A Toolkit for Algorithmic Equity and Community Empowerment

Michael Katell[1][0000-0003-2200-6246], Meg Young[1][0000-0002-9300-8575], Bernease Herman[1][0000-0002-5453-4994], Dharma Dailey[1], Corinne Binz[2][0000-0001-9487-2069], Vivian Guetler[3][0000-0001-5256-0104], Daniella Raz[4][0000-0001-6259-9715], Aaron Tam[1], and P. M. Krafft[5][0000-0002-9300-8575]

[1] University of Washington, Seattle, WA 98195, USA
[2] Middlebury College, Middlebury, VT 05753, USA
[3] University of West Virginia, Morgantown, WV 26505, USA
[4] University of Michigan, Ann Arbor, MI 48104, USA
[5] University of Oxford, OX1 2JD, Oxford, UK
mkatell@uw.edu

Abstract. A wave of recent scholarship documenting the discriminatory harms of algorithmic systems has spurred widespread interest in algorithmic accountability and regulation. Yet effective accountability and regulation is stymied by a persistent lack of resources supporting public understanding of algorithms and artificial intelligence. We present a toolkit for algorithmic legibility developed using participatory design methodologies. Through interactions with a US-based civil rights organization and their coalition of community organizations, we identify a need for (i) "street level" heuristics that aid stakeholders in distinguishing between types of analytic and information systems in lay language, and (ii) risk assessment tools for such systems that begin by making algorithms more legible. The present work delivers a toolkit to achieve these aims.

Keywords: Participatory Design, Surveillance, Regulation, Algorithmic Equity, Fairness, Accountability, Transparency.

1 Introduction

This Extensive evidence demonstrates that the harms of algorithmic and information technologies are significant. Demonstrated harms exist across highly varied applications. Automated pretrial and sentencing risk assessment systems used in courts of law are racially biased [1–3], facial recognition is racially and gender biased [4], algorithmically supported hiring decisions are gender biased [5], automated license plate readers lead to unwarranted police stops [6], sensitive financial information has been stolen in major privacy breaches [7], and much more.

Community organizations and civil rights groups, concerned about the discriminatory risks of public sector technology adoption, have pushed for algorithmic equity (accountability, transparency, fairness) through the implementation of municipal surveillance ordinances in several U.S. cities. These ordinances manage the acquisition and use of surveillance technology. Berkeley, Cambridge, Nashville, Seattle, and others have passed ordinances that differ in their scope, processes, and degree of oversight for regulating government technologies [8]. The City of Seattle passed a surveillance ordinance in 2017 that requires the publication of a “master list” of government surveillance technologies and requires “surveillance impact reports” (SIR) that includes input from both city personnel and designated community representatives [9, 10].

Yet existing legislation does not go far enough to address the risks at hand. Policymakers and street-level stakeholders alike find algorithmic systems to be inscrutable and illegible [11]. Risks that are already subject to existing legislation are not being recognized because the risks are algorithmic in nature. Little or no legislation has been attempted to regulate technologies at the algorithmic level. One result is that public officials may not be paying close attention to the algorithmic features of their technical systems. For example, recent ethnographic research in a major U.S. city found that city personnel tasked with implementing that city’s surveillance ordinance did not consider any of the surveillance technologies in their portfolio to be algorithmic systems, focusing instead on their data collection functions [12]. This finding suggests a “crisis of legibility” in algorithmic regulation. In this paper, we present the Algorithmic Equity Toolkit, a set of heuristic tools that responds to problems of “legibility” in public sector algorithmic systems. The Toolkit is intended to help community organizers and non-experts better identify surveillance and automated decision-making system (ADS) technologies.

2 Methods

We iteratively developed the Algorithmic Equity Toolkit through a participatory design process that engaged data science experts, community partners, and policy advocates, while also drawing upon an array of prior literature [13, 14] and similar toolkit efforts [15]. Initially, based on the regulatory focus of prior academic research, we envisioned that the primary users of the Algorithmic Equity Toolkit would be employees in state and local government seeking to surface the potential for algorithmic bias in existing systems. We thought advocacy and grassroots organizations could also find the Toolkit useful for understanding the social justice implications of public sector systems.

Through our participatory design process [16], we refined our audience and design goals to focus on helping civil rights advocates and community activists—rather than state employees—identify and audit algorithmic systems embedded in public-sector technology, including surveillance technology. We achieve this goal through three Toolkit components: (1) A flowchart designed for lay users for identifying algorithmic systems and their functions; (2) A Question Asking Tool (QAT) for surfacing the key issues of social and political concern for a given system. These tools together reveal a system’s technical failure modes (i.e., potential for not working correctly, such as false positives), and its social failure modes (i.e. its potential for discrimination when working correctly); and (3) An interactive web tool that illustrates the underlying mechanics of facial recognition systems, such as the relationship between how models are trained and adverse social impacts. In creating this Toolkit, we followed a weekly prototyping schedule interspersed with stakeholder feedback and co-design sessions.

3 A Toolkit for Algorithmic Equity and Community Empowerment

At the time of writing, the Algorithmic Equity Toolkit has three components:

1. A flowchart for distinguishing surveillance and ADS's and their different functions.

2. A question-asking tool (QAT) for surfacing the social context of a given system, its technical failure modes (i.e., potential for not working correctly, such as false positives), and its social failure modes (i.e. its potential for discrimination when working correctly)

3. An interactive demo of facial recognition that reveals the underlying harms and mechanics of facial recognition technology.

A key goal of this toolkit is to overcome power asymmetries between individuals and systems of authority, such as government agencies who should be held accountable for the technologies they implement in their communities. The Toolkit can be used when engaging with policymakers and other public officials, or in other contexts where individuals and groups want to learn more about surveillance and ADS technologies and their potential harms.

3.1 Flowchart/guide for identifying a machine learning or AI system

Unmet need. Information technologies are an increasing part of our everyday lives. Some technologies are more impactful than others, potentially affecting individual and group autonomy, civil rights, and safety. Our work with community groups and civil rights activists suggests that a means of ensuring that the effects of information technologies are mainly positive, or that their negative aspects are minimized, begins at recognizing and understanding the technologies in our midst. This is particularly true of public-sector technologies, where the principles of democratic governance require that state actors be accountable to the public for the tools and technologies they use to manage and control the population. Research by [redacted for review] suggests that the public, including policy makers, need assistance in identifying the opaque algorithmic aspects of public sector systems so that technology implementations can be sufficiently transparent and publicly accountable.

Meeting the need of helping community organizers understand: Where is the algorithm in this system—what is the algorithm doing? As described by [redacted for review], lay observers, including professionals who should know, often do not recognize that a system is "algorithmic". At other times, people may know a technology is algorithmic, but they don't know how the algorithm is coming into play. In still more cases, there are systems that can be understood as algorithmic but their harms are not necessarily of concern (e.g. simple calculators, thermostats). The goal of the flowchart tool is to signal the likely presence of algorithms that likely pose harms, especially harms that correspond to marginalized identities and histories of discriminatory state action.

The tool represents a set of definitional criteria, which, when applied to algorithmic systems, help to scope which technologies should be part of the conversation.

Form. The tool we developed guides users through a process for identifying components of technical systems that are algorithmic. Many technological artifacts are ambiguous as to their inner functionality leaving observers, including users, unaware of what kind of work the artifact does over and above its most obvious functions. To make the embedded features more salient and open to questioning, our flowchart tool offers a decision tree for contemplating what has been disclosed or can be observed about a technology, providing a verdict about whether it might be an AI system. While some systems are relatively straightforward, either because their functions are obvious, publicized, or fully disclosed, there are other technologies that are more challenging to unpack. An example of the former is booking photo comparison software (BPCS), which employs an algorithmic system that has already faced considerable public scrutiny, facial recognition. Many other artifacts contain algorithmic features that are much harder to detect simply by encountering them or even by having them explained by a public official or software vendor.

The flowchart differentiates algorithmically-enhanced systems from systems that are merely surveillant (i.e. only a data collection tool and not a tool that performs, say, an analysis and/or renders action-guiding judgements, or takes its own actions). An automated license plate reader (ALPR) may appear at first to be merely surveillant—basically a device that captures license plate images. But embedded within are AI components such as computer vision and algorithms for recognizing alpha-numeric sequences and matching the results to lists of license plates of interest. It is helpful to understand these features because, over and above whatever functionality is most obvious (e.g. a camera), embedded systems have their own failure modes, design constraints, and social valences that can contribute to the artifact's impact on individuals and communities. For example, some ALPR systems do not detect the issuing state of a license plate suggesting that a driver from Arizona could be misidentified as a driver from Pennsylvania whose license plate contains a similar alpha-numeric sequence. Even when such a system accurately identifies a license plate of interest, there are questions about the social conditions that lead to drivers being subjects of detection, such as the correlation between unpaid parking tickets and racialized poverty, that cannot be asked without peeling back the layers of technology to the sociotechnical imaginaries bundled within.

3.2 Asking the right questions.

Unmet need. Having identified an algorithmic system, the next step is to pose questions about it; about its functions and features, about the claims made about its efficacy, and about its potential to harm those to whom it is applied. Armed with a narrowly tailored set of questions, community organizers and activists can contest the narratives provided to them by authority figures and product vendors, proposing richer shared meanings onto the technologies in question. Given a camera with facial recognition capabilities, for example, Toolkit users will be able to address concerns about this technology, such as issues of race and gender detection parity and the potential for the tool contribute to oppressive feedback loops in which systemic discrimination is reproduced through the use of the tool by institutions with a history of discriminatory action. In creating this tool, we set some baseline standards, including: (i) it must be intuitive and legible to non-technical users; and (ii) questions should employ familiar language to the extent possible.

Form. The Question Asking Tool (QAT) is a tool for guiding users through the salient issues presented by an algorithmic system. Its goal is to surface social contexts and technical failure modes and to prompt questions that reveal potential harms, particularly harms to particular communities and identities. The QAT could also contribute to algorithmic impact assessments required by local and international laws (e.g. the General Data Protection Regulation) and recommended by legal experts and other scholars [17, 18], including the public accountability processes required by municipal surveillance ordinances in the United States. The tool can also be used by individual community members in dialogue with public officials and other authority figures. The tool distills known harms from the Fairness, Accountability, and Transparency literature and translates them for non-specialist audience.

The QAT prompts Toolkit users to identify the socio-ethical issues community advocates and civil rights activists should be concerned with in regards to algorithmic systems. In what ways does a particular type of algorithmic system reinforce bias and discrimination? What should individuals and groups with little or no technical expertise understand about the impacts of algorithmic tools? What answers should they demand from public officials and other authority figures implementing management and control technologies in their communities? The QAT contains a series of questions sorted into categories designed to assess an algorithmic system's potential harms in regard to social impact, appropriate use, transparency and accountability, data security and privacy, and interpretability or operability.

3.3 Interactive demo of intersectional failures of facial recognition

Unmet need. Observers may have heard that algorithmic systems are problematic but may have difficulty envisioning and internalizing what those problems are. The interactive demo makes at least some issues of algorithmic sorting and decision making salient to the user.

Form. The interactive demo tool demonstrates the problem of algorithmic harms such as bias in machine learning due to technical limitations and model representation, among other problems. Our demo involved running ten celebrity photos in Open Face's model using a database of 60 celebrity photos collected from Labeled Faces in the Wild and Google image searches. We then selected the top 8 closest images for each of the ten celebrity photos to include in our demo. Of all the ten celebrity photos, the minimum similarity score of the top 8 closest images was 0.15, between a photo of Aaron Peirsol and Ai Sugiyama, and the maximum similarity score was 1.384, between two different photos of LeBron James. Overall, celebrities with lighter skin tones had lower similarity scores than celebrities with darker skin tones. Our demo showing differences in similarity scores along the lines of skin tone are consistent with the literature surrounding facial recognition software and accuracy according to skin tone [9].

4 Conclusion

Community organizers and civil rights activists throughout the U.S. are concerned about surveillance technologies being implemented in their communities. There is concern that these technologies are being used by law enforcement and other public officials for profiling and targeting historically marginalized communities. Activists and advocates have pushed for algorithmic equity (accountability, transparency, fairness) through the implementation of legislation like municipal surveillance ordinances that regulate and supervise the acquisition and use of surveillance technology. Major cities, including Seattle, Berkeley, Nashville, Cambridge, and others have implemented ordinances that differ in their scope, process, and power in regulating government technologies. However, most technology policy legislation in the U.S. fails to manage the growing use of automated decision systems such as facial recognition and predictive policing algorithms. Despite its limitations, the Algorithmic Equity Toolkit is a vital tool that community civil rights advocates can use to voice their concerns about these technologies during the decision-making process for the acquisition of these technologies.

References

1. Angwin, J., Larson, J., Mattu, S., Kirchner, L.: Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And it's Biased Against Blacks., https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing, (2016).
2. Desmarais, S.L., Lowder, E.M.: PRETRIAL RISK ASSESSMENT TOOLS. 12 (2019).
3. Dressel, J., Farid, H.: The accuracy, fairness, and limits of predicting recidivism. Sci. Adv. 4, (2018). https://doi.org/10.1126/sciadv.aao5580.
4. Buolamwini, J., Gebru, T.: Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. In: Proceedings of Machine Learning Research. p. 15. , New York, NY (2018).
5. Dastin, J.: Amazon scraps secret AI recruiting tool that showed bias against women, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, (2018).
6. Lecher, C.: Privacy advocate held at gunpoint after license plate reader database mistake, lawsuit alleges, https://www.theverge.com/2019/2/21/18234785/privacy-advocate-lawsuit-california-license-plate-reader, (2019).
7. Cowley, S.: Equifax to Pay at Least $650 Million in Largest-Ever Data Breach Settlement, https://www.nytimes.com/2019/07/22/business/equifax-settlement.html, (2019).

8. American Civil Liberties Union: Community Control Over Police Surveillance, https://www.aclu.org/issues/privacy-technology/surveillance-technologies/community-control-over-police-surveillance, last accessed 2019/09/23.
9. González, M.L.: Seattle: Surveillance Ordinance (Seattle). (2017).
10. Harrell, B.: Seattle: Surveillance Ordinance Amendment. (2018).
11. Krafft, P.M., Young, M., Katell, M., Huang, K., Bugingo, G.: Defining AI in Policy versus Practice. SSRN Electronic Journal. J. 14.
12. Young, M., Katell, M., Krafft, P.M.: Municipal surveillance regulation and algorithmic accountability. Big Data Soc. 6, (2019). https://doi.org/10.1177/2053951719868492.
13. Green, B.: Data Science as Political Action: Grounding Data Science in a Politics of Justice. (2018).
14. Dillahunt, T.R., Erete, S., Galusca, R., Israni, A., Nacu, D., Sengers, P.: Reflections on Design Methods for Underserved Communities. In: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion. pp. 409–413. ACM Press, Portland, Oregon, USA (2017). https://doi.org/10.1145/3022198.3022664.
15. David Anderson, Joy Bonaguro, Miriam McKinney, Andrew Nicklin: Ethics & Algorithms Toolkit (beta), https://ethicstoolkit.ai/, last accessed 2019/09/18.
16. Katell, M., Young, M., Herman, B., Dailey, D., Tam, A., Guetler, V., Binz, C., Raz, D., Krafft, P.M.: Toward Situated Interventions for Fairness, Accountability, and Transparency: Lessons from the field. In: Proceedings of Machine Learning Research. , Barcelona, Spain (2020).
17. Article 29 Data Protection Working Party: Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679 (wp251rev.01). (2018).
18. Edwards, L., Veale, M.: Slave to the Algorithm? Why a "right to an explanation" is probably not the remedy you are looking for. Duke Law Technol. Rev. 16, 18–84 (2017). https://doi.org/10.31228/osf.io/97upg.