THE GREENLINING INSTITUTE

MAY 2023

Race Aware Algorithms

Principles and policies for a more equitable future

Alice Lee
MPA 2023, UC Berkeley

ABOUT THIS REPORT

This report was produced in partnership with The Greenlining Institute (GLI). My gratitude goes to Vinhcent Le and Caroline Siegel-Singh at GLI for bringing me on to lead this work and guiding the literature review and qualitative interviews. The eight expert interviewees we spoke with at the beginning of this project, who will remain anonymous, represent tech industry, academic, and civil society perspectives. The insights from these interviews built the foundation of this report. Throughout the duration of this capstone project, I also took a course at the UC Berkeley School of Information called "Data, Power, and Infrastructure" taught by Dr. Alex Hanna, which greatly influenced my perspective on topics relevant to this report. Ongoing conversations with Todd Achilles, my capstone advisor, also brought cohesion and rigor to my thinking around these complex issues. Finally, the folks who reviewed draft sections and versions of this report also contributed greatly to the development of this report: Ziad Obermeyer (UC Berkeley School of Public Health), Eliza McCullough (Partnership on AI), Alex Hanna (UC Berkeley School of Information), Vinhcent Le (GLI), and Todd Achilles (UC Berkeley, Goldman School of Public Policy). I am sincerely grateful to everyone named here for the time, trust, and collaboration they generously provided in producing this work.

Disclaimer: The author conducted this study as part of the program of professional education at the Goldman School of Public Policy, University of California at Berkeley. This paper is submitted in partial fulfillment of the course requirements for the Master of Public Affairs degree. The judgments and conclusions are solely those of the author, and are not necessarily endorsed by the Goldman School of Public Policy, by the University of California, by The Greenlining Institute, or by any other agency.

TABLE OF CONTENTS

INTRODUCTION

We are living in "the age of algorithms."[1] From ad delivery to job recruiting to welfare allocation, algorithms are the backbone of the artificial intelligence (AI) and automated decision systems (ADS) that shape our daily lives. These algorithms depend on ever-growing big data sets that know our personal identities and preferences, and are often used to make predictions about our future decisions – for example, predict if a first-time homebuyer will repay a mortgage loan. In the age of algorithms, different forms of algorithmic discrimination are a known fact, including racial discrimination through ADS. The use of ADS in almost every industry – including civil rights protected areas – has brought to light the embedded racism across all of our systems. As an expert interviewed for this report shared, "Race is always part of the system, whether or not it is ignored or explicit or implicit. The implicit biases built into our systems are becoming increasingly explicit as they are written into code and executed at scale."[2]

At the same time, the United States has remained in an "age of colorblindness."[3] In particular, our legal and policy frameworks offer race blind solutions in the name of anti-discrimination and equal opportunity. Despite the progress of the Civil Rights movement, the historical biases and systemic racism stemming from Jim Crow and chattel slavery remain embedded in our physical and digital infrastructure today. Our race blind policy framework is incapable of addressing the racial discrimination embedded in our digital infrastructure.

In 2021, The Greenlining Institute published "Algorithmic Bias Explained: How Automated Decision-Making Becomes Automated Discrimination,"[4] where we detailed how algorithms exacerbate biases in healthcare, employment, government programs, and additional industries. This white paper builds on this work, focusing on racial discrimination in ADS in the civil rights protected areas of housing, employment, and banking. We drew from the breadth of research covering the topic of racial discrimination in algorithms, aided by new reports released at the beginning of 2023 in light of the rise of large language models. Our desk research was supported by interviews with experts in academia, the tech industry, and civil society who are referenced anonymously throughout this paper. We synthesize this research and make the case for a race aware policy framework characterized by five core principles of race aware algorithms and propose five key policy recommendations for making race aware algorithms a reality. Finally we offer a high-level, illustrative timeline for how these recommendations may be carried out in practice, recognizing that this timeline is likely to change once more information is learned from greater algorithmic transparency and reliable race data.

1 Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, & Cass Sunstein. (2018). *Discrimination in the Age of Algorithms.* Journal of Legal Analysis, 10, 113–174.

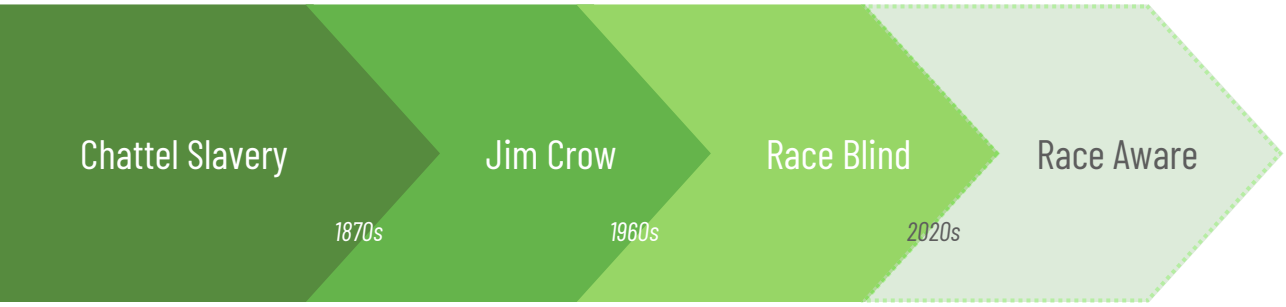
2 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

3 Michelle Alexander. (2012). *The New Jim Crow: Mass Incarceration in the Age of Colorblindness.* The New Press.

4 Gissela Moya & Vinhcent Le. (2021). *Algorithmic Bias Explained.* The Greenlining Institute. https://greenlining.org/publications/algorithmic-bias-explained/

Notably, there are many related topics this paper does not cover that are necessary to explore in turning the policy recommendations shared here into practice. First, this paper focuses exclusively on racial discrimination, but many of the same concerns pertain to other dimensions of identity such as gender. Second and most importantly, algorithms amplify racial discrimination because racial discrimination is embedded in our socioeconomic systems. Combatting algorithmic discrimination will not solve racial discrimination at large. As a racial equity organization, The Greenlining Institute works to address the root causes of racial inequity across our socioeconomic systems and we present this white paper as one component of a holistic suite of programs and partnerships advancing more equitable futures for communities of color.

Phases of US Race Policy



A new policy era

In the United States, the Black-White Paradigm is a powerful lens through which to view the history of US policy. Angela Glover Blackwell defines the Black-White Paradigm as "the composite of the economic, legal, institutional, social, and psychological structures forged from slavery that have systemized and codified oppression…[this paradigm] illuminates the interconnectedness of all people who have experienced racial violence, bigotry, and limited opportunity, and it exposes the biases and beliefs at the root of oppression."[5] Through the Black-White Paradigm, we can analyze US racial policy in three phases. First was the era of chattel slavery that encompasses enslavement, Reconstruction, all the way through to the Compromise of 1877 which marked the end of Reconstruction and removed any hope for legal protection for formerly enslaved people.[6] Instead, the Compromise of 1877 made way for the second policy era: Jim Crow. In the Jim Crow policy era,

5 Angela Glover Blackwell. (2023, Spring). *How We Achieve a Multiracial Democracy.* Stanford Social Innovation Review. https://ssir.org/articles/entry/how_we_achieve_a_multiracial_democracy

6 MADEO. (n.d.). Apr. 24, 1877 | *President Hayes Withdraws Federal Troops from South, Ending Reconstruction.* Retrieved April 15, 2023, from https://calendar.eji.org/racial-injustice/apr/24

segregation was written into law and redlining became the basis for overtly racist policies, like the implementation of the Federal Highly Act of 1956 which further segregated neighborhoods and created more barriers to employment and economic mobility along racial lines.[7] These overtly racist policies lasted for nearly a century and left a lasting impact on the nation's physical, social, and economic infrastructure.

The Civil Rights Act of 1964, the Fair Housing Act of 1968, and the Equal Credit Opportunity Act of 1974 are landmark anti-discrimination laws that outlawed Jim Crow era policies by countering racism with race blindness. This third phase of racial policy – what we are calling the race blind era – was characterized by laws which were designed based on the belief that by leaving race out of a decision-making process, racial discrimination would be mitigated. This policy era championed equality of opportunity in financial access, education, employment, and housing regardless of sex, race, color, national origin, or religion. Yet, nearly 60 years after the Civil Rights Act was passed, we know such equality is far from the reality. The race blind policy approach of the past half century has enabled racial discrimination not only to persist, but to flourish. Behind the veil of equal opportunity, racism continued to strengthen its hold on every facet of daily life. Today, the racial wealth gap has remained exceedingly wide: white households have, on average, ten times the wealth of Black households.[8] "Black [mortgage] loan applicants are 80 percent more likely to be denied than their White counterparts."[9] Reports of racial discrimination in employment remain common, and in Silicon Valley, the hub of the technology industry, less than 8 percent of the workforce is either Black or Latinx across 177 companies based in Silicon Valley.[10] Race blind policies like zoning laws restrict economic mobility, effectively reinforcing redlining and keeping poor, majority Black and majority people of color neighborhoods, vastly underserved with critical public infrastructure. Today, zip code remains a highly accurate proxy for race.[11]

It comes as no surprise, then, that as a result of this history, our digital infrastructure – born in a race blind policy era – follows a similarly race blind architecture. It is a common argument that ADS are objective, logical equations, incapable of committing harm and therefore better than human decision-making. Yet, ADS are designed by humans, and like any other system designed by humans, algorithms are embedded with human bias. Unlike all other systems, however, algorithms automate and amplify human bias at scale. The outcomes produced by ADS bring to light the embedded racism across all of

7 Solomon Greene, Margery Austin Turner, & Ruth Gourevitch. (2017). *Racial Residential Segregation and Neighborhood Disparities.* US Partnership on Mobility from Poverty. https://www.mobilitypartnership.org/publications/racial-residential-segregation-and-neighborhood-disparities

8 Economic Equity—The Greenlining Institute. (n.d.). Retrieved April 15, 2023, from https://greenlining.org/work/economic-equity/

9 Emmanuel Martinez & Lauren Kirchner. (2021, August 25). *The Secret Bias Hidden in Mortgage-Approval Algorithms – The Markup.* https://themarkup.org/denied/2021/08/25/the-secret-bias-hidden-in-mortgage-approval-algorithms

10 Tech Workforce – The Leaky Tech Pipeline. (n.d.). Retrieved April 15, 2023, from https://leakytechpipeline.com/pipeline/tech-workforce/

11 Alexandra George. (2018, December 11). *Thwarting bias in AI systems.* Carnegie Mellon University College of Engineering. https://engineering.cmu.edu/news-events/news/2018/12/11-datta-proxies.html

our systems, from automated soap dispensers that cannot "see" darker skin tones,[12] to employment algorithms trained on historical data that effectively replicate biases against applicants of color and women. In the age of algorithms, ADS expose the systemic racism embedded in human decisions. As Jon Kleinberg, esteemed computer science professor at Cornell University, aptly asserts, "algorithms can reveal our own biases" and make apparent that which has been hidden.[13]

Without ADS, it was possible for racial discrimination to happen at the individual level without systematic detection, even if discrimination was clearly a product of systemic racism. For example, a loan officer could individually determine a Black small business applicant with a healthy financial profile to be too high risk to approve for a loan by gathering what some might consider basic information – full name, credit history, previous residences, education. Even without making a judgment about the applicant's race through visual or auditory perception, the officer would likely make a racialized judgment based on these presumably benign – but actually highly racialized – factors. Racial discrimination is not immediately apparent when a decision comes down to a series of rote forms and a loan officer's opaque decision-making process. But when this process is codified into an ADS, the bias embedded in the factors becomes obvious as a result of the ADS' large-scale discriminatory outcomes. By nature of the algorithm's design, the racial discrimination embedded in society is uplifted in a way that race blind policy is incapable of addressing. As ADS rapidly advance, it is time that we enter a fourth policy era: race awareness.

Race aware policy acknowledges the historical and systematic subjugation of people based on race and, as per Ruha Benjamin, "takes seriously and address[es] the matter of how racism structures the social and technical components of design."[14] It also takes as fact the role of technology in the systems that impact our daily lives, such as in housing, employment, and banking. Because racial bias is embedded into all of our systems, a race aware policy framework rejects neutrality in policy and neutrality in technology.

Here, we will discuss race aware policy in the context of its intrinsic relationship with the algorithms that power ADS. It is these systems that are rapidly proliferating and amplifying the systemic biases that race blind policy attempts to hide.

12 Ruha Benjamin. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code* (1st ed.). Polity.

13 Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, & Cass Sunstein. (2018). *Discrimination in the Age of Algorithms.* Journal of Legal Analysis, 10, 113–174.

14 Ruha Benjamin. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code* (1st ed.). Polity.

Current algorithmic policy environment

Presently, algorithms are completely unregulated. To date, the private sector – specifically Big Tech – has dominated the technology policy environment, implementing voluntary methods and safeguards to build ethical credibility and curb any potential government regulation that might stifle innovation.[15] Big Tech firms like Microsoft, Google, and Meta house research teams that produce cutting-edge work on algorithmic fairness. They also fund significant nonprofit and academic leadership in this space.[16] Human-Centered AI, Responsible AI, and AI Ethics are now common terms in the tech industry, representing the fields of study that are focused on minimizing the potential harms of AI and identifying pathways to maximize societal benefit. But in the unregulated environment of ADS today, what does it mean to maximize societal benefit when the leading actors are corporations, ultimately driven to maximize profit?

Rodrigo Ochigame's 2019 essay in The Intercept entitled "The Invention of 'Ethical AI'" calls out Big Tech's investment in "ethical AI" as a "strategic lobbying effort" in which tech companies voluntarily follow self-created ethical guidelines in an effort to ward off future, more stringent regulation. This suggests that the corporate incentive to maximize societal benefit has its limits, and those limits are set by firms' ultimate profit motive. This is not to say that there is no societal benefit. Groundbreaking insights have come from research funded, in part or in full, by Big Tech (including many sources cited in this paper). Further, there are cases in which tech firms have responded to social pressure and voluntarily implemented initiatives that, though motivated by profit, demonstrate valuable precedent for tech firms' capabilities to implement a more race aware approach. For example, in 2020, #AirbnbWhileBlack started trending on social media, revealing stories of Airbnb hosts who faced book discrimination as a result of their race. In response to this public outcry, Airbnb partnered with Color of Change to produce Project Lighthouse, a platform-wide effort to minimize racial discrimination. Airbnb also formed a permanent anti-discrimination product team whose specific focus is to ensure greater equity and belonging across Airbnb's product experiences.[17]

Finally, in expert interviews supporting the development of this paper, we heard industry leaders share the urgent need for greater government regulation. One expert emphasized the need for regulatory "air cover," as tech firms are unwilling to do anything unconventional with race data. Although in most cases this is likely for valid reasons, it also means that tech firms who are aware of racial discrimination in their systems are unlikely to voluntarily address this discrimination without strong public pressure.

15 Rodrigo Ochigame. (2019, December 20). *How Big Tech Manipulates Academia to Avoid Regulation.* The Intercept. https://theintercept.com/2019/12/20/mit-ethical-ai-artificial-intelligence/

16 *Id.*

17 Airbnb anti-discrimination team. (2020). *Measuring discrepancies in Airbnb guest acceptance rates using anonymized demographic data.* Airbnb. https://news.airbnb.com/wp-content/uploads/sites/4/2020/06/Project-Lighthouse-Airbnb-2020-06-12.pdf

Given the rapid rise of AI systems and the AI policy advancements in other countries – such as the EU AI Act set to pass at the end of 2023[18] – we can be confident that algorithmic regulation is coming in the US. Soft policy frameworks like the Blueprint for an AI Bill of Rights[19] and the NIST AI Risk Management Framework[20] indicate what is possible with stronger government leadership on technology policy, but passing enforceable policies will be much more difficult. Race aware algorithms can only be realized with a stronger, race aware policy environment. As we look forward into what this could look like, we outline roles for the two most powerful stakeholders responsible for making this future a reality: private firms and government agencies. We recommend that private firms begin implementing the five principles of race aware algorithms outlined in this report. These principles are immediately adoptable and would be advantageous to get ahead of impending algorithmic regulation. However, these principles would not reach their full potential without the government also implementing significant policies that get at the root of algorithmic bias and build public infrastructure to support discrimination in the age of algorithms, as described through five race aware policy recommendations.

Finally, we want to call out the power of civil society in advancing a race aware future. It was only through public pressure that Airbnb created Project Lighthouse, just one of many examples in which civil society has built successful movements to combat technological inequities and government ineffectiveness. But we echo AI Now's sentiments in their 2023 report "Confronting Tech Power" – we must "employ strategies that place the burden on companies to demonstrate that they are not doing harm."[21] In particular, Big Tech is the most well resourced and most knowledgeable stakeholders to immediately address issues of algorithmic discrimination on their platforms. Rather than placing any responsibility on civil society, particularly minority racial demographics who are the most burdened by racial discrimination, we focus on actions the private sector must take to improve racial equity. We also place responsibility on the federal government to create a race aware policy environment that protects civil rights and enables race aware algorithms to meet their full potential.

18 Futurium | European AI Alliance - The EU AI Act's Risk-Based Approach: High-Risk Systems and What They Mean for Users. (n.d.). Retrieved April 22, 2023, from https://futurium.ec.europa.eu/en/european-ai-alliance/document/eu-ai-acts-risk-based-approach-high-risk-systems-and-what-they-mean-users

19 Office of Science and Technology Policy. (2022). *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People.* The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

20 US Department of Commerce, National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* https://doi.org/10.6028/NIST.AI.100-1

21 Amba Kak & Sarah Myers West. (2023). *AI Now 2023 Landscape: Confronting Tech Power.* AI Now Institute. https://ainowinstitute.org/2023-landscape

RECOMMENDATIONS

	RACE AWARE PRINCIPLES (PRIVATE SECTOR)	RACE AWARE POLICIES (GOVERNMENT)
1	Start by articulating what an algorithm should do	Apply a "rights-then-risk-based" framework
2	Contextualize fairness	Set standards for race data collection and related privacy safeguards
3	Track and report the impact of race on an algorithm	Require algorithmic audits in civil rights protected contexts, including auditing for racial discrimination
4	Optimize the algorithm for racial equity	Assign and equip government institutions to regulate ADS with ongoing multistakeholder consultation
5	Use race data to advance racial equity	Update anti-discrimination law for the age of algorithms

5 PRINCIPLES OF RACE AWARE ALGORITHMS

Principle 1

Start by articulating what an algorithm should do

Principle 2

Contextualize fairness

Principle 3

Track and report the impact of race on an algorithm

Principle 4

Optimize the algorithm for racial equity

Principle 5

Collect race data to test for racial discrimination

Principle 1: Start by articulating what an algorithm should do

In 2018, The Philadelphia Inquirer introduced us to Kat Payne, a housekeeper at the Philadelphia Marriott Downtown, the largest hotel in the city. After eight years on the job, Payne knew how to maximize the efficiency in her day based on room location and whether the guest was checking out or leaving for the day. When hotel management introduced a service optimization app, she could no longer follow her own schedule, crafted based on years of housekeeping experience. Instead, she had to follow the app's algorithm-based task prioritization system, which assigned her only a few tasks at a time, forcing her to blindly "zigzag" across the Marriott's 23 floors, pushing her cart back and forth through hallways spanning an entire block.[22] In Payne's experience, the algorithm should have helped housekeepers work in a more efficient manner, thereby earning the hotel higher profits. However, in reality the Marriott's task prioritization algorithm did a much better job of helping hotel management maintain stronger oversight and control of employee productivity, effectively creating an employee surveillance tool. Had the algorithm really been solving for task prioritization, the algorithm should

22 Juliana Feliciano Reyes. 2018. *"Hotel Housekeeping on Demand: Marriott Cleaners Say this App Makes their Job Harder."* The Philadelphia Inquirer, July 2, 2018. https://www.inquirer.com/philly/news/hotel-housekeepersschedules- app-marriott-union-hotsos-20180702.html

have been designed in consultation with the people most knowledgeable about housekeeping tasks – hotel housekeepers. Although we don't have information about the profitability of the hotel since utilizing the algorithm to prove this, it is evident that the algorithm failed to make Payne more productive.

Consider a different example – facial recognition algorithms. In 2020, the New York Times introduced us to Robert Julian-Borchak Williams, a Black man living in Detroit, Michigan. Williams was arrested and held in a detention center overnight after a facial recognition algorithm wrongly identified him as a shoplifter in a luxury goods store, prompting detectives to come to his home and arrest him in front of his family. Even after detectives realized they had arrested the wrong person, Williams continued to be held for a total of 30 hours, "released on a $1,000 personal bond."[23]

In the case of facial recognition algorithms, it seems that at least the conceptualization of the algorithm operates as intended – matching up faces in a database with faces in a visual. However, everything else is wrong. Facial recognition algorithms are reportedly wrong more than 90 percent of the time.[24] This inaccuracy disproportionately harms minority demographic groups whose data is unsurprisingly less accurate than that of white men. Researchers have proven that facial recognition technology "has a history of misidentifying young girls with dark skin" as well as "people of Asian descent" and "those who do not conform to gender norms."[25] Yet, improving the accuracy of facial recognition algorithms for law enforcement is also unjust. Law enforcement outright should not rely on automated systems to make life-altering decisions – including life or death decisions – about people. Further, the risk associated with the use of an algorithm is extremely high: once you are identified as high-risk on the algorithm, this high-risk label follows you wherever you go, whether it is accurate or not. Ultimately, law enforcement's use of facial recognition algorithms to criminalize individuals is an inherently racist objective, and in this case, articulating what an algorithm should do should result in concluding not to deploy an algorithm at all.

Starting by articulating what an algorithm should do requires designers to ask questions including: What objective is an algorithm meant to achieve? Are the variables and data accurate for the outcome an algorithm is intended to predict? Will using an ADS add value, or are the risks of harm too high? These questions are worthy first steps in the process of designing an algorithm because they encourage greater specificity and awareness of the societal implications of ADS.

23 Kashmir Hill. (2020, June 24). *Wrongfully Accused by an Algorithm.* The New York Times. https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html

24 Khari Johnson. (2022, March 7). *How Wrongful Arrests Based on AI Derailed 3 Men's Lives*. Wired. https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/

25 *Id.*

Principle 2: Contextualize fairness

"If we reduce fairness to a specific pinpoint such as a mathematical definition, we're missing the real conversation." [26]

What is fairness? In the algorithmic context, fairness is a loaded, divisive term. Deirdre Mulligan et al. describe the conflicting definitions of fairness in their paper "This Thing Called Fairness," where legal fairness refers to protecting individuals and groups from discrimination, social science considers fairness in terms of power dynamics and relationships, and quantitative fields see fairness as a mathematical equation.[27] Yet, as one interviewee shared, even "common mathematical definitions of fairness cannot be satisfied simultaneously."[28] "Equally accurate across groups" is a common term used in algorithmic fairness discussions, but its meaning is diluted by a range of mathematical applications. Does it mean equal opportunity, in which there is "equal false negative rates between groups?" Or could it mean counterfactual fairness, in which the "outcome probability remains the same if you change a sensitive feature?" Or does it simply refer to demographic parity, in which all groups have "equal probability of being assigned by the model to the positive class?"[29] Or, is debating mathematical fairness a distraction from truly achieving societal fairness overall?

Our expert interviews indicated that fairness is highly contextual, but what drives fairness across contexts is transparent reporting and diverse, contextually representative training data. One technologist we spoke with summarized, "Ultimately, so many algorithmic fairness questions come down to the data that you have."[30] We know that all datasets are unfair in some way because they are trained on biased, real world information. As another interviewee shared, "even tools you don't think interact with social constructs like pulse oximeters show differences in performances across racial groups because of the sociological design of the study and training data."[31]

Transparency tools like the Dataset Nutrition Label – a documentation tool that "enhances context, contents, and legibility of datasets"[32] – address these fairness tools by requiring dataset creators to articulate important information about the dataset including its intended use cases, the demographic groups represented in the data, and any potential risks associated with the data. Addressing fairness in this way allows designers to interrogate datasets and call out risks with the dataset upfront that

26 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

27 Deirdre K. Mulligan, Joshua A. Kroll, Nitin Kohli, & Richmond Y. Wong. (November 20019). *This Thing Called Fairness: Disciplinary Confusion Realizing a Value in Technology*. Association for Computing Machinery, 3. https://doi.org/10.1145/3359221

28 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

29 *Id.*

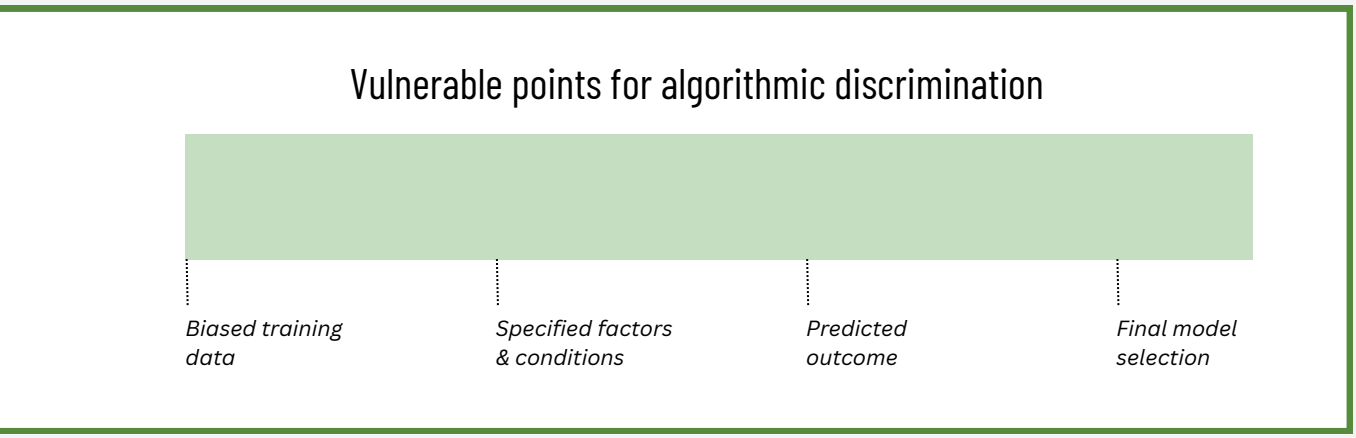
30 *Id.*

31 *Id.*

32 Data Nutrition Project. (n.d.). *The Dataset Nutrition Label.* Retrieved April 23, 2023, from https://labelmaker.datanutrition.org/

will inherently be replicated in the ADS trained on the dataset. It will also enable further testing and auditing of ADS at later stages of algorithmic design using more context-specific measures of fairness like testing for inclusion and accuracy across demographic groups. Although testing for fairness alone is not enough to achieve equity, contextualizing fairness in a transparent way upfront is critical to enable a shared understanding of a datasets' potential and limitations for advancing racial equity.

Principle 3: Track and report the impact of race on an algorithm



Algorithms that determine housing, employment, or loan eligibility are what Kleinberg calls "screening algorithms," in which applicants are evaluated for a specified outcome based on a set of independent variables. Far from a "black box," Kleinberg breaks down the three components of an algorithm that are vulnerable to discrimination.[33] First, algorithms require training data and yet, regardless of intent, all training data is biased in some way because the data is produced in a biased social environment. Importantly, this includes bias in the data not fed to the trainer. For example, if we feed the trainer data about executive leaders in the technology industry, it will inherently predict outcomes that advantage white, male, highly educated professionals because that is by far the dominant profile of executive tech leadership. In contrast, data about people of color in leadership positions in the tech industry is limited because it is a limited observed reality. As you disaggregate racial categories, these gaps are likely to become increasingly apparent. Ruha Benjamin says, "To the extent that machine learning relies on large, "naturally occurring" datasets 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."[34]

33 Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, & Cass Sunstein. (2018). *Discrimination in the Age of Algorithms.* Journal of Legal Analysis, 10, 113–174.

34 Ruha Benjamin. (2019). *Race After Technology: Abolitionist Tools for the New Jim Code* (1st ed.). Polity.

Second, the factors and conditions that make up the algorithm and the weight given to each factor also make way for discrimination. For example, credit score algorithms include the length of a borrower's credit history into its credit prediction. Yet, taking this factor into account disproportionately disadvantages borrowers who do not come from families with good credit scores or strong US financial literacy – commonly borrowers of color and immigrants.[35] Borrowers with this advantage can start building up a good credit score from a young age, gain credit history by being co-signed onto credit cards as a child through their parents, and begin adulthood with an advantageous credit score. Borrowers without this advantage may not be less likely to repay loans, but will be scored as such because they opened up a line of credit in their twenties with limited personal wealth. That said, the factors included in an algorithm can also be a way of correcting for bias. One could imagine diminishing the weight of the length of a borrower's credit history in an algorithm or removing it entirely to account for this bias. Additionally, alternative credit scoring systems are already addressing this challenge by adding in additional factors to the algorithm, such as repayment rate of utility bills and rent.

The third way algorithms can be biased by design is in the outcome the algorithm is instructed to predict. According to Ziad Obermeyer, this is the single most important determinant of bias in an algorithm's design. Obermeyer's study on algorithmic bias in health care systems is a prime example of why. When analyzing a common algorithm used to determine the magnitude of health care need, Obermeyer and his coauthors found that using health care cost as a proxy for health care need resulted in significantly under-providing care for Black patients because the algorithm assumed that White and Black patients seek out and receive health care at the same rate. In Obermeyer's Algorithmic Bias Playbook, he further adds that the predicted outcome is a reflection of our value system. In this health care example, the algorithm "values people who get health care more than people who need health care." He then concludes, "Algorithms are literal genies - they give us exactly what we ask for, even if we meant something very different." [36] Kleinberg would agree with Obermeyer, summarizing eloquently that "...the training algorithm can only optimize whatever outcome, candidate predictors and training data are given to it. The flip side is that this is all the algorithm does. It has no ulterior motives or hidden agenda. And much of what it does is transparent to us." [37]

35 *Unscoreable: How The Credit Reporting Agencies Exclude Latinos, Younger Consumers, Low-Income Consumers, and Immigrants.* (2019). National Fair Housing Alliance. https://www.congress.gov/116/meeting/house/108945/witnesses/HHRG-116-BA00-Wstate-BrownJ-20190226.pdf

36 Ziad Obermeyer, Rebecca Nissan, Michael Stern, Stephanie Eaneff, Emily Joy Bembeneck, & Sendhil Mullainathan. (2021). *Algorithmic Bias Playbook.* Chicago Booth Center for Applied Artificial Intelligence. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias/playbook

37 Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, & Cass Sunstein. (2018). *Discrimination in the Age of Algorithms.* Journal of Legal Analysis, 10, 113–174.

Breaking down the opportunities for bias into these three areas allows algorithm designers to systematically track and report the impact of race on an algorithm. Transparency tools like Timnit Gebru's Datasheets for Datasets[38] or Data Nutrition Labels[39] have already been created to address this need in training datasets, including guidance around intended use cases for datasets and specific use cases in which the dataset should not be used.[40] Greater transparency is also needed to track the impact of race on the factors chosen in the algorithm itself as well as the predictive variable. Using a similar reporting method, designers can make informed decisions about the systemic impact of race on the factors included in the algorithm and better tailor the predictive variable to predict the outcome it is intended to predict. The systemic nature of race in society means that reducing the impact of race on an algorithm may not always be feasible, but transparently tracking and reporting on race in this systematic way will encourage more informed deployment of ADS. It may also slow down the design process of an algorithm, hopefully resulting in more thoughtfully selected variables.

Principle 4: Optimize the algorithm for racial equity

Once you have identified racial bias in an algorithm, the natural next step is to solve for bias. In some cases, this might mean getting more or better training data. In others, different variables or variable weights may be selected to ensure the algorithm is adequately accounting for the impact of race on an algorithm and predicting the correct outcome. You may also consider the specific use cases to deploy an ADS and specific use cases where the same ADS may be harmful. However, in many cases, running through these solutions still results in a highly biased algorithm.

One final technical solution is in the model selection process. Algorithms are often described as having one optimal model that is selected based on accuracy. Yet, research points to the emerging concept of model multiplicity, in which "there often exist multiple models for a given prediction task with equal accuracy that differ in their individual-level predictions or aggregate properties…The existence of model multiplicity presents exciting opportunities because it offers model developers the flexibility to prioritize, and optimize for, desirable properties at no cost to accuracy, contrary to some conventional wisdom."[41]

This research on model multiplicity led by Emily Black emphasizes that there is not one single "best" model, particularly not when optimizing for accuracy. This supports the findings of Kit Rodolfa et al's 2021 paper entitled "Empirical observation of negligible fairness-accuracy trade-offs in machine learning for public policy," in which Rodolfa and co-authors leverage empirical evidence from machine learning models that have an "extensive impact on society…[such as] bail determination decisions,

38 Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, & Kate Crawford. (2021). *Datasheets for Datasets.* arXiv. https://arxiv.org/abs/1803.09010

39 Data Nutrition Project. (n.d.). *The Dataset Nutrition Label.* Retrieved April 23, 2023, from https://labelmaker.datanutrition.org/

40 *Id.*

41 Emily Black, Manish Raghavan, & Solon Barocas. (2022). *Model Multiplicity: Opportunities, Concerns, and Solutions.* Association for Computing Machinery, 850–863. https://doi.org/10.1145/3531146.3533149

hiring, healthcare delivery, and social service interventions."[42] Using contextualized mathematical definitions of fairness, their research found that "explicitly focusing on achieving equity...improved the equity of predictions with only a very modest decrease in accuracy."[43] The key conclusion of Rodolfa's work is that it is critical to contextually define fairness upfront and set it as a goal – as described above – that then influences the final model selection that is accurate, fair, and equitable. Black's research takes it one step further by suggesting that the flexibility model multiplicity allows should motivate legal pressure to ensure firms take adequate upfront care to select a model that reduces the risk of avoidable discrimination.[44]

Although technical solutions alone will not solve for equity, Rodolfa and Black's work illustrate that opportunity for equity can be maximized in the technical design of an algorithm with, at worst, a "negligible" fairness tradeoff. Firms should be expected to go through a careful model selection process, integrating a fairness and/or equity definition into the final model selection criteria. The Blueprint for an AI Bill of Rights also alludes to this, stating that firms should "evaluate multiple models and select the one that has the least adverse impact, modify data input choices, or otherwise identify a system with fewer disparities. If adequate mitigation of the disparity is not possible, then the use of the automated system should be reconsidered."[45]

Principle 5: Use race data to advance racial equity

"Researchers have noted that to the extent there are true differences across relevant groups in whether a data point is equally predictive, like if an SAT score is more predictive of college grades for one group than for another because of social circumstances, then including the demographic characteristic in the model is actually important to making sure the model is equally accurate across those groups."[46]

Race data is needed to advance racial equity in the age of algorithms. Examples from the healthcare industry, which has historically been more open to collecting and using race data "to address racial

42 Kit T. Rodolfa, Hemank Lamba, & Rayid Ghani. (2021). *Empirical observation of negligible fairness-accuracy trade-offs in machine learning for public policy*. Nature Machine Intelligence, 3, 896–904. https://doi.org/10.1038%2Fs42256-021-00396-x

43 *Id.*

44 Emily Black, Manish Raghavan, & Solon Barocas. (2022). *Model Multiplicity: Opportunities, Concerns, and Solutions.* Association for Computing Machinery, 850–863. https://doi.org/10.1145/3531146.3533149

45 Upturn. (2022, November 21). *Re: Advance Notice of Proposed Rulemaking on Commercial Surveillance and Data Security (Commercial Surveillance ANPR, R111004)* [FTC Public Comments].

46 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

and ethnic disparities in health outcomes, rather than compliance with anti-discrimination laws alone" [47] are helpful to illustrate this point. In "Awareness in Practice," Miranda Bogen cites a 2003 study that found "the presence of data on race and ethnicity does not, in and of itself, guarantee any subsequent actions...to identify disparities or any actions to reduce or eliminate disparities that are found. The absence of data, however, essentially guarantees that none of those actions will occur." [48] Recent research on examining family history and cancer risk also supports the impact of race aware algorithms on reducing racial discrimination. Anna Zink et al. first designed a study that found that family history for self-reported White participants is "strongly predictive," but much less predictive for self-reported Black participants because there is much less robust recording of family history of cancer for Black patients as a result of "historic disparities in access to care." They then compared two screening algorithms to model risk: one that is race blind (omits race as a variable) and another that is race aware (accounts for racial effects on the algorithm). The race aware algorithm was more accurate. [49]

One common solution to addressing racial discrimination in algorithms is not only to avoid using race as a variable, but also to remove variables that might be proxies for race. However, in practice, this is very difficult to do because race, by nature, impacts everything. It is impossible to remove race's impact on an algorithm's variables, and it is also impossible to know the degree to which race is correlated with a proxy variable. [50] As a result, there is no way to systematically approach removing proxy variables to reduce the influence of race on an algorithm's design. Barocas and Selbst also argue that removing proxies for race negatively impacts the accuracy of an algorithm. "Simply withholding these variables from the data mining exercise often removes criteria that hold demonstrable and justifiable relevance to the decision at hand." [51] Removing proxies is thus neither a feasible solution, nor desirable, because it does not serve the purpose of minimizing racial discrimination.

A better solution, then, is to do the opposite of removing race and its proxy variables – effectively attempting to blind an algorithm to race. As per the theme of this paper, we propose race aware algorithms by explicitly utilizing race data. The use cases of race data to advance racial equity in algorithms are twofold: first, to evaluate an algorithm for racial discrimination and second, to include

47 Miranda Bogen, Aaron Rieke, & Shazeda Ahmed. (2020). *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 492–500. https://doi.org/10.1145/3351095.3372877

48 *Summary: Race, Ethnicity, and Language Data: Standardization for Health Care Quality Improvement. (2018, October).* [Department of Health and Human Services]. Agency for Healthcare Research and Quality. https://www.ahrq.gov/research/findings/final-reports/iomracereport/reldatasum.html
as cited in Miranda Bogen, Aaron Rieke, & Shazeda Ahmed. (2020). *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 492–500. https://doi.org/10.1145/3351095.3372877

49 Anna Zink, Ziad Obermeyer, & Emma Pierson. (2023). *Race Corrections in Clinical Models: Examining Family History and Cancer Risk.* MedRxiv, 2023.03.31.23287926. https://doi.org/10.1101/2023.03.31.23287926

50 Solon Barocas & Andrew Selbst. (2016). *Big Data's Disparate Impact.* California Law Review, 104(3), 671–732.

51 *Id.*

race as a factor within an algorithm to increase racially equitable outcomes. On the former, our interviews with experts across civil society, industry, and academia revealed that race data is crucial for racial bias testing. As one interviewee shared, "It is incredibly important and useful to use race data in [algorithmic] evaluation."[52] Similarly, McKane Andrus' research with Partnership on AI interviewed 38 practitioners and found that "almost every participant described access to demographic data as a significant barrier to implementing various fairness techniques."[53] There is consensus that race data collection is a "prerequisite to progress," at minimum to understand existing disparities and evaluation. Having race data would almost enable future algorithmic audits to test for racial bias and enable designers to better account for biased variables in an algorithm's design.

Secondly, race data could be used as a factor within an algorithm to increase racially equitable outcomes. Among expert interviewees, this was an area of apprehension. Although most interviewees saw potential value in using race as a factor within an algorithm, in general, our interviews concluded that this is an area for greater exploration. There are many ways to improve the race awareness of an algorithm without including race as a factor in the model itself and these methods should be deployed first. Only in civil rights protected contexts, and only after extreme caution has been applied in the designing and testing process, should race be used as a factor in the algorithm itself to advance racial equity.[54] Pauline Kim's paper "Race Aware Algorithms" describes a hypothetical example from Cynthia Dwork's research: consider a model that predicts the most talented students by using proficiency in finance as a factor. If one racial group is more likely to encourage engineering for high performing students while another is more likely to steer high performing students toward finance, this model becomes systematically biased against the racial group that encourages high performers to study engineering. Without race as a factor in the algorithm in this case, the algorithm equalizes the impact of studying finance on all racial groups, rather than accounting for critical racial differences. In contrast, if race is in the model itself, it can be integrated in a way to correct for racial differences in finance proficiency to "improve accuracy and fairness for all individuals."[55]

Currently, race data in ADS is very limited and inconsistent in how it is collected, for what purpose, and how it is stored. In general, technology companies choose not to collect race data because the legal inconsistencies of race data collection in anti-discrimination law create ambiguity that breeds legal risk. For example, the Equal Credit Opportunity Act (ECOA) of 1974, which prevents discrimination in lending, explicitly bans the collection of race data among other protected characteristics except "in

52 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

53 McKane Andrus, Elena Spitzer, Jeffrey Brown, & Alice Xiang. (2021, March 1). *What We Can't Measure, We Can't Understand: Challenges to Demographic Data Procurement in the Pursuit of Fairness.* Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, NY, USA. https://dl.acm.org/doi/10.1145/3442188.3445888

54 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

55 Cynthia Dwork, Nicole Immorlica, Adam Tauman Kalai, Max Leiserson. (2018). *Decoupled Classifiers for Group-Fair and Efficient Machine Learning.* Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in Proceedings of Machine Learning Research 81:119-133 Available from https://proceedings.mlr.press/v81/dwork18a.html as cited in Pauline Kim. (2022). *Race-aware algorithms: Fairness, nondiscrimination and affirmative action.* California Law Review, 110, 1539-1596.

connection with a self-test" or "to determine an applicant's eligibility for a special purpose credit program."[56] In practice, very few companies have utilized self-tests or special purpose credit programs, although it is unclear why.[57] In contrast, the Home Mortgage Disclosure Act does allow for race data collection, including "using visual observation and surname analysis" when an individual declines to self-report.[58] Title VII of the Civil Rights Act similarly requires employers to collect race data in a "highly standardized" manner.[59] This data is then used to assess an employer's racial discrimination through a disparate impact (discriminatory outcomes) and disparate treatment (discriminatory intent) framework at any point in the recruitment or employment process.[60] Anti-discrimination law tends to be broad in its guidance beyond these specific legal requirements, causing companies to only collect race data where it is explicitly required. Outside of the healthcare setting, companies are generally reticent of voluntarily collecting race data because they do not want to be liable for non-compliant data use or discrimination. Importantly, companies also do not need explicit race data to discriminate based upon race. Whether intentional or not, there is a breadth of other data sources that provides a broad selection of proxies to use to provide an accurate picture of race is enough for their purposes.

Case Study: Meta Variance Reduction System

In 2022, the Department of Justice reached a landmark settlement with the social media giant Meta, formerly known as Facebook. The settlement determined that Meta's advertising delivery algorithms were discriminatory in a way that violated the Fair Housing Act. Meta's 2023 report "Toward fairness in personalized ads" thus details the measures the firm took to improve their housing, employment, or credit opportunity ad delivery algorithms, namely through the Variance Reduction System (VRS). One of the methods they deployed was removing features in the algorithm that could even remotely be used as a proxy for a protected characteristic (not only race). This resulted in the removal of over 50 variables, including everything from legal marijuana use, political affiliation, and years of driving experience.[61]

Where Meta removed any variables that could be potentially correlated with race within the algorithm itself, Meta then used aggregate demographic measurements to analyze the demographic distribution of ad delivery. If variance is detected, the VRS will implement an "adjusted strategy" to correct for the variance.[62] Meta's VRS is an elegant solution that demonstrates the capacity big tech firms have to address algorithmic discrimination in their systems, including through using race data.

56 Federal Reserve. (n.d.). Federal Fair Lending Regulations and Statutes: Equal Credit Opportunity (Regulation B). Consumer Compliance Handbook.

57 *Id.*

58 Miranda Bogen, Aaron Rieke, & Shazeda Ahmed. (2020). *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 492–500. https://doi.org/10.1145/3351095.3372877

59 *Id.*

60 *Id.*

61 Miranda Bogen, Pushkar Tripathi, Timmaraju, A. S., Mehdi Mashayekhi, Qi Zeng, Rabyd Roudani, Sean Gahagan, Andrew Howard, & Isabella Leone. (2023). *Toward fairness in personalized ads.* Meta. https://about.fb.com/wp-content/uploads/2023/01/Toward_fairness_in_personalized_ads.pdf

62 *Id.*

Note: Race data collection necessitates racial classification

It is important to acknowledge that race awareness necessitates classifying individuals into racial categories. The concept of classification has been well-criticized for its limitations in the field of sociology. Notably, the design of classification systems themselves are embedded with a range of biases and power constructs. First and foremost, racial classification begs the question, "Who decides one's race?"

In the policy environment, US census categories are the most common way of classifying race. The US census is updated every decade, and although the racial categories (along with data collection practices) have historically contributed to greater disenfranchisement of already disenfranchised groups, President Biden's Equitable Data Working Group[63] is advocating for more equitable, disaggregated data that is representative of the changing demographics of the nation. A 2022 Equitable Data Working Group report calling for updated, more inclusive federal race and ethnicity categories.[64] Yet, we acknowledge that this is an initial step in the right direction, not a comprehensive solution.

In her 2020 paper "Towards a Critical Race Methodology in Algorithmic Fairness," Alex Hanna and co-authors note that "the question of how best to operationalize race for the purposes of studying or mitigating different aspects of algorithmic unfairness has received little attention." Further, how to "conceptualize…the unique oppressions encountered by each group?"[65] By way of solutions, Hanna et al. suggests "adopting a multidimensional view" of race, acknowledging the contextual and fluid societal construct of race in which practices like self-identification and phenotype might have varying utility in different contexts. Hanna recommends collecting "multiple measures of race" where possible.[66] Similar to the recommended transparency approach of defining fairness upfront and providing rationale for model selection that maximizes opportunity for equity, we could integrate racial classification as well. Practitioners should define how race is classified in a certain context upfront and demonstrate critical analysis for how that particular choice of racial classification optimizes for more equitable outcomes.[67]

Racial classification also expands the risk of surveillance on minority racial identities. The contradiction of algorithms is that they need more data to be more fair, yet collecting more data is

63 The White House. (2022, April 22). *FACT SHEET: Biden-Harris Administration Releases Recommendations for Advancing Use of Equitable Data.* The White House. https://www.whitehouse.gov/briefing-room/statements-releases/2022/04/22/fact-sheet-biden-harris-administration-releases-recommendations-for-advancing-use-of-equitable-data/

64 Equitable Data Working Group. (2022). *A Vision for Equitable Data: Recommendations from the Equitable Data Working Group.* The White House. https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf

65 Alex Hanna, Emily Denton, Andrew Smart, & Jamila Smith-Loud. (2020). *Towards a Critical Race Methodology in Algorithmic Fairness.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 501–512. https://doi.org/10.1145/3351095.3372826

66 *Id.*

67 *Id.*

often extractive and reinforces societal inequities. Potential solutions for this are detailed in Policy Recomendation 2 (Set standards for race data collection and related privacy safeguards), drawing from research on data cooperatives[68] and data justice practices,[69] but we acknowledge that these solutions are imperfect. There is still much to be discovered about operationalizing race in the context of ADS, and more attention must be given to the field of racial classification "as an empirical problem[70] in its own right." As a starting point, we support (1) updating racial categories in the US census and (2) transparently reporting racial classification methodology enabling multiple measures of race are helpful starting points for acknowledging the tensions posed by racial classification.

OUR VISION FOR ALGORITHMIC GREENLINING

Given the voluntary policy environment characteristic of technology regulation, the idea of slowing down innovation through government intervention and even banning high-risk technologies can often sound radical and sometimes impossible. It is not. Slowing down technological systems that infringe on our ability to access civil rights is the basic role of government. What does require radical change is the shift in policy and government infrastructure required to meet the needs of our digital environment. This includes designing policies in a way that enables adaptability to rapidly changing technologies and greater interagency collaboration in government systems that, as a result of big tech platforms, need to regulate the same technology in different policy domains (ie: Facebook algorithms determining who sees ads for certain loan products, housing offers, and employment opportunities).

In "Algorithmic Bias Explained," we quote Cathy O'Neill's writing in her book, "Weapons of Math Destruction":

"Big Data processes codify the past. They do not invent the future. Doing that requires moral imagination, and that's something only humans can provide. We have to explicitly embed better values into our algorithms, creating Big Data models that follow our ethical lead. Sometimes that will mean putting fairness ahead of profit." [71]

68 McKane Andrus & Sarah Villeneuve. (2022, May 4). *Demographic-Reliant Algorithmic Fairness: Characterizing the Risks of Demographic Data Collection in the Pursuit of Fairness.* 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), Seoul, Republic of Korea. https://doi.org/10.1145/3531146.3533226

69 *Advancing research and practice on data justice—GPAI.* (n.d.). Retrieved April 22, 2023, from https://gpai.ai/projects/data-governance/data-justice/

70 Alex Hanna, Emily Denton, Andrew Smart, & Jamila Smith-Loud. (2020). *Towards a Critical Race Methodology in Algorithmic Fairness.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 501–512. https://doi.org/10.1145/3351095.3372826

71 Cathy O'Neil. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (1st ed.). Crown.

Although we would replace "fairness" with "equity," O'Neill's quote is still highly relevant. The principles of race aware algorithms and the five policy recommendations we share below envision an equitable future that looks nothing like the past. We advocate for "explicitly embedding" race to advance this vision so that our policy, legal, and sociotechnical systems will not only minimize racial discrimination, but tackle the root causes of racism. We propose an approach to regulating ADS that embeds equitable values into our systems – what we call algorithmic greenlining.

What is algorithmic greenlining?

Algorithmic greenlining is the ultimate goal of the race aware policies and principles proposed here. We envision a future in which algorithms build a digital infrastructure that corrects historical harms and actively contributes to greater racial equity. Once race data is more widely collected and available in protected environments, there is great opportunity to explore the use of this data to address historical biases by increasing minority racial representation in training data, exploring the equitable outcomes resulting from using race as a variable in an algorithm, and purposely advantaging historically disadvantaged groups in ways that actively contributes to greater racial equity. We agree with Pauline Kim's argument that race data collection and testing on its own is not affirmative action,[72] but our vision is for race aware algorithms to support the advancement of racially just policies including affirmative action and reparations. Race aware algorithms could be a tool to increase access to loans for historically disadvantaged groups. Race aware algorithms could be used to better target public health services or increase the uptake of welfare benefits among those who need it the most. One interviewee even suggested an adversity scoring system in which race is one variable among others that more directly captures the intersectionality of race with other factors such as class and gender. If our sociotechnical and policy systems can see the intersectionality and multidimensionality of race, the possibilities for achieving racial equity are endless.

Criteria for policy analysis

One interviewee shared, "Algorithms are inherently unfair and a more realistic goal is to identify the specific instances in which "we are okay" with them being used through a lens of advancing racial justice."[73] We recommend pursuing the following five race aware policy recommendations to create a policy environment and framework that helps us analyze when we are "okay" with the use of algorithms and puts in place the proper policy, legal, and technical safeguards to support rapid AI advancement. We believe policies that advance racial equity, confront private power, and meet the political moment will be the most strategic approach to bringing about algorithmic greenlining.

72 Pauline Kim. (2022). *Race-aware algorithms: Fairness, nondiscrimination and affirmative action*. California Law Review, 110, 1539–1596.

73 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

CRITERIA FOR POLICY ANALYSIS

1 **Advances racial equity**

Addresses the root causes of racial discrimination and challenging systemic inequities to bring about race awareness in pursuit of algorithmic greenlining

2 **Confronts private power**

Reinforces the role of government in protecting civil rights and removes power from big tech corporations to influence civil liberties

3 **Meets the political moment**

Politically feasible enough to be pursued with relative immediacy

Criteria	Policy 1	Policy 2	Policy 3	Policy 4	Policy 5
Advances racial equity					
Confronts private power					
Meets the political moment					

5 RACE AWARE POLICY RECOMMENDATIONS

Policy 1
Apply a "rights-then-risk-based" framework

Policy 2
Set standards for race data collection and related privacy safeguards

Policy 3
Require algorithmic audits in civil rights protected contexts, including auditing for racial discrimination

Policy 4
Assign and equip government institutions to regulate ADS with ongoing multistakeholder consultation

Policy 5
Update anti-discrimination law for the age of algorithms

Policy 1: Apply a "rights-then-risk-based" framework

Advancing technology policy should always consider human rights first. Because this paper is focused exclusively on areas protected by civil rights, this is assumed, but it's worth calling out. Human rights must be protected by law and enforced by all government institutions. Only once our rights are protected can we then apply a risk-based framework for regulation. This "rights-then-risk-based" framework models the European Union's General Data Protection Regulation (GDPR), which operationalizes citizens' digital rights and then includes a risk-based analysis on privacy as part of regulatory implementation.[74]

74 Fanny, Daniel Leufer, & Estelle Massé. (2023, January 13). *The EU should regulate AI on the basis of rights, not risks.* Access Now. https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

This "rights-then-risk-based" framework is also common in US policy. In the Voting Rights Act of 1965 (VRA), there are special provisions for certain geographic regions of the country where racial discrimination is particularly prevalent and embedded in the system. As a response, the VRA established a "coverage formula" that "identif[ies] those areas and provide[s] for more stringent remedies where appropriate."[75] The coverage formula has continued to be extended since its creation, most recently in 2006 for 25 years. Included in the formula is the use of a different "test or devices" with potential to result in disparate impact on the opportunity for different racial populations to register to vote, such as the availability of non-language voting information. States can seek a "bailout" of the special provision if they believe it is applied "overly inclusively."[76]

We propose applying a concept similar to the coverage formula to ADS used in areas protected by civil rights. Higher risk ADS should be assigned "special provisions" depending on the algorithm's potential impact on people's lives. Under this approach, more stringent remedies might look like a more comprehensive or more frequent algorithmic audit or an outright ban. Many cities, starting with San Francisco, have already banned government use of facial recognition systems.[77] Reflecting back to the first principle of race aware algorithms (Start by determining if an algorithm is appropriate), we want to ensure policy makes space for contexts in which algorithms may not be appropriate, whether as a result of the sensitive context or their risk to human rights. The EU AI Act, expected to become law by the end of 2023, is a worthy example of this. The EU AI Act groups AI systems into four risk categories ranging from low-risk to unacceptable risk. AI systems in the unacceptable risk category are not allowed on the EU market.[78]

Policy 2: Set standards for race data collection and related privacy safeguards

> **"Without clear guidance on how to go about race data collection and its implications, [tech companies] are extremely cautious because we need to meet stakeholder expectations of legitimacy."**[79]

There is consensus that race data collection is needed, at minimum, to understand and evaluate existing racial discrimination in algorithms. This information will enable the government to be a regulatory body that more effectively counters racial discrimination. It will also increase the

75 *Section 4 Of The Voting Rights Act.* (2015, August 6). https://www.justice.gov/crt/section-4-voting-rights-act

76 *Id.*

77 Nathan Sheard and Adam Schwartz. (2022, May 5). *The Movement to Ban Government Use of Face Recognition.* Electronic Frontier Foundation. https://www.eff.org/deeplinks/2022/05/movement-ban-government-use-face-recognition

78 *Futurium | European AI Alliance - The EU AI Act's Risk-Based Approach: High-Risk Systems and What They Mean for Users.* (n.d.). Retrieved April 22, 2023, from https://futurium.ec.europa.eu/en/european-ai-alliance/document/eu-ai-acts-risk-based-approach-high-risk-systems-and-what-they-mean-users

79 Anonymous. (2023, February). *Expert Interview with The Greenlining Institute on Race Aware Algorithms* [Personal communication].

effectiveness of algorithmic impact audits in measuring the impact of algorithms on race. However, our policy recommendation is not simply to make it easier for companies to collect race data. Because of the high sensitivity of race data, we urge the government to set clear, context-specific standards for race data collection and play a heavy role in ensuring the proper privacy safeguards are in place. The government should also detail the entity that is best positioned to collect, store, and use this data. Companies would then comply with these standards, in part, by implementing greater documentation of their bias testing efforts, including reporting tools like Timnit Gebru's Datasheets[80] for Datasets or Data Nutrition Labels.[81]

The three most common methods of race data collection are self reporting, BISG, and perceived race, described in greater detail in the table below.

METHODS OF RACE DATA COLLECTION	WHAT	EXAMPLE	BENEFITS	CHALLENGES
SELF-REPORTING	Respondents are provided the option to self-identify their race/ethnicity.	In the US, most employment applications will include an optional demographic survey that includes a question about race/ethnicity.	Enables individuals to self-identify their race, often in a format that offers standard classification alongside an option for self-classification.	Historically generates a low response rate, often making the data unusable.
BISG	Methodology developed by the RAND Corporation that uses address and surname from Census data to infer race. [82]	Consumer Financial Protection Bureau, Federal Trade Commission, Meta's Variance Reduction System	Reports a high accuracy rate across major US demographic groups. Able to be immediately implemented.	Replicates systemic bias and marginalization of respondents whose last names are not obviously indicative of race; ie: mixed race individuals, Native Americans. Ethical concerns with inferential methodology applied to race.
PERCEIVED RACE	Utilizes third-party providers to assign perceived race classifications based on photos and first name (Airbnb).	Airbnb Lighthouse Project, HMDA data	Has reported positive outcomes for Airbnb and seems to accurately describe peer-to-peer racial bias.	Risk of inaccuracy Relies on visual perception which duplicates harms of historical racism Reliance on harmful data labor practices

80 Timnit Gebru, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé III, & Kate Crawford. (2021). *Datasheets for Datasets.* arXiv. https://arxiv.org/abs/1803.09010

81 Data Nutrition Project. (n.d.). *The Dataset Nutrition Label.* Retrieved April 23, 2023, from https://labelmaker.datanutrition.org/

82 Marc N. Elliott & Steven C. Martino. (n.d.). *Bayesian Indirect Surname Geocoding (BISG).* RAND. Retrieved May 5, 2023, from https://www.rand.org/health-care/tools-methods/bisg.html

In reality, fitting the multidimensionality and fluidity of racial identity into a classification system is an imperfect solution. The act of collecting race data also requires collecting more data from disproportionately surveilled communities. Community trust in the government and private sector, particularly in terms of data collection, is extremely low. There is already public aversion to tech companies asking for race data, and adding in transparency around why race data is being collected and how the data will be used is almost meaningless – public perception does not trust tech companies or the government to use their data for only responsible purposes.

We recommend setting a long-term goal of collecting race data through self-reporting, generating a higher response rate through the proliferation of data trusts or data cooperatives, data governance approaches that strengthen the power people have on deciding how their data is provided, used, and stored. McKane Andrus and Sarah Villeneuve define data cooperatives as a decentralized data governance approach where data subjects pool their data together. They define data trusts as a more centralized governance mechanism that relies on a data "trustee" as a steward of pooled data.[83] These are two examples of participatory, collective approaches Andrus and Villeneuve recommend in their 2022 paper "Demographic-Reliant Algorithmic Fairness: Characterizing the Risks of Demographic Data Collection in the Pursuit of Fairness." The Global Partnership on AI has also published a suite of data justice guides that address what this looks like in a policy context.[84]

In the more immediate term, we recommend race data collection standards to support self-reported race where possible, and BISG to fill in gaps. We recommend that BISG be updated to follow the White House Equitable Data Working Group's recommended disaggregated race data categories,[85] and for the BISG to be updated more frequently than every 10 years to accommodate the rapidly evolving racial demographics of the United States. The government can also encourage greater accepted applications of race data collection by encouraging self-tests and special purpose credit programs, legally permissible contexts in which ECOA Regulation B allows race data collection.[86]

Additional race data standards should be context specific, require ongoing testing, and require the highest levels of data privacy and security. We recommend that this guidance be provided in greater detail by regulatory bodies after consulting with communities that are disproportionately surveilled and generating momentum toward actualizing long-term participatory data governance approaches.

83 McKane Andrus & Sarah Villeneuve. (2022, May 4). *Demographic-Reliant Algorithmic Fairness: Characterizing the Risks of Demographic Data Collection in the Pursuit of Fairness.* 2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), Seoul, Republic of Korea. https://doi.org/10.1145/3531146.353322

84 *Advancing research and practice on data justice—GPAI.* (n.d.). Retrieved April 22, 2023, from https://gpai.ai/projects/data-governance/data-justice/

85 Equitable Data Working Group. (2022). *A Vision for Equitable Data: Recommendations from the Equitable Data Working Group.* The White House. https://www.whitehouse.gov/wp-content/uploads/2022/04/eo13985-vision-for-equitable-data.pdf

86 Federal Reserve. (n.d.). Federal Fair Lending Regulations and Statutes: Equal Credit Opportunity (Regulation B). Consumer Compliance Handbook.

This would not only serve as a safeguard against the harm of less represented groups, it would also improve the consistency of race data collection methodologies and increase the accuracy of discrimination analyses, a key concern that has prevented race data from being collected under ECOA in the past. [87]

Policy 3: Require algorithmic audits in civil rights protected contexts, including auditing for racial discrimination

We support the spirit of the Algorithmic Accountability Act of 2022 currently before the Senate, particularly Section 4 which discusses impact assessments on ADS inclusive of racial discrimination analyses." [88] Audits have become one of the most commonly promoted policies to regulate algorithmic systems, potentially because it is a politically feasible solution that would have sweeping implications on private firms' use of ADS. Audit requirements would automatically slow down the process of developing and deploying an algorithm, putting the responsibility on private firms to test for bias from a socio-technical perspective.[89] New York City has been a leader in supporting algorithmic audits, passing the AI Bias Audit Act in 2021, although its enforcement has continued to be delayed, with the latest enforcement date set in July 2023. The NYC AI Bias Audit would put the responsibility on employers to audit ADS used in employment prior to deploying the ADS.[90]

We recommend improving the Algorithmic Accountability Act of 2022 by adding greater specificity to auditing for racial discrimination and creating industry-specific policy guidelines for what this audit looks like. The audit should have specific requirements for reporting on the ways in which private firms have evaluated the training data, variables and associated weights, and predicted outcome, essentially ensuring that Race Aware Principle 3 was followed. Algorithms should be tested against a standard set of guidelines for the relevant sector, as well as against a specified fairness metric pre-defined in the firm's transparency documentation. The audit should also test for multiple models and ensure the least discriminatory model(s) are selected.

Auditing for racial discrimination should apply the standard disparate impact and disparate treatment framework from Title VII of the Civil Rights Act which relates to employment discrimination, as well as a negligence framework, implying that companies are expected to uphold specific standards for minimizing racial discrimination outlined in Race Aware Principle 3, with proper documentation and processes to demonstrate thorough bias testing, transparency, and data privacy. Applying a

87 Miranda Bogen, Aaron Rieke, & Shazeda Ahmed. (2020). Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 492–500. https://doi.org/10.1145/3351095.3372877

88 S.3572—Algorithmic Accountability Act of 2022, S.3572, 117th Congress, 2D Session (2022). https://www.congress.gov/bill/117th-congress/senate-bill/3572/text

89 Amba Kak & Sarah Myers West. (2023). *AI Now 2023 Landscape: Confronting Tech Power.* AI Now Institute. https://ainowinstitute.org/2023-landscape

90 Roy Maurer. (2023, April 6). *NYC AI Bias Law's Enforcement Date Postponed Again.* SHRM. https://www.shrm.org/resourcesandtools/hr-topics/talent-acquisition/pages/nyc-ai-bias-law-enforcement-date-postponed-again.aspx

negligence framework also puts the responsibility on firms to properly and more transparently document the algorithmic design process, more deeply consider the societal impacts of their technologies, and explicitly build anti-discrimination practices into ADS.

As we have seen with the NYC AI Bias Audit and the California Consumer Privacy Act, it may be more politically feasible to require audits at the local or state level, and perhaps at the sector level, before implementing it across all civil rights protected areas at the federal level. Doing so would also put pressure on the federal government to more urgently create standardized federal regulations.

Policy 4: Assign and equip government institutions to regulate ADS with ongoing multistakeholder consultation

As algorithms, ADS, and more broadly AI/ML systems become increasingly integrated into our daily lives, we need to clearly assign and equip responsible government institutions to lead on AI governance. Countries including the UK and Singapore have established central AI governing bodies to create cohesive AI strategies and govern AI research and development. The US government has a growing number of federal institutions with offices dedicated to AI development.[91] The NIST AI Risk Management Framework is housed in the Department of Commerce[92] and the Blueprint for an AI Bill of Rights is under the White House Office of Science and Technology Policy.[93] The Federal Trade Commission (FTC) has also created an Office of Technology.[94] Under Chair Lina Khan's leadership, the FTC has made perhaps the most pivotal strides cracking down on Big Tech power and regulating algorithmic harms, including charging Facebook $5 billion for privacy violations in 2019.[95] In May 2023, Khan released a public statement urging policy enforcers and regulators to be "vigilant" against the harms of AI tools, adding that "the FTC is well equipped with legal jurisdiction to handle the issues brought to the fore by the rapidly developing AI sector, including collusion, monopolization, mergers, price discrimination and unfair methods of competition."[96]

As is detailed in the Algorithmic Accountability Act of 2022, we recommend urgently increasing the technical capacity of government staff and assigning responsible bodies to have the authority to lead

91 Niklas Berglind, Ankit Fadia, & Tom Isherwood. (2022, July 25). *AI in government: Capturing the potential value.* McKinsey & Company. https://www.mckinsey.com/industries/public-and-social-sector/our-insights/the-potential-value-of-ai-and-how-governments-could-look-to-capture-it

92 US Department of Commerce, National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* https://doi.org/10.6028/NIST.AI.100-1

93 Office of Science and Technology Policy. (2022). *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People.* The White House. https://www.whitehouse.gov/ostp/ai-bill-of-rights/

94 *FTC Launches New Office of Technology to Bolster Agency's Work.* (2023, February 16). Federal Trade Commission. https://www.ftc.gov/news-events/news/press-releases/2023/02/ftc-launches-new-office-technology-bolster-agencys-work

95 *FTC Imposes $5 Billion Penalty and Sweeping New Privacy Restrictions on Facebook.* (2019, July 24). Federal Trade Commission. https://www.ftc.gov/news-events/news/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy-restrictions-facebook

96 Khan, L. M. (2023, May 3). *Opinion | Lina Khan: We Must Regulate A.I. Here's How.* The New York Times. https://www.nytimes.com/2023/05/03/opinion/ai-lina-khan-ftc-technology.html

data collection and auditing.[97] This would require expanding the scope of institutions like the FTC who have been leading on AI regulation, or creating a new federal AI agency working in collaboration with existing agencies like the Department of Housing and Urban Development (HUD) and the Consumer Financial Protection Bureau (CFPB) who bring expertise on the historical context of algorithms in housing and banking. Through collaboration, this AI agency would set contextual standards in each industry where algorithms are used and enforce audit requirements that specifically test for racial bias as it might arise in each unique context.

Further, part of the expertise required of algorithmic systems lies outside of the government and private sector. Alondra Nelson, former Deputy Director of the Office of Science and Technology Policy at the White House who spearheaded the Blueprint for an AI Bill of Rights, argues for the role of non-expert consultation with government institutions. She says,"it is tremendously important that people who are not experts understand that they can have a role and a voice here."[98] Thus, in addition to prioritizing government technical expertise and clearly assigning regulatory responsibility for AI technologies, it is also critical for the assigned governing body to develop a consultative process that continually engages with multistakeholder representatives of civil society as experts in the societal implications of technology, centering most impacted communities.[99] This multistakeholder unit could function like an oversight board and should include developers, statistics experts, and representatives of impacted communities. This policy recommendation would require a fundamental shift in government infrastructure, both in terms of interagency collaboration and in developing a more effective process for civil society consultation. Although this recommendation does not quite meet the current political moment, we include it here because more immediate policy actions to regulate algorithms should strategically work toward this ultimate outcome.

Policy 5: Update anti-discrimination law for the age of algorithms

Our final policy recommendation is admittedly the least politically feasible recommendation. Yet across our interviews and literature review we found that algorithmic fairness is often analyzed through the lens of anti-discrimination law, yet anti-discrimination law is inadequate to address the harms of algorithmic discrimination. Anti-discrimination law must be more broadly adapted to capture the many ways discrimination can occur through algorithms.

Although we ultimately recommend a holistic reconsideration of anti-discrimination law, we uplift one specific update that does meet the political moment: legalizing race data collection under ECOA Regulation B, with greater data collection guidance that is consistent across sectors. This means applying learnings from special purpose credit programs and self-tests to develop context-specific

97 S.3572—Algorithmic Accountability Act of 2022, S.3572, 117th Congress, 2D Session (2022). https://www.congress.gov/bill/117th-congress/senate-bill/3572/text

98 Ezra Klein. (2023, April 11). *Opinion | What Biden's Top A.I. Thinker Concluded We Should Do.* The New York Times. https://www.nytimes.com/2023/04/11/opinion/ezra-klein-podcast-alondra-nelson.html

99 Anonymous. (2023, February). Expert Interview with The Greenlining Institute on Race Aware Algorithms [Personal communication].

standards for race data collection, data privacy and storage safeguards, and race classification methods. We believe this is politically feasible because the Home Mortgage Disclosure Act of 1975 (HMDA), passed just one year after ECOA, does allow for race data collection. HMDA was enacted as a response to redlining and requires comprehensive public data reporting on mortgage loans. "HMDA's initial reporting requirements involved publicizing geographic data about lending patterns…[and was later amended] to call for reporting of sensitive attribute data on borrowers' gender, race, income, and other categories."[100] Race data collection in mortgage loans has not demonstrated any positive or negative impact on racial discrimination as a result of race data collection. Using HMDA as an example, the Federal Reserve Board (FRB) has reconsidered allowing race data collection twice since ECOA was enacted, most recently in 2003 with support from the Department of Justice, the Department of Housing and Urban Development, small businesses, community organizations, and more.[101] Ultimately, "the FRB reasoned that making this a voluntary action could result in incomplete data collection and inconsistent data formatting that would hinder cross-market comparison between creditors."[102]

Two decades have passed since ECOA Regulation B was last considered, and although the reasoning for maintaining the ban in 2003 may have been warranted, the age of algorithms creates a very different setting. Private firms can now see race through proxies and ECOA's ban on race data collection now runs counter to its objective of protecting civil rights. Responding to the FRB's concerns around incomplete or inconsistent data collection, anti-discrimination law should instead provide greater guidance on race data collection and classification methodologies, as detailed in Policy Recommendation 2. In addition, where race data collection is legal, we believe there must be greater guidance and standard setting from the government to both encourage race data collection and ensure it is done in a protected manner that advances racial equity. In short, we encourage leveraging learnings from algorithmic transparency improvements to inform updates to anti-discrimination law around race data collection. One of the most immediate ways race data collection would improve the effectiveness of anti-discrimination law is by creating a clear, transparent data source to evaluate racial discrimination in ADS and prosecute with evidence when discrimination has occurred.

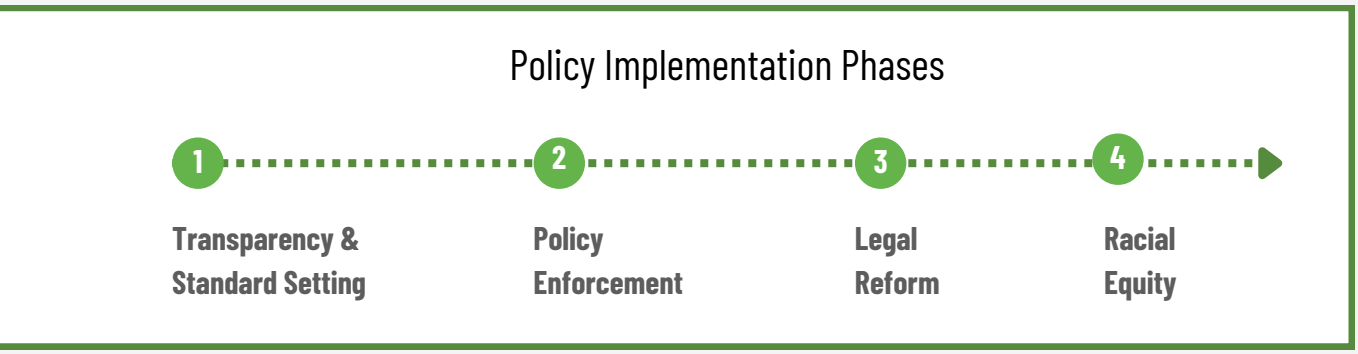
100 Miranda Bogen, Aaron Rieke, & Shazeda Ahmed. (2020). *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 492–500. https://doi.org/10.1145/3351095.3372877

101 Board of Governors of the Federal Reserve System, Proposed Rule re: Equal Credit Opportunity, 64 FR 44582, 44585 (Aug. 16, 1999), available at https://www.govinfo.gov/content/pkg/FR-1999-08-16/pdf/99-20598.pdf as cited in Upturn. (2023, March 6). *Re: Privacy, Equity, and Civil Rights Request for Comment (NTIA-2023-0001) [Public Comment to NTIA].* https://www.govinfo.gov/content/pkg/FR-1999-08-16/pdf/99-20598.pdf

102 Miranda Bogen, Aaron Rieke, & Shazeda Ahmed. (2020). *Awareness in Practice: Tensions in Access to Sensitive Attribute Data for Antidiscrimination.* Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 492–500. https://doi.org/10.1145/3351095.3372877

IMPLEMENTATION TIMELINE

We propose a phased implementation timeline, starting with greater transparency and standard setting, including the immediate voluntary adoption of the five principles of race aware algorithms. As argued above, these principles cannot reach their full potential until they are enforced by policy that pushes private firms to step beyond the bounds of profit maximization. Policy would then be written into law, and all three of these steps combined would make way for true racial equity. Although the current political environment challenges equitable policies like affirmative action and reparations, our hope is that by the time we write race aware policy into law, we will have created a more conducive environment to systematically advance racial equity. This is the thriving future The Greenlining Institute is building today. Below, we share an illustrative implementation timeline following four phases described above:



Phase 1: Transparency & Standard Setting

#	Action	Lead
1	**Increase public pressure for greater algorithmic transparency and greater adoption of race aware algorithms.**	Private sector
2	**Set transparency standards by making all government screening algorithms transparent and publicly accessible.**	All levels of government

Phase 1: Transparency & Standard Setting (Cont.)

#	Action	Lead
3	**Publish targeted, industry-specific guidance on how to collect race data, methods of race data classification, and the required privacy safeguards around collecting it.**	State and/or federal government
4	**Incentivize financial institutions to increase the use of self-tests and special purpose credit programs to collect race data in a highly regulated, protected context to gather more information about how race data can minimize lending discrimination**	Federal government
5	**Advance legislation around algorithmic transparency, ADS disclosures, and data privacy. Similar to CCPA and the NYC AI Bias Audit, starting at the local or state level may generate momentum more quickly and put pressure on federal regulation.**	Local and/or state government
6	**Convene a multi-stakeholder oversight board to support the development of government regulation of ADS in the context of racial discrimination.**	State and/or federal government
7	**Assign and equip the responsible governing bodies for algorithmic audits and regulatory enforcement.**	Federal government
8	**Pursue greater efforts to determine the contexts in which ADS should be banned and build that into a race aware policy framework.**	Local, state, and/or federal government

Phase 2: Policy Enforcement

#	Action	Lead
1	**Pass legislation that requires rigorous algorithmic audits in civil rights protected areas. Audits should clearly specify their role in auditing for racial bias in (1) training data and (2) testing multiple outcomes in the screener.**	Federal government
2	**Fund and staff responsible governing bodies for algorithmic audits and regulation.**	Federal government
3	**Update federal guidance on permissible methods and contexts of race data collection and privacy safeguards based on learnings from Phase 1.**	Federal government
4	**Pass legislation that bans ADS in specific contexts and applies a "rights-then-risk-based" approach to the use of ADS in civil rights protected areas.**	Local, state, and/or federal government
5	**Facilitate an ongoing feedback loop between the multi-stakeholder oversight board and governing bodies to support the rapidly evolving implementation of AI policy.**	State and/or federal government

Phase 3: Legal reform

#	Action	Lead
1	**Update anti-discrimination law to adapt to a race aware policy framework and algorithmic context.**	Federal government
2	**Update legal analysis of critical civil rights frameworks including affirmative action to better specify what is considered affirmative action under ADS**	Supreme Court

Phase 4: Racial Equity

#	Action	Lead
1	**Explore algorithmic greenlining, such as using race as a variable in ADS in contexts where it has been proven to reduce racial discrimination.**	Federal government
2	**Explore the use of race aware algorithms to purposely advantage historically disadvantaged groups. In other words, race aware algorithms could be a form of affirmative action and/or reparations.**	Local, state, and/or federal government

CONCLUSION

When we talk about regulating screening algorithms that impact our civil rights, we are not regulating black boxes, we are regulating ourselves. In this white paper, we attempt to break down screening algorithms to illustrate the role of human bias in ADS and guide the policies and principles to design algorithmic systems for racial equity.

ADS are making real decisions on real people's lives now, and to summarize the sentiment heard among our interviews, greater regulation, guidance, and consistency of practice is urgently needed. The government and private sector must take leading roles to allocate its wealth of resources and power to advance greater race awareness. We share the five principles of race aware algorithms and five race aware policy recommendations in hopes of spurring rapid progress where there is consensus and productive conversation where action requires more thorough debate. There is still much to be discovered and defined about ADS – this is the start.