversight bodies and what form of data would be reviewed is still an open question within the FDA framework. Currently, performance monitoring is done on a voluntary basis with device manufacturers. Additionally, FDA has not yet addressed how to regulate adaptive or continuously learning AI devices. For these algorithms, lifecycle oversight and management will be vital to ensure that patient safety is protected.

Beyond FDA, other federal bodies have expressed their interest in regulating or guiding AI development. The U.S. Department of Health and Human Services (HHS) has released a wide-ranging artificial intelligence strategy to promote innovation while encouraging the trustworthy development and use of AI.[40] HHS says it will promulgate federal expectations for equitable, safe, and ethical AI tools. The National Institute for Standards and Technology has also released a proposal for identifying and managing bias in AI, which provides industry non-specific recommendations on preventing bias from the pre-design stage through deployment.[41]

The day-to-day use of AI tools also raises questions about product liability and the attribution of harm associated with AI use. In the case of continuously learning algorithms, products may adapt in non-desirable or unpredicted ways, potentially contributing to avoidable harm. In such cases, the developer or producer could potentially be held at fault, even if the tool's process has changed since it was introduced into use. Users, or AI operators, might also be held at fault if they are using AI tools incorrectly or without proper care.[42] Device operators may not be aware of issues with built-in bias. This "black box problem" inherent to complex AI tools may also make determining the cause of harm challenging, again highlighting the importance of understandable and clear AI.

V. CONCLUSION

When developed thoughtfully and used correctly, AI has shown real promise to advance the delivery of health care. Ensuring these tools are safe is critical, especially

[38] AHLA, *supra* note 36, at 3, 10.

[39] NAT'L ACAD. OF MED., ARTIFICIAL INTELLIGENCE IN HEALTH CARE: THE HOPE, THE HYPE, THE PROMISE, THE PERIL 224 (2019).

[40] U.S. DEPT. OF HEALTH & HUM. SERVS., ARTIFICIAL INTELLIGENCE (AI) STRATEGY 7 (2021).

[41] SCHWARTZ ET AL., *supra* note 9, at 5–12.

[42] Herbert Zech, *Liability for AI: Public Policy Considerations*, 22 ERA FORUM 147, 154–56 (2021).

in the health care context. This requires a clear structure for oversight from development to deployment. However, the unique ability to learn and adapt that fuels AI's promise also presents a challenge to current regulatory frameworks. Additionally, the pattern-recognition capabilities key to AI's success also poses a risk of encoding racial bias from historical practice into these new tools. Current regulation of AI tools is limited to a relatively small portion of software and remains largely undefined. As regulatory bodies and stakeholders come together to flesh out guidelines for the safe development and use of AI tools, regulators must place a strong emphasis on ensuring that AI tools do not further entrench racial bias.