

What Is the Point of Fairness?

Insights

- Disabled lives and lifeworlds must be considered in AI ethics.
- Considering disabled lives troubles the idea of fairness (equal opportunity) as a sufficient ethical value for those seeking better technology design.
- Instead, we propose that AI ethics research focus on justice (equitable outcomes), which accounts for the historic and ongoing discrimination of marginalized people.

As machine learning becomes more ubiquitous, questions of AI and information ethics loom large. Much concern has been focused on promoting AI that results in more fair outcomes that do not discriminate against protected classes, such as those marginalized on the basis of gender and race. Yet little of that work has specifically investigated disability. Two notable exceptions, both from within the spaces of disability studies and assistive technology (AT), are Shari Trewin’s statement on “AI Fairness for People with Disabilities” [1] and the World Institute on Disability’s comments on AI and

accessibility [2]. Together they argue that making disability explicit in discussions of AI and fairness is urgent, as the quick, black-boxed nature of automatic decision making exacerbates the disadvantages that people with disabilities already endure and creates new ones. Though low representation in datasets is blamed, increasing representation will be complex, given disability politics. For example, disabled people strategically choose whether and how to disclose their disabilities (if they even identify as having disabilities), likely leading to inconsistent datasets even when disability information is

intentionally collected. Additionally, disabilities present themselves (or not) in myriad ways, destabilizing (category-dependent) machine learning as an effective way of correctly identifying them.

We are encouraged by the nascent engagement between disability studies, AT, and AI ethics, and agree with many of the concerns outlined in both documents. For example, healthcare and employment remain out of reach for many disabled people despite policies that prohibit discrimination on the basis of disability, and we would be remiss in denying AT's role in increasing quality of life for some people with disabilities, however incrementally. At the same time, fairness is not an uncontested concept; ethicists have troubled the notion that it can produce justice in and of itself. A recent paper by Anna Lauren Hoffmann, for example [3], pointing to the way that fairness is modeled on U.S. anti-discrimination law, surfaces the gaps and injustices on which a fairness framing remains silent, including its failure to dismantle and rework structural oppression. In fact, without addressing the hierarchies that disadvantage people with disabilities in the first place, Hoffmann and disability justice activists argue, fairness may reproduce the discrimination it seeks to remedy. Justice, on the other hand, guides recovery by aiming at repairing past harm. In doing so, it may scaffold more accountable and responsible AI that is equitable in its handling of data as well as deployment (or withholding). Therefore, we argue for a reframing from fairness to justice in the realm of AI ethics and disability.

For the rest of this article, we will present two case studies—one on the use of AI to diagnose neurodiversity, including autism,

and the second about computer vision that provides information for blind people. After introducing a case, we will offer an overview of some concerns that might be raised through a fairness lens and then some concerns which might be raised with a justice lens. We will show that the application of a principle of fairness, while an improvement over inaction, does not prevent the harms for which the technology opens space. Through these cases, we hope to concretize differences between the two lenses and demonstrate how justice can situate and pluralize our conversations on AI, ethics, and disability to address societal, structural oppression beyond improving automatic decision making and datasets themselves.

AI FOR DIAGNOSIS

A body of research within computer vision attempts to create systems that, using facial recognition, can automatically diagnose certain neurodiverse states—including autism [4]. Using already recognized and diagnosed autistic children, researchers rely on examining facial expressions, degrees of emotiveness, and repetitive behaviors to provide diagnostic tools, arguing that doing so may reduce the delay of diagnosis in a child's life.

Concerns raised through a fairness lens. With diagnosis, researchers are confronted with biases in the preexisting framework of autism—particularly the widely studied gender bias in symptoms, and the consequential discrepancies in diagnostic rates, and less-studied but firmly established biases around race and ethnicity, class and geography. Dependence on diagnostic tools that are based on the experiences of *those already diagnosed* thus risks replicating these biases, providing

seemingly objective rigor to determinations that a child presenting inconsistently with (white, assigned male at birth) autistic children cannot be autistic and should be gatekept out of support systems. With a fairness metric, we might suggest diversifying datasets so marginalized genders and races can be correctly diagnosed. But this solution may not adequately consider what it means to have the power to diagnose, and who might endure what consequences as a result.

Concerns raised through a justice lens. In the case of diagnostic tools for autism, we run into concerns around medicalization and gatekeeping: the distinct power that comes with diagnostic authority given the institutionalization of a medical model of disability into the power structures of society.

Tools to “help” autistic people in the model of existing computer vision prototypes do not just provide diagnosis—they also reinforce the notion that the formal diagnostic route is the only legitimate one for autistic existences, in turn reinforcing the power that psychiatrists hold. Examinations of medicalization—the process by which this notion of formal gatekeeping becomes legitimized—have already identified it within autism diagnostics, simultaneously finding little validity to the diagnostic systems that computer vision researchers are using as their baseline. By adding technical and scientific authority to medical authority, people subject to medical contexts are even further disempowered, with the patient's voice getting even less legitimacy. Once again, fairness is not a solution; the issue is not one of discrimination against the patient for being autistic but rather for being a *patient*. Just outcomes in this area, in other words, require a consideration of power, not fairness, and of the wider social context into which technical systems are placed.

Finally, and more cut and dried, there is the question of consequences and implications in the case of an autism diagnosis. AI systems in this domain are built on the premise that an early diagnosis is a good outcome, and that diagnosis leads to

Without addressing the hierarchies that disadvantage people with disabilities in the first place, fairness may reproduce the discrimination it seeks to remedy.

possibilities for treatment, support, and consideration. Notwithstanding the already discussed biases in who can access diagnosis (and how diagnostic tests are constructed), there are serious questions raised in psychiatry and critical disability studies about whether an earlier diagnosis is a better one. Rather than helping people, earlier diagnoses may harm them.

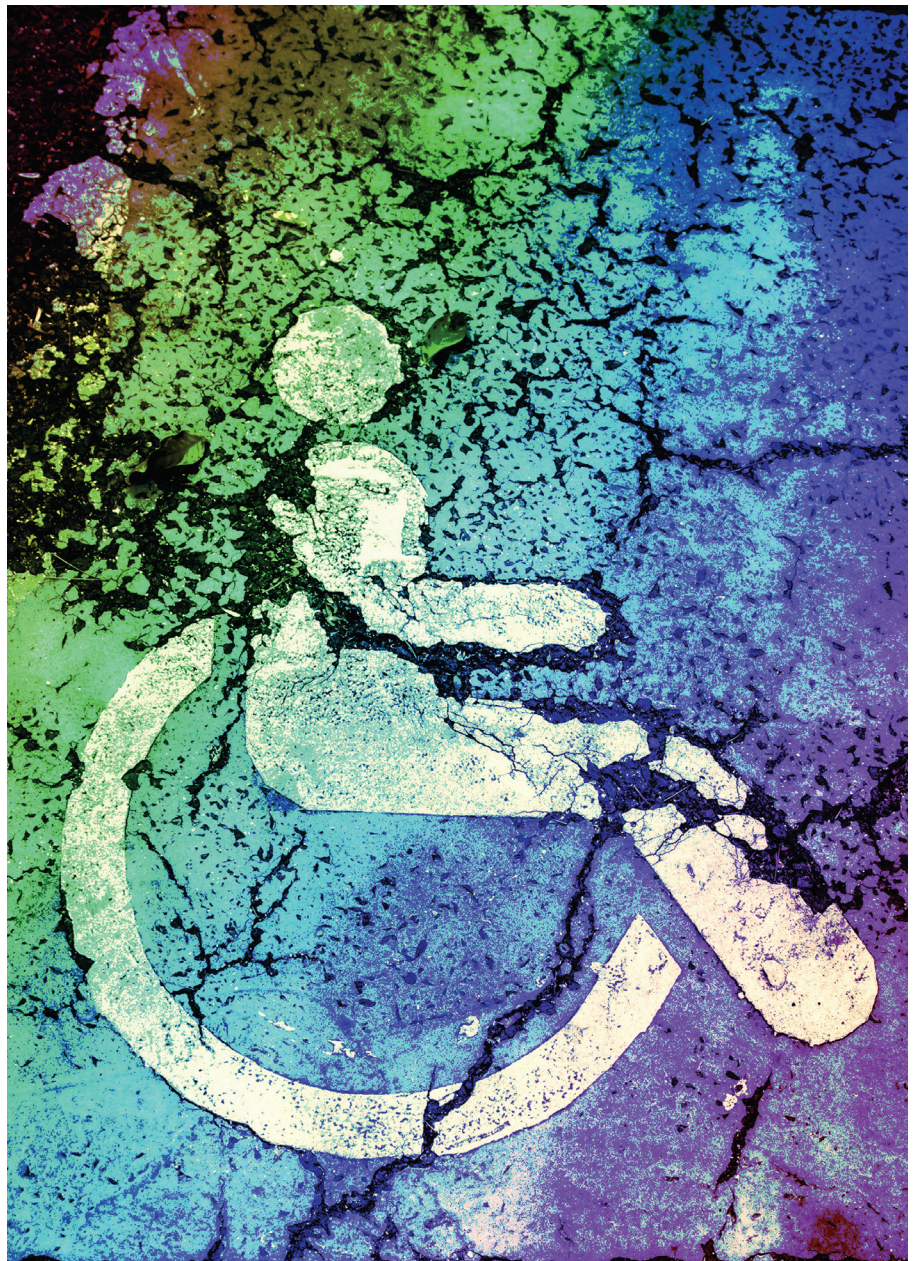
Even worse consequences stem from the fact that autism is not “just” a diagnostic label, whatever computer vision researchers may think. It is also a label that carries with it certain associations about financial cost, incapability, and risk—associations that have led to myriad harmful behavior-change therapies and autistic children being murdered as “mercy killings” [5]. As Mitzi Waltz puts it, “autism = death.” Morally and ethically, computer vision systems that provide that label, if designed without attendance to the wider societal contexts in which autistic people live, might well be considered death too.

Autism diagnosed with AI *is* an issue of fairness—an issue of the unfair treatment of autistic people—but it cannot be solved simply through examining the immediate algorithmic inputs and outputs of the computer-vision system. Instead, we need models that consider holistic, societal implications, and the ways in which technologies alter the life chances of those they are used by or on.

AI FOR “SIGHT”

Our second case concerns a longstanding area of research—engaged in by AI researchers, health researchers, and HCI researchers—that of using computer vision (AI that “sees”) to assist vision-impaired people. These include, for example, haptic/vision-based systems for facial recognition for communicating conversational partners’ identity [6], and for object and scene recognition [7].

Concerns raised through a fairness lens. First, we must ask: Sight for whom, and what gets seen? There is a longstanding recognition of biases within computer vision systems, and limitations in their ability to represent the complexity of the world—biases that often impact



those already marginalized. In the case of object recognition, for example, a recent paper demonstrates that such systems are developed largely in a white, Western, and middle-class context, failing to recognize common household objects that are more often found in poor or non-Western environments [8]. The centering of such systems in AT design risks further harm to people already marginalized within both societies widely and the disability community. And improving algorithms to recognize more genders, races, and objects still predisposes futures where surveillance technologies may be justified for their utility for blind people, while ignoring their ongoing documented misuse.

Concerns raised through a justice lens. Unlike AI for diagnosis, computer vision to help people see seems to put more control in the disabled users’ hands. They are not the focus of the gaze: They are the ones gazing. But this inversion does not necessarily redistribute power in a positive fashion; it can still promote asymmetric and harmful power distributions. Whereas tools like a white cane assume the brain as the analytical unit, computer vision may transfer such judgment to automatic decision making. Though developers of many identifying technologies clarify that their use is meant to support, not replace, human decision making, we know that technology is often pedestalized; that technological



Association for
Computing Machinery

2018 JOURNAL IMPACT
FACTOR: 6.131

ACM Computing Surveys (CSUR)

ACM Computing Surveys (CSUR) publishes comprehensive, readable tutorials and survey papers that give guided tours through the literature and explain topics to those who seek to learn the basics of areas outside their specialties. These carefully planned and presented introductions are also an excellent way for professionals to develop perspectives on, and identify trends in, complex technologies.



For further information
and to submit your
manuscript,
visit csur.acm.org



and scientific ways of knowing are treated as superior to the alternative, and frequently deferred to even in the presence of contradictory information or assumed to be more accurate than they are [9]. The result is that a computer vision system for accessibility, while rendering things more accessible, does so by shifting

the center of analysis and judgment away from the user and toward the (frequently expensive, black-boxed, and commercially shaped) technology in hand.

Finally, computer vision, even deployed fairly, cements vision as a superior sense and legitimizes surveillance. Much research,

Computer vision to help people see seems to put more control in the disabled users' hands. They are not the focus of the gaze: They are the ones gazing.

IMAGE BY HALFPPOINT / SHUTTERSTOCK.COM

including that cited to inform AI for accessibility, acknowledges and even praises nonvisual sensemaking. Accessibility researchers hardly advocate substituting this knowledge with technology. Yet these gestures would be more substantive if the same rigor and enthusiasm were applied to the development of technologies that train in or privilege nonvisual sensemaking. Next, surveillance technologies are controversial, and disability studies scholars have critiqued the ubiquity and inaccuracy of technology-savior narratives that hail automation for increasing the quality of life of people with disabilities. Here we risk glorifying surveillance without questioning its misuse. How could technology to assist a blind person be kept from integration into policing technologies? And who's to say blind people aren't among the users of policing technologies? Instead, until significant work is done to correct and bring nuance to stories about disability, those who question the use of surveillance technologies even when they are used with the intention of assisting disabled people may be shamed.

These are not issues that notions of fairness can surface, articulate, and tackle, because the issue is not only that disabled subpopulations may be treated unequally among each other or compared to normative society, but also that the technologies' model of liberation is liberation that neglects to challenge wider structures of power.

CONCLUSION

We have presented two case studies of AI interventions in disabled lives and the issues they raise around and with fairness. As we have made clear, we believe that fairness—a concept that critical data studies is already shifting away from—is highly dangerous for conversations around disability and AI to center. Rather, we advocate that everyone interested in questions of disability and AI critically examine the overarching social structures we are participating in, upholding, and creating anew with our work. Doing so requires and results in centering our work not on questions of fairness, but instead on questions of justice.

There are many places to draw from in doing that. Technology has always been a part of the construction of disability, and of the nature of disabled lives. Consequently, disability studies has long considered questions of technology. Just as Mankoff et al. urged the integration of disability studies into assistive technology [9], we urge a similar integration of AI and disability conversations with disability studies conversations around technology, justice, and power—conversations that are already taking place [10,11].

Similarly, though disability itself leads to unique life experiences and oppression, there is myriad scholarship on AI and black lives, trans lives, poor lives—and many of those lives are disabled lives too. As such, it is imperative that efforts concerning just developments and deployments of AI for people with disabilities center multiply marginalized disabled people, or we risk helping only the most privileged. Additionally, we need to carve out space in AI ethics programs that are not considering disability, calling in the disability forgetting that has gone on in many purportedly justice-oriented conversations. AI is new—but the systems of oppression that stigmatize disability are very old. They will not be unraveled piecemeal, or separate from recognizing and reckoning with the structural inequalities that have made unjust AI possible.

ACKNOWLEDGMENTS

Our thanks to Margret Wander, Nikki Stevens, and Adam Hyland for their feedback, help, and support.

ENDNOTES

1. Trewin, S. AI fairness for people with disabilities: Point of view. 2018. arXiv preprint arXiv: 1811.10670
2. World Institute on Disability. AI and Accessibility. June 12, 2019; <https://wid.org/2019/06/12/ai-and-accessibility/>
3. Hoffmann, A.L. Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication, & Society* 22, 7 (2019), 900–915.
4. Thevenot, J., Bordallo López, M., and Hadid, A. A survey on computer vision for assistive medical diagnosis from faces.

- IEEE Journal of Biomedical and Health Informatics* 22, 5 (Oct. 2017), 1497–1511.
5. Waltz, M. Autism= death: The social and medical impact of a catastrophic medical model of autistic spectrum disorders. *Popular Narrative Media* 1, 1 (2008), 13–24.
 6. Astler, D., Chau, H., Hsu, K., Hua, A., Kannan, A., Lei, L., Nathanson, M., Paryavi, E., Rosen, M., Unno, H., Wang, C., Zaidi, K., Zhang, X., and Tang, C-M. Increased accessibility to nonverbal communication through facial and expression recognition technologies for blind/visually impaired subjects. *Proc. of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, 2011, 259–260; <https://doi.org/10.1145/2049536.2049596>
 7. Mulhari, D. A TensorFlow-based assistive technology system for users with visual impairments. *Proc. of the Internet of Accessible Things*. ACM, New York, 2018, Article 11; <https://doi.org/10.1145/3192714.3196314>
 8. DeVries, T., Misra, I., Wang, C., and van der Maaten, L. Does object recognition work for everyone? *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2019, 52–59.
 9. Mankoff, J., Hayes, G.R., and Kasnitz, D. Disability studies as a source of critical inquiry for the field of assistive technology. *Proc. of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, 2010, 3–10; <https://doi.org/10.1145/1878803.1878807>
 10. Banner, O. Technopsycence and Afro-Surrealism's cripistemologies. *Catalyst: Feminism, Theory, Technoscience* 5, 1 (2019); <https://doi.org/10.28968/cftt.v5i1.2961>
 11. Williams, R.M. and Gilbert, J.E. Cyborg perspectives on computing research reform. *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, 2019, Paper alt13; <https://doi.org/10.1145/3290607.3310421>

📍 **Cynthia Bennett** is a postdoctoral researcher examining human-centered design methods with a disability studies lens. She has been awarded both the National Science Foundation Graduate Research Fellowship and a Microsoft Research Dissertation Grant. → bennec3@uw.edu

📍 **Os Keyes** is a Ph.D. student at the University of Washington, where they study gender, infrastructure, and power. They have published at venues including CHI, CSCW, and ASSETS, and written for outlets including *Vice*, *Logic*, and *Scientific American*. They are the inaugural recipient of an Ada Lovelace Fellowship. → okeyes@uw.edu