DATA STEWARDSHIP
FOR GOOD

Power and Progress in Algorithmic Bias

By Kristine Gloria, PhD, Aspen Digital
July 2021

ASPEN DIGITAL
aspen institute

Center for
Inclusive Growth

Foreword

Algorithmic bias has real-world implications that manifest in issues from employment, to housing, to healthcare. It is often described as "systematic and repeatable errors in a computer system that create unfair outcomes."[1] It is a pernicious challenge that at minimum exposes decades of discriminatory practices against certain communities and at worst has actively harmed specific groups (e.g. facial recognition systems). History underscores that bias in technology systems is not a new phenomenon. Instead, we are at a moment in which we have both the tooling and critical capacity to recognize, analyze, and potentially address how automated decision-making systems have negatively affected specific communities.

Aspen Digital recently launched the Data Stewardship for Good Initiative as part of its work with the Global Inclusive Growth Partnership (GIGP), a collaboration between the Aspen Institute and the Mastercard Center for Inclusive Growth. The Data Steward Initiative is focused on empowering industry decision makers to take action around discriminatory data practices. The first part of this initiative is centered on understanding the impact of algorithmic bias through analysis and in-depth discussions. The overall work is motivated by a larger need to understand how technology, and information culture, can help eradicate social inequalities.[2] This effort highlights and seeks to address the harmful impacts of algorithmic bias on historically excluded communities.

The mission of this effort is to elevate the voices and ideas that have long pursued inclusion, equity, and justice, particularly in the digital realm. We are not the first to highlight digital inequities in large scale systems. Research and advocacy on this issue spans decades. Our process encompasses an extensive literature review, in-depth interviews with scholars and experts from across disciplines and convening key cross-sector stakeholders to discuss current findings and potential solutions. We applied a values-based lens focused on fairness, accountability, and transparency. Our North Star throughout this process was the following: *"How can we (data and algorithmic operators) recast our own models to forecast for a different future, one that centers around the needs of the most vulnerable?"* The result is the beginning articulation of a digital bill of rights in support of low-income, marginalized communities that may be impacted by such automated systems.

This report is a synthesis of the various discussions and learnings we have uncovered along the way. It outlines past and current efforts to tackle this question through concrete programs and initiatives, reviews the practical considerations of politics and power, and industry

[1] Algorithmic Bias. (n.d.) Wikipedia. https://en.wikipedia.org/wiki/Algorithmic_bias

[2] Offered by Noble, S. U. as a key goal in reimagining how information online can and should function; from (2018). *Algorithms of Oppression.* New York University Press.

willingness to address these issues, and discusses the need for greater civil society involvement and oversight. Our hope is that these findings will help guide legislation, industry guidelines, and civil society in addressing algorithmic bias, and the harm it causes for historically excluded and marginalized communities. We also offer a beginning framework that captures the rights and privileges not yet afforded to these communities. We see this as a work in progress and invite others to participate in its development and adoption. Specifically, we submit this report and its recommendations in hopes to shift the focus from highlighting problems towards pragmatic solutions.

What is algo bias?

Algorithmic bias is — at its core — about power. It is the literal codification of what knowledge, which values, and whose truth is most valued. It calls into question: who has the authority and privilege to designate pieces of knowledge; whose narrative is left out; and to what extent are these human decisions embedded across various technical systems?

More directly, algorithmic bias describes a repeatable error that creates unfair outcomes for certain groups and often manifests within large-scale technology systems. As this piece outlines, there are several entry points by which bias may be interjected within a technical system and/or mitigated. The primary takeaway, however, is that such bias reflects not just a single component but the role a set of actors play in shaping the algorithm and the system in which they operate.

A brief history

Algorithmic bias, whether economic, political, or cognitive, influences how we interact with technology every day. It is also reinforced by the technology itself. This is further complicated by the growing reliance on — and blind faith in — such automated technologies to manage additional aspects of our daily lives more accurately, and more objectively, than any human might accomplish. Therefore, the more flexible and powerful learning systems become, the greater the need to understand what, exactly, they are learning to do on our behalf.[3]

It is clear that bias can be introduced at various stages in the lifecycle of any technical system. Whether it can be completely prevented, avoided, and/or deleted remains unknown. However, one entry point from which much of the criticisms arise is on the data level.

[3] Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values.* W.W. Norton & Company, Inc.

Bias in data

The literature identifies five main stages in data development: *Define and Collect; Label and Transform; Analysis and Insight; Modeling; and Deployment*.[4] To illustrate, in the first stage, organizations define a problem and begin to discover and collect internal or external sources of data to solve that problem. Data collection is a systematic approach that is grounded in solving for the problem statement. This first stage is crucial for any attempt towards responsible data use that is principled and moral, particularly as it relates to personal information.[5] Unfortunately, bias may enter at this stage for a variety of reasons such as a lack of representative data, data labeling,[6] and/or cost prohibitive access to data.

Bias at the datal level is vexing but not insurmountable. In statistics, quality data will present features such as completeness and comprehensiveness, accuracy and precision, timeliness and relevance, etc. But, statistical correctness is not enough to solve for an entire system's bias. The real challenge, as Princeton Associate Professor of Computer Science Arvind Narayanan posed in his lecture, is "how do we make algorithmic systems support human values?"[7] With this reframing, we can begin to recognize the complexity and the critiques of algorithmic systems to include issues beyond just data, but of societal bias as well.

Embedded biases in technical systems are not a new phenomenon. In 1950s Britain, the largest of the government's computer installations was housed within the Ministry of Pensions and National Security. By the mid-1950s, hundreds of trans Britons had petitioned the Ministry of Pension to correct the gender on their benefits cards in order to gain legal employment. These requests for change, noted by historian Mar Hicks, resulted in a centralized list, making it possible to track every transgender citizen in the country. The list was shared, unbeknownst to the individuals, with government medical institutions for longitudinal medical studies on the nature of the trans identity and the process of transitioning.[8] When the system changed and required that records be punched into the mainframe, the Ministry changed course and did not honor requests for gender identity changes, describing the new system as one blind to gender. As Hicks notes, this resulted in

[4] This is informed by the "Data Science Value Chain" from data.org

[5] *The Global Data Responsibility Imperative*. (2019, October). Mastercard. https://www.mastercard.us/content/dam/mccom/en-us/documents/global-data-responsibility-whitepaper-customer-10232019.pdf

[6] Obermeyer, Z. Powers, B., Vogeli, C. & Mullainathan, S. (Oct. 2019). "Dissecting racial bias in algorithm used to manage the health populations." *Science*. https://science.sciencemag.org/content/366/6464/447

[7] Narayanan, A. [Arvind Narayanan]. (2018, March 1). *Tutorial: 21 fairness definitions and their politics* [video]. YouTube. https://www.youtube.com/watch?v=jIXIuYdnyyk&ab_channel=ArvindNarayanan

[8] Hicks, M. (2019). Hacking the Cis-tem. *IEEE Annals of the History of Computing*, 41(1), 20-33. https://ieeexplore.ieee.org/document/8634814

discriminatory practices: “In essence, the Ministry had decided to use the power of the new computer system to resubmerge trans citizens and their requests for recognition.”[9]

Bias in algorithms

Fast forward to today, we find numerous studies on the effects of algorithmic bias. Safiya U. Noble’s groundbreaking 2018 research, *Algorithms of Oppression: How Search Engines Reinforce Racism*, demonstrates how commercial engines like Google enable the act of “technological redlining.” Through an analysis of media searches and online paid advertising, Noble exposes the impact that classification and the organization of information has over certain communities, especially those that have no direct redress for how they are represented, categorized, and indexed through searches. For example, Noble points to the use of stereotypes of Asian, Latina, and Black women to inform search engine results, returning over-sexualized images and porn.

The key lesson history has shown us is that the actors, beneficiaries, and stakeholders of algorithms are not equally present in the process of developing algorithms. The primary actors, *algorithmic creators and deployers,* typically do not represent the individuals who would be negatively impacted by the system’s decision. To address these challenges will require a multi-prong, multi-stakeholder approach that expands beyond just internal proactive interventions but to additional oversight, regulatory action, and, perhaps most importantly, public pressure.

What are the major issues with fairness, accountability, and transparency?

Over the past decade, researchers from various disciplines have critiqued technical systems against three social values: fairness, accountability, and transparency. This is a matter of utmost urgency as algorithmic creators and deployers continue to design, develop, use, and distribute systems built on biased models and data. This is particularly pressing as governments move towards digitizing operations and processes from public healthcare to public education and credit scoring. While there are other common threads between the major issues with fairness, accountability, and transparency in algorithmic models, this section will outline the issues unique to each value.

[9] Hicks, M. (2019). Hacking the Cis-tem. *IEEE Annals of the History of Computing*, 41(1), 20-33. https://ieeexplore.ieee.org/document/8634814

Fairness

What does fairness mean when producing algorithmic systems? Here the key issues are how we define fairness and for whom the system is meant to be fair. "How we should start talking about fairness is with justice, which is the moral and ethical treatment of people," suggested one expert. Approaches to fairness range from the political, philosophical, social, and economic. These varied approaches combined with human shortcuts (e.g. use of a convenience sample or off-the-shelf machine learning tools) taken in the machine learning lifecycle result in unwanted bias within the system. This point was underscored by several researchers who noted that even if group or statistical parity can be achieved, the outcome may not be fair, indicating that parity is not sufficient. As one participant shared, "It's not a math problem, it's not a statistics problem, it is not a computer science problem. It is a political problem."

One salient example, discussed by roundtable participants, is in determining fairness in credit scoring algorithms. Sian Townson wrote in 2020, increasing deployment of AI in credit decisions sheds light on a system that has a long history of discriminatory practices against protected characteristics, such as race, gender and sexual orientation.[10] Participants noted that instead of training models on historical, incomplete, and/or biased data and optimizing for greater return for the lender, these models need to reframe the problem statement. Instead, algorithmic systems should solve for lending for a more equitable world.

While reconfiguring the technical system is one piece of the puzzle, norms and regulations should be considered in parallel. For example, in a 2017 *University of Pennsylvania Law Review* article, authors found that "there may be cases where allowing an algorithm to consider protected class status can actually make outcomes fairer" requiring a "doctrinal shift" in the consideration of protected status as presumptively a legal harm.[11] Moreover, in forming regulatory guidelines, experts cautioned against the risk of generating "rudimentary checks" that are no more fair than flipping a coin. Fairness is therefore only as useful as the oversight mechanisms that are in place to evaluate whether a system accomplishes what it is designed to do.

[10] Towson, S. (November 2020). AI Can Make Bank Loans Fair. *Harvard Business Review*. https://hbr.org/2020/11/ai-can-make-bank-loans-more-fair

[11] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, 165(3), 633-705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

Accountability

Even if a system is deemed to be fair, experts caution the need for its continuous monitoring and iteration, particularly as context changes. This brings to bear the question: "who does the algorithmic decision-making system need to be accountable to?" Is it enough for such systems to be held accountable to only its creators and deployers? Perhaps in a few narrow use cases. However, in systems where data is used predictively to automate decisions on an entire class of people (especially classes historically excluded), then accountability requires further consideration. Current recommendations include the need for third-party oversight, such as a regulatory regime, and/or the use of algorithmic auditing. Unfortunately, as one participant noted, "the accountability part of 'fairness, bias, accountability, and transparency,' is always the least developed of the approaches."

Currently, civil society plays a tremendous role as third-party oversight and arbiter. Frameworks such as *algorithmic hygiene*[12] have enabled others to clearly articulate specific causes of biases and to identify mitigation tactics. Take for example targeted advertising, which relies on algorithms to customize content we come across on our computers, or in scrolling social media platforms. These algorithms allow advertisements to target specific audiences for certain sales or transactions. While some targeted ads provide needed services to key communities, such as employment opportunities in a certain area, bias in data can result in differential treatment. For example, online retail advertising which relies on micro-targeting tactics, have been tied to price discrimination, racial profiling, and exploitation.[13]

One potential accountability solution, which also offers ordinary users more direct agency, is to develop an active feedback loop mechanism between customers and companies. Benefits of this approach include exposing harmful advertising and illuminating those specific companies that must be held accountable for their algorithmic harm. Participants also discussed other agency-based tactics, such as data obfuscation[14] or data strikes, which empower consumers to push back and protest against pervasive digital surveillance. A third potential option is the adoption and employment of a personal AI, which could represent the

[12] Turner-Lee, N., Resnick, P., & Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

[13] Nathan, N. (2013). How Big Data Enables Economic Harm to Consumers, Especially to Low-Income and Other Vulnerable Sectors of the Population. FTC.gov. https://www.ftc.gov/system/files/documents/public_comments/2014/08/00015-92370.pdf

[14] Brunton, F., and Nissenbaum, H. (2011). Vernacular Resistance to Data Collection and Analysis: A Political Theory of Obfuscation. FirstMondays.org. https://firstmonday.org/article/view/3493/2955. *See also* Finn Brunton and Helen Nissenbaum, Obfuscation: A User's Guide for Privacy and Protest (2015).

end users' interests by mediating directly with the platform-based biased AI systems.[15] While these methods are useful, some experts highlighted that it places at least some of the burden back onto an individual consumer to defend themselves against potential harm.

Transparency

Echoing the conversation around fairness and accountability, transparency relies on addressing the multi-dimensions of "why"—"why is the model the way it is?" (backward causation) and "what was this model being optimized for? And for whom?" (forward causation). These questions seem rudimentary, yet increasingly difficult to answer. Why? Because the power and flexibility of neural network based systems, which is a subset of machine learning and core to deep learning models, are often opaque and unintelligible, meaning even its creator may not be able to explain how the model came to its conclusion.

Transparency is a popular mitigation strategy, equated with the need to see and observe a phenomenon in order to bring about accountability and governance. Transparency can apply at differing levels such as platform design down to a system's source code. However, as Kroll et. al, cautions: "source code alone teaches a reviewer very little, since the code only exposes the machine learning method used and not the data-driven decision rule"[16] Instead, Ananny and Crawford suggest reimagining transparency in an algorithmic system "not just as code and data but as an *assemblage* of human and non-human actors." In doing so, we must reconstruct both transparency and accountability to effect across a system, rather than inside a particular context.[17]

Participants discussed this issue in the context of hiring. Algorithms for employment have been trained to highlight candidates who may meet certain qualifications, privilege those that have had access to elite universities and work experiences, and are typically opaque. Applying transparency to algorithms used to filter candidates can ensure the hiring pool is not a result of bias towards or against certain communities. For example, in New York City, a new law is being deliberated that would ensure that when a vendor sells an employment or hiring tool it shares a "bias audit" and that people who have been screened using the tool are

[15] *See, e.g.,* Richard Whitt, *Democratizing AI (Part 3): Action plans for creating Personal AIs* (July 1, 2019), at https://whitt.medium.com/democratizing-ai-part-3-efb8da5d956a.

[16] Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable Algorithms. *University of Pennsylvania Law Review*, 165(3), 633-705. https://scholarship.law.upenn.edu/penn_law_review/vol165/iss3/3

[17] Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989. https://doi.org/10.1177/1461444816676645

notified as to what characteristics were used.[18] This level of transparency could be groundbreaking, not only to expose the bias in hiring models, but to ensure models are improved to avoid such bias in the first place.

Current landscape of efforts to counter bias

As outlined above, efforts to counter algorithmic bias can come from the implementers, regulators, policymakers, and civil society. Strategies include legislation, audits, ethical frameworks, and grassroots activism and lobbying. For example, there are now multiple toolkits, such IBM's AI Fairness 360[19] (AIF360-open source Python toolkit for algorithmic fairness) or Amazon's SageMaker Clarify,[20] and/or Microsoft's InterpretML[21] that help facilitate evaluations of algorithms, identify bias in datasets, and explain predictions. We also see growing interest in additional efforts on explainable AI, AI system factsheets,[22] datasheets for datasets,[23] bias impact statements,[24] and many others. Additionally, power and progress in the data context must include a robust approach to data literacy, which is vital to consumers and citizen advocates. As we learned, this lack of knowledge may hinder a group's ability to acknowledge and challenge algorithmic decision-making outcomes. This section will expand on current interventions from the spaces of public policy and algorithmic auditing.

Public policy and governance

Policymakers and regulators serve a critical function in countering bias in algorithmic systems to make them more fair, accountable, and transparent. For example, the General Data Protection Regulation (GDPR) poses several actions that require a better formulation for explainability, transparency, and accountability of automated decision systems. In the U.S.,

---

[18] Simonte, T. (2021). New York City Proposes Regulating Algorithms Used in Hiring. Wired.com https://arstechnica.com/tech-policy/2021/01/new-york-city-proposes-regulating-algorithms-used-in-hiring/?comments=1

[19] Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Majsilovic, A. Nagar, S., Ramamurthy, K. N., Richards, J., Saha, D., Sattigeri, P., Singh, M., Varshney, K. R., & Zhang, Y. (2018). AI Fairness 360: An Extensible Toolkit for Detecting, Understanding, and Mitigating Unwanted Algorithmic Bias. IBM Research. https://arxiv.org/pdf/1810.01943.pdf

[20] Amazon SageMaker Clarify. (n.d.). Amazon Web Services. https://aws.amazon.com/sagemaker/clarify/

[21] Microsoft. InterpretML. https://interpret.ml/#why-interpret

[22] Richards, J., Piorkowski, D., Hind, M., Houde, S., & Mojsilović, A. (2020, June). *A Methodology for Creating AI FactSheets*. arXiv. https://arxiv.org/abs/2006.13796

[23] Gebru, T., Morgenstern, J., Vecchione, B., Vaughn, J. W., Wallach, H., Daumé III, H., & Crawford, K. (2018). *Datasheets for Datasets*. Fairness, Accountability, and Transparency in Machine Learning. https://www.fatml.org/media/documents/datasheets_for_datasets.pdf

[24] Turner, N. L., Resnick, P., & Barton, G. (n.d.). *Algorithmic bias detection and mitigation: best practices and policies to reduce consumer harms*. Brookings Institute. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/

several state-level and municipal-level regulators have placed moratoriums on the use of facial recognition software in response to rising civil liberty concerns. In addition to establishing ethical principles, legal scholars are calling upon companies to follow an approach grounded in the Universal Declaration of Human Rights (UDHR). Human rights law provides a clear, legally binding strategy against discriminatory impacts, which may include algorithms.[25]

Alternative legal concepts, such as information fiduciaries and data trusts, are also taking shape. Positioned by some as "anticipatory regulation," this approach is considered to be more nimble, inclusive, collaborative, and future-facing—all characteristics missing in today's regulatory environment.[26] These approaches also rest on giving users the trustworthy support that currently is lacking in the Web environment. In general, these legal concepts call for new data governance models—whether in the form of a collective data trust, personal data stores, or individualized digital fiduciaries— each of which can designate how one's data should be collected, stored, processed, and handled. Regardless of which model, the goal is to rebalance power and control between the individual end user and those (e.g. companies) that collect, manage, and use our data.[27]

The realization of any of these emergent strategies requires both capacity and resources for additional research and piloting.[28] One suggestion is to create an alliance to develop a Fiduciary Tech or "FidTech" sector. FidTech would consist of companies and organizations that would be necessary to "offer protection and mediating support based on fiduciary principles, in balance to the increasing power of institutional AI."[29]

Algorithmic auditing

An algorithmic audit is another strategy to expose systematic biases embedded in software platforms. These audits serve as a bridge between the concerns raised in ethics discussions and tangible, actionable solutions. Christian Sandvig et. al, point to the use of "audit studies," which are commonly used in social science research, to ascertain whether algorithms result

---

[25] Raso, F. A., Hilligoss, H., Krishnamurthy, V., Bavitz, C., & Kim, L. (2018, September 26). *Artificial Intelligence & Human Rights: Opportunities & Risks*. Berkman Klein Center Research Publication 2018-6, https://ssrn.com/abstract=3259344

[26] *Data Trusts: A new tool for data governance*. (n.d.). ElementAI. https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

[27] Whitt, R., *Old School Goes Online: Exploring Fiduciary Obligations of Loyalty and Care in the Digital Platforms Era*, 36 Santa Clara High Tech. L.J. 75 (2020), at https://digitalcommons.law.scu.edu/chtlj/vol36/iss1/3.

[28] Whitt, R., *Hacking the SEAMs: Elevating Digital Autonomy and Agency for Humans*, 19 Colo. Tech. Law Journal 720 (January 29, 2021), at https://ctlj.colorado.edu/?p=720.

[29] *Data Trusts: A new tool for data governance*. (n.d.). ElementAI. https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf

in harmful discrimination.[30] A number of algorithm audits have been conducted successfully over the years, including detecting unfairness in online advertising and online search and price discrimination on e-commerce websites.[31] It should be noted that while audits can be powerful instruments, it has its limitations. As Mona Sloane notes the question: "'Does the algorithm do what we say it does?' That question strategically precludes a focus on broader issues of representation and bias across the whole algorithm life cycle, including bias in the dataset, representation in the design team, the context in which the algorithmic tool gets deployed, and the maintenance of the tool."[32]

In their 2019 paper on publicly disclosing biased performance results of commercial AI products, Raji and Buolamwini present an audit design and disclosure procedure that engages companies to reevaluate and make model improvements to address classification biases in their systems.[33] This approach, informed by the information security practice of "coordinated vulnerability disclosures" (CVD) and bug bounties, requires a neutral third-party to conduct the audit. The critical lever, however, is the public disclosure of these performance vulnerabilities which applies external pressure necessary to motivate targeted companies to change.

Recently, Google and the Partnership for AI created and published a framework for auditing AI systems targeted at engineering teams, named *Scoping, Mapping, Artifact Collection, Testing, and Reflection* (SMACTR). Unlike previous audit processes, this model provides an end-to-end framework to be applied throughout the internal organization development life-cycle. The SMACTR audit is informed by other fields "where safety is critical to protect human life, such as aerospace and health care, which now carry out audits as part of the design process."[34] Another example is the Algorithmic Bias Playbook created by the Center for Applied AI at the Booth School of Business at University of Chicago, which outlines key steps and specific examples for execution. Finally, it should be noted that while these toolkits for bias are useful in detecting cases of an algorithm's poor performance, not one can automatically evaluate or check whether a human choice in building the system contributes to its bias.

---

[30] Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). An Algorithm Audit. In Seeta Peña Gangadharan (Eds.), *Data and Discrimination: Collected Essays* (pp. 6-10). New America Foundation.
[31] Auditing Algorithms. (n.d.). Algorithm Audits by Researchers. https://auditingalgorithms.science/?p=153
[32] Sloane, M. (2021). The Algorithic Auditing Trap. OneZero.com. https://onezero.medium.com/the-algorithmic-auditing-trap-9a6f2d4d461d
[33] Raji, I., & Buolamwini, J. (2019). *Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products*. Conference on Artificial Intelligence, Ethics, and Society, Honolulu, HI, United States. https://dam-prod.media.mit.edu/x/2019/01/24/AIES-19_paper_223.pdf
[34] Johnson, K. (2020, January 30). *Google researches release audit framework to close AI accountability gap*. VentureBeat. https://venturebeat.com/2020/01/30/google-researchers-release-audit-framework-to-close-ai-accountability-gap/

What comes next?

As this initiative is in service to the needs of the most vulnerable, our problem set is the pursuit of digital equity in algorithmic decision-making systems. To bring clarity to what this means, we submit for consideration a digital bill of rights in support of impacted communities that organizations should recognize and act on in their commitment to data stewardship for good.

Digital Bill of Rights

In Support of Impacted Communities

The Right To

1. Clear, Transparent, and Accessible Information.
2. Offerings Designed to Be Inclusive and Avoid Inappropriate Bias.
3. Not Be Subjected to Automated Algorithmic Decision-Making When It Relates to Life-Changing Effects on Consumers' Financial Status, Employment, Health and/or Education.
4. Easily Correct Inaccurate and/or Incomplete Information Used by Automated Decision-Making Systems when Creating User Profiles.
5. Privacy, with Minimal Data Collection Limited Only to Information Necessary to Provide Goods or Services Sought.
6. Know When and How Personal Data is Being Gathered and Used.
7. Influence Algorithmic Impact Assessments and Audits.

From principles to implementation

In defining the ends, we can then begin to unpack the means by which we can operate. The challenge is ensuring practices and solutions move downstream and are implemented throughout an entire ecosystem. This is where the Data Stewardship for Good Initiative seeks to inform and impact industry behavior. We also acknowledge much of the resources already available for those interested in examining algorithmic bias in their own organizations. For example, one starting point is to explore the following steps as outlined by the Center for Applied AI at the University of Chicago Booth School of Business:

- **STEP 1: INVENTORY LIVE ALGORITHMS:** Output a comprehensive list of all the algorithms used by your organization.
- **STEP 2: SCREEN FOR BIAS:** Clearly articulate what each algorithm is doing, the problem it is intended to address, and a diagnostic chart that illustrates bias in the context of your organization.
- **STEP 3: RETRAIN BIASED ALGORITHMS**: Analyze feasibility and efficacy of possible fixes or need to suspend use of biased algorithms.
- **STEP 4: PREVENT FUTURE BIAS:** Set up internal structures to implement best practices for preventing algorithmic bias.[35]

From our own exploration, we have identified two fundamental approaches an organization can take to mitigate unwanted bias in automated systems. First, is to address the challenges posed by already deployed systems. A sample intervention for this includes algorithmic or systems auditing. The second approach poses the question of how we may design new systems that begin with equity at the center. Here the entire algorithmic ecosystem must begin with the question: who is defining the rules?

Our framework identifies two categories of actors within the algorithmic decision-making ecosystem: *System Originators* and *System Deployers.* System Originators are defined as the original designers, creators, and developers of an algorithmic decision-making system (e.g. Google AdSense or HireVue). System Deployers are defined as entities who purchase and deploy algorithmic decision-making systems as part of their business offering (e.g schools using automated grading systems). We recognize that these categories are fluid, and in some cases, will include impacted communities. However, we argue this distinction is necessary in support of actions towards accountability and transparency. To address the concerns articulated above, the next phase of the Initiative will be to workshop a set of principles for algorithmic originators and deployers to adopt and a framework for next steps. We anticipate

[35] Center for Applied AI. (June 2021)."Algorithmic Bias Playbook." UChicago Booth School of Business. https://www.chicagobooth.edu/research/center-for-applied-artificial-intelligence/research/algorithmic-bias

this phase to be a critical juncture in developing pragmatic solutions that serve both industry and vulnerable communities. Thus, we invite others to join us in this pursuit.

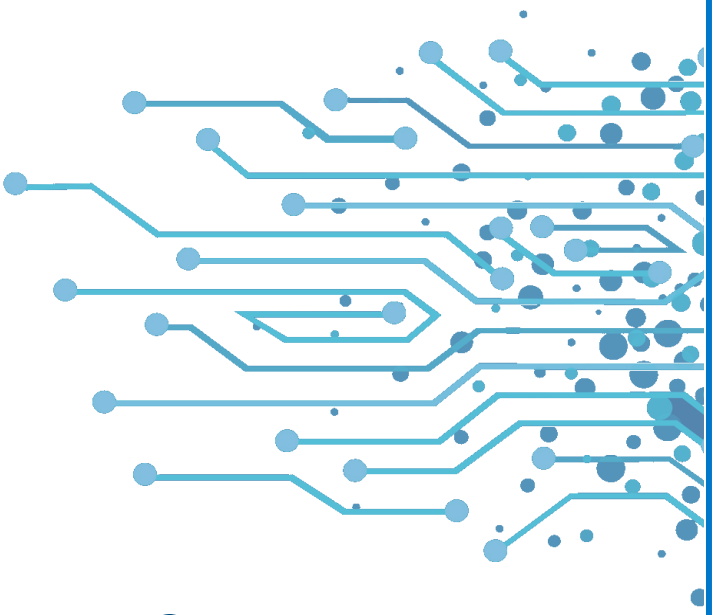
Conclusion

Bias in algorithms and AI systems is a complex issue. The solution is not as simple as adding more diverse data sets. As illustrated in the previous sections, bias can and will be introduced at every step of a system's design and development. And as a result, algorithms can amplify existing systemic discrimination. This critique is powerful in illuminating the deep entrenchment of human- and societal- bias that exists within our systems (technical or not). Simultaneously, we recognize the difficulties in addressing what seems to be an intractable problem with the right amount of recognition, resources, and political will.

But progress is being made. And, as a silver lining, we now have the language necessary to name and identify potential harms and the call to address it. It is a clear directive to acknowledge and mitigate the negative consequences of algorithmic decision-making—from incomplete and unrepresentative data; to a lack of diversity among decision makers, developers, and funders of AI tools; to a need for more inclusive design; and, for a re-evaluation of current legal and regulatory mechanisms.

Pointing back to Naraynan's original provocation at the beginning of this piece, the problem is not purely about mathematical (or statistical) correctness. Instead, the challenge is how to make algorithmic systems support human values. As Christian suggests, perhaps it's an alignment problem.[36] This initiative and its outputs are our contribution towards that better alignment.

[36] Christian, B. (2020). *The Alignment Problem: Machine Learning and Human Values*. W.W. Norton & Company, Inc.



DATA STEWARDSHIP FOR GOOD

Individuals are encouraged to cite this report and its contents. In doing so, please include the following attribution:

Kristine Gloria. "Power & Progress in Algorithmic Bias."
Washington, D.C.
Aspen Digital, a program of the Aspen Institute. July 2021.

ASPEN DIGITAL
aspen institute

Center for Inclusive Growth