

Racism in, Racism out

A Primer on Algorithmic Racism



ACKNOWLEDGMENTS

This report was written by Jane Chung, Big Tech Accountability Advocate at Public Citizen, with assistance from Scott Hulver. It was edited by Alan Zibel, Robert Weissman, and Rick Claypool. The author is indebted to external experts who generously reviewed the report, including Amba Kak, Amy Traub, Erin Simpson, and Frank Pasquale. Any mistakes or omissions are the author's own.

"Racism in, racism out" is a phrase first used by Hamid Khan, an organizer with Stop LAPD Spying Coalition, to describe predictive policing, an algorithmic system discussed later in this review.

ABOUT PUBLIC CITIZEN

Public Citizen is a national non-profit organization with more than 500,000 members and supporters. We represent consumer interests through lobbying, litigation, administrative advocacy, research, and public education on a broad range of issues including consumer rights in the marketplace, product safety, financial regulation, worker safety, safe and affordable health care, campaign finance reform and government ethics, fair trade, climate change, and corporate and government accountability.

Contact Public Citizen

Main Office
1600 20th Street NW
Washington, D.C. 20009

Capitol Hill
215 Pennsylvania Avenue SE, #3
Washington, D.C. 20003

Texas Office
309 E 11th Street, Suite 2
Austin, Texas 78701

Phone: 202-588-1000

Phone: 202-546-4996

Phone: 512 477-1155

For more information, please visit www.citizen.org.

Table of Contents

Executive Summary	2
Summary of Review	4
Introduction	5
Part 1: Review of harms of algorithmic racism	10
Auto insurance premiums are more expensive	10
Credit scores are lower	11
Mortgages are more expensive or altogether inaccessible	13
Students get rejected from better schools and assigned worse grades	14
Patients are denied life-saving care	17
Criminal justice system is more punitive	18
Communities are over surveilled and policed	20
Part 2: Questions about existing principles	26
Part 3: Recommendations for next steps	35
Learn by expanding on principles and conducting deeper audits	35
Address questions posed in Part 2 of this review	35
Include race, gender, and other protected classes in algorithmic auditing	36
Enable and mandate public and third-party audits	38
Create new rules for algorithmic design, regulation, and governance	39
Enact principles of design justice to create just algorithms	39
Litigate unsafe products and false claims under products liability law	40
Consider disparate impact and discrimination per se frameworks	42
Update existing regulation to promote transparency	43
Tailor accountability standards to different sectors and uses	44
Proactively enforce through the federal government	46
Enforce algorithmic accountability in the federal government	46
Empower existing agencies for enforcement	48
Create a new data protection agency to fill in gaps	49
Conclusion	52

Executive Summary

Background

Algorithmic bias leads to discrimination and harm for communities of color, women, trans people, and gender non-conforming individuals. Yet, while the algorithmic perpetuation of discrimination is relatively new, the harms are not. Issues of discriminatory and racist housing policy, credit terms, and policing are as old as mortgages, banks, and law enforcement themselves.

That is why some refer to algorithmic discrimination as “algorithmic redlining”,¹ “algorithmic Jim Crow”,² or the “new Jim Code”,³ appropriately tracing the now technology-driven discrimination to a history of biased and inequitable policy. Specifically, this review will focus on how predictive algorithms, which try to predict the likelihood of certain social and behavioral outcomes, exacerbate racial discrimination and bias leading to serious economic, physical, and social harm.

Addressing “algorithmic racism”, a phrase coined in a recent paper by Demos and Data 4 Black Lives,⁴ not only requires an analysis of the technology that executes decisions, but also the context of pervasive economic inequity in which the technology was created and is operated. Fully addressing algorithmic racism is likely impossible without fully addressing the broader context of racism. Nonetheless, algorithms pose problems and harms that demand focus in addition to broader economic and social remedies.

Government policy is falling further and further behind in developing appropriately scaled solutions to address algorithmic racial bias and discrimination.⁵ This review aims to help kickstart the search for policy solutions

¹ James A. Allen, [The Color of Algorithms: An Analysis and Proposed Research Agenda for Deterring Algorithmic Redlining](#), 46(2) FORDHAM URBAN L. REV. 219 (2019).

² Margaret Hu, [Algorithmic Jim Crow](#), 86(2) FORDHAM L. R. 633 (2017).

³ Ruha Benjamin, *Race After Technology* (2019).

⁴ Yeshimabeit Milner & Amy Traub, [Data Capitalism + Algorithmic Racism](#) (2021).

⁵ This review is focused on issues of racial injustice. In this review, references to “algorithmic bias,” “algorithmic discrimination” and “algorithmic justice,” generally refer to racial bias, discrimination and justice. Growing reliance on algorithms raise many other issues of bias and discrimination which are not the focus of this review.

that seek to achieve algorithmic justice. We do so by posing questions that will expand the conversation around algorithmic racism and help disentangle policy implications, as well as centralizing and condensing policy proposals that already exist. The goal of this review is to bring organizers, advocates, academics, and policymakers together toward a framework for action.

There will not be a single silver bullet to address the wide-ranging harms of algorithmic racism, not only because it is a complex policy problem, but also because algorithmic systems are designed in, by, and for a world ordered on racial and economic injustice. It will take a long-standing, concerted, and collaborative effort among policymakers, enforcers, organizers, advocates, and technologists to implement many different remedies toward rectifying algorithmic racism; but the fight doesn't end there. The fight to end algorithmic racism necessarily requires a continued toward a fight for economic and social justice at large.

Algorithms may reflect or worsen bias and discrimination based on gender, religion, ability/disability, age, and nationality, among other attributes.

Summary of Review

Algorithmic decision systems have been demonstrated to replicate and exacerbate racial bias in the following ways:

Auto insurance is more expensive.

Communities of color pay 30% more for auto insurance premiums than whiter communities with similar accident costs.

Credit scores are lower.

White homebuyers have credit scores 57 points higher than Black homebuyers, and 33 points higher than Latinx homebuyers.

Mortgages are more expensive or altogether inaccessible.

Higher, discriminatory mortgage prices cost Latinx and Black communities \$750 million each year. At least 6% of Latinx and Black applications are rejected but would be accepted if the borrower were not a part of these minority groups.

Students get screened out of better schools and assigned worse grades.

In New York City, Black and Latinx students are admitted to top schools at half the rate of white and Asian students. At some universities, Black students are up to 4 times as likely to be labeled 'high risk' as white students.

Patients are denied life-saving care.

White patients with the same level of illness were assigned higher algorithmically determined risk scores than Black patients. As a result, the number of Black patients eligible for extra care was cut by more than half.

Criminal justice system is more punitive.

Black defendants are 45% to 77% more likely to be assigned higher risk scores than white defendants.

Communities are over surveilled and policed.

Black individuals were targeted by predictive policing for drug use at twice the rate of white individuals. Non-Black people of color were targeted at a rate 1.5 times that of white individuals. Notably, the actual pattern of drug use by each race is comparable across the board.

Introduction

Algorithmic systems are sets of rules used along with data and statistical analyses in calculations for decision making, or to aid decision making. People and organizations have always developed rules and criteria to make decisions, and many times, in discriminatory ways.⁶ But algorithms do not autocorrect for human bias and discrimination. In many cases, unless algorithms are intentionally designed to account for the legacy of and ongoing systems of discrimination, inequality, and bias, they will replicate and exacerbate racial inequity.

Below we review a few of the many sources of algorithmic discrimination, including human bias and biased training data. Other sources of discrimination include biased algorithmic models themselves.⁷

First, the technologies humans create inevitably reflect the biases that individuals or groups carry. The problem of algorithmic bias and discrimination is exacerbated by the fact that computer technology industries, especially in the field of artificial intelligence (AI), overrepresent men, and underrepresent Black and Latinx communities⁸. Technologists in these fields may thus be particularly less attuned to the potential for bias.⁹ As data scientist Cathy O’Neil has written, “models are opinions embedded in mathematics,” and are thus not immune to encoding the opinions and biases of the technologists designing them.¹⁰

Researchers at New York University’s AI Now Institute have written at length about how lack of diversity in the AI industry and discrimination in AI systems are deeply intertwined. They conclude that addressing bias in algorithmic systems requires addressing AI workforce biases and discrimination.¹¹

⁶ Allen, *supra* note 1.

⁷ Ansgar Koene, Chris Slifton, Yohko Hatada, et. al., [A Governance Framework for Algorithmic Accountability and Transparency](#), EUROPEAN PARLIAMENT: SCIENTIFIC FORESIGHT UNIT, PANEL FOR THE FUTURE OF SCIENCE AND TECHNOLOGY, 20 (April 2019); @rajiinio, TWITTER (March 27, 2021).

⁸ Sarah Myers West, Meredith Whittaker, & Kate Crawford, [Discriminating Systems: Gender, Race, and Power in AI](#), NEW YORK UNIVERSITY AI NOW INSTITUTE (April 2019).

⁹ *Id.*

¹⁰ Cathy O’Neil, *Weapons of Math Destruction* (2016).

¹¹ West, *supra* note 9.

Another potential source of algorithmic bias is incomplete or unrepresentative training data used to create algorithms.¹² If the training data is biased, the model that learned from that data can be biased as well.¹³ For example, if medical trials only focus on results from wealthy, white patients, algorithmic systems designed using this training data may not accurately predict medical outcomes for non-wealthy, non-white patients.

These biases are at times unintentional, but a reflection of human bias all the same. An E.U. Parliament review of algorithmic accountability aptly summarizes, “Human values are (often unconsciously) embedded into algorithms during the process of design through the decisions of what categories and data to include and exclude.”¹⁴

Third, bias and discrimination can be caused by algorithmic decision systems that are complete and representative of real-world conditions—but those conditions are inevitably shaped by historical racist policy. When these biases are not corrected for, algorithmic systems can exacerbate racist outcomes.¹⁵ Amy Traub, a researcher at Demos, explains how this happens in the realm of credit scoring:

“Credit scores never formally take race into account. Yet these metrics cannot be race-neutral because they draw on data about personal borrowing and payment history that is shaped by generations of discriminatory public policies and corporate practices.

“From the American economy’s earliest roots in chattel slavery that treated Black people as property to more recent policies like redlining, Black and brown families have been systematically excluded from wealth-building opportunities that benefitted white families...

¹² [ALGORITHMIC JUSTICE LEAGUE](#).

¹³ Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. U. L. REV. 1, 4 (2014) (“Because human beings program predictive algorithms, their biases and values are embedded into the software’s instructions, known as the source code and predictive algorithms. Scoring systems mine datasets containing inaccurate and biased information provided by people.”); Mary Madden et al., *Privacy, Poverty, and Big Data: A Matrix of Vulnerabilities for Poor Americans*, 95 WASH. U. L. REV. 53, 86 (2017) (“While data analytics is touted for its ability to reduce human biases, it often merely replicates them.”); Kate Crawford, [Artificial Intelligence’s White Guy Problem](#), N.Y. TIMES (June 25, 2016).

¹⁴ *Id.* at 22.

¹⁵ For additional potential sources of bias and discrimination, see [@rajiinio](#), TWITTER (March 27, 2021).

“It is little surprise that, locked out of wealth passed down across generations, Black and brown consumers now disproportionately show up in the data as worse credit risks. And because consumers with poor credit are more likely to be denied loans or charged higher interest rates, the cycle of disadvantage reinforces itself.”¹⁶

To illustrate how these challenges apply to real-world decisions, we can assess how these biases take hold in the process of another common financial transaction: mortgage lending. Several decisions must be made before making a mortgage loan: Who gets a loan? At what rate? What information must a customer provide to get a loan?

In the past, bank employees may have made these decisions by comparing data points about customers like household income or employment status against pre-determined criteria. There may have been bias embedded in these decisions, as humans are all biased, but the entry points of biases may have been more traceable because decisions were made on paper, with fewer steps.

Yet with the introduction of decision-making algorithms, computers now make more complex decisions—oftentimes with more data points, more complex calculations, and more complex criteria. No longer are decisions written out on paper, easy to decipher, audit, and share with others. Instead, computers gather countless data points from around the web that go beyond the financial transactions that previously informed a credit score, and in turn, lending terms.¹⁷

As a result, the algorithmic decision-making process is less transparent, less accessible, and more difficult to explain or understand—and thus, more difficult to audit and regulate. This lack of transparency is the reason many call decision-making algorithms, ‘black box’ algorithms.¹⁸ There is no way to look inside and see how the sausage is made.

¹⁶ Amy Traub, [How Biden’s Plan for a Public Credit Registry Will Build Economic Power for Black and Brown Communities](#), DEMOS (November 16, 2020); Amy Traub, [Establish a Public Credit Registry](#), DEMOS (March 2019).

¹⁷ Jason Scott Johnston, *The Freedom to Fail: Market Access as the Path to Overcoming Poverty and Inequality*, 40 HARV. J.L. & PUB. POL’Y. 41, 44 (2017); Jonelle Marte, [Here’s How Much Your Credit Score Affects Your Mortgage Rate](#), WASH. POST (Nov. 17, 2016); Julius Adebayo & Mikella Hurley, *Credit Scoring in the Era of Big Data*, 18 YALE J.L. & TECH. 148, 159, 179 (2016).

¹⁸ Jenna Burrell, [How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms](#), BIG DATA & SOCIETY (2016); Elin Wihlborg, Hannu Larsson, Karin Hedström, [“The Computer Says No!” – A Case Study on Automated Decision-Making in Public Authorities](#) INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS 2903 (2016).

In principle, algorithmic decision making in an area like mortgages offers the promise of correcting for human bias by making more “objective” assessments; in practice, that aspiration has not borne out. In this review, we summarize instances where researchers and journalists have audited algorithms, either from outside the black box, or with permission to look inside. We focus on instances where algorithmic systems have failed to correct for human bias, resulting in feedback loops of racist bias and discrimination.

We focus on predictive algorithms, which try to predict the likelihood of certain social and behavioral outcomes: how likely crime is in a certain neighborhood, or how likely it is that a customer will default on a loan.¹⁹ For example, law enforcement might use these algorithmic systems to make decisions about which communities are prioritized for patrolling, or banks might use them to decide which customers are eligible for a loan. Below, we detail decisions made by predictive algorithmic systems for insurance policies, credit scores, mortgage loans, education, health care, criminal justice, and policing that discriminate against Black and Brown communities.

Primarily, this review covers issues of racial discrimination and bias and looks for solutions to advance racial justice. But there is a much larger field of unfairness and bias of algorithms, where discrimination can occur based on gender, socioeconomic status, religion, immigration status, ethnicity, nationality, sexuality, ability, and other characteristics. Investigating these biases is much needed, but outside of the scope of this review.

Discussions about algorithmic accountability have historically recommended solutions around self-governance: in other words, that those who create algorithmic systems should design them while attempting to correct for bias. For example, some proposals call for technologists to design algorithmic systems that are more transparent, or explainable.²⁰ Others recommend that technologists

¹⁹ Stephanie K. Glaberson, [Coding Over the Cracks: Predictive Analytics and Child Protection](#), 46(2) FORDHAM URBAN L. J. 307 (2019).

²⁰ [Universal Guidelines for Artificial Intelligence](#), THE PUBLIC VOICE (October 23, 2018); Margot E. Kaminski, [Understanding Transparency in Algorithmic Accountability](#), 20(34) U. COLO. L. LEGAL STUDIES (July 1, 2020); Lorna McGregor, Daragh Murray & Vivian Ng, [International Human Rights Law as A Framework for Algorithmic Accountability](#), 68 INT'L AND COMPAR. LAW Q. 309 (April 2019).

conduct internal audits that can detect algorithmic bias, or open the ‘black box’, allowing third parties to audit the algorithms.²¹

In addition to these proposals relying on self-governance, we aim to also direct attention to government and system-change solutions that codify and impose standards and requirements on those designing and using algorithms. In recent years, a growing movement of activists and researchers is pushing for government action, including banning faulty facial surveillance technology,²² banning faulty predictive policing technology,²³ as well as implementing new impact assessments models that promote the public interest.²⁴ We are eager to see more governance in specifically in the realm of predictive algorithmic systems discussed below.

In Part 1 of this review, we detail a ledger of harms from racist algorithms. Then, in Part 2, we look at the existing landscape of principles and proposed remedies and pose questions to be answered that may help develop policy. Finally, in Part 3, we recommend next steps to evaluate potential policy solutions.

²¹ Christine Henry et al., [Examining the Black Box: Tools for Assessing Algorithmic Systems](#), ADA LOVELACE INSTITUTE (2020).

²² Kate Conger, Richard Fausset & Serge F. Kovalski, [San Francisco Bans Facial Recognition Technology](#), N.Y.TIMES (May 14, 2019); David Gutman, [King County Council Bans Use of Facial Recognition Technology by Sheriff’s Office, Other Agencies](#), SEATTLE TIMES (June 1, 2021); [Ban Facial Recognition](#), FIGHT FOR THE FUTURE.

²³ Avi Asher-Schapiro, [In a U.S. First, California City Set to Ban Predictive Policing](#), REUTERS (June 17, 2020); Todd Feathers, [More Cities Are Moving to Drop Automated Gunshot-Detection Tech](#), VICE (August 3, 2021).

²⁴ Emanuel Moss et al., [Assembling Accountability: Algorithmic Impact Assessment for the Public Interest](#), DATA & SOCIETY (June 29, 2021).

Part 1: Review of harms of algorithmic racism

In our research, we found extensive empirical, quantitative, and qualitative studies, as well as journalistic investigations, that reveal ways that predictive algorithmic systems disproportionately harm Black and Brown communities by restricting and withholding opportunity and access, raising prices, and increasing surveillance and law enforcement.

Auto insurance premiums are more expensive

Even prior to algorithmic use, Black and Brown households often were charged more for car²⁵ and life²⁶ insurance policies. These discriminatory prices are consequential. Expensive car insurance policies, as one example, can pose a significant burden on car owners, who are required to have auto insurance in 48 states.

While the introduction of algorithmic systems held out the promise of eliminating unfair pricing, discriminatory pricing has remained a problem. A nationwide study by the Consumer Federation of America in 2015 found that predominantly Black neighborhoods pay 70 percent more, on average, for auto insurance premiums than other areas do.²⁷ In a statement rebutting the Consumer Federation of America's report, American Property Casualty Insurance Association, a trade association representing private insurance companies, justified their rates with the explanation that insurance rates are in part based on credit scores. The Association declined to reveal that credit scores, too, are generally algorithmically determined.²⁸

In 2017, a ProPublica study also found that insurers such as Allstate and Geico were charging algorithmically determined premiums that were, depending

²⁵ Associated Press, [Auto Loan Rates Higher for Blacks](#), CBS News (May 7, 2007, 8:16 PM).

²⁶ Mary L. Heen, [Ending Jim Crow Life Insurance Rates](#), 4(2) NW. J. L. SOC. POLICY 360 (2009).

²⁷ Tom Feltner & Douglas Heller, [High Price of Mandatory Auto Insurance in Predominantly African American Communities](#), CONSUMER FEDERATION OF AMERICA (November 2015).

²⁸ [Auto Insurance Rates are Based on Cost Drivers, Not Race](#), AMERICAN PROPERTY CASUALTY INSURANCE ASSOCIATION (November 18, 2015).

on the jurisdiction, as much as 10 to 30 percent on average higher in zip codes where most residents are minorities than in whiter neighborhoods with similar accident costs.²⁹ More recently, in 2020, The Markup found that a proposed rate setting algorithmic system from Allstate would have disproportionately affected people living in communities that were 75 percent or more nonwhite.³⁰

Communities of color pay 30% more for auto insurance premiums than whiter communities with similar accident costs.

In addition to perpetuating bias that existed prior to algorithmic use, insurers now use new types of data that lead to discrimination. For example, it is increasingly common for life insurers to use non-traditional sources of public data like court documents and motor vehicle

records.³¹ This practice, which tends to penalize Black and Brown communities more than others, requires more scrutiny and regulation.³²

Credit scores are lower

Credit scores, widely accepted as the modern basis of creditworthiness decisions, are predictive algorithmic systems in and of themselves. In calculating credit scores, the three major credit reporting agencies in the U.S., Equifax, Experian, and TransUnion, consider a range of factors including:

- Payment history (e.g., Are there any late payments on car loans, mortgage loans, or retail credit accounts?)
- Account balances (i.e., What are your total balances and debt?)
- Credit use (i.e., How much of your credit do you use?)
- Credit history (e.g., How long have you had credit? How long has it been since you last applied for credit?)

²⁹ Jeff Larson, Julia Angwin, Laurie Kirchner & Surya Mattu, [How We Examined Racial Discrimination in Auto Insurance Prices](#), PROPUBLICA (April 5, 2017).

³⁰ Maddy Varner, Aaron Sankin, Andrew Cohen & Dina Haner, [How We Analyzed Allstate's Car Insurance Algorithm](#), THE MARKUP (February 25, 2020).

³¹ Angela Chen, [Why the Future of Life Insurance May Depend on your Online Presence](#), VERGE (February 7, 2019).

³² [Systemic Racism in Auto Insurance Exists and Must Be Addressed by Insurance Commissioners and Lawmakers](#), CONSUMER FEDERATION OF AMERICA (June 17, 2020).

- Credit mix (i.e., Which of the following types of credit do you have: Credit cards? Retail accounts? Mortgage loans?)³³

Like the other predictive algorithm decision systems described in this review, credit scores disproportionately harm people of color.³⁴ A Brookings report published in 2020 found that white prospective homebuyers have an average credit score that is 57 points higher than Black homebuyers, and 33 points higher than Latinx homebuyers.³⁵

According to the same report, low credit scores are more common among communities of color. One in five Black individuals have FICO credit scores below 620, as do one in nine Latinx people. Yet, the same is true for only one out of every 19 white individuals.³⁶ At the same time, high credit scores favor white individuals. In 2019, the Urban Institute found that while only 21% of Black households had a FICO credit score above 700, more than 50% of white households did.³⁷

White homebuyers have credit scores 57 points higher than Black homebuyers; 33 points higher than Latinx homebuyers.

There is good reason to believe that these credit score disparities reflect economic inequity and discrimination rather than true “credit-worthiness”—i.e., the ability and likelihood of paying back debts. The primary reason for biased credit scores is historic economic inequity from policies like redlining, income disparity, educational segregation, biased employment, and mass incarceration that has “baked” itself into modern financial systems.³⁸

³³ Michelle Black, [What Makes Up Your Credit Score](#) FORBES (May 28, 2021); Bev O’Shea, [What Factors Affect Your Credit Scores?](#) NERDWALLET (January 28, 2020); [What’s In My FICO Scores?](#) FICO.

³⁴ Lisa Rice & Deidre Swesnik, [Discriminatory Effects of Credit Scoring on Communities of Color](#), 46 SUFFOLK U.L. REV. 935 (2013).

³⁵ Aaron Klein, [Reducing Bias in AI-Based Financial Services](#), BROOKINGS (July 10, 2020).

³⁶ Id.

³⁷ Jung Hyun Choi, [Breaking Down the Black-White Home Ownership Gap](#), URBAN INSTITUTE (February 21, 2020).

³⁸ [Past Imperfect: How Credit Scores and Other Analytics “Bake In” and Perpetuate Past Discrimination](#), NATIONAL CONSUMER LAW CENTER (May 2016); also see Amy Traub, [How Biden’s Plan for a Public Credit Registry Will Build Economic Power for Black and Brown Communities](#), DEMOS (November 16, 2020); Amy Traub, [Establish a Public Credit Registry](#), DEMOS (March 2019).

Some Black and Brown households are not eligible for credit scores at all, instead deemed “credit invisible” or “unscorable.” “Credit invisible” individuals lack credit history with nationwide credit reporting companies thereby making them ineligible for credit scoring; “Unscorable” individuals lack sufficient credit history with those companies. This means “unscorable” individuals are deemed ineligible for credit scoring, or are scored but with high likelihood of errors. In 2016, 27 percent of Black and Latinx adults, according to the Consumer Financial Protection Bureau, were credit invisible or unscorable, as compared to 16 percent of white adults.³⁹

Mortgages are more expensive or altogether inaccessible

Prior to use of algorithms, lenders often charged Latinx and Black borrowers higher rates for refinancing mortgages. These higher, discriminatory prices cost Latinx and Black communities a total of \$750 million each year, according to another University of California-Berkeley (UC Berkeley) research team.⁴⁰ Moreover, the research shows at least 6 percent of Latinx and Black applications are rejected—but would have been accepted if the borrower were not a part of these minority groups.

More advanced algorithmic systems based on credit history and scores have perpetuated rather than remedied these harms. The same study from UC Berkeley found that both human-determined and algorithmically driven lenders charge Latinx and Black borrowers 6 to 9 basis points higher interest rates.⁴¹

Higher mortgage prices cost Latinx and Black communities \$750 million each year.

There is a potentially optimistic story to be told here: applying predictive algorithmic systems to the lending process appears to reduce rates of discrimination in mortgage lending. In fact, the same study found that algorithmic systems discriminated 40 percent less on average than face-to-

³⁹ [Who Are the Credit Invisibles?](#) CONSUMER FINANCE PROTECTION BUREAU (December 2016).

⁴⁰ Robert Bartlett, Adair Morse, Richard Stanton, & Nancy Wallace, [Consumer-Lending Discrimination in the FinTech Era](#), FEDERAL DEPOSIT INSURANCE CORPORATION (February 2019).

⁴¹ *Id.*

face lenders in loan pricing, and did not discriminate at all in accepting and rejecting loans.⁴² This reduction is attributed to the decreased involvement of human judgment, which may introduce bias and discrimination.

But even among algorithmically driven lenders, the disparity in prices suggests that discrimination persists. It is not enough to simply be less discriminatory than previous systems. We must aim for no discrimination at all, or even anti-discrimination to correct for societal inequities. To this end, more effort is needed to intentionally reduce opportunity for bias and discrimination in algorithmic lending.⁴³

Students get rejected from better schools and assigned worse grades

An investigation in 2021 by The Markup found that Black and Latinx students are systematically screened out of the top-performing high schools across New York City. These discriminatory results are not attributable to differences in application rates; in fact, when Black and Latinx students apply for admission to these schools, they are consistently admitted at about half the rate of white and Asian students (at 4.4 percent and 4.9 percent acceptance rates, versus 9.2 percent and 8.6 percent respectively).⁴⁴

Some individual schools' racial discrepancies in acceptance are much greater: the Scholars' Academy, a top-ranked school in Queens, had an acceptance rate of 35 percent for white students, while for Black students, the acceptance rate was 8 percent. (More Black students applied to the Scholars' Academy than white students.)

These screening processes are driven by algorithms, but the details of the algorithmic systems are opaque. The Markup reported that the algorithmic systems factor test scores, attendance, and behavioral records—all of which tend

⁴² Id.

⁴³ Jennifer Miller, [Is an Algorithm Less Racist Than a Loan Officer?](#), N.Y. TIMES (September 18, 2020).

⁴⁴ Colin Lecher and Maddy Varner, [NYC's School Algorithms Cement Segregation. This Data Shows How](#), MARKUP (May 26, 2021).

to disadvantage Black and Brown students.⁴⁵ The algorithmic systems vary from school to school, with each weighing these and additional factors differently.

Algorithmic racism takes hold not only in admissions decisions, but also in grading. During the COVID-19 pandemic, schools around the world that administer the International Baccalaureate (IB) program, a standardized international curriculum⁴⁶, canceled in-person exams, and instead assigned high-stakes final grades using algorithms.⁴⁷ (The IB program is offered at more than 3,000 schools and educates over 90,000 students just in the U.S.)

In New York City, Black and Latinx students are admitted to top schools at half the rate of white and Asian students.

The International Baccalaureate has yet to disclose the factors considered in its grading algorithm, but they reportedly included coursework, teachers' predictions of how a student might have performed in an exam (even though teachers often have lower expectations of Black and Brown students⁴⁸), and a student's school

district's historical performance⁴⁹ (which could disadvantage high-performing students in low-income school districts and communities of color).

In fact, a top-performing, low-income, native Spanish-speaking student in Colorado failed her Spanish course when assigned a grade by the IB⁵⁰—ostensibly because the algorithmic system was designed to predict that the students in her

⁴⁵ On test scores: In almost all of the country's 100 largest school districts, Black students' scores are far below the national average, while white students' scores are above the national average. S. F. Reardon, [What explains White-Black differences in average test scores?](#) THE EDUCATIONAL OPPORTUNITY PROJECT AT STANFORD UNIVERSITY (September 2019); On attendance: "Compared to their white peers, American Indian and Pacific Islander students are over 50 percent more likely to lose three weeks of school or more, Black students 40 percent more likely, and Hispanic students 17 percent more likely." [Chronic Absenteeism in the Nation's Schools](#), U.S. DEPARTMENT OF EDUCATION (January 2019); On behavioral records: Black students' rates of days of instruction lost due to out-of-school suspension per 100 students was 51 days more than white students'. Daniel J. Losen and Paul Martinez, [Lost Opportunities](#), THE CENTER FOR CIVIL RIGHTS REMEDIES AT UCLA (January 2020).

⁴⁶ Avi Asher-Shapiro, [Global Exam Grading Algorithm Under Fire for Suspected Bias](#), THOMAS REUTERS FOUNDATION (July 21, 2020).

⁴⁷ Id.

⁴⁸ Nicholas Papageorge and Seth Gershenson, [Do Teacher Expectations Matter?](#) BROOKINGS (September 16, 2016).

⁴⁹ Ethical AI, [Bad Robots: Global Exam-Grading Software In Trouble For Algorithm Bias](#) (July 25, 2020).

⁵⁰ Id.

school district, mostly low-income students of color, would have done poorly on their final exams.⁵¹ Yet, as a native Spanish speaker, it seems likely that this student would have not only passed, but scored well on her Spanish final exam.

And for the 94 schools that implemented the IB program for the first time and lacked sufficient historical data to feed these algorithms, students' grades were determined by feeding the algorithmic system performance from other schools.⁵²

At the University of Massachusetts Amherst, Black students were up to 4 times as likely to be labeled 'high risk' as white students.

Algorithmically driven grades that are lower than expected have meaningfully harmed students and their opportunities: conditional offers to universities⁵³ and scholarships⁵⁴ were withdrawn, thousands of dollars were needed to pay for college credits they would have earned through better exam scores,⁵⁵ and a \$540 fee was required for a request for IB grades to be re-reviewed.⁵⁶

A third recent example of algorithmic racism in education takes place once students matriculate into higher education. This year, The Markup investigated predictive algorithmic systems from a company called EAB's Navigate, which assigns college students risk scores for retention. Over 500 universities in the U.S. use this technology today.

The Markup found that the EAB's Navigate systematically assigns higher risk scores to Black students over white students. Black men were up to 4 times as likely to be labeled "high risk" as white men; Black women were up to 3 times more

⁵¹ Meredith Broussard, [When Algorithms Give Real Students Imaginary Grades](#), N. Y. TIMES (September 8, 2020).

⁵² Hye Jung Han, [An Algorithm Shouldn't Decide a Student's Future](#), POLITICO (August 13, 2020).

⁵³ Cathy O'Neil, [This Grading Algorithm is Failing Students](#), BLOOMBERG OPINION (July 16, 2020).

⁵⁴ Tom Simonite, [Meet the Secret Algorithm That's Keeping Students Out of College](#), WIRED (July 20, 2020).

⁵⁵ Ethical AI, *supra* note 49.

⁵⁶ O'Neil, *supra* note 53.

likely to be labeled “high risk” as white women. Four out of seven schools The Markup observed even explicitly utilized race as a prediction factor of risk.⁵⁷

The impacts of these algorithms, while not currently used for admissions decisions, have the potential to meaningfully shape—positively or negatively—students’ educational experiences.⁵⁸ The justification for these risk ratings, including use of race as a factor, is that they should help identify students to whom universities should offer extra support. However, there are evident risks, as well; for example, faculty advisors may be more likely to advise “high risk” students to “easier” or “less risky” courses and majors.

Patients are denied life-saving care

A recent *Science* publication has shown that an algorithmic system to help large health providers and insurers determine who should be enrolled in “high-risk care management” programs, which provide extra resources to patients like more nurses, specialists, and coordinated care, was biased against Black patients compared to white patients.⁵⁹

Upon auditing the algorithm, researchers observed that the results of the algorithmic system were racially biased—even though the algorithmic system specifically excluded race as a factor—because the algorithmic system relied too heavily on total historical health costs accumulated by patients to predict future need for treatment. Because Black patients in the U.S. receive less care than white patients of equal levels of health (due to inequitable accessibility⁶⁰)—and thus generate lower levels of health care expenditures—the algorithmic system predicted less need for high-risk care for Black patients.

⁵⁷ Todd Feathers, [Major Universities Are Using Race as a “High Impact Predictor” of Student Success](#), MARKUP (March 2, 2021).

⁵⁸ Another example of algorithmic systems harming students is in the application of racial recognition technology. Black and Brown students can be disproportionately surveilled and disciplined through inaccurate and biased facial recognition systems. See: Stefanie Coyle & Rashida Richardson, [Bottom-Up Biometric Regulation: A Community’s Response to Using Face Surveillance in Schools](#), AI NOW INSTITUTE.

⁵⁹ Ziad Obermeyer, Brian Powers, Christine Vogeli, Sendhil Mullainathan, [Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations](#), 366(6464) SCIENCE 447 (October 25, 2019).

⁶⁰ Inequitable access to health care in the U.S. persists because poor communities and communities of color generally use fewer health care services, even when insured, due to geography and differential access to transportation, competing demands from jobs or child care, or knowledge of reasons to seek care.

As a result of the biased algorithmic systems, white patients with the same level of illness were offered much more treatments than Black patients—a clear marker of racial disparity in access to care. Furthermore, this racial bias reduced the number of Black patients identified for extra care by more than half.

These predictive algorithmic systems are widespread in the healthcare industry, and similar systems are used to allocate resources for over 200 million patients in the U.S. each year.

Biased algorithms reduce the number of Black patients eligible for extra care by more than half.

Criminal justice system is more punitive

Criminal justice and court systems make consequential decisions every day: which prisoners to release on parole, how to set bail for inmates awaiting trial, or to determine sentencings.⁶¹ In an attempt to help make these decisions more accurately, social scientists developed “risk assessments,” or calculations that use data points about the defendant to predict the risk the defendant will commit an offense again in the future.

Since the 2000s, criminal justice systems across the country have adopted over 60 new risk assessment tools to help make decisions at each stage of the legal process.⁶² These technologies use information like age, employment history, and prior criminal record to try to predict the risk of a defendant committing future offenses. As of 2019, 46 of 50 states in the U.S. had at least one pretrial risk assessment tool in use.⁶³

The impacts of these technologies are significant: A low score designating a defendant as “low risk,” could result in lower bail, less prison time or less restrictive probation or parole terms; a high score for defendants designated as “high risk” could lead to longer sentences or tighter monitoring. Judges are

⁶¹ Vincent Southerland, [The Intersection of Race and Algorithmic Tools in the Criminal Legal System](#), MARYLAND L. REV. (forthcoming).

⁶² Koene et al., *supra* note 7, at 17.; Anna Maria Barry-Jester, Ben Casselman & Dana Goldstein, [The New Science of Sentencing](#), THE MARSHALL PROJECT (August 4, 2015).

⁶³ [Mapping Pretrial Injustice: A Community-Driven Database](#), MOVEMENT ALLIANCE PROJECT & MEDIAJUSTICE (2019).

provided these risk assessments to inform their decision, but not generally obligated to make decisions solely based on the results.

Apart from concerns about bias, there is reason to question how effective these risk assessments are at predicting future crime. One law review describes that “somehow, criminal justice risk assessment has gained the near-universal reputation of being an evidence-based practice despite the fact that there is virtually no research showing that it has been effective.”⁶⁴

Furthermore, even though race is not a factor considered in risk assessments, these algorithmic systems often lead to worse outcomes for Black defendants when using data that often correlate with race (e.g., past arrest record, zip code, parents’ criminal record).⁶⁵

Black defendants are 45 to 77% more likely to be assigned higher risk scores than white defendants.

In 2016, ProPublica analyzed a risk assessment algorithmic system called COMPAS and found that Black defendants were indeed often predicted to be at a higher risk of recidivism than they were, while white defendants were often predicted to be less at risk for recidivism than they were. Adjusting for prior crimes and other indicators, Black defendants were 45 percent more likely to be assigned higher risk scores than white defendants. And when looking at violent recidivism, Black defendants were 77 percent more likely to be assigned higher risk scores than white defendants.⁶⁶ (The company that created COMPAS, Northpointe, disputed these findings.)⁶⁷

One cause of this type of discrimination is the usage of biased data points. For example, predictive policing algorithms may use data points like *past arrest record*,

⁶⁴ Megan T. Stevenson, [Assessing Risk Assessment in Action](#), 103 MINN. L. R. 303 (2018).

⁶⁵ Barry-Jester et al., *supra* note 69.

⁶⁶ Jeff Larson, Surya Mattu, Lauren Kirchner & Julia Angwin, [How We Analyzed the COMPAS Recidivism Algorithm](#), PROPUBLICA (May 23, 2016).

⁶⁷ Robust academic and legal debates have played out since the release of this analysis. For legal arguments on how to challenge algorithmic decision systems like COMPAS, see Anne L. Washington, [How to Argue with an Algorithm: Lessons from the COMPAS ProPublica Debate](#), 17(1) THE COLORADO TECHNOLOGY LAW JOURNAL (February 2019).

rather than *prior conviction record* (a record of ‘guilty’ verdicts administered by a jury or judge). Including past arrest records disadvantages Black defendants, who are more likely to be arrested than white individuals. This is both because law enforcement has historically overpoliced Black and Brown neighborhoods at higher rates than white neighborhoods⁶⁸—and because when policing, police are more likely to arrest Black and Brown people than white.⁶⁹

What results from these biased algorithmic systems is a compounding problem, or feedback loop. Future risk assessments may rate defendants with similar data points to defendants already designated as high risk—who live in the same neighborhood, whose parents have been arrested a similar number of times—with higher risk scores. This kicks off a feedback loop that ultimately assigns higher risk scores, and thus more punitive outcomes, for those that are deemed to be most at-risk of recidivism when criminal justice systems first implement these technologies. And we know that, for several reasons, these defendants are more likely to be Black.

Communities are over surveilled and policed

Predictive policing is an application of algorithmic systems that aims to help law enforcement detect crime.⁷⁰ Place-based predictive policing, sometimes also called predictive crime mapping, uses historical data on crime to predict where crime could occur next. These algorithmic systems define “hot spots” where crime may occur to direct policing resources.

Another common type of predictive policing, person-based predictive policing, aims to direct policing resources to people most likely to commit future crimes. As Vincent Southerland, the Executive Director at the Center on Race, Inequality, and the Law at New York University puts it, person-based predictive policing is a shift

⁶⁸ Elizabeth Moore, [Numbers Show Over-Policing in Historically Black Areas. And History Tells More](#), THE DAILY TAR HEEL (September 3, 2020); Robin Smyton, [How Racial Segregation and Policing Intersect in America](#), TUFTS NOW (June 17, 2020); Andrea Cipriano, [‘Overpolicing’ Still Common in NYC Black Neighborhoods. Report Finds](#), THE CRIME REPORT (September 23, 2020).

⁶⁹ Tammy Kochel, David Wilson & Stephen Mastrofski, [Effect of Suspect Race on Officers’ Arrest Decisions](#), 49(2) CRIMINOLOGY 473 (May 25, 2011); [How Do Arrest Trends Vary Across Demographic Groups?](#), VERA INSTITUTE OF JUSTICE.

⁷⁰ David Robinson & Lojan Koepke, [Stuck in a Pattern: Early Evidence on “Predictive Policing” and Civil Rights](#), UPTURN (August 2016).

from focusing on “hot spots” where crime might occur to “hot people” who may engage in (or be victims of) violence.⁷¹

Person-based policing itself may lead to feedback loops of bias and discrimination. “Once [a defendant has] been arrested once, they are more likely to be arrested a second or a third time—not because they’ve necessarily done anything more than anyone else has, but because they’ve been arrested once or twice beforehand,” one public defender explained.⁷² While person-based policing thus warrants further study, we focus on place-based predictive policing in this review.

A survey by Upturn in 2016 found that of the 50 largest police forces in the U.S., at least 20 used a predictive policing system, while at least an additional 11 were actively exploring the option.⁷³ PredPol (recently rebranded as Geolotica⁷⁴), one such policing technology used by local police in at least seven states, has been found to be a basic equation, not a complex algorithm. Simply put, the software directs police to places where arrests have already happened.⁷⁵ Despite its advertising as such, PredPol did not use any proprietary or mathematically complex algorithms.

Nevertheless, like risk assessments, predictive policing algorithmic systems (or, more accurately described as equations) initiate a “runaway feedback loop.”⁷⁶ Danielle Ensign, an AI expert, and her co-authors of a paper explain, “Once a decision has been made to patrol a certain neighborhood, crime discovered in *that* neighborhood will be fed into the training apparatus for the next round of decision-making.”

Accordingly, if a neighborhood is not frequently patrolled, less crime will be discovered in that neighborhood to feed into the algorithm, making the neighborhood less likely to be recommended to police for patrolling. The overall impact of these systems is oversurveillance and over policing of communities of

⁷¹ Southerland, *supra* note 60.

⁷² Anna Maria Barry-Jester, Ben Casselman & Dana Goldstein, [The New Science of Sentencing](#), MARSHALL PROJECT (August 4, 2015).

⁷³ Robinson, *supra* note 70.

⁷⁴ [Geolotica: A New Name, A New Focus](#), PREDPOL (March 2, 2021).

⁷⁵ Caroline Haskins, [Academics Confirm Major Predictive Policing Algorithm Is Fundamentally Flawed](#), VICE (February 14, 2019).

⁷⁶ Danielle Ensign, Sorella A. Friedler, et. al., [Runaway Feedback Loops in Predictive Policing](#), 81 PROCEEDINGS OF MACHINE LEARNING RESEARCH 1 (2018).

color, since they have been over surveilled and policed throughout this country's history.⁷⁷

Researchers who replicated PredPol's model found that in Oakland, California, Black individuals indeed would be targeted by predictive policing for drug use at twice the rate of white individuals. Individuals of color that are not Black would be targeted at a rate 1.5 times that of white individuals. Notably, the actual pattern of drug use by each race is comparable across the board.⁷⁸

Part of this bias may be attributed to the fact that predictive policing algorithmic systems use data like arrests, citations, and stops, which are not accurate measures of crime⁷⁹, and are often biased and discriminatory against Black and Brown communities.⁸⁰ As Rashida Richardson, lawyer and expert on algorithmic accountability notes, biases take the form of data missing from predictive policing (unreported crimes, deprioritized crimes, under-investigated crimes, and corruption), as well as errors and misrepresentations in data due to poor recordkeeping, corruption, police practice/policy, and misleading aggregate data.⁸¹

Of course, some of these limitations are unavoidable. Police departments cannot know the full universe of crime that has occurred, like crimes that are not reported. But if using data from police departments to inform decisions, we must be clear about the distinction between arrests (and other enforcement actions) and crime: namely, that predictive policing models work more to predict the most likely location of *future recorded crimes*, rather than *future crimes*. In other words, PredPol instructs law enforcement to look for people to arrest in areas they have arrested people before.

⁷⁷ Over surveillance in communities of color is not limited to predictive policing. "In San Diego, for example, police have used face recognition technology and license-plate readers up to two and a half times more on people of color than expected by population statistics," page 88 from Jameson Spivack & Clare Garvie, [A Taxonomy of Legislative Approaches to Face Recognition in the United States](#), in *Regulating Biometrics, Global Approaches and Urgent Questions*, edited by Amba Kak for AI Now (September 2020).

⁴⁵ Kristian Lum & William Isaac, [To Predict and Serve?](#), 14 SIGNIFICANCE (October 2016).

⁷⁹ David A. Harris, [The Reality of Racial Disparity in Criminal Justice: The Significance of Data Collection](#), 66 L. & CONTEMP. PROB. 71, 80 (2003).

⁸⁰ Southerland, *supra* note 60.

⁸¹ Rashida Richardson, [How Criminal Justice Data Reproduces Racialized Outcome](#) (April 1, 2021).

If we are not careful with how we label our data, Rashida Richardson, Jason Schultz, and Kate Crawford warn that predictive policing actually leads to data production, in a process they call “dirty data”.⁸² This “dirty data” process then initiates a feedback loop: When algorithmic systems dispatch police to locations predicted to have high rates of crime, any crimes they encounter will be added to law enforcement record, further validating the initially biased predictions of high crime rates.⁸³ As noted, this feedback loop leads to further biases in algorithmic decision making, and further harm on Black and Brown communities.

Black individuals are targeted by for drug use at 2x rate of white individuals; non-Black individuals of color were targeted at 1.5x the rate.

David Robinson, a scholar of algorithmic accountability, cites this as one of the core tensions in criminal justice prediction⁸⁴— but the principle can apply to many different areas of algorithmic accountability.

Other harms not included in this review

Of course, this is not an exhaustive list of the harms algorithmic systems perpetuate among communities of color. Among many other examples, emerging issues include algorithmic bias⁸⁵ in: tenant screening that wrongfully rejects prospective tenants,⁸⁶ discriminatory algorithmic systems in housing marketing and advertisement,⁸⁷ risk assessments for child welfare services,⁸⁸ facial recognition that aims to prevent potential repeat shoplifters from entering stores,⁸⁹ faulty facial recognition technology that has led to multiple wrongful arrests of

⁸² Rashida Richardson, Jason Schultz, Kate Crawford, [Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice](#) 94 N.Y.U. L. REV. ONLINE 192 (March 5, 2019).

⁸³ William Isaac & Kristian Lum, [To Predict and Serve? Predictive Policing with Bias Training Data](#), HUMAN RIGHTS DATA ANALYSIS GROUP (October 2016).

⁸⁴ David G. Robinson, [The Challenges of Prediction: Lessons from Criminal Justice](#), 14(2) I/S: J. L. POLICY 151 (2018).

⁸⁵ [Algorithms Behaving Badly: 2020 Edition](#), THE MARKUP (December 15, 2020).

⁸⁶ Lauren Kirchner & Matthew Goldstein, [Access Denied: Faulty Automated Background Checks Freeze out Renters](#), MARKUP (May 28, 2020).

⁸⁷ Allen, *supra* note 1.

⁸⁸ Emnet Almedom, Nandita Sampath & Joanne Ma, Algorithms and Child Welfare: [The Disparate Impact of Family Surveillance in Risk Assessment Technologies](#), BERKELEY PUBLIC POLICY JOURNAL (February 2, 2021).

⁸⁹ Matt Burgess, [Co-op Is Using Facial Recognition Tech to Scan and Track Shoppers](#), WIRED (December 10, 2020).

Black men,⁹⁰ algorithmic management of workers,⁹¹ to personalized online content (i.e. Facebook News Feed),⁹² as well as search and other associations.⁹³

And of course, a review of algorithmic bias must highlight the groundbreaking work of Joy Buolamwini and Timnit Gebru, who found that face-based gender classification technologies were least accurate on darker-skinned women.⁹⁴

Outside of algorithmic discrimination, other social harms stem from algorithmic technologies, including market segmentation and price discrimination⁹⁵, reinforcement of extremist views,⁹⁶ promulgating misinformation and disinformation,⁹⁷ invasive data collection and privacy breaches justified by the need for more data to feed algorithms,⁹⁸ and more.⁹⁹ A condensed taxonomy of harms of algorithmic systems from the Future of Privacy Forum is reproduced below:

⁹⁰ Kashmir Hill, [Wrongfully Accused by an Algorithm](#), N.Y. TIMES (June 24, 2020).

⁹¹ Josh Dzieza, [How Hard Will the Robots Make Us Work](#), VERGE (February 27, 2020).

⁹² Koene et al, *supra* note 7, at 13-14.

⁹³ The Markup found that searches for ‘black girls’ most commonly led to results related to pornography. Leon Yin & Aaron Sankin, [Google Ad Portal Equated “Black Girls” with Porn](#), THE MARKUP (July 23, 2020); As a computer teaches itself English, it becomes prejudiced against black Americans and women. Brian Resnick, [Yes, Artificial Intelligence Can Be Racist](#), VOX (January 24, 2019).

⁹⁴ Joy Buolamwini and Timnit Gebru, [Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification](#), Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91 (2018).

⁹⁵ Oren Bar-Gill, [Algorithmic Price Discrimination: When Demand Is a Function of Both Preferences and \(Mis\)Perceptions](#), 86(2) U. CHI. L. REV. 217 (2019).

⁹⁶ Mark Ledwich & Anna Zaitsev, [Algorithmic Extremism: Examining YouTube’s Rabbit Hole of Radicalization](#), ARXIV (December 24, 2019).

⁹⁷ Karen Hao, [How Facebook Got Addicted to Spreading Misinformation](#), MIT TECHNOLOGY REV. (March 11, 2021).

⁷¹ Aaron Holmes, [533 Million Facebook Users’ Phone Numbers and Personal Data Have Been Leaked Online](#), BUSINESS INSIDER (April 3, 2021).

⁹⁹ [Unfairness By Algorithm: Distilling the Harms of Automated Decision-Making](#), FUTURE OF PRIVACY FORUM (December 2017); Jack Bandy, [List of Algorithm Audits](#) (June 9, 2021); Joanna Redden, Jessica Brand & Vanesa Terzieva, [Data Harm Record](#), DATA JUSTICE LAB (August 2020).

Table 1: Potential harms from automated decision-makingSource: Adapted by author from [Future of Privacy Forum](#)

Individual harms	Collective and societal harms
Loss of opportunity	
Employment discrimination	Differential access to job opportunities
Insurance & social benefit discrimination	Differential access to insurance & benefits
Housing discrimination	Differential access to housing
Education discrimination	Differential access to education
Economic loss	
Credit discrimination	Differential access to credit
Differential pricing	Differential access to goods & services
Narrowing of choice	Narrowing of choice for groups
Social detriment	
Network bubbles	Filter bubbles
Dignitary harms	Stereotype reinforcement
Constraints of bias	Confirmation bias
Loss of liberty	
Constraints of suspicion	Increased surveillance
Individual incarceration	Disproportionate incarceration

Part 2: Questions about existing principles

In part 2 of this review, we propose questions about several principles that have been previously established by algorithmic accountability experts,¹⁰⁰ with an eye toward solutions through government legislation and enforcement.

Transparency

One of the most cited principles for algorithmic accountability is transparency.¹⁰¹ Transparency is also, by far, the most common approach in federal and state legislation on algorithmic accountability that has been introduced to date. Yet, more work needs to be done to determine what ‘transparency’ really means in the context of algorithmic accountability, and how technologists can implement it.

First, crucially, we must ask for details on what should be transparent. Should we require transparency in training data that is used to create algorithms? The algorithmic model itself? And/or the predictions—and differential results—generated as outputs when new data is fed into the algorithm? The answer is likely all of this, and more.

But as Hannah Bloch-Wehba at the Texas A&M University School of Law points out, defining what should be transparent cannot be left to technology companies. Rather, Bloch-Wehba writes, it must be a public process considering not only technical specifications of the algorithm systems, but also the social, political, and legal principles the systems engage in.¹⁰² Similarly, some algorithmic accountability experts suggest that in addition to the mechanisms and data that make up algorithms, we should also demand transparency about goals, outcome,

¹⁰⁰ Several examples include the Algorithmic Bill of Rights, Sigal Samuel, [10 Things We Should All Demand from Big Tech Right Now](#), VOX (May 29, 2019); an international human rights law framework, McGregor, *supra* note 20; principles of equitable and accountable AI, [The Algorithmic Justice League’s 101 Overview](#), THE ALGORITHMIC JUSTICE LEAGUE (2020); civil rights principles for era of big data, [Civil Rights Principles for the Era of Big Data](#), THE LEADERSHIP CONFERENCE ON CIVIL & HUMAN RIGHTS (February 27, 2014).

¹⁰¹ Kaminski, *supra* note 19 (posing additional questions).

¹⁰² Hannah Bloch-Wehba, [Transparency’s AI Problem](#), KNIGHT FIRST AMENDMENT INSTITUTE AT COLUMBIA UNIVERSITY (June 2021).

compliance, influence, and usage of the algorithms,¹⁰³ as well as information about the practical application of algorithms, which would allow us to evaluate the efficacy of an algorithmic system in comparison to its intent and goal.

This information should, to be effective, be publicly searchable, sortable, downloadable, and understandable so they are auditable, and housed by an independent body with expertise in algorithms, like the Federal Trade Commission, or a forthcoming agency with expertise in digital platforms and data.¹⁰⁴ The body would need to set clear definitions on what level of transparency we demand of which algorithms.

Robust transparency poses substantial technical challenges, expertly laid out by numerous scholars and advocates.¹⁰⁵ For example, some algorithmic systems are iterative, meaning they “repeatedly run a sequence of steps until the algorithmic system converges to a stable outcome.”¹⁰⁶ This makes it hard to determine what (if anything) will be sufficient information to provide for transparency to be useful. Similarly, randomized algorithms, as their name suggest, do not run the same way each time. This also presents a unique challenge to auditing.

Another important question to be answered is how to simultaneously demand transparency from technologists, while allowing technologists to protect their trade secrets. Algorithmic systems are the backbone of the most valuable technologies today (e.g., Google Search, Facebook News Feed, YouTube recommendation engines). Demanding full transparency of the proprietary innovations that make these technologies profitable may be impractical.

However, we might draw examples from other industries. For example, the Coca-Cola recipe is a highly guarded and valuable trade secret. But to inform consumers about its contents, the company lists its ingredients in order of predominance, in addition to a nutritional facts label. This level of detail protects the Coca-Cola Company’s trade secrets, while giving consumers enough information to make informed choices.

¹⁰³ Koene et al., *supra* note 7, at 5.

¹⁰⁴ More on this type of proposed agency is in Part 3.

¹⁰⁵ Koene et al., *supra* note 7, at 28.

¹⁰⁶ Koene et al., *supra* note 7, at 28.

Drawing on this example, how might we require transparency in algorithmic systems on a similar level of detail: general enough to protect trade secrets, but specific enough to allow the public to make informed decisions, third-party auditors to assess the algorithmic systems meaningfully, and regulators to protect against bias, discrimination, and harm?

Additionally, how can we ensure transparency not just in the data and formulae of algorithms, but also on algorithmic systems including their intended purpose and application? Unlike the Coca-Cola's nutrition labels, transparency in algorithmic decision systems requires additional contextual information in addition to the technical specifications, like who created the algorithm, how it has been trained, what audits and impact assessments it has undergone, the results of such assessments, and so on. One model of this has been put forward by one of the Big Tech companies themselves. Apple recently introduced "privacy labels", which require apps on its app store to disclose their data collection, storage, and sharing practices.¹⁰⁷

Beyond the substance of transparency requirements is the question: to whom should algorithmic systems be transparent to?¹⁰⁸ Is there any benefit in opening this information to the public? What might they be? Are there any potential harms of doing so? Is the information only of practical use to small circles of government, journalists, academics, and civil society? What information is most important and essential to provide to those being impacted by the algorithm?

Lastly, and perhaps most importantly, we must investigate whether transparency will promote the outcome we desire algorithmic systems that are just and accountable. It is short-sighted to believe that requiring transparency will automatically make corporate actors and technologists more accountable. Identifying biases and problems in design and effect will not necessarily translate into remedies. Thus transparency is a necessary but insufficient tool to facilitate accountability in algorithms. This is important to remember when evaluating the algorithmic accountability legislation that has made its way to introduction in the

¹⁰⁷ Lily Hay Newman, [Apple's App 'Privacy Labels' Are Here – and They're a Big Step Forward](#), WIRED (December 14, 2020).

¹⁰⁸ Jakko Kemper & Daan Kolkman, [Transparent to Whom? No Algorithmic Accountability without a Critical Audience](#), 22(14) INFORMATION, COMMUNICATION & SOCIETY 2081 (2019).

U.S. Congress, which leans heavily on transparency as an accountability mechanism.¹⁰⁹

Explainability

Another foundational principle of algorithmic accountability is explainability.¹¹⁰ A focus on ‘explainability’ highlights that whatever is made ‘transparent’ about algorithmic systems must then also be ‘explainable’ to third parties, so they can understand, question, and audit them. To determine what level of explanation is required, we must first answer the question: to whom should the algorithmic system be explainable to?

After the appropriate audience is defined, we should determine how we want to build systems for technologists to explain their algorithms. For example, we might require explanations to conform to a consistent template of terms and organization, not unlike U.S. Securities and Exchange Commission (SEC) filings.

Sandra Wachter at the University of Oxford and colleagues have sought to detail what is necessary in an explanation to make it useful. She suggests that the individuals and communities being impacted by algorithmic systems should be provided:

- *“explanations to understand a given decision*
- *“grounds to contest it, and*
- *“advice on how the data subject can change his or her behaviour or situation to possibly receive a desired decision (e.g. loan approval) in the future”¹¹¹*

But even when explainability bolsters transparency, there are limitations. Like with transparency, the case for explainability as a path toward algorithmic accountability on its own is weak. Explainability is a useful tool to advance accountability, but on its own is insufficient.

¹⁰⁹ [Algorithmic Accountability Act of 2019](#), H.R. 2231, 116th Cong. (2019); Grace Dille, [Sen. Wyden to Reintroduce AI Bias Bill in Coming Months](#), MERITALK (February 19, 2021).

¹¹⁰ Nicholas Diakopoulos et al., [Principles for Accountable Algorithms and a Social Impact Statement for Algorithms](#), FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY IN MACHINE LEARNING; Tae Wan Kim & Bryan Routledge, [Algorithmic Transparency, A Right to Explanation and Trust](#), CARNEGIE MELLON UNIVERSITY (June 2017).

¹¹¹ Sandra Wachter, Brent Mittelstadt & Chris Russell, [Counterfactual Explanations Without Opening the Black Box: Automated Decisions and the GDPR](#), 31(2) HARVARD JOURNAL OF LAW & TECHNOLOGY (2018).

Furthermore, we should not let explainability as a principle be weaponized by corporate power. When technology companies argue that algorithmic systems are too complex and unknowable, they should not be given a pass on the explainability standard. Instead, algorithmic systems that are not explainable simply should not be allowed for use in sensitive domains, like the ones discussed in this review (e.g., loan determinations, educational decisions, criminal justice sentences). Algorithmic systems that are not explainable or understandable should not be used to make life-altering decisions for people, and especially those already marginalized in our society.

Fairness

A third foundational principle of algorithmic accountability is fairness. The definition of fairness is subjective, wide-ranging, and nuanced in the context of algorithms.¹¹² There are a host of values that relate to fairness: equality of opportunity, equality of outcome, equity, freedom of choice, justice, truth, autonomy, consent, and privacy. Which of these values should we prioritize when regulating fairness in algorithms?¹¹³ A seminal paper on fairness concludes that some of these may even be at odds with one another.¹¹⁴

Relatedly, who do we want systems to be fair for? Is an algorithmic decision-making system fair when beneficial to a specific individual? Or is it fair when beneficial to a certain disadvantaged group or groups?¹¹⁵ Sometimes, decisions based on “fairness” might conflict with one another, which is why we might consider identifying an arbiter of these questions. Who should this arbiter be?

Consent

For this discussion, we borrow from of the longstanding debate around privacy, wherein “consent” implies a person should have full knowledge of how their data is being used and algorithmic systems are being applied to them—and with this

¹¹² Koene, *supra* note 7.

¹¹³ Koene, *supra* note 7, at 19.

¹¹⁴ Sorelle A. Friedler, Carlos Scheidegger & Suresh Venkatasubramanian, [On the \(Im\)possibility of Fairness](#) CoRR (September 2016).

¹¹⁵ Koene, *supra* note 7, at 18.

knowledge, the individual must proactively and explicitly agree to the terms of use.¹¹⁶

Though simple in theory, consent becomes complicated when applied to the realm of algorithmic accountability, considering how many individuals' data are being collected and used by algorithms, how many individuals are affected by algorithmically generated decisions, as well as the lack of transparency and legibility of the algorithmic systems themselves.

First, as algorithmic systems become ubiquitous in our daily lives, we must decide which types of uses should require consent. In what scenarios do we require consent from individuals whose data is being used to train algorithmic systems? And in which do technologists and practitioners need consent from those being impacted by the decisions the algorithmic systems inform?

Thinking about the big data sets of anonymized information that might go into a health-related algorithm, for example, raises relevant questions: Does every single individual in a data set of 10,000 patients need to consent to having their anonymized data used to train algorithms? What about one with data from 1,000,000 patients?

Further, what happens if an individual does not consent? Will they be locked out of opportunity that they otherwise would have had access to? Or will they be subject to a degraded form of the product or service? We must ensure that individuals' private data are not a bargaining chip that they can negotiate or trade away in a false choice between privacy and access to goods and services.

And what does it mean for an individual to be "fully informed" as a condition of consent? Typically, it means each individual must have all the relevant information they need to know before making a decision. But in algorithmic systems as that may be as complicated as those that predict the likelihood of illness, how much can, and should a patient know before they can make a decision? What is the necessary and appropriate level of detail that should be provided about algorithms? And what does it mean for a person to opt out? What duty if any would that impose on a service provider otherwise relying on algorithmic systems for decision making?

¹¹⁶ Helen Nissenbaum, [A Contextual Approach to Privacy Online](#), 140(4) DAEDALUS 32 (February 2015).

Lastly, the question of enforcement persists. How can we keep track of all the algorithmic systems being used at one time, and ensure that they are operating with informed consent of required parties? If an organization fails to follow these principles of consent, what consequences will it face? Who will levy relevant penalties, and who will bring the cases forth for adjudication?

Not all consent models are designed equally. Some design principles offer stronger protections than others. For example, instead of designing a system that automatically defaults individuals into “opting in,” meaning they are informed about the use of their data and have to take proactive action if they want to opt themselves out, designers should consider “opt out” as the default setting. In an “opt out” system, individuals are initially opted out, and can then opt in if, after fully informed, they choose to do so.

Redress

If an individual is harmed by an algorithmic system without their informed consent, what can they do in response? Who can they turn to? What government agency would enforce the process of redress? Would the individual have a private right of action, meaning they could bring the case to court? In practice, how likely is it that the average individual, or importantly, an individual from an already marginalized community, will be able to navigate this process or even understand the root cause of the harm they experienced?

Recent interviews reveal that even legal professionals do not have the bandwidth or resources to know how to seek redress on behalf of clients, nor about the algorithmic decision-making systems that they aim to challenge.¹¹⁷ Advocates have created informal policy guides to challenge algorithmic systems that exacerbate poverty, for example, but the scope and reach of these resources are limited.¹¹⁸ Who, if anyone, should be responsible for disseminating relevant information?

¹¹⁷ Karen Hao, [The Coming War on the Hidden Algorithms that Trap People in Poverty](#), MIT TECHNOLOGY REVIEW (December 4, 2020); Rashida Richardson, Jason M. Schultz & Vincent M. Southerland, [Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems](#), AI NOW (September 2019).

¹¹⁸ Michele Gilman, [Poverty Lawgorithms](#), DATA & SOCIETY (September 2020).

Even once an individual determines who they would ask for help, and starts the arduous process of litigation or legal remedy, there remain many unanswered questions: How do we quantify the damages of algorithmic bias? Should redress only cover actual damages incurred, or add punitive damages? How do we quantify loss of opportunity, as well as emotional distress from biased practices? Should these issues be settled in court? Or is there a simpler, less resource intensive and more accessible path for relief?

Independent oversight

Without oversight and enforcement, none of the legislation or regulation levied on technology companies will have teeth, and the companies will then have no incentive to comply with these principles. To avoid this outcome, should there be a single, governmental oversight body for algorithms? Would this body be a part of the Federal Trade Commission? A new data protection agency? Or another enforcer? How would we ensure the body is equipped with the expertise, authority, and budget to effectively execute the program?

Once a body is established, we must define its jurisdiction, and establish criteria for its programs. Of the countless algorithmic systems currently in use, how will the body prioritize which algorithmic systems to audit? A commonsense approach might be to define a floor of the number of people impacted by the algorithmic system and prioritize overseeing algorithmic systems with impacts above that threshold. If so, what would an appropriate threshold be? 10,000 individuals? 100,000? 1 million?

Or the enforcement agency might prioritize algorithmic systems affecting particularly sensitive decisions, like ones in this review that impact economic and social opportunity. Lastly, after an appropriate threshold is established, how will the body's oversight keep up with the volume and pace of new algorithmic decision-making systems?

Federal and global governance

Even after determining roles and processes for federal regulation of algorithms, we must determine roles and processes for international regulation. Today's algorithmic systems are not limited to development and application within national borders; rather, many of the algorithmic systems used in the U.S. are also

being used internationally, and algorithmic systems developed and hosted outside of the U.S. may be used within the U.S. How will the federal government navigate the commerce and use of algorithms, and related data and decision-making processes, across borders? How will we coordinate with other countries to ensure we are promoting safety and protection not only for individuals in the U.S., but also internationally?

Even in the field of privacy, which has a longer history of advocacy and policy than algorithmic accountability, we have yet to establish a sustainable system to govern data transfer across national borders. The E.U.-U.S. Privacy Shield, which gave registered companies legal protection to authorize transatlantic transfer of E.U. citizens' data, was struck down just in the past few months.¹¹⁹ There is a vacuum in its place, as the U.S. and E.U. have yet to negotiate a system to replace it. Establishing a global governance mechanism for algorithmic justice may prove equally or even more difficult than establishing a successful global governance mechanism for privacy.

¹¹⁹ Natasha Lomas, [EU-US Privacy Shield Is Dead. Long Live Privacy Shield](#), TECHCRUNCH (August 11, 2020).

Part 3: Recommendations for next steps

This review aims to pose questions that help build towards algorithmic accountability legislation and regulation. Crucially, we don't need answers to all the questions raised in this paper to proceed to adopt new measures to combat algorithmic bias. Below are ideas for investigating next steps toward building on this work.

Learn by expanding on principles and conducting deeper audits

Address questions posed in Part 2 of this review

Dozens of questions posed in part 2 of this review aim to tease out the details of well-established and agreed upon principles like transparency and explainability. Answering these questions can help develop a clearer vision of the policy and enforcement mechanisms needed to start remedying the harms.

Examining specific cases of discrimination can help generate ideas to tackle underlying biases—for example, in credit risk assessment, or in mortgage rate determination. One previously underutilized federal regulatory tool is section 6(b) of the Federal Trade Commission Act, which authorizes the FTC to conduct studies that do not have a specific law enforcement purpose.¹²⁰ We suggest the FTC consider using section 6(b) to commission studies that investigate discriminatory commercial practices underlying algorithms.

Additionally, the White House can lead on algorithmic accountability. In June, President Biden announced the launch of an Artificial Intelligence Research

¹²⁰ 15 U.S.C. §46(b) (“To require, by general or special orders, persons, partnerships, and corporations, engaged in or whose business affects commerce, excepting [banking institutions and common carriers], to file with the Commission... reports or answers in writing to specific questions, furnishing to the Commission such information as it may require as to the organization, business, conduct, practices, management, and relation to other corporations, partnerships, and individuals of the respective persons, partnerships, and corporations filing such reports or answers in writing.”).

Resource Task Force as a first step toward improving AI governance.¹²¹ Yet, there is more to do to center civil rights in the conversation. A group of civil rights, civil liberties, and human rights groups, led by the ACLU, The Leadership Conference on Civil and Human Rights, and Upturn sent a letter to the White House Office of Science & Technology Policy urging its leadership to commit to promotion of civil rights and racial equity in its work.¹²² Public Citizen supports and echoes this call.

All these actions may help us gather more information on the applications of predictive algorithmic decision systems and hone an approach toward accountability and justice in each.

Include race, gender, and other protected classes in algorithmic auditing

Rather than removing race, gender, or other protected classes from data sets to be used in algorithmic decision systems, we should consider including the protected class in the auditing process.¹²³ By including the protected class, algorithmic system designers can assess (to the extent it is at all possible) whether the decision system's results demonstrate bias against the protected class and try to correct it if they do.¹²⁴

This idea stems from the fundamental question: How would we be able to evaluate whether a predictive algorithmic system is causing racially biased outcomes without knowing the race of the people whose data the algorithmic system is processing? The answer is, according to Ignacio Cofone of McGill University, "We need to collect information on race in order to see impact on race, but we must also prevent information on race from producing discrimination based on race."¹²⁵

Indeed, one study that tried to predict students' college performance to simulate admissions decisions found that a race-aware algorithmic system allowed them to substantially to increase the share of racial minority admitted students. The authors propose that in this way, race-aware algorithmic systems can be used not

¹²¹ [The Biden Administration Launches the National Artificial Intelligence Research Resource Task Force](#), WHITE HOUSE (June 10, 2021).

¹²² [Centering Civil Rights in Artificial Intelligence and Technology Policy](#), ACLU, THE LEADERSHIP CONFERENCE ON CIVIL AND HUMAN RIGHTS, UPTURN (July 13, 2021).

¹²³ [@mathbabedotorg](#), TWITTER (March 30, 2021).

¹²⁴ Southerland, *supra* note 60, at 55-57.

¹²⁵ Ignacio N. Cofone, [Algorithmic Discrimination Is an Information Problem](#), 70 HASTINGS L.J. 1389 (2019).

only to de-bias algorithmic decisions, but also to promote proactive equity within them.¹²⁶

The question of what to then do with the information about protected class, including how to keep those details private and secure, remains to be addressed. Though legal scholars offer that training data that informs the algorithmic system can be pre-processed to remove bias, the effort will undoubtedly prove to be complex.¹²⁷ However, we should not let the importance of privacy and security of this data cause us to shy away from collecting and using it, as it is imperative to testing algorithmic systems for bias and discrimination.

As mathematician, data scientist, and founder of an auditing firm Cathy O’Neil suggests, a post-outcome evaluation can be used to audit for bias. This process seems it would be more straightforward than efforts to pre-process data before it enters the algorithm. Algorithmic system designers could run the outputs of the algorithmic system against the protected classes and see whether there are meaningful correlations. If there are, the designers can then take action to retrace their steps, investigate where bias may have been introduced, and work to scrub the data set and algorithmic system until a post-outcome evaluation proves results are no longer correlated with a protected class. Of course, this process is immensely simpler to describe than it would be for algorithmic system designers to implement.

This approach surely faces significant challenges, foremost among them avoiding unlawful disparate treatment, which “forbids unequal treatments based on protected characteristics (race, gender, age, etc.)”,¹²⁸ especially in areas covered under Title VII of the Civil Rights Act, like employment. The challenge notwithstanding, what this review makes clear is that simply adopting “race neutral” algorithmic systems does not solve—and may even exacerbate—racial inequity.

¹²⁶ Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan & Cass R Sunstein, [Discrimination in the Age of Algorithms](#), 10 JOURNAL OF LEGAL ANALYSIS 113-174 (2018).

¹²⁷ Cofone, *supra* note 125.

¹²⁸ Kay Li, [Anti-Discrimination Laws and Algorithmic Discrimination](#), MICHIGAN TECHNOLOGY LAW REVIEW (January 2021).

Enable and mandate public and third-party audits

One proposal to consider deepening accountability is to mandate a pre-release, public version of new algorithmic systems for those that will impact a significant population. By way of example, lawmakers or regulators could set a theoretical threshold of 1,000,000 people impacted. Each algorithmic system that potentially impacts 1,000,000 or more people would be required to release a pre-release, public version.

Third parties could then test and audit the algorithmic system by submitting sample inputs and observing the decision that the algorithmic system generates. Rumman Chowdhury, Director of the ML Ethics, Transparency & Accountability team at Twitter, and Deb Raji, a fellow with Mozilla Foundation and leading scholar in algorithmic accountability and tech ethics, have led projects that provide precedent in third-party algorithmic bias detecting.¹²⁹ They have created “bug bounties” for algorithmic bias, incentivizing the public to identify and report instances of algorithmic bias. Using this type of model, trade secrets would be protected, while journalists, activists, and civil society at large would be able to test algorithmic systems for bias.

One codification of such a potential policy lies within the Algorithmic Accountability Act, introduced in 2019 by Senators Cory Booker and Ron Wyden, and Representative Yvette Clarke, which mandates algorithmic impact assessments for entities above a certain revenue, above a certain user base (1 million), or that function as a data broker.¹³⁰ These assessments need not be available to the public.

There is already precedent within the federal government for these types of audits. Andrew Tutt compares this model to the mandate of the Food & Drug Administration (FDA) and the responsibility of a new government agency, in his paper, “An FDA For Algorithms.”¹³¹ Clearly this approach, while it has merits, would require tremendous resources and administrative process, and thinking on who would govern this process, and how.

¹²⁹ Rumman Chowdhury & Jutta Williams, [Introducing Twitter’s First Algorithmic Bias Bounty Challenge](#), TWITTER (July 30, 2021); Daphne Leprince-Ringuet, [The New Weapon in the Fight Against Biased Algorithms: Bug Bounties](#), ZDNET (March 17, 2021).

¹³⁰ [Algorithmic Accountability Act of 2019](#), H.R. 2231, 116th Cong. (2019).

¹³¹ Tutt, Andrew, [An FDA for Algorithms](#) 69(83) ADMIN. L. REV. (March 15, 2016).

Another question to consider is the tradeoffs between this type of public and transparent auditing system, versus a private industry that is already developing quickly.¹³²

Yet in public systems, there is more transparency and potentially more accountability. Deb Raji and other scholars suggest that the U.S. can look to the E.U.'s Digital Services Act as an example. Articles 28 and 31 of the Digital Services Act, for example, mandates the participation of independent, external reviewers.¹³³ Legal scholars Frank Pasquale and Gianclaudio Malgieri suggest that the U.S. can also learn from the European Union's proposal for AI regulation for further ideas.¹³⁴

Create new rules for algorithmic design, regulation, and governance

Enact principles of design justice to create just algorithms

Pursuing algorithmic justice requires that we scrutinize algorithmic systems not only at the application and impact stages, but from the very beginning, in the design phase. Sasha Costanza-Chock, Director of Research and Development at the Algorithmic Justice League, has written a book titled, "Design Justice", which offers recommendations to ensure that the process of creating and applying new algorithmic systems is democratically led by the communities most impacted by the technologies.

The first steps toward design justice are simple. As discussed in the introduction of this review, today's algorithmic decision-making systems are biased in part due to the biases of the technologists who design these algorithms, who especially in the field of artificial intelligence, overrepresent men, and underrepresent Black and Latinx communities.¹³⁵ These technologists may be particularly less attuned to

¹³² Alfred Ng, [Can Auditing Eliminate Bias from Algorithms?](#), MARKUP (February 23, 2021).

¹³³ @rajinio, TWITTER (May 28, 2021); [Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services \(Digital Services Act\) and amending Directive 2000/31/EC](#), EUROPEAN COMMISSION (December 15, 2020).

¹³⁴ Frank Pasquale & Gianclaudio Malgieri, [If You Don't Trust A.I. Yet, You're Not Wrong](#), N.Y.TIMES (July 30, 2021).

¹³⁵ Sarah Myers West, Meredith Whittaker, & Kate Crawford, [Discriminating Systems: Gender, Race, and Power in AI](#), NEW YORK UNIVERSITY AI NOW INSTITUTE (April 2019).

the potential for bias due to their limited lived experience of biases. A simple step to consider toward improving technologists' ability to create equitable algorithms, as well as anticipate potential biases, is to ensure diverse perspectives lead the design process. Incorporating traditionally excluded groups in the process of designing algorithmic systems can help to ensure equity, fairness, and justice are translated into technology.¹³⁶

Beyond the question of who leads and is involved in algorithmic development and design, we should also adopt principles of design justice in the design process itself. Within a design justice framework, technologists view the development, auditing, and revision of algorithmic systems as “an accountable, accessible, and collaborative process”¹³⁷—or in other words, a living technology with room to grow and change. This new type of design process would require relevant stakeholders to be involved at each step of development, not just at the beginning or end.¹³⁸ Further research can be conducted to understand tactically what this engagement looks like.

Litigate unsafe products and false claims under products liability law

Tech exceptionalism has led users to assume that algorithmic decision systems products do, indeed, do what they are advertised to do, and in a safe manner. We can take these assumptions to task with products liability law. Simply put, products liability law holds product creators accountable for their products working as described, and not causing harm. As an example, if a car manufacturer sells cars with defective brakes that causes accidents, the car manufacturer could be held accountable under products liability law.

Though algorithmic systems are different from physical products like cars, we can consider them as products as well. And the first question we need to ask is: Do these products even work at all? Or are they, as computer scientist Arvind Narayanan puts it, mere “AI snake oil”?¹³⁹ Narayanan suggests that for predicting

¹³⁶ [Algorithmic Accountability: Applying the Concepts to Different Country Contexts](#), WEB FOUNDATION (July 2017).

¹³⁷ [Design Justice Network Principles](#), DESIGN JUSTICE NETWORK (last updated 2018).

¹³⁸ Julia Dressel & Hany Farid, [The Accuracy, Fairness, and Limits of Predicting Recidivism](#), 4(1) SCIENCE ADVANCES 1 (January 17, 2018); David Robinson & Logan Koepke, [Civil Rights and Pretrial Risk Assessment Instruments](#), UPTURN (December 2019).

¹³⁹ Arvind Narayanan, [How to Recognize AI Snake Oil](#).

social outcomes, algorithmic systems are in most cases, snake oil—not substantially better than manual scoring using just a few features.

John Villasenor of the Brookings Institute lays out a guide to litigating faulty or dangerous algorithmic systems under products liability law, even offering potential defenses from algorithmic system creators, as well as counterarguments to address them.¹⁴⁰ If we were to apply the products liability framework, examples abound of both unsafe and defective (or, not performing as described) algorithmic systems as places to start. Many algorithmic systems in the market, including but not limited to predictive algorithmic systems discussed in this paper, simply do not work. Below are some examples:

Table 2: Potential products liability cases in algorithmic decisions systems

Liability type	Vendor(s)	Purpose	Liability case
Unsafe	InterRAI	Allocating health resources	Home care hours as state benefit for patient were cut by 43%; patients have been hospitalized, and even died as result ¹⁴¹
Unsafe	DataWorks Plus	Facial recognition	Faulty facial recognition match led to wrongful arrest ¹⁴²
Defective	Microsoft, IBM	Facial recognition	Facial recognition had error rates up to 35% for darker-skinned women ¹⁴³
Defective	Experian, Trans Union, Equifax	Credit score	Credit scores do not act as predictors of true risk of default ¹⁴⁴
Defective	COMPAS	Pretrial risk assessments	Risk assessments were only 65% accurate at predicting recidivism, no better than a non-expert person ¹⁴⁵

¹⁴⁰ John Villasenor, [Products Liability Law as a Way to Address AI Harms](#), BROOKINGS (October 31, 2019).

¹⁴¹ Colin Lecher, [What Happens When an Algorithm Cuts Your Health Care](#), VERGE (March 21, 2021). Erin McCormick, [What Happened When a ‘Wildly Irrational’ Algorithm Made Crucial Healthcare Decisions](#), GUARDIAN (July 2, 2021).

¹⁴² Kashmir Hill, [Wrongfully Accused by an Algorithm](#), N.Y. TIMES (June 24, 2020).

¹⁴³ Joy Buolamwini & Timnit Gebru, [Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification](#), 81 PMLR 77-91 (2018).

¹⁴⁴ Klein, *supra* note 34.

¹⁴⁵ Julia Dressel & Hany Farid, [The Accuracy, Fairness, and Limits of Predicting Recidivism](#), 4(1) SCIENCE ADVANCES (January 17, 2018).

Of course, products liability law is not a silver bullet to the problem of algorithmic racism. Amba Kak, Rashida Richardson, and Roel Dobbe of the AI Now Institute at New York University provided comments to the European Commission that discuss the limitations of a products safety framework for algorithmic systems. They suggest products liability is limiting in its treatment of algorithmic systems as discrete products, rather than with consideration of the following important contextualizing details: safety culture, human-machines interactions, inadequate specifications in the engineering development process, or a lack of empirically verified safety assurances.¹⁴⁶

Consider disparate impact and discrimination per se frameworks

Another potential remedy is to shift the burden of proof of harm of algorithmic systems. A way we can do this is through the a legal framework known as “disparate impact”. Traditional disparate impact cases, which aim to prove that there is an unintentional yet discriminatory outcome of decisions, shifts the burden of proof three times:

- First, the plaintiff, or the party who brings a case to court, must show that a decision procedure causes a disproportionate harmful effect on a protected class.
- Then, the defendant is required to show that the decision procedure serves a legitimate business purpose.
- Finally, the burden shifts back to the plaintiff who must produce evidence of an available alternative that would achieve the purpose with a less harmful impact on the protected class.¹⁴⁷

While adopting a disparate impact case has its advantages, for example by not requiring an intention to discriminate, there are several challenges to applying the disparate impact legal framework to mitigate algorithmic racism. A recent Brookings report outlines some of these challenges in algorithmic hiring processes, but which can also be generalized to the broad applications discussed in this report as well.¹⁴⁸

¹⁴⁶ Amba Kak, Rashida Richardson & Roel Dobbe, [Submission to the European Commission on “White Paper on AI - A European Approach](#), AI NOW (June 14, 2020).

¹⁴⁷ Mark MacCarthy, [Fairness in Algorithmic Decision-Making](#), BROOKINGS (December 6, 2019).

¹⁴⁸ Manish Raghavan and Solon Barocas, [Challenges for Mitigating Bias in Algorithmic Hiring](#), BROOKINGS (December 2019).

For one, as the authors of the Brookings report suggest, it would be difficult for the average plaintiff to have enough information to be able to demonstrate disparate impact. To bring a disparate impact case to litigate, the plaintiff must be able to show that an algorithmic system caused a disproportionate harmful effect on a whole class of people, not just herself. And because these algorithmic systems are commonly black boxes, someone disproportionately impacted by them might not even know that they were being evaluated using an algorithm, or that their result was potentially part of a disproportionately discriminatory pattern.¹⁴⁹

Furthermore, the threshold for proving discriminatory impact is high in some applications. For example, according to the Equal Employment Opportunity Commissions' Uniform Guidelines on Employee Selection, disparate impact can be identified if a selection procedure accepts candidates from one protected group at a rate 80% lower than that of another, otherwise known as the "four-fifths rule".¹⁵⁰ Proving this level of variance in candidate acceptance rate will likely be difficult.

While evaluating potential frameworks like disparate impact, we must take care to remember that what we are aiming for, overall, is not only litigating algorithmic racism that can pass the four-fifths rule, but algorithmic racism overall. How can we work to mitigate or prevent cases of discrimination that are less than 80%, but still greater than zero?

Update existing regulation to promote transparency

To address algorithmic racism in housing, James Allen in the Fordham Law Review suggests updating existing law and regulation created to prevent analogue methods of discrimination, so that they extend to modern methods of both unintended and intended discrimination.¹⁵¹

Specifically, Allen suggests updates to the Fair Housing Act (FHA), the Community Reinvestment Act (CRA), and the Fair Credit Reporting Act (FCRA) to regulate algorithmic systems to ensure they provide fair and adequate housing opportunity. According to Allen, the original intent of the FHA, CRA, and FCRA,

¹⁴⁹ *Id.*

¹⁵⁰ 15 U.S.C. § § 1607.1-1607.13

¹⁵¹ Allen, *supra* note 1.

respectively, included: introducing prohibitions against housing discrimination on the basis of protected categories, including race and national origin; establishing lending test policies that required banks to disclose information about their effort to lend in low-income communities; and ensuring banks and credit reporting bureaus were not using incorrect or biased information in evaluations.

Using principles of algorithmic accountability, Allen thus suggests the government consider the following improvements:

- The CRA could be updated to apply beyond banks and their locale. Allen proposes amendments to the act to apply to all online lending institutions, and not just geographic locale.
- The ECOA could be amended to include a provision that requires lenders to disclose the exact metrics or data points they use to generate scores or interest rates.
- Credit card companies, credit reporting bureaus, and mortgage lenders could be required to disclose the data inputs they use to formulate credit scores and mortgage rates.

Tailor accountability standards to different sectors and uses

Since there is no single silver bullet solution to the problem of algorithmic accountability, consideration should be given to decentralized regulatory approaches to more narrowly tailor policies, standards, and rules to consider factors such as the type of algorithmic system being used, what sector it is being used in, and how it is being used.

This approach could build on frameworks and approaches from existing policy and law. For example, the Federal Trade Commission, which enforces the Equal Credit Opportunity Act, may be able to make and enforce rules around algorithmic systems that inform credit decisions. The Department of Housing and Urban Development, which enforces the Fair Housing Act, could address issues of algorithmic discrimination in housing sales and rentals. Similarly, the Department of Labor could use the Equal Employment Opportunity Act to enforce algorithmic fairness in employment decision and opportunity.

Ideas abound for enhancing algorithmic accountability in medicine;¹⁵² risk assessments and predictive policing;¹⁵³ public housing;¹⁵⁴ health;¹⁵⁵ and hiring,¹⁵⁶ among many others. (See table below for examples.) Policymakers with expertise in these respective fields can consider evaluating each of these proposals to create industry standards, and regulators can then build corresponding enforcement structures.¹⁵⁷ When deployed concurrently, each of these policies can work together to ensure algorithmic accountability in several different areas.

Table 3: Sector-specific proposals toward achieving algorithmic accountability

Application	Proposals to improve algorithmic accountability
Credit scores	Establish a public credit agency with new algorithmic systems that draw on alternative data sources, which also excludes credit data that causes disparate racial impact ¹⁵⁸
Financial services	Improve transparency and explainability by setting a “floor” of industry-wide standards, through collaboration at the state and federal levels ¹⁵⁹
Lending	Shift fair lending law to outcome-focused analysis, as opposed to scrutinizing inputs ¹⁶⁰ ; encourage agencies to use alternative data for underwriting in loan application assessments ¹⁶¹
Housing	Expand Department of Housing and Urban Development’s 2016 guidance on using criminal records in tenant screening, and eviction records and credit reports in rental applications ¹⁶²

¹⁵² W. Nicholson Price II, [Regulating Black-Box Medicine](#), 116(3) MICH. L. REV. 421 (2017).

¹⁵³ Anthony Potts, [Implementation of Risk Assessment Tools in the Criminal Justice System: What Is a Fair Approach?](#), ROOSEVELT INSTITUTE (October 2020); Andrew Guthrie Ferguson, [Policing Predictive Policing](#), 94(5) WASH. U. L. REV. 1115 (2017); John Koepke & David Robinson, [Danger Ahead: Risk Assessment and the Future of Bail Reform](#), 93 WASH. L. REV. 1725 (2018); Ben Green, [“Fair” Risk Assessments: A Precarious Approach for Criminal Justice Reform](#), FAIRNESS, ACCOUNTABILITY AND TRANSPARENCY (2018).

¹⁵⁴ Rebecca Heilweil, [Tenants Sounded the Alarm on Facial Recognition in Their Buildings. Lawmakers Are Listening](#), VOX (December 26, 2019).

¹⁵⁵ [Ethics and Governance of Artificial Intelligence for Health: WHO Guidance](#), WORLD HEALTH ORGANIZATIONS (2021).

¹⁵⁶ Miranda Bogen & Aaron Rieke, [Help Wanted: An Examination of Hiring Algorithms, Equity, and Bias](#), UPTURN (December 2018).

¹⁵⁷ Koene et al., *supra* note 7, at 63-64.

¹⁵⁸ Amy Traub, [Establish a Public Credit Registry](#), DEMOS (April 3, 2019); Mikella Hurley & Julius Adebayo, [Credit Scoring in the Era of Big Data](#), 18 YALE J.L. & TECH. (2016).

¹⁵⁹ Kristin Johnson, Frank Pasquale & Jennifer Chapman, [Artificial Intelligence, Machine Learning, and Bias in Finance: Toward Responsible Innovation](#), 88(2) FORDHAM LAW REVIEW (2019).

¹⁶⁰ Talia Gillis, [False Dreams of Algorithmic Fairness: The Case of Credit Pricing](#) (2020).

¹⁶¹ [Addressing Technology’s Role in Financial Services Discrimination](#), ACLU 4 (July 13, 2021).

¹⁶² [Addressing Technology’s Role in Housing Discrimination](#), ACLU 5-7 (July 13, 2021).

Application	Proposals to improve algorithmic accountability
Hiring	Require algorithmic transparency reports through the U.S. Equal Employment Opportunity Commission, ¹⁶³ or through states or municipalities ¹⁶⁴ strengthen audits and rules through the Office of Federal Contract Compliance Programs ¹⁶⁵
Medicine	Regulate moderately upfront, while maintaining robust post market surveillance to monitor real-world performance through the Food and Drug Administration ¹⁶⁶
Risk assessments	Publish details on justice system’s algorithmic tools, and limit applications only to high-level offenses ¹⁶⁷ ; never recommend detention; instead, when a tool does not recommend immediate release, recommend a pretrial release hearing ¹⁶⁸

Proactively enforce through the federal government

Enforce algorithmic accountability in the federal government

While private sector businesses that use algorithmic systems may resist algorithmic accountability, the federal government can adopt best practices to ensure the accountability of algorithmic systems it uses within its own agencies and offices.

International, state, and local municipalities have been leading this charge: in 2020, the cities of Helsinki and Amsterdam in Europe launched the world’s first AI register to ensure citizens have access to information about: why and how algorithmic systems are being used; who is affected by the algorithms; what data is being used; how the data is processed; how discrimination is prevented; and how much human oversight there is; and how risks have been mitigated.¹⁶⁹ Also in 2020, New York City published a report of algorithmic tools being used by city government that impact city residents.¹⁷⁰

While recent announcements suggest greater interest in federal self-regulation of algorithms, early attempts have fallen short. Recently, the Government

¹⁶³ Max Langenkamp, Allan Costa & Chris Cheung, [Hiring Fairly in the Age of Algorithms](#) (April 2020).

¹⁶⁴ Tom Simonite, [New York City Proposes Regulating Algorithms Used in Hiring](#), WIRED (January 8, 2021).

¹⁶⁵ [Addressing Technology’s Role in Hiring Discrimination](#), ACLU 3 (July 13, 2021).

¹⁶⁶ W. Nicholson Price II, [Regulating Black-Box Medicine](#), 116 MICH. L. REV. 421 (2017).

¹⁶⁷ Anthony Potts, [Implementation of Risk Assessment Tools in the Criminal Justice System: What is a Fair Approach?](#) ROOSEVELT INSTITUTE (October 2020).

¹⁶⁸ [The Use of Pretrial “Risk Assessment” Instruments: A Shared Statement of Civil Rights Concerns](#), LEADERSHIP CONFERENCE ON CIVIL AND HUMAN RIGHTS.

¹⁶⁹ [Amsterdam and Helsinki Launch Algorithm and AI Register](#), AI REGULATION (October 13, 2020).

¹⁷⁰ Jeff Thamkittikasem, [Summary of Agency Compliance Reporting](#), NYC ALGORITHMS MANAGEMENT AND POLICY OFFICER (2020).

Accountability Office (GAO) released the results of an audit they conducted of 42 federal agencies that employ law enforcement officers about their use of facial recognition technology.¹⁷¹ The report revealed that the great majority of agencies who use facial recognition for criminal investigations have no process to track the technology's use.¹⁷² Not only this; recent reporting suggests that federal agencies who used Clearview AI, a facial recognition technology, failed to report that fact in the GAO report altogether.¹⁷³

Weeks later, the GAO published an accountability framework for federal agencies using artificial intelligence.¹⁷⁴ This is a good start toward federal governance, and a good guide to encourage better behavior, but more controls and accountability mechanisms need to be formalized.

To this end, research groups like AI Now at New York University and Data + Society have outlined clear recommendations for the federal government. For example, the president can take executive action that requires federal agencies to publish algorithmic impact assessments. AI Now's framework for the government includes the following recommendations:

- Self-assessments in which agencies catalogue all existing and proposed automated decision systems;
- Evaluations of potential impacts of each of these systems on fairness, justice, and bias;¹⁷⁵
- External researcher review processes as an additional check on agencies;¹⁷⁶
- Public disclosure of all of the above information;
- Open comment periods to solicit feedback or clarify questions on the above information, and

¹⁷¹ [Facial Recognition Technology: Federal Law Enforcement Agencies Should Better Assess Privacy and Other Risks](#), U.S. GAO (June 3, 2021).

¹⁷² Dave Gershgorn, [Federal Agencies Use Facial Recognition From Private Companies But Almost Nobody Is Keeping Track](#), VERGE (June 29, 2021).

¹⁷³ Caroline Haskins & Ryan Mac, [A Government Watchdog May Have Missed Clearview AI Use By Five Federal Agencies In A New Report](#), BUZZFEED (June 30, 2021).

¹⁷⁴ [Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities](#), U.S. GAO (June 2021).

¹⁷⁵ Nicol Turner Lee, Paul Resnick & Genie Barton, [Algorithmic Bias Detection And Mitigation: Best Practices And Policies to Reduce Consumer Harm](#), BROOKINGS INSTITUTE (May 22, 2019).

¹⁷⁶ See e.g., [Algorithmic Audits](#), ETICAS CONSULTING.

- Due process mechanisms for impacted individuals or communities from the use of these algorithmic decision systems.¹⁷⁷

Federal agencies that utilize algorithmic decision-making systems also could be required to purchase AI programs or services only from firms that have already conducted impact assessments.

Impact assessments may work to push policymakers to engage seriously with the complexity of algorithmic biases.¹⁷⁸ They can be designed to be public-facing, and to encourage public participation from communities, experts, researchers, journalists, and policymakers. Open accountability, transparency, and collaboration are all crucial to success in a nascent field filled with nuances and potential unintended consequences.¹⁷⁹ As one example, Data 4 Black Lives and Demos demand that the government “require algorithmic systems used for public purposes... to make decisions about housing, policing, or public benefits—to be open, transparent, and subject to public debate and democratic decision-making.”¹⁸⁰

Empower existing agencies for enforcement

The executive branch of the U.S. government is already equipped with the authority to address algorithmic discrimination and bias. For example, the Federal Trade Commission (FTC) has seldom-used authority in rulemaking, and underutilized authority in enforcement, on issues of algorithmic accountability.

To start, the FTC can consider interpreting Federal Trade Commission Act Section 5, Unfair or Deceptive Acts or Practices,¹⁸¹ to apply to cases of algorithmic unfairness. In his dissenting comment on the case of Bronx Honda,¹⁸² FTC Commissioner Rohit Chopra details how the Commission can use its unfairness authority to combat discrimination caused by algorithmic systems across the economy.

¹⁷⁷ Dillion Reisman, Jason Schultz, Kate Crawford & Meredith Whittaker, [Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability](#), AI NOW INSTITUTE (April 2018).

¹⁷⁸ Koene et al., *supra* note 7, at 53.

¹⁷⁹ Koene et al., *supra* note 7, at 53.

¹⁸⁰ Milner & Traub, *supra* note 4.

¹⁸¹ [CONSUMER COMPLIANCE HANDBOOK § 5](#) (Board of Governors of the Federal Reserve System 2016).

¹⁸² [Statement of Commissioner Rohit Chopra In the Matter of Liberty Chevrolet, Inc. d/b/a Bronx Honda](#), FEDERAL TRADE COMMISSION FILE NO. 1623238 (May 27, 2020).

Indeed, a 2016 study by the FTC¹⁸³ details the legal basis for this authority. By definition, an act or practice is unfair where it causes or is likely to cause substantial injury to consumers, cannot be reasonably avoided by consumers, and is not outweighed by countervailing benefits to consumers or to competition. Most of the examples of algorithmic harms listed earlier in this review would meet all these criteria.

Second, the FTC can consider engaging in rulemaking to combat algorithmic discrimination and bias. As the Electronic Privacy Information Center (EPIC) has petitioned,¹⁸⁴ the Commission also has authority to create broad rules that help defend against algorithmic discrimination. Under 15 U.S.C. § 57a of the FTC Act, the Commission is empowered to issue rules “which define with specificity acts or practices which are unfair or deceptive acts or practices in or affecting commerce,” which “may include requirements prescribed for the purpose of preventing such acts or practices.”¹⁸⁵ The Commission could use this authority to initiate rulemaking on accountability, transparency, and fairness in algorithmic decision-making systems.

A detailed potential pathway to use each of these authorities is outlined by FTC Commissioner Rebecca Slaughter in a recent whitepaper at the Yale Law School Information Society Project.¹⁸⁶

Create a new data protection agency to fill in gaps

Many democratic nations have a dedicated data protection agency with independent authority, oversight operations, and enforcement capability. While the FTC helps to safeguard consumers and promote competition, it does not have the expertise nor resources that are needed in a dedicated data protection agency.

¹⁸³ [Petition for Rulemaking](#). In Re: Petition for Rulemaking Concerning Use of Artificial Intelligence in Commerce, FEDERAL TRADE COMMISSION.

¹⁸⁴ Electronic Privacy Information Center, [Petition for Rulemaking Concerning Use of Artificial Intelligence in Commerce](#) (February 2020).

¹⁸⁵ 15 U.S.C. § 57a(a)(1)(B).

¹⁸⁶ Rebecca Kelly Slaughter & Janice Kopec & Mohamad Batal, [Algorithms and Economic Justice: A Taxonomy of Harms and a Path Forward for the Federal Trade Commission](#), ISP DIGITAL FUTURE WHITEPAPER & YALE JOURNAL OF LAW & TECHNOLOGY (August 2021).

Another solution to consider is establishing a data protection agency with resources, rulemaking authority, and effective enforcement powers. A data protection agency (DPA), not unlike the one Senator Kirsten Gillibrand has proposed through legislation,¹⁸⁷ in addition to governing privacy, security, and digital rights, can serve as the government enforcer of algorithmic accountability.

If a DPA were to be established, the agency could build a team of experts within it that works cross functionally with other relevant agencies (e.g., Federal Trade Commission, Consumer Financial Protection Bureau, Department of Housing and Urban Development) to create new rules and enforce existing ones related to algorithmic justice.

For example, James Allen suggests that the new agency could work with the Department of Housing and Urban Development (HUD), as well as state and local housing agencies, to ensure that algorithmic systems in affordable housing lotteries and rental tenancy applications are being administered fairly.

Similarly, the centralized agency could work with banking regulators like the Federal Deposit Insurance Commission (FDIC) and FTC to spearhead reforms for “algorithmic compliance,” so that the algorithmic systems used in credit ratings and lending evaluations satisfy the standards of the CRA.¹⁸⁸ This team of experts could then coordinate among the distributed areas of application, as noted above.

To ensure that its enforcement has teeth, Congress would need to allocate meaningful budget and authority for the DPA to levy significant penalties for organizations using or producing algorithmic systems without following these rules.

In addition, Ansgar Koene et. al and Allen suggest the responsibilities of such an agency could also include:

- Creating risk assessment matrices to classify algorithmic system types by potential for harm;
- Establishing standards for design, testing, and performance in algorithmic safety;

¹⁸⁷ [Data Protection Act of 2021](#), S.2134, 117th Cong. (2021).

¹¹⁷ Allen, *supra* note 1.

- Developing liability standards among coders, implementers, distributors, and end users
- Imposing requirements of transparency and explainability; and
- Advising other government agencies on their use of algorithmic systems.¹⁸⁹

¹⁸⁹ Koene, *supra* note 7, at 74.; Tutt, *supra* note 130.

Conclusion

In this review, Public Citizen attempts to highlight many of the harms that predictive algorithmic racism pose, ask questions to get closer to finding solutions, and identify next steps. We hope it serves to inspire further study, auditing, legislation, regulation, and enforcement toward algorithmic justice.

Yet, as mentioned in the introduction of this paper, issues of discriminatory and racist housing policy, credit terms, and policing are as old as mortgages, banks, and law enforcement themselves. Karen Hao of the MIT Technology Review describes the limitations of algorithmic justice, writing, “Algorithms cannot fix broken systems. They inherit the flaws of the systems in which they’re placed.”¹⁹⁰

To drive transformative, systemic change, we must couple work on algorithmic accountability with the pursuit of transformative justice that tackles the economic injustice that lie under algorithms. Rashida Richardson offers a framework for starting this process of transformative justice. She writes,

“Transformative justice requires a broader examination of collective responsibility in society for creating structural conditions and social practices that enable and perpetuate systemic harms and injustices.

“Such comprehensive and shrewd analysis can fashion radical social changes as well as a variety of technical and non-technical interventions that can adequately confront the intersectional and intergenerational nature of technology-mediate problems and withstand the current pace of innovation.”¹⁹¹

One way to do this is to consider not only how to ensure algorithms are not worsening inequity, but also how they might be used to improve equitable outcomes. In this way, we can start creating not only not-racist algorithms, but also proactively anti-racist algorithms.

¹⁹⁰ Karen Hao, [The UK Exam Debacle Reminds Us That Algorithms Can’t Fix Broken Systems](#), MIT TECHNOLOGY REVIEW (August 20, 2020).

¹⁹¹ Rashida Richardson, [Racial Segregation and the Data-Driven Society: How Our Failure to Reckon with Root Causes Perpetuates Separate and Unequal Realities](#), 36(3) BERKELEY TECHNOLOGY LAW JOURNAL (2022).

References

In this review, we build on scholarship and advocacy on algorithmic accountability that others have led for decades. Below is a list of organizations, academics, advocates, artists, and articles that have informed the perspectives in this review.

Publications

Aaron Rieke, David Robinson & Harlan Yu, Civil Rights, Big Data, and Our Algorithmic Future: A September 2014 Report on Social Justice and Technology, UPTURN (2014).

Algorithmic Accountability Policy Toolkit, AI NOW INSTITUTE (October 2018).

The Algorithmic Justice League's 101 Overview, ALGORITHMIC JUSTICE LEAGUE (2020).

Big Data and Differential Pricing, OFFICE OF THE PRES. OF THE U.S. (February 2015).

CATHY O'NEIL, WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY (Crown ed., 2016).

CHARLTON D. MCLWAIN, BLACK SOFTWARE: THE INTERNET AND RACIAL JUSTICE, FROM THE AFRONET TO BLACK LIVES MATTER (Oxford U. Press ed., 2019).

ED FINN, WHAT ALGORITHMS WANT: IMAGINATION IN THE AGE OF COMPUTING (MIT Press ed., 2017).

FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (Harvard U. Press ed., 2016).

Frank Pasquale, The Second Wave of Algorithmic Accountability, LAW & POLITICAL ECONOMY (November 25, 2019).

Ifeoma Ajunwa, The Paradox of Automation as Anti-Bias Intervention, 41 CARDOZO, L. REV (2020).

Ifeoma Ajunwa, An Auditing Imperative for Automated Hiring, 34 HARV. J.L. (2021).

JARON LANIER, WHO OWNS THE FUTURE? (Simon & Schuster ed., 2014).

Joseph Blass, Algorithmic Advertising Discrimination, 114(2) NW. U. L. REV. 415 (2019).

KARTIK HOSANAGAR, *A HUMAN'S GUIDE TO MACHINE INTELLIGENCE: HOW ALGORITHMS ARE SHAPING OUR LIVES AND HOW WE CAN STAY IN CONTROL* (Wharton School Press ed., 2019).

KARTIK HOSANAGAR, *A HUMAN'S GUIDE TO MACHINE INTELLIGENCE: HOW ALGORITHMS ARE SHAPING OUR LIVES AND HOW WE CAN STAY IN CONTROL* (Wharton School Press ed., 2019).

Civil Rights Principles for the Era of Big Data, THE LEADERSHIP CONFERENCE ON CIVIL AND HUMAN RIGHTS (February 2014).

LOUISE AMOORE, *CLOUD ETHICS: ALGORITHMS AND THE ATTRIBUTES OF OURSELVES AND OTHERS* (Duke U. Press ed., 2020).

Lorna McGregor, Daragh Murray, & Vivian Ng, International Human Rights Law as a Framework for Algorithmic Accountability, 68 INT'L. HUM. RTS. L. 309 (2009).

MARY F.E. EBELING, *HEALTHCARE AND BIG DATA: DIGITAL SPECTERS AND PHANTOM OBJECTS* (MacMillan ed., 2016).

Mary Madden, Michele E. Gilman, Karen Levy & Alice E. Marwick, Privacy, Poverty and Big Data: A Matrix of Vulnerabilities for Poor Americans, 95 WASH. U. L. R. 52 (2017).

MAR HICKS, *PROGRAMMED INEQUALITY: HOW BRITAIN DISCARDED WOMEN TECHNOLOGISTS AND LOST ITS EDGE IN COMPUTING* (MIT Press ed., 2017).

MEREDITH BROUSSARD, *ARTIFICIAL UNINTELLIGENCE: HOW COMPUTERS MISUNDERSTAND THE WORLD* (MIT Press ed., 2018).

Nicholas Diakopoulos, Algorithmic Accountability Reporting: On the Investigation of Black Boxes, COLUMBIA JOURNALISM SCHOOL: TOW CENTER FOR DIGITAL JOURNALISM (2013)/

Rashida Richardson, Jason M. Schultz & Kate Crawford, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice, N.Y.U.L. REV. ONLINE 15 (2019).

Rashida Richardson, Jason M. Schultz & Vincent M. Southerland, Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems, AI NOW INSTITUTE (September 2019).

RUHA BENJAMIN, *RACE AFTER TECHNOLOGY: ABOLITIONIST TOOLS FOR THE NEW JIM CODE* (Polity Press ed., 2019).

SASHA COSTANZA-CHOCK, *DESIGN JUSTICE: COMMUNITY-LED PRACTICES TO BUILD THE WORLDS WE NEED* (MIT Press ed., 2020).

SARAH ESTHER LAGESON, *DIGITAL PUNISHMENT: PRIVACY, STIGMA, AND THE HARMS OF DATA-DRIVEN CRIMINAL JUSTICE* (Oxford U. Press ed., 2020).

SAFIYA UMOJA NOBLE, *ALGORITHMS OF OPPRESSION: HOW SEARCH ENGINES REINFORCE RACISM* (NYU Press ed., 2018).

SIMONE BROWNE, *DARK MATTERS: ON THE SURVEILLANCE OF BLACKNESS* (Duke U. Press ed., 2015).

VIRGINIA EUBANKS, *AUTOMATING INEQUALITY: HOW HIGH-TECH TOOLS PROFILE, POLICE, AND PUNISH THE POOR* (St. Martin ed., 2018).

Advocates, researchers, and organizers

Action Center on Race and the Economy (ACRE); 21st Century Policing Report

Amba Kak (Director, Global Policy & Programs, AI Now Institute)

Andrea M. Matwyshyn (Associate Dean, Penn State Law)

Cathy O’Neil (Founder, O’Neil Risk Consulting & Algorithmic Auditing)

Chris Gilliard (Fellow, Harvard Kennedy School Belfer Center for Science and International Affairs)

Data & Society; Data & Civil Rights: Criminal Justice Primer, Algorithmic Accountability: A Primer

Data 4 Black Lives and Demos; Data Capitalism + Algorithmic Racism

Deb Raji (Fellow, Mozilla)

Frank Pasquale (Professor of Law, Brooklyn Law School)

Ifeoma Ajunwa (Professor, University of North Carolina School of Law)

Joy Buolamwini (Founder, Algorithmic Justice League)

Karen Hao (Senior Reporter on AI at MIT Technology Review)

MediaJustice and Movement Alliance Project; Mapping Pretrial Injustice

Meredith Broussard (Research Director, NYU Alliance for Public Interest Technology)

Meredith Whittaker (Founder, AI Now Institute)

Mona Sloane (Fellow, NYU Institute for Public Knowledge)

Ruha Benjamin (Professor, Princeton University)

Rumman Chowdhury (Director, ML Ethics, Twitter)

Rashida Richardson (Assistant Professor, Northeastern University)

Sasha Costanza-Chock (Steering Committee, DesignJustice.org)

Safiya Noble (Associate Professor, UCLA)

Suresh Venkatasubramanian (Professor, University of Utah)

Timnit Gebru (Computer scientist)

Upturn; Recent work including amicus briefs, reports, and testimony

Virginia Eubanks (Associate Professor, University of Albany SUNY)