

Deprogramming Implicit Bias: The Case for Public Interest Technology

Darren Walker

New technologies have fundamentally transformed the systems that govern modern life, from criminal justice to health care, housing, and beyond. Algorithmic advancements promise greater efficiency and purported objectivity, but they risk perpetuating dangerous biases. In response, the field of public interest technology has emerged to offer an interdisciplinary, human-centered, and equity-focused approach to technological innovation. This essay argues for the widespread adoption of public interest technology principles, including thinking critically about how and when technological solutions are deployed, adopting rigorous training to educate technologists on ethical and social context, and prioritizing the knowledge and experiences of communities facing the disproportionate harms or uneven benefits of technology. Tools being designed and deployed today will shape our collective future, and collaboration between philanthropy, government, storytellers, activists, and private-sector technologists is essential in ensuring that these new systems are as just as they are innovative.

Three years ago, Robert Julian-Borchak Williams, a Detroit office worker, received a call from the Detroit Police Department. He assumed it was a prank, but when he pulled into his driveway, police officers were waiting in his front yard. They handcuffed Robert in front of his wife and daughters, and refused to answer his family's panicked questions. Williams spent the night in a crowded jail cell. The next afternoon, the day before his forty-second birthday, the police brought him to an interrogation room. Stone-faced detectives showed him photographs of a robbery suspect. "Is this you?" they demanded. Williams held the photograph next to his face. The image clearly displayed a different man. The reason for Williams's unjustified arrest was not a witness statement or a botched DNA match. Instead, Williams had been falsely identified by law enforcement officers who used a faulty facial recognition algorithm to ensnare the wrong man in the criminal legal system.¹ While Robert Williams's story is alarming, it is not an anomaly. Since the Detroit Police Department began using facial recognition, at least two other Black men in the same city have been falsely arrested, destroying their job prospects and fracturing a marriage.² One of these men even considered

accepting a plea deal for a crime he did not commit. In fact, Detroit's facial algorithm misidentifies suspects more than 90 percent of the time.³ Yet it is still used widely across the department, nearly exclusively against Black people. In Detroit, as elsewhere across the country, technology replicates, reinforces, and indeed masks human bias on a scale we have never encountered before, a scale only accessible in the language of machines. Algorithms, artificial intelligence, and technology pervade our criminal legal system, often with little oversight. Judges use risk-assessment technology to determine parole and probation terms.⁴ Compared with white defendants, some of these tools are 77 percent more likely to predict that Black defendants will commit a violent offense.⁵

These harmful algorithms extend beyond the criminal legal system, to the services that determine health and safety. An algorithm used to manage health care for two hundred million people in the United States was found to refer disproportionately few Black people to programs providing personalized care, even though Black patients were often substantially more ill than their white counterparts.⁶ Meanwhile, landlords across the country increasingly rely on artificial intelligence to screen applicants, including with algorithms that can penalize applicants for criminal accusations that are later dropped.⁷ Even issues as mundane as the photos we see on our screens are affected by biased technology. In one widely cited example, a Google Photos algorithm falsely identified Black people as gorillas.⁸ Technologies that once seemed confined to science-fiction novels are now embedded in our democracy, and with them, a host of algorithmic biases at a colossal and concerning scale. These examples, among many others, indicate a recursive problem. Our algorithms are embedded with the biases of the humans who create them; and with each additional algorithm built atop an unjust foundation, the initial bias recurs, repeats, and worsens, to devastating effect.

When privatized, without oversight and careful regulation, this self-sustaining cycle of algorithmic bias will continue unabated, not only exacerbating existing inequality but creating new inequalities altogether. As Latanya Sweeney, head of Harvard's Public Interest Tech Lab and former chief technology officer of the U.S. Federal Trade Commission, rightly noted, "Once a design or business practice works, it gets replicated just as it is. The design of the technology really does dictate the rules that we have to live by."⁹ Those of us invested in a more just and equitable future face an urgent question: How do we address this mounting crisis of algorithmic injustice?

Some argue that the project of reforming technology is best left in the hands of programmers and specialists: the technical experts who designed these systems. As technology advances, this logic contends, its consequences will reveal themselves, and then be corrected by the forward march of new technology. Certainly, these groups have crucial expertise and insight needed to understand the algo-

rhythms that define our lives. But the growth-at-any-cost mindset that pervades the tech industry often overlooks the realities of race, gender, and disability inequities, and risks repeating a vicious cycle *ad infinitum*.¹⁰

On the other end of the spectrum, a coalition of industry leaders and technologists recently signed a letter calling for an AI development moratorium.¹¹ This short-term solution would do little to address the structural issues that shape the development of artificial intelligence. For instance, while it might tackle discrete safety concerns, it is unlikely to fundamentally shift the training that computer scientists and engineers receive to grapple with technology's unintended consequences for marginalized groups. A tech-imposed temporary stoppage also problematically implies that the industry is self-governed, which is simply not true. Existing federal regulatory schemes, from product liability statutes to civil rights protections, already apply to artificial intelligence.¹² The answer is not to ask for a proverbial time out, but rather to bring in the referees: the advocates and regulators who carry the capacity and technical expertise to enforce laws and correct violations at scale. Moving forward, we should address this recursive problem the way we would any other: by breaking it down into a series of smaller subproblems and solving them one at a time. We might start by investing in the excellence of a new generation of talented technologists with the technical expertise, interdisciplinary training, and lived experience to deploy strategies that end algorithmic bias, once and for all.

The good news is academics, advocates, and technologists have been engaged in this work for years, building the new field of public interest technology together. This interdisciplinary approach calls for technology to be designed, deployed, and regulated in a responsible and equitable manner.¹³ It goes beyond designing technology for good, asking and answering: “Good for *whom*?” Public interest technologists center *people*, not innovation for its own sake. They focus on those most affected by new innovations: the historically marginalized groups who have experienced the most harms or the uneven benefits of technology. At the same time, public interest technologists understand that technology is not, and never has been, neutral. The dangers of technology, they argue, cannot be resolved with one product or program. Instead, these technologists evaluate and address potential inequalities at every stage of innovation, from design and development to the real-world impact in the hands of users. The field includes leading technical experts, researchers, and scientists. And it invites those outside of technology – storytellers, activists, artists, and academics – to offer their crucial expertise and hold designers and decision-makers accountable. As celebrated filmmaker Ava DuVernay noted about the artist’s role in addressing these harms: “The idea that the story that technology is telling about us could possibly not be our true story, makes it just as important as any crime thriller I might be

covering.”¹⁴ Simply put, public interest technology is a multisector effort. It calls everyone to consider how we use, encourage, and adopt technology in our lives, our fields, and our broader institutions.

From academics to funders to private-sector innovators, we can all benefit from taking a public interest technology approach to our work. First, we can and must question the gospel of tech solutionism.¹⁵ Instead of assuming new technology will inevitably correct a social ill, we must think more critically about how and when technology is deployed. Being more intentional about the technology we adopt can move us from *reacting* to unforeseen consequences to *preventing* these negative effects. For example, the Algorithmic Justice League, an organization devoted to “unmasking AI harm,” and other advocates recently prevented the Internal Revenue Service from implementing a controversial plan forcing taxpayers to use facial-recognition software to log in to their IRS accounts.¹⁶ The change would have exposed millions to privately owned software with limited oversight.

Second, we must also embed rigorous public interest technology training in computer science, engineering, and data science curriculums. Such training will ensure that talented technologists graduate with both technical expertise and an extensive understanding of the social context in which technology is deployed. These efforts may also include funding or pursuing research and projects that interrogate how technology furthers systemic bias.¹⁷ Such revelations have come from resource hubs like those at Harvard’s Public Interest Tech Lab.¹⁸ Researchers and scientists at the lab have unmasked biased Facebook advertising algorithms that targeted Black users and exposed the proliferation of deepfake comments in U.S. public comment sites.¹⁹ And educational institutions nationwide are building the next generation of public interest technologists – together. The Public Interest Technology University Network unites sixty-three universities in connecting public interest students and faculty with resources and institutional support.²⁰

Of course, any attempt to correct technology’s ills will fall short if we do not center the knowledge and experiences of the marginalized people most vulnerable to its inherent risks. So, technologists can and must partner with marginalized communities to repair the damage caused by bias and prevent it from the outset. For instance, after studies revealed that non-white Airbnb hosts were earning less money than their white counterparts, Airbnb partnered with civil rights organizations to create Project Lighthouse, an initiative to reduce discrimination for hosts and travelers on the platform.²¹ These efforts drew on the experiences of Black hosts and guests, who shared their struggles with securing housing under the hashtag #AirbnbWhileBlack.²²

Finally, public interest technologists themselves can and must draw on their own intersectional experiences, with support from funders and academic institutions alike. At the Ford Foundation, our commitment to public interest technol-

ogy arose out of a strategy to promote internet rights and digital justice. Through our Technology and Society program, Ford has committed more than \$100 million to fostering the field of public interest technology since 2016 – all to build an ecosystem that will lead to a more just technological future for all. Many researchers affiliated with the program have personally experienced the harms of biased algorithms or inaccessible technology. They bridge specialized expertise with a rich personal background, advocating for structural and long-term solutions like an AI Bill of Rights, which would ensure that a shared set of norms and values shape technology to better serve the public good.²³

Technology's ever-changing landscape presents a daunting challenge. Nevertheless, I am hopeful for a future in which technology empowers us to serve the public good, because I know we've solved these problems before. Indeed, the ideological ancestor of the public interest technology field exists. It is called public interest law.

Six decades ago, during the early 1960s, there was no such thing as public interest law. Law schools focused on academic and corporate issues to the detriment of addressing social inequities. Legal aid groups struggled to survive. But the Ford Foundation set out to change that and to train a new generation of lawyers who would work in the best interest of the public to provide legal representation to low-income and marginalized groups, engage in advocacy more broadly, and expand rights throughout society. By the time I graduated from law school in the mid-1980s, the once-nascent field was flourishing. Today, public interest law is so prominent that many take it for granted. Low-income tenants who have been evicted can join a class-action lawsuit, free of charge. Young people fleeing discriminatory anti-LGBTQ+ legislation can access entire organizations dedicated to supporting their legal rights. The field is far from perfect but it's a prescient reminder that time, investment, and collaboration can turn a sore lack into a surplus. Those who have long driven the field of public interest law – people of color, people with disabilities, low-income people, and LGBTQ+ people – are best equipped to fight a barrage of implicit bias-based challenges. If we support them, we can build a parallel public interest field anew.

The technology that determines our housing, health, and safety cannot and must not be the protected intellectual property of a few. It is a public good for the many. And people from every sector can contribute to a more just vision of tech by extending support and funding for crucial research, welcoming public interest technologists to nontechnical fields, and advancing solutions that reject the philosophy of “move fast and break things” by instead calling us all to fix what is broken.²⁴ By embarking on this mission to center people in the technology that is supposed to help us, we move toward justice for the millions of people who face algorithmic bias in their everyday lives, including Robert Williams, who is still reckoning with the consequences of his false arrest. It has been three years since Williams was wrongly

handcuffed on his front lawn, but his seven-year-old daughter still cries when she sees his arrest footage.²⁵ And still the recursive loop circles.

On November 25, 2022, Randal Reid, a Black man, was driving in Georgia to a late Thanksgiving celebration with his mother. Police pulled him over, announcing there was a warrant for his arrest for a theft that had occurred in Louisiana.²⁶ Reid pleaded that he had never spent a day in Louisiana. Yet he was booked and spent six days in jail based on an incorrect facial recognition match claiming he was a man forty pounds heavier and without a mole on his face. Let us learn with humility from the shattering experiences endured by too many families and break this recursive loop before it's too late.

ABOUT THE AUTHOR

Darren Walker, a Fellow of the American Academy since 2015, is President of the Ford Foundation. He previously served as Vice President of the Rockefeller Foundation, where he oversaw the Rebuild New Orleans initiative after Hurricane Katrina. He is the author of *From Generosity to Justice: A New Gospel of Wealth* (2023) and has published in journals such as *JAMA* and *Foreign Affairs*.

ENDNOTES

- ¹ Kashmir Hill, "Wrongfully Accused by an Algorithm," *The New York Times*, June 24, 2020, <https://www.nytimes.com/2020/06/24/technology/facial-recognition-arrest.html>.
- ² Khari Johnson, "How Wrongful Arrests Based on AI Derailed 3 Men's Lives," *Wired*, March 7, 2022, <https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives>.
- ³ Jason Koebler, "Detroit Police Chief: Facial Recognition Software Misidentifies 96% of the Time," *Vice*, June 29, 2020, <https://www.vice.com/en/article/dyzykz/detroit-police-chief-facial-recognition-software-misidentifies-96-of-the-time>.
- ⁴ Michael Brenner, Jeannie Suk Gersen, Michael Haley, et al., "Constitutional Dimensions of Predictive Algorithms in Criminal Justice," *Harvard Civil Rights-Civil Liberties Law Review* 55 (1) (2020): 267–310.
- ⁵ Julia Angwin, Jeff Larson, Surya Mattu, and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; and Jeff Larson, Surya Mattu, Lauren Kirchner, and Julia Angwin, *ProPublica*, May 23, 2016, "How We Analyzed the COMPAS Recidivism Algorithm," <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.
- ⁶ Heidi Ledford, "Millions of Black People Affected by Racial Bias in Health-Care Algorithms," *Nature* 574 (7780) (2019): 608–609, <https://doi.org/10.1038/d41586-019-03228-6>.

- ⁷ Valerie Schneider, “Locked Out by Big Data: How Big Data, Algorithms, and Machine Learning May Undermine Housing Justice,” *Columbia Human Rights Law Review* 52 (1) (2020): 251–305.
- ⁸ Jenna Burrell, “How the Machine ‘Thinks’: Understanding Opacity in Machine Learning Algorithms,” *Big Data & Society* 3 (1) (2016): 1–12, <http://doi.org/10.2139/ssrn.2660674>.
- ⁹ Dave Gershgorin, “A Harvard Professor Thinks that Tech’s True Power Comes from Design,” *Quartz*, February 24, 2018, <https://qz.com/1214645/latanya-sweeney-explains-why-tech-companies-are-so-powerful>.
- ¹⁰ Greta Byrum and Ruha Benjamin, “Disrupting the Gospel of Tech Solutionism to Build Tech Justice,” *Stanford Social Innovation Review*, June 16, 2022, https://ssir.org/articles/entry/disrupting_the_gospel_of_tech_solutionism_to_build_tech_justice.
- ¹¹ Cade Metz and Gregory Schmidt, “Elon Musk and Others Call for Pause on A.I., Citing ‘Profound Risks to Society,’” *The New York Times*, March 29, 2023, <https://www.nytimes.com/2023/03/29/technology/ai-artificial-intelligence-musk-risks.html>.
- ¹² See Charlotte A. Burros, Rohit Chopra, Kristen Clarke, and Lina M. Khan, “Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems,” Federal Trade Commission, April 25, 2023, <https://www.ftc.gov/legal-library/browse/cases-proceedings/public-statements/joint-statement-enforcement-efforts-against-discrimination-bias-automated-systems>.
- ¹³ Katharine Lusk, “Public Interest Technology University Network: Understanding the State of the Field,” Boston University Initiative on Cities, May 31, 2022, <https://open.bu.edu/handle/2144/44469>.
- ¹⁴ See Public Interest Technology University Network, “Using Tech for Good: A New Generation of Civic-Minded Technologies,” filmed October 7, 2019, at Georgetown University, Washington, D.C., video, [38:45], <https://www.youtube.com/watch?v=cP1IMsFQMDQ&t=2s>.
- ¹⁵ Byrum and Benjamin, “Disrupting the Gospel of Tech Solutionism to Build Tech Justice.”
- ¹⁶ Rachel Metz, “Activists Pushed the IRS to Drop Facial Recognition. They Won, but They’re Not Done Yet,” *CNN Business*, March 7, 2022, <https://amp.cnn.com/cnn/2022/03/07/tech/facial-recognition-activists-irs/index.html>; and Joy Buolamwini, “The IRS Should Stop Using Facial Recognition,” *The Atlantic*, January 27, 2022, <https://www.theatlantic.com/ideas/archive/2022/01/irs-should-stop-using-facial-recognition/621386>.
- ¹⁷ Lusk, “Public Interest Technology University Network.”
- ¹⁸ See Harvard University, “Public Interest Tech Lab,” <https://techlab.org>.
- ¹⁹ Jinyan Zang, “How Facebook’s Advertising Algorithms Can Discriminate By Race and Ethnicity,” *Technology Science*, October 19, 2021, <https://techscience.org/a/2021101901>; and Max Weiss, “Deepfake Bot Submissions to Federal Public Comment Websites Cannot Be Distinguished from Human Submissions,” *Technology Science*, December 17, 2019, <https://techscience.org/a/2019121801/#Citation>.
- ²⁰ See New America, “The Public Interest Technology University Network,” <https://pitcases.org>.
- ²¹ Benjamin Edelman, Michael Luca, and Dan Svirsky, “Racial Discrimination in the Sharing Economy: Evidence from a Field Experiment,” *American Economic Journal: Applied Economics* 9 (2) (2017): 1–22, <https://doi.org/10.1257/app.20160213>; and Airbnb, “A New Way

We're Fighting Discrimination on Airbnb," June 15, 2020, <https://www.airbnb.com/resources/hosting-homes/a/a-new-way-were-fighting-discrimination-on-airbnb-201>.

- ²² Maggie Penman, Shankar Vedantam, and Max Nesterak, "#AirbnbWhileBlack: How Hidden Bias Shapes the Sharing Economy," NPR, April 26, 2016, <https://www.npr.org/2016/04/26/475623339/-airbnbwhileblack-how-hidden-bias-shapes-the-sharing-economy>.
- ²³ The White House, "Blueprint for an AI Bill of Rights," October 2022, <https://www.whitehouse.gov/ostp/ai-bill-of-rights>.
- ²⁴ Hemant Taneja, "The Era of 'Move Fast and Break Things' Is Over," *Harvard Business Review*, January 22, 2019, <https://hbr.org/2019/01/the-era-of-move-fast-and-break-things-is-over>.
- ²⁵ Johnson, "How Wrongful Arrests Based on AI Derailed 3 Men's Lives."
- ²⁶ Kashmir Hill and Ryan Mac, "'Thousands of Dollars for Something I Didn't Do,'" *The New York Times*, March 31, 2021, <https://www.nytimes.com/2023/03/31/technology/facial-recognition-false-arrests.html>.