Toward Situated Interventions for Algorithmic Equity: Lessons from the Field

Michael Katell*
Information School
University of Washington
mkatell@uw.edu

Meg Young*
Information School
University of Washington
megyoung@uw.edu

Dharma Dailey
Human Centered Design and
Engineering
University of Washington
ddailey@uw.edu

Bernease Herman
eScience Institute
University of Washington
bernease@uw.edu

Vivian Guetler
Department of Sociology
West Virginia University
vfg0002@mix.wvu.edu

Aaron Tam
Evans School of Public Policy and
Governance
University of Washington
tama2@uw.edu

Corinne Binz
Department of Computer Science
Middlebury College
cbintz@middlebury.edu

Daniella Raz
School of Information
University of Michigan
drraz@umich.edu

P. M. Krafft[†]
Oxford Internet Institute
University of Oxford
p.krafft@oii.ox.ac.uk

ABSTRACT

Research to date aimed at the fairness, accountability, and transparency of algorithmic systems has largely focused on topics such as identifying failures of current systems and on technical interventions intended to reduce bias in computational processes. Researchers have given less attention to methods that account for the social and political contexts of specific, situated technical systems at their points of use. Co-developing algorithmic accountability interventions in communities supports outcomes that are more likely to address problems in their situated context and re-center power with those most disparately affected by the harms of algorithmic systems. In this paper we report on our experiences using participatory and co-design methods for algorithmic accountability in a project called the Algorithmic Equity Toolkit. The main insights we gleaned from our experiences were: (i) many meaningful interventions toward equitable algorithmic systems are non-technical; (ii) community organizations derive the most value from localized materials as opposed to what is "scalable" beyond a particular policy context; (iii) framing harms around algorithmic bias suggests that more accurate data is the solution, at the risk of missing deeper questions about whether some technologies should be used at all. More broadly, we found that community-based methods are important inroads to addressing algorithmic harms in their situated contexts.

*Authors contributed equally.

[†]Work conducted at the University of Washington.

1 INTRODUCTION

Existing scholarship on algorithmic fairness often begins with a recognition of real-world harms [7, 11, 49], yet many contributions to date offer technical solutions that precede application or evaluation of systems in use. A situated investigation is one that forefronts the details of the context of specific people and places at particular points in time, rather than trying to study a system or question with an abstract approach removed from social context. We invoke this use of the term *situated* from feminist philosopher of science Donna Haraway [27], who describes situated knowledge as the product of embodied vantage points, located historically (see also [28, 29]). Situated investigations offer opportunities to reveal not only the technical features and limitations of a system, but also the different ways a system affects different people in its context of use. Many recent studies have pointed to the need to incorporate sociotechnical context in design of algorithmic systems. Selbst et al. call for scholars in this area to mind the traps associated with abstract analysis [58]. Green & Chen [25] demonstrate that human decision-making interacts with algorithmic processing in the ultimate disparate impacts of algorithmic systems. Other scholars have pursued empirical case studies [19, 50, 52], field evaluations [10, 65], context-aware uses of datasets [46, 63], and questions of institutional access to real-world data [67]. In each case, attending to end-use provides a richer design space and new opportunities for improvements and safeguards. In our work we move towards a deeper engagement with systems' social and political context. We report on a process of using a participatory design and action

research approach to craft interventions supporting the agendas of people most affected by the disparate impacts of algorithmic systems.

We look to the field of human-computer interaction (HCI) to deepen our commitment to engaging communities. As Paul Dourish [16] notes, HCI joins distinct scholarly traditions that are often viewed as at odds with each other: a positivist engineering approach, which seeks to achieve precisely defined objectives through system design; and a phenomenological, interpretivist approach [1, 61], which understands the meaning of both social action and technical systems as constructed, contingent, and constantly in flux—in defiance of the rigid operationalization of an instrumentalist view. At the intersection of these two traditions, HCI has fostered a wide array of participatory design methods. Here, we report on our process drawing on participatory methods and a distinct tradition (sharing similar goals) known as participatory action research in support of algorithmic accountability and community decision-making.

Participatory methodologies are valuable complements to the practice of data science more broadly. The work of data science has a normative dimension, but there is little agreement about which norms or values should be prioritized [5]. Research methods that center the people disparately affected by the practice of data science can guide design choices and clarify the implications of these choices by situating the work where the social costs and benefits can be more directly assessed [6, 13, 18]. Researchers in critical data studies have explored similar themes in the context of mixed methods data science [44].

Our work joins others in understanding algorithmic discrimination as a product of societal inequity rather than as solely a result of inaccurate performance by models or under-representative training data. When systems are deployed in contexts historically shaped by discrimination in policy and practice, such as policing [33, 47] or medical care [50], data employed as inputs for decision systems contributes to feedback loops and amplified discrimination [33, 47, 51]. As a result, it is not enough to attend to bias in systems from the perspective of accuracy; interventions must acknowledge the historical, political, and institutional contexts in which systems are used, and by whom.

Our Contributions: The imperative to account for situated context in algorithmic accountability requires a move beyond the laboratory and academy. In this work we ask, *What are the opportunities and challenges in using participatory design and action methods for research on fairness in algorithmic systems?* We approach this question by drawing on insights from our use of these methods in a project aimed at empowering community organizers and community members in advocacy and public comment efforts in state and local technology policy. In work described in more detail in a companion paper [36], we engaged in a co-design project that included local community groups, advocacy campaigns, and policy stakeholders, resulting in the Algorithmic Equity Toolkit [36], a set of heuristic tools and an interactive demo that assists users in recognizing algorithmic systems, understanding potential algorithmic harms, and holding policymakers accountable to public input.

Our central contributions here are lessons learned from the development of the Algorithmic Equity Toolkit for the practice of situated investigations into fairness, accountability, and transparency

in algorithmic systems. Our interactions with community partners re-framed our objectives and highlighted the opportunities and challenges of clearly communicating concepts from artificial intelligence and scholarship on algorithmic bias. We also note existing frictions to further incorporation of community input in settings where data science and machine learning research conventionally happens. The algorithmic fairness community has a unique opportunity to foster the intersection of data science, machine learning, and community-based work. This shift will require us to learn from other fields as to how to build reciprocal partnerships with community organizations, elicit their input, and attend to the practical dimension of managing trust and power dynamics in these relationships. A more inclusive data science practice will result in novel conventions of work which attend to fairness, accountability, transparency, and equity as an explicit part of research *method and practice*– rather than as topic alone.

2 RELATED WORK

Research in human-computer interaction has drawn on and developed methods for participatory design for decades. Participatory design is the process of enlisting intended user groups in the development of information technologies in hope of better supporting these users' interests and goals. Historically, it is explicitly concerned with the politics of system design with respect to worker power; early participatory design in Scandinavia aimed to support workers' self-determination of their local working conditions, as well as to affect policy [20]. More recent calls for an emancipatory HCI also affirm these value commitments [4, 15, 21, 39].

Participatory design approaches can intervene at individual, organizational, and political scales [2, 37]. All participatory methods recognize users as authorities over their own social context and hold that active cooperation with designers can make meaningful input or control possible. Participatory approaches to design fall across a spectrum from eliciting information from users to active collaboration with them in the design process; this latter approach is known as co-design. Co-design describes any collaborative design process undertaken by designers, end-users, and stakeholders [56]. Key strategies to make critiques and alternative visions for sociotechnical systems more accessible include the use of material prompts, such as photos and card decks that provide questions and themes for further reflection [23, 24, 64]. Other methods that aim to empower stakeholders as designers include the use of metaphors [42], design fiction [3], and futuristic scenarios [38].

Participatory action research (PAR) is a wholly distinct and long-standing tradition of community-embedded research that involves partnering with affected communities to achieve social change [40, 45]. The tradition was formalized in the late 1970s by an international network of researchers via a set of principles which specify, for example, that the approach requires a research "problem [to be] defined, analyzed and solved by the community" [54]. While its historical development defies easy attribution, PAR draws on a range of influences including work of Paulo Freire [22], feminist epistemology and praxis, and grassroots civil rights and workers' movements from the U.S., Latin America, and South Asia. It has been used in a broad a range of settings and fields, including research in health, education, and social work. The resulting tradition is

committed to advancing social change in active collaboration with affected communities, particularly by restructuring power relationships between researchers and participants to collectively reflect on intermediate findings and co-determine action in an "action-reflection" cycle [9]. Participatory action research operates within a epistemological paradigm that is constructivist, recognizing knowledge as partial and situated; critical, interrogating knowledge as an artifact of power relations; and participatory, as forming through an interaction between researchers and participants [26]. Its ultimate aim is to re-center power in communities outside academia and to support those communities in taking action.

Both participatory design and action research approaches have informed generations of work in HCI [31]. In machine learning and AI scholarship their value is becoming more apparent, as a growing contingent holds that AI has "ignored ... the voices, opinions, and needs of disempowered stakeholders ... in favor of [those] with power, money, and influence." [1] Academics interested in addressing such concerns have turned to activist and advocacy groups to learn from them and support their work. For example, Seeta Peña Gangadharan and Virginia Eubanks' Our Data Bodies project is centered on empowering communities to address digital data collection and human rights through a series of workshops and activities. A key strength of the project is close collaboration with activists Tamika Lewis, Tawana Petty, Kim M. Reynolds, and Mariella Saba, who are embedded within networks of grassroots organizations in cities across the US. A number of labs, research institutes, and meetings have also committed their missions to community engagement (c.f. [8, 12, 14, 53]). Often such efforts emphasize the importance of equitable *process* as determinative of equitable outcomes. As the principles of the media activism organization Allied Media Projects state, "The strongest solutions happen through the process, not in a moment at the end of the process." [2] As part of outlining this new paradigm, a notable recent call for greater partnership with and service to communities comes from the Design Justice Network [13] initially formed in 2016 by community organizers, which includes among its principles to "center the voices of those who are directly impacted by the outcomes of the design process." [3] In our work, we report on lessons learned from our own efforts to draw on participatory and co-design approaches.

3 BACKGROUND

Our work takes place in the context of Seattle, Washington. In September 2017, the Seattle City Council passed a significant revision of the Seattle Surveillance Ordinance, producing landmark municipal legislation that mandated the disclosure of surveillance technologies in use by city departments and subjecting those technologies to a political process that includes community input and oversight. As indicated in the declarations that introduce it, the Ordinance was explicitly motivated to "incorporate racial equity principles...to ensure that surveillance technologies do not perpetuate institutionalized racism or race-based disparities" (Seattle Surveillance Ordinance 118930, 2017). The law responds in part to a troubled

---

[1] Ali Alkhatib, Blog. "Anthropological/Artificial Intelligence & the HAI" 26 March 2019. <https://ali-alkhatib.com/blog/anthropological-intelligence> (Accessed August 22, 2019).
[2] https://alliedmedia.org/about/network-principles
[3] http://designjusticenetwork.org/network-principles

history of discriminatory policing in Seattle [17] as well as to a consensus among local civil rights activists and key policymakers that surveillance systems disparately harm people of color and other marginalized communities. [65].

The Seattle Surveillance Ordinance joins a growing cadre of similar ordinances enacted in response to concerns about the expanding use of surveillance technologies by government agencies. Civil rights activists have pressed for these ordinances with the intention of making public-sector technologies subject to greater transparency and to political oversight. While the laws differ, many share common attributes. They require the disclosure of systems that meet the definition of a "surveillance technology," the production of reports about their capabilities and intended uses, and authorization for such uses by lawmakers. Importantly, many of these laws provide opportunities for public engagement and comment prior to lawmaker determinations on the acquisition and use of a particular technology.

In a case study of Seattle's surveillance ordinance, three of this paper's authors found that although the ordinance was aimed at assessing surveillance systems' disparate impact, its efficacy was potentially limited by a lack of explicit language directing public officials to consider the algorithmic bias dimension of each technology [65]. A key finding of this prior work was that the analytic capabilities inside systems, such as automated license plate readers, were not recognized by interviewed city employees as relying on machine learning. This finding reflects a larger definitional disconnect affecting AI policy oversight efforts between researchers and policymakers in what technologies are considered to be AI, or are potential sources of algorithmic bias [41]. As a result, the potential for algorithmic harm from everyday technologies may be overlooked, at least in existing surveillance regulations. The incipient effort leading to the Algorithmic Equity Toolkit was inspired by a desire to assist lawmakers in closing this definitional disconnect and to aid the policy implementation process.

4 METHODS

4.1 Participatory Action Research

Our work on the Algorithmic Equity Toolkit relied on participatory design methods and a participatory action research (PAR) approach. In the tradition of PAR, our work involved partnering with organizations engaged in civil rights advocacy, eliciting those partners' priorities, creating prototypes, seeking input, and reflecting on what we learned. Our main partner organization was the American Civil Liberties Union of Washington (ACLU-Washington). We also worked closely with two member organizations of a coalition assembled by the ACLU-Washington, the Tech Fairness Coalition, which advocates for fairness and equity in technology policy. The first coalition organization was Densho, an organization dedicated to preserving the history of World War II incarceration of Japanese Americans and advocating to prevent state violence in the future. The second organization was the Council on American-Islamic Relations of Washington (CAIR-Washington), a prominent Islamic civil liberties group who defends the rights of American Muslims. After a four-month planning process with ACLU-Washington, our co-design work kicked off in June 2019 and involved meeting several times with all three organizations. The project conducted Diverse

Voices panels [43] in August 2019, presented to the broader Tech Fairness Coalition in September, and is planned to be shared for member organizations' use in March 2020.

We also note that ACLU-Washington encouraged early and continuous engagement with community organizations, given underrepresentation of historically marginalized communities in surveillance policy decision-making. This input was a key takeaway of an early scoping discussion about the Toolkit in June, 2019. Through prior engagements with ACLU-Washington, we learned that their work was motivated by a desire to increase public input on surveillance and automated decision-making system (ADS) technologies from those who are typically not consulted: "[We] recognize that some communities have a greater stake in this than others...[they're] really not talking about the safety of everyone. We're talking about some communities being posed as threats to safety and not talking about the ones we're trying to keep safe" (interview with Shankar Narayan, August 9th, 2018). Drawing on our partner's input and larger motivation, we revisited and expanded the co-design dimension of the Toolkit to center the perspectives of communities with the most at stake in decisions about surveillance and ADS technologies.

4.2 Participatory Design Methods

The techniques we adopted from participatory design [57] included informal elicitation sessions, as well as design methods more formally applied. We drew on contextual inquiry [34] to learn more about partners' work and needs, asking, "How do you currently advocate in the area of algorithmic decision systems?" and, "What would support your work?" We also used iterative heuristic evaluation [48] to identify specific ways to improve draft prototypes, asking, "Given the goals that we identified in the last meeting, how could these drafts better meet those goals?" "What should this do differently?" We documented each session and revised the Toolkit in turn. After an open-ended needs assessment, co-development proceeded more quickly once we began sharing draft prototypes as part of eliciting feedback [55].

We also brought in additional civil rights organizations and advocates for three panel sessions using the Diverse Voices method [43] to get input from groups focused on particular lived experiences. Diverse Voices panels are a means of enlisting experiential experts to identify potential sources of disparate impact in the way technology policy and other artifacts are designed for a specific stakeholder group. The purpose of the panels is not to speak on behalf of a particular group nor in a representative way, rather—like heuristic evaluation—to surface specific problems with the draft or artifact-in-development that can be changed to respond to matters of concern for this group. In each 90 minute panel session, we used 10 minutes to ask for introductions, provide context, and describe our co-design process to date. We planned 20 minutes for open discussion of the topic of automated decision systems as they relate to their impact on the stakeholder group for which the panel had been convened. In the remaining 60 minutes we asked for edits and critiques of the Toolkit, asking questions like, "How would these tools be more useful to you in your work?" "What do these tools not do that you wish they did?"

The experiential experts convened for our three panels addressed the lived experiences of three vantage points: immigrants, race and social justice activists, and formerly incarcerated people. The panel on immigrants featured representatives from the Northwest Immigrant Rights Project, a legal services nonprofit; Mijente, a grassroots organization for Latinx movement building; La Resistencia, an organization led by undocumented people to stop deportation and detention; and Community 2 Community Development, an organization dedicated to food sovereignty and immigrant rights. The panel on formerly incarcerated people featured (and took place at) Pioneer Human Services, a nonprofit organization that assists people reentering society from prison or jail as well as those who are overcoming substance use disorders and mental health issues; as well as unaffiliated active community members who were formerly incarcerated. The panel on race and social justice activists featured representatives from Black Lives Matter Seattle, which organizes for the empowerment and liberation of people of color; and the City of Seattle Office of Civil Rights, a municipal office dedicated to oversight toward equal access to housing, employment, public accommodations, and contracting. Panelists were paid honoraria for their time. This measure, along with deliberateness about holding panel sessions in convenient locations, was intended to make participation more accessible and reciprocal to those involved.

4.3 Institutional Context

Other major stakeholders in our work included data scientists and administrative staff at the University of Washington's eScience Institute. Much of the work leading to the present state of the Toolkit happened within the eScience Institute's 2019 Data Science for Social Good (DSSG) summer program and was carried out by DSSG fellows.[4] This program provided important resources and support to increase the project capacity and provided ready access to technical expertise on data science and machine learning. At the same time, the programmatic goals of a data science skills training program introduced within the DSSG context also presented at times competing expectations, which we explore in our lessons learned.

5 CASE STUDY

We present the Algorithmic Equity Toolkit project as a case study of the value of participatory design and action research toward greater fairness, accountability, and transparency in algorithmic systems. We embrace a wider design space inclusive of systems, laws, community organizations, and advocacy campaigns that affords a new perspective to achieve these aims. In our case, we support public efforts to push for greater community control over public-sector technology. Specifically, the Algorithmic Equity Toolkit assists public understanding of algorithmic systems for more empowered advocacy and decision-making.

5.1 Toolkit Description

The co-design project we report on here is an ongoing effort. We are aiming toward a final version of the Toolkit in early 2020. We provide a detailed description of the present state of the Toolkit in a companion paper documenting the choices we made in rendering

---

[4]https://escience.washington.edu/2019-data-science-for-social-good-projects/

concepts from algorithmic bias and fairness for policy advocacy, the substance of partners' input, and the details of the changes we made in response [36]. Here, we provide a brief summary of these details for the purposes of grounding our lessons learned from this work.

The intended users of the Algorithmic Equity Toolkit are community members, including civil rights advocacy, grassroots organizations, and members of the general public. Through planning conversations with the ACLU-Washington we set out to address power asymmetries between the public and government agencies in determining what technologies should be implemented in communities in Seattle and Washington. In order to integrate into ACLU-Washington and their Tech Fairness Coalition's efforts towards technology oversight laws, the Toolkit was aimed for use in organizing and outreach, participating in public comment sessions and meetings, and assessing the impact of particular technologies.

At the time of writing, the Toolkit had three components (illustrated in Figure 1), supporting a user to:

(1) *Determine whether a system is artificial intelligence (AI):*
A flowchart for assessing whether a given system relies on machine learning or automated decision-making capabilities.

(2) *Ask tough questions:*
A list of questions for advocates to surface the social context of a given system, its technical failure modes (i.e., potential for not working correctly, such as producing false positives), and its social failure modes (i.e. its potential for discriminatory harms when working correctly). This tool is intended for communities to demand answers from policymakers and vendors as to salient risks in algorithmic systems and to use those answers as an input to their policy positions.

(3) *Better understand how machine learning (ML) works and how it fails:*
An interactive demo of facial recognition that allows the user to make adjustments to the match threshold of a facial recognition system in order to demonstrate how false positive identifications are unevenly distributed with respect to race and gender [11].

Through conversations with community organizations, we iteratively co-designed and refined the Algorithmic Equity Toolkit. As a result of community input, we came to reflect on the shortcomings of a narrow focus on "bias," made materials more accessible to community organizers and community members, and provided additional context and examples alongside the tools to support users to reach their own assessments.

5.2 Project Action and Reflection

We reflected on this work as it unfolded using the norms of participatory design, in which the quality of work is evaluated by users. This process is akin to "member checking" in ethnographic investigations wherein members of the community are asked to assess an ethnographer's work [59]. In our case, we evaluated the Toolkit continuously in consultation with community partners by asking them if it would be of practical use to them in their existing advocacy efforts. In the areas where the prototype of a tool fell short, we responded by asking how it could be made more informative and (most importantly) useful to their own objectives.
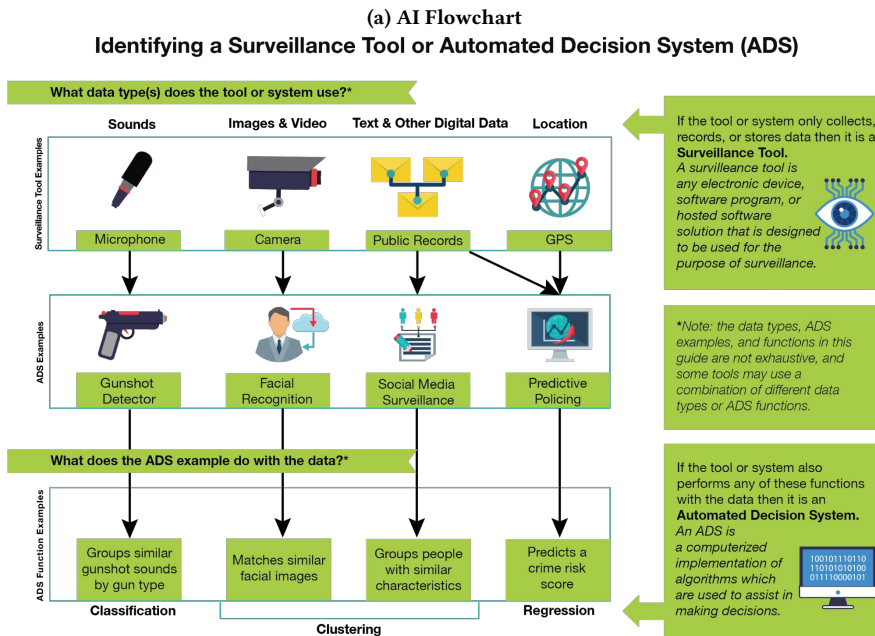
This form of evaluation is drawn from a tradition of participatory work that derives its validity through long-term, robust interactions and relationships; it values practical and experiential knowledge as authoritative [26]. Our other goals included (i) supporting community partners' stated priorities by reflecting learning through our collaboration and (ii) foregrounding partners' voices.

We worked with multiple partners who had different priorities and therefore different success criteria. For instance, one organization was interested in the Toolkit producing ways to amplify messages through social media to other grassroots organizations. Meanwhile, technical advisors to the project emphasized the importance of translating complex technical details about how machine learning algorithms work to a broader audience. To navigate these contrasting priorities, we worked to identify, clarify, and co-articulate shared priorities and forge common understanding. Ultimately we found a shared interest in empowering community organizations during public comment meetings and other coordinated campaigns for algorithmic accountability. A shared focus on improving draft Toolkit materials also gave a common focal point to integrate different perspectives. While we leveraged HCI evaluation techniques, such as user evaluations of prototypes to meet these aims, the metrics that flow directly from those evaluations (e.g. did a user correctly interpret x?) only partially capture what is instrumental to project success at this stage. Proof of success for our ongoing work is our partners' continued willingness and interest to collaborate.

5.2.1 Diverse Voices Panels. As another part of our evaluative strategy and co-design process, we presented the Toolkit to three panels using the Diverse Voices method [66]. During these panels, activists and members of historically marginalized communities evaluated the Toolkit and provided feedback about what would make it more useful and accessible. In addition to targeted feedback on draft prototypes, members of these panels asked for additional features, such as organizational charts specific to the local municipal government departments to support community implementation of the Toolkit in practice. As we discuss in our lessons learned, this type of evaluative feedback made an important fact salient — that the most meaningful measures towards fairness, accountability, and transparency are not necessarily technical but are instead informed through contextualized understanding about how a given tool or technology is implemented and contested in practice.

6 LESSONS LEARNED

Our work has provided us key lessons that may be useful to others adopting participatory methods for fairness, accountability, and transparency work. We arrived at these lessons by assessing the Algorithmic Equity Toolkit project to date against two goals; (i) supporting community advocates in algorithmic accountability campaigns by explaining the risks of automated decision-making systems (ADS) and making them more legible, and (ii) performing this work in a manner that centered the voices and interests of historically marginalized communities, particularly those who have been the primary targets of surveillance. We report both successes and challenges in service of these goals.

(a) AI Flowchart

Identifying a Surveillance Tool or Automated Decision System (ADS)

What data type(s) does the tool or system use?*

Surveillance Tool Examples

Sounds	Images & Video	Text & Other Digital Data	Location
Microphone	Camera	Public Records	GPS

If the tool or system only collects, records, or stores data then it is a **Surveillance Tool.**
A surveillance tool is any electronic device, software program, or hosted software solution that is designed to be used for the purpose of surveillance.

ADS Examples

Gunshot Detector	Facial Recognition	Social Media Surveillance	Predictive Policing

*Note: the data types, ADS examples, and functions in this guide are not exhaustive, and some tools may use a combination of different data types or ADS functions.

What does the ADS example do with the data?*

ADS Function Examples

Groups similar gunshot sounds by gun type	Matches similar facial images	Groups people with similar characteristics	Predicts a crime risk score
Classification			**Regression**
	Clustering		

If the tool or system also performs any of these functions with the data then it is an **Automated Decision System.**
An ADS is a computerized implementation of algorithms which are used to assist in making decisions.

(b) Question-Asking Tool

Transparency & accountability

The extent to which the algorithms (codes, data) used are available to community members/ data subjects.

Example Case: Residents of Idaho with developmental disabilities saw their Medicaid assistance slashed by thousands of dollars. The formula, data, and code that the Medicaid program was using to determine how much funding individuals were due, was not available to the public or interested parties. Thus, there was no explanation for why an individual's Medicaid spending allowance was cut. A judge ruled that the Medicaid program had to release this information, and upon further inspection a team of experts found serious flaws with the statistical methods and data used. These flaws ultimately resulted in individuals' medicaid allowances being incorrectly and baselessly cut.
https://www.aclu.org/blog/privacy-technology/pitfalls-artificial-intelligence-decisionmaking-highlighted-idaho-aclu-case

Does the tool or system provide documentation about its design and functions, such as:

- When was the training data collected?
- How was the data collected?
- Who was included in the sample population?
- If sampled, what was the sampling strategy?
- How was it analyzed?
- Was the data public or private?
- What were the variables used?
- How were the weights of the variables determined?
- Was the data tested for bias?
- Does the data have identifiable personal information?
- Does the data identify any subpopulation, such as by age, gender or race?
- What types of personal information are used?
- Will there be a system maintenance?
- If so, how will the system maintenance and monitoring be conducted?

(c) Interactive Demo

Current subject

- LeBron James
- Lisa Leslie
- Paris Hilton
- Jennifer Lopez
- Aaron Peirsol
- Jacqueline Edwards
- ● Kalpana Chawla
- Jason Campbell
- Katie Couric
- Vicki Zhao Wei

Faye Wong 0.306	Kelly Santos 0.342	Joy Bryant 0.372
Jason Campbell 0.476	Amy Brenneman 0.691	Kalpana Chawla 1.268

0.0 0.15 0.3 0.45 0.6 0.75 0.9 1.05 1.2 1.35 1.5

Threshold: You have selected a minimum similarity score to qualify for a match as "0"

Case Studies:

ICE Uses Facial Recognition To Sift State Driver's License Record: In July of 2019, researchers at Georgetown University Law Center found that Immigration and Customs Enforcement (ICE) agents mined millions of driver license photographs in search of facial recognition matches to target undocumented migrants who have legally obtained driver licenses. ICE did this illegally, as they did not have congressional approval to access DMV databases of driver license photos. In this scenario, the use of facial recognition technology clearly put undocumented migrants at risk, and increased accuracy of the facial recognition software would have heightened the danger towards undocumented migrants. Not only does facial recognition software work less well on people of color, even if it is accurate, it's used to target communities of color. Read the NPR news coverage of this case study here.

Figure 1: The three primary components of the Algorithmic Equity Toolkit as of an August 2019 version, consist of: (a) an identification guide for automated decision systems, (b) a questionnaire on potential harms, and (c) an interactive tool demonstrating algorithmic bias in a face recognition algorithm. (Reproduced from Katell et al. [36])

6.1 Participatory design requires a means of handling competing input

We first note the effectiveness of the participatory design approach toward our goals. Our work began with a relatively loosely defined project scope, which provided room for us to respond to the priorities of community partners who participated in the project. Our interactions with partners were overwhelmingly positive. Our key partners expressed enthusiasm about the project early on, which encouraged us and helped garner institutional support for the project.

At the same time, our partnering organizations sought features that sometimes contradicted other stakeholder input to make lightweight and ready-at-hand tools. For example, one organization asked for information about ongoing activist campaigns on ADS, and for updated developments to be included with the Toolkit. Another organization pointed out the importance of legal resources to be included as a tool:

> "There needs to be included in the Toolkit not just an understanding of how algorithms work, but what to do if you think your rights—or the rights of your family members—have been violated. Because I think frequently the time people come to us for information is when harm's already been done. So understanding the tool isn't helpful at that moment; they need to know what they can do to find justice for the person that was mistreated. So having a coordinated effort with civil rights lawyers or whomever who have a background in this kind of law that can give people information about how to get help." (Panel with immigrant experiential experts, August 15, 2019).

In some cases these ideas presented competing visions and goals, such as a trade-off between completeness and lightweight accessibility. The difficulty of managing the contributions of design participants is an expected result of adopting a participatory approach, but it also created uncertainty during the development process when competing input would be translated into shareable prototypes. As a result, our attempts to be responsive to input also impacted the development timeline by re-defining design specifications as they emerged.

6.2 Partners valued more local, contextualized materials for their work

Another tension at play in our work was between our desire to conduct situated research and the potential for a broader usefulness and constituency for the Toolkit in other jurisdictions. While we committed to delving deeply into the Seattle context, the frame of surveillance ordinances as operating in multiple municipalities shaped our impression in the background that the Toolkit would be useful in other cities. While some conversations with partner organizations did highlight the value in having a Toolkit that could be widely useful in many locales, other conversations emphasized that advocates would derive the most value from highly localized materials as opposed to what is "scalable" beyond a particular policy context.

Experiential experts pointed out that additional localization to Seattle would make it more useful in their advocacy campaigns.

In particular, the idea arose in multiple panels that the Toolkit should come with a list of technologies known to be in use, an organizational chart listing personnel within local government who procure and oversee these technologies, their phone numbers, and their email addresses.

> "To make this more concrete, I want to see a diagram of who are the key players? Who's making the decision, like who's determining, who's allocating the money for the departments? Because then that helps you a little more to target who these folks are ask questions in terms of racialized outcomes." (Panel with formerly incarcerated experiential experts, August 16, 2019).

> "Starting with Seattle, ...We are investing all this money, who is making that decision? If City Council wants to vote on every piece of software or hardware that's being brought in, how are they being educated on... its racial implications and impacts?" (Panel with race and social justice activist experiential experts, August 13, 2019).

> "Within this diagram, almost like a diagram of who are the power players? Who is making the decision? Who is allocating the money? Who are the departments? Who are the agencies? It helps you a little bit more to target who these folks are."(Panel with race and social justice activist experiential experts, August 13, 2019).

Here, these panelists point out that localization would streamline the process of identifying the correct point of contact for public questioning and campaigning. For the Toolkit to address this need would potentially require additional resources to keep a highly localized version of the Toolkit updated and easily editable. In responding to partners' and participants' suggested changes, we find increased responsiveness to community needs and increased likelihood of its adoption for those users would merit reducing its potential to scale.

6.3 Non-technical measures are powerful steps toward algorithmic accountability

We also find that many meaningful interventions toward equitable algorithmic systems are non-technical. Where many efforts consider the design and development of a specific algorithm or system, our work highlights the role of community organizing, public engagement, and policy oversight in addressing system failures. This lens opens challenging problems, such as AI safety and explainability, to non-technical approaches. For example, an approach that highlights known risks and failures could prepare the public (and government agencies) to better address them. For example, on another component of the Toolkit, a panelist suggested resources on the history of automated systems.

> "This would maybe be something that accompanies it— even thinking about what has been the history of similar technologies? Even recently, how have these technologies been used in the past to target and harm communities of color? Would be helpful. And I think part of that too, when we ask where does race come

in— it's not just in the increased police presence, it is in determining that gunshots are what we are going to spend our money on. Why aren't we targeting other crimes? So it's also a question of what's the data we are valuing, and that's telling us how these are racialized... at every level" (Panel with race and social justice activists, August 13, 2019).

Here, the panelist highlights how race permeates the history and use of public sector technologies, going beyond technical failures. These considerations further highlight how organizing work itself can be a crucial non-technical contribution to the fields of machine learning and data science, as organizations such as AI Now have pointed out [62]. Notably, a wave of recent scholarship promotes the ban of face recognition systems (e.g. Hartzog & Selinger, Stark [30, 60]); the increasing uptake of this policy position in U.S. municipalities (like Somerville, MA; San Francisco, Oakland, and Berkeley, CA; and Portland, OR) provides a strong example of the power of non-technical interventions.

Nor are these approaches limited to policy interventions in a strict sense. At its inception, our focus had been on how we could develop materials responsive to our previous findings on the definitional gaps we saw in city government [65], for instance by recommending specific changes to existing municipal government technology procurement and reporting processes. Whereas initially we had approached ACLU-Washington about the value we saw at the time of promoting the legibility of public-sector ADS to municipal government employees, our partners recommended the primary audience instead be community organizers pursuing ongoing work in this space. As a result of this meeting, we pivoted the project focus to supporting community advocacy.

At the same time, panelists point out that asking the right questions is resource-intensive for community groups to do alone:

> "These are very helpful questions. They are so helpful that I'm like, how can we use them even more? It's very useful for activist groups to be asking these questions and documenting the results of it. But— can we offset that work to a routine government work item? So there will be a report, and an auditing office cares a lot about their credibility, so we can wave that report and say 'Look at what this found.' It will have a strong impact. Community groups have ups and downs, they are not always around. People are busy and it's harder. So that's why I'm thinking from that angle" (Panel with race and social justice activists, August 13, 2019).

Here, the panelist conceived of asking the right questions for algorithmic accountability as a form of labor, which should ultimately be included as part of conventional government reporting or auditing functions.

6.4 Making systems more "fair" at the expense of making them more just

A key lesson learned came out of conversations about one component of the Toolkit; the interactive demonstration of facial recognition and its disproportionate false positive rate for people of color. This component illustrated a single dimension of algorithmic harm— erroneous results. Conversations with our partner organizations

and participants emphasized that while such demonstrations are important, they risk promoting "single axis thinking" [32] about algorithmic systems in which a focus on technical errors in a problematic technology diverts attention from the social systems that produce both the technology's inequitable effects and the narratives that justify its use. Surveillance systems that work as intended still produce undesirable effects and reproduce patterns of discrimination [35]. This point is particularly salient for technologies in the context of law enforcement; as Moy explains:

> Police technologies are not adopted in a vacuum, or by brand new police agencies with no history of racial inequity. Rather, police technologies are adopted by existing agencies colored by the same inequities that permeate American society. [47]

Given these issues, we were attentive to feedback we received from our colleagues and community members alike who worried that illustrating technical failures and inaccuracies in algorithmic systems has the potential to distract from the fundamental problems with the deployment of sophisticated surveillance technologies by powerful actors such as government officials. We recognize that we must take care not to present algorithmic equity as a problem that can be solved with more accurate surveillance systems.

6.5 Machine learning skills gaps did not present a barrier to receiving meaningful feedback

As succinctly stated by the common "black box" metaphor for algorithmic systems, it is difficult (even for specialists) to understand the reasoning driving operations in machine learning and artificial intelligence. These domains defy easy translation to a non-specialist audience. However, a specialist machine learning background was not necessary for providing feedback on the Toolkit:

> "[Person 1:] "I'm not very clear on predictive policing and what that does. What I see here is that the output is a score; it's troubling to me because it furthers the idea of a good, bad— good or bad neighborhood, and that impacts people who are poor.
> [Person 2:] When trying to draw the border around the 'system' itself, there's also how the system feeds into itself— predictive policing leading to certain neighborhoods and more crimes are found there because more eyes are there. That seems like it's an important piece to capture in your tool for how to recognize these systems... the human element of the inputs to these systems belongs with the other inputs in this diagram" (Panel on race and social justice activist experiential experts, August 13, 2019).

This panelist went on to add that the Toolkit should also highlight what a non-automated alternative would be to the use of a system. For example, rather than the use of automated gunshot detectors in neighborhoods, community members hearing a gun would call 9-1-1.

A key component of the Toolkit is intended to help non-specialists identify algorithmic systems; creating this tool required decisions

about what aspects of these systems needed to be explained. For example, at the time of the Diverse Voices panels, the prototype of this tool explained regression, classification, and clustering methods in machine learning algorithms. Through community engagement, we learned that these distinctions were not as useful as material that would highlight the salient failure modes of these systems, and what community organizations could ask for as an alternative.

> "On the bottom it says classification, clustering, regression, right?... You follow this diagram and you're moving through it, and that is the thing you are left with....It's like, okay, then what am I to do with 'Classification?' ...I found myself wanting examples of how these different data and systems have been used to make decisions...So, going from the inputs to, 'This is the racialized outcome'" (Panel with race and social justice activist experiential experts, August 13, 2019).

The same panelist pointed out that the questions in the Toolkit should emphasize the users of the systems:

> "Is there something we can add about who gets to interpret the information and make decisions about it?...Even if this data was accurate for age, gender, and race, who is using it still determines how they use the information, right?"(Panel with race and social justice activist experiential experts, August 13, 2019).

Overall, these comments helped us recognize that the Toolkit should emphasize sociotechnical context of systems' adoption and use rather than a narrow focus on how the algorithms themselves work. We also found that technical background is not required for community participation in technical decisions in a policy setting.

6.6 Traditional data science practice benefits from qualitative and participatory methods

A substantial portion of our work on the Algorithmic Equity Toolkit took place within the Data Science for Social Good program, which brings together student fellows and domain experts for data intensive projects in the public interest. As a result, we also reflect on our project as an intervention into traditional data science practice. This DSSG program was hosted within the University of Washington's data science studio at the eScience Institute and was supported by hands-on data science mentors. During the year of our work, the DSSG program was expanding its vision of what constituted data science and hired a mentor to be an on-site Interpretive Data Scientist. This colleague became an important resource for our team, deepening the qualitative and reflective dimensions of this work and supporting us in enlarging the scope of who should be considered essential to the practice of data science.

This arrangement provided invaluable insights and essential resources, but at times it was attended by competing visions of what data science is or could be. Given that the field conventionally valorizes engineering techniques to a greater degree than understanding the social domain in which data arise and are used, our approach does not meet conventional data science learning outcomes. In the 10-week summer program that formed a core component of the development timeline, our team spent several weeks in conversations doing background research, understanding

partner needs, and revisiting the design specifications of the work. Only the demo component of the Toolkit required software development or machine learning model training—and while the demo is modest as an attempt at technological innovation (which is the aim of many computational approaches), it primarily responds to an unmet need for a particular audience. The time spent reflecting on how to provide technical communication for advocacy use on complex and hard-to-explain technologies competed with time that might have spent doing conventional data science. Nevertheless, we ultimately found that the deliberative process undergirding this project allowed for a deeper engagement, critique, and problematization of data practices that support the pedagogical objectives of the program.

7 CONCLUSION

Our work presents lessons learned from a case study of participatory design and action research for fairness, accountability, and transparency in algorithmic systems. To create the Algorithmic Equity Toolkit, we worked with community groups to produce materials promoting awareness of algorithmic systems and equip users with strategies to advocate for their communities in the face of surveillance and automated decision systems. In so doing, we navigated competing input from community partners as to what interventions would be most effective. The co-design process pointed toward materials responsive to local needs, increasing the likelihood the intervention would be used by project partners. Where technical literacy and approaches would at first appear to present barriers to meaningful engagement, we find that non-technical perspectives and approaches are uniquely valuable contributions to algorithmic accountability outcomes. Most importantly, the process of developing Toolkit artifacts engendered questions about the ultimate aim of efforts to make systems more fair. We argue for future interventions (and the practice of data science more broadly) to understand algorithmic systems in their situated context. A research process centered on the priorities, goals, and direct feedback of community partners and experiential experts moves toward fairness, reciprocity, and accountability in research method itself.

REFERENCES

[1] Philip E Agre. 1997. Lessons learned in trying to reform AI. *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide* (1997), 131.
[2] Mariam Asad, Christopher A Le Dantec, Becky Nielsen, and Kate Diedrick. 2017. Creating a sociotechnical API: Designing city-scale community engagement. In

Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. ACM, 2295–2306.

[3] Stephanie Ballard, Karen M Chappell, and Kristen Kennedy. 2019. Judgment call the game: Using value sensitive design and design fiction to surface ethical concerns related to technology. In *Proceedings of the 2019 on Designing Interactive Systems Conference.* ACM, 421–433.

[4] Shaowen Bardzell. 2010. Feminist HCI: Taking stock and outlining an agenda for design. In *Proceedings of the SIGCHI Conference on Human Factor in Computing Systems.* ACM, 1301–1310.

[5] Solon Barocas and Danah Boyd. 2017. Engaging the ethics of data science in practice. *Commun. ACM* 60, 11 (2017), 23–25.

[6] Ruha Benjamin. 2019. *Captivating Technology: Race, Carceral Technoscience, and Liberatory Imagination in Everyday Life.* Duke University Press.

[7] Ruha Benjamin. 2019. *Race After Technology: Abolitionist Tools for the New Jim Code.* John Wiley & Sons.

[8] Davide Beraldo and Stefania Milan. 2019. From data politics to the contentious politics of data. *Big Data & Society* 6, 2 (2019), 2053951719885967.

[9] O Fals Borda. 2006. Participatory (action) research in social theory: Origins and challenges. *Handbook of Action Research* (2006), 27–37.

[10] Dimitrios Bountouridis, Jaron Harambam, Mykola Makhortykh, Mónica Marrero, Nava Tintarev, and Claudia Hauff. 2019. SIREN: A Simulation framework for understanding the effects of recommender systems in online news environments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, 150–159.

[11] Joy Buolamwini and Timnit Gebru. 2018. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency.* 77–91.

[12] Sasha Costanza-Chock. 2018. Design Justice: Towards an intersectional feminist framework for design theory and practice. *Proceedings of the Design Research Society* (2018).

[13] Sasha Costanza-Chock. 2020. *Design Justice: Community-led Practices to Build the World We Need.* MIT Press.

[14] Lina Dencik, Arne Hintz, Joanna Redden, and Emiliano Treré. 2019. Exploring Data Justice: Conceptions, Applications and Directions.

[15] Carl DiSalvo, Jonathan Lukens, Thomas Lodato, Tom Jenkins, and Tanyoung Kim. 2014. Making public things: How HCI design can express matters of concern. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems.* ACM, 2397–2406.

[16] Paul Dourish. 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8, 1 (2004), 19–30.

[17] Jenny Durkan. 2012. EXHIBIT A (Settlement Agreement and Stipulated [Proposed] Order of Resolution, Office of the Mayor of Seattle) - 1 12-CV-1282. , 76 pages.

[18] Catherine D'Ignazio and Lauren Klein. 2020. Data Feminism.

[19] Severin Engelmann, Mo Chen, Felix Fischer, Ching-yu Kao, and Jens Grossklags. 2019. Clear sanctions, vague rewards: How China's social credit system currently defines good and bad behavior. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, 69–78.

[20] Christine Floyd, Wolf-Michael Mehl, Fanny-Michaela Resin, Gerhard Schmidt, and Gregor Wolf. 1989. Out of Scandinavia: Alternative approaches to software design and system development. *Human–Computer Interaction* 4, 4 (1989), 253–350.

[21] Sarah Fox, Mariam Asad, Katherine Lo, Jill P Dimond, Lynn S Dombrowski, and Shaowen Bardzell. 2016. Exploring social justice, design, and HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 3293–3300.

[22] Paulo Freire. 2018. *Pedagogy of the Oppressed.* Bloomsbury Publishing USA.

[23] Batya Friedman and David Hendry. 2012. The envisioning cards: A toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1145–1148.

[24] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural probes. *interactions* 6, 1 (1999), 21–29.

[25] Ben Green and Yiling Chen. 2019. Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, 90–99.

[26] Egon G Guba and Yvonna S Lincoln. 2005. Paradigmatic controversies, contradictions, and emerging confluences. (2005).

[27] Donna Haraway. 1988. Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies* 14, 3 (1988), 575–599.

[28] Sandra G Harding. 2004. *The Feminist Standpoint Theory Reader: Intellectual and Political Controversies.* Psychology Press.

[29] Nancy CM Hartsock. 2017. The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In *Karl Marx.* Routledge, 565–592.

[30] Woodrow Hartzog and Evan Selinger. 2018. Facial recognition is the perfect tool for oppression. *Medium* (2018).

[31] Gillian R. Hayes. 2011. The relationship of action research to human-computer interaction. *ACM Transactions on Computer-Human Interaction* 18, 3 (2011), 1–20.

[32] Anna Lauren Hoffmann. 2019. Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication & Society* 22, 7 (2019), 900–915.

[33] Elizabeth E Joh. 2014. Policing by numbers: Big data and the Fourth Amendment. *Wash. L. Rev.* 89 (2014), 35.

[34] Holtzblatt Karen and Jones Sandra. 2017. Contextual inquiry: A participatory technique for system design. In *Participatory Design.* CRC Press, 177–210.

[35] Michael Katell. 2018. Adverse detection: The promise and peril of body-worn cameras. In *SURVEILLANCE, PRIVACY AND PUBLIC SPACE.*, Bryce Clayton Newell, Tjerk Timan, and Bert-Jaap Koops (Eds.). Taylor & Francis, Abingdon, 99–118. OCLC: 1009072661.

[36] Michael Katell, Meg Young, Bernease Herman, Dharma Dailey, Aaron Tam, Vivian Guetler, Corinne Binz, Daniella Raz, and PM Krafft. 2019. An Algorithmic Equity Toolkit for Technology Audits by Community Advocates and Activists. *arXiv preprint arXiv:1912.02943* (2019).

[37] Finn Kensing and Jeanette Blomberg. 1998. Participatory design: Issues and concerns. *Computer Supported Cooperative Work (CSCW)* 7, 3-4 (1998), 167–185.

[38] Finn Kensing and Kim Halskov Madsen. 1992. *Generating visions: Future workshops and metaphorical design.* L. Erlbaum Associates Inc.

[39] Os Keyes, Josephine Hoy, and Margaret Drouhard. 2019. Human-computer insurrection: Notes on an anarchist HCI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM, 339.

[40] Sean A Kidd and Michael J Kral. 2005. Practicing participatory action research. *Journal of Counseling Psychology* 52, 2 (2005), 187.

[41] PM Krafft, Meg Young, Michael Katell, Karen Huang, and Ghislain Bugingo. 2020. Defining Artificial Intelligence in Policy versus Practice. *Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES).* (2020).

[42] Nick Logler, Daisy Yoo, and Batya Friedman. 2018. Metaphor cards: A how-to-guide for making and using a generative metaphorical design toolkit. In *Proceedings of the 2018 Designing Interactive Systems Conference.* ACM, 1373–1386.

[43] Lassana Magassa, Meg Young, and Batya Friedman. 2017. Diverse Voices: A how-to guide for creating more inclusive tech policy documents. *Tech Policy Lab* (2017).

[44] Momin M Malik, Katja Mayer, Hemank Lamba, and Claudia Müller-Birn. [n. d.]. Workshop on Critical Data Science. ([n. d.]).

[45] Robin McTaggart. 1991. Principles for participatory action research. *Adult Education Quarterly* 41, 3 (1991), 168–187.

[46] Margaret Mitchell, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, 220–229.

[47] Laura Moy. 2019. How police technology aggravates racial inequity: A taxonomy of problems and a path forward. *Available at SSRN 3340898* (2019).

[48] Jakob Nielsen. 1994. Heuristic evaluation. In *Usability Inspection Methods.* John Wiley & Sons, Inc., 25–62.

[49] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press.

[50] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366, 6464 (2019), 447–453.

[51] Frank Pasquale. 2015. *The Black Box Society.* Harvard University Press.

[52] Samir Passi and Solon Barocas. 2019. Problem formulation and fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT* '19.* ACM Press, Atlanta, GA, USA, 39–48.

[53] Seeta Peña Gangadharan and Jędrzej Niklas. 2019. Decentering technology in discourse on discrimination. *Information, Communication & Society* 22, 7 (2019), 882–899.

[54] Md Anisur Rahman. 2008. Some trends in the praxis of participatory action research. *The Sage Handbook of Action Research: Participative Inquiry and Practice* (2008), 49–62.

[55] Marc Rettig. 1994. Prototyping for tiny fingers. *Commun. ACM* 37, 4 (1994), 21–27.

[56] Elizabeth B-N Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *Co-design* 4, 1 (2008), 5–18.

[57] Douglas Schuler and Aki Namioka. 1993. *Participatory design: Principles and practices.* CRC Press.

[58] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, 59–68.

[59] Clay Spinuzzi. 2005. The methodology of participatory design. *Technical Communication* 52, 2 (2005), 163–174.

[60] Luke Stark. 2019. Facial recognition is the plutonium of AI. *XRDS: Crossroads, The ACM Magazine for Students* 25, 3 (2019), 50–55.

[61] Lucy A Suchman. 1987. *Plans and situated actions: The problem of human-machine communication.* Cambridge University Press.

[62] Meredith Whittaker, Kate Crawford, Roel Dobbe, Genevieve Fried, Elizabeth Kaziunas, Varoon Mathur, Sarah Mysers West, Rashida Richardson, Jason Schultz, and Oscar Schwartz. 2018. *AI Now Report 2018.* AI Now Institute at New York University.

[63] Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, HV Jagadish, and Gerome Miklau. 2018. A nutritional label for rankings. In *Proceedings of the 2018 International Conference on Management of Data.* ACM, 1773–1776.

[64] Daisy Yoo, Alina Huldtgren, Jill Palzkill Woelfer, David G Hendry, and Batya Friedman. 2013. A value sensitive action-reflection model: Evolving a co-design space with stakeholder and designer prompts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 419–428.

[65] Meg Young, Michael Katell, and P. M. Krafft. 2019. Municipal surveillance regulation and algorithmic accountability. *Big Data & Society* 6, 2 (2019).

[66] Meg Young, Lassana Magassa, and Batya Friedman. 2019. Toward inclusive tech policy design: A method for underrepresented voices to strengthen tech policy documents. *Ethics and Information Technology* 21, 2 (2019), 89–103.

[67] Meg Young, Luke Rodriguez, Emily Keller, Feiyang Sun, Boyang Sa, Jan Whittington, and Bill Howe. 2019. Beyond open vs. closed: Balancing individual privacy and public accountability in data sharing. In *Proceedings of the Conference on Fairness, Accountability, and Transparency.* ACM, 191–200.