

WHAT IS THE POINT OF FAIRNESS? DISABILITY, AI AND THE COMPLEXITY OF JUSTICE

Cynthia L. Bennett & Os Keyes
Human Centered Design & Engineering, University of Washington
{benne3, okeyes}@uw.edu

Abstract

Work integrating conversations around AI and Disability is vital and valued, particularly when done through a lens of fairness. Yet at the same time, analysing the ethical implications of AI for disabled people solely through the lens of a singular idea of "fairness" risks reinforcing existing power dynamics, either through reinforcing the position of existing medical gatekeepers, or promoting tools and techniques that benefit otherwise-privileged disabled people while harming those who are rendered outliers in multiple ways. In this paper we present two case studies from within computer vision - a subdiscipline of AI focused on training algorithms that can "see" - of technologies putatively intended to help disabled people but, through failures to consider structural injustices in their design, are likely to result in harms not addressed by a "fairness" framing of ethics. Drawing on disability studies and critical data science, we call on researchers into AI ethics and disability to move beyond simplistic notions of fairness, and towards notions of justice.

Introduction

As machine learning becomes more ubiquitous, questions of AI and information ethics loom large in the public imagination. 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" [30], and the World Institute on Disability's comments on AI and accessibility [36]. 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 disadvantages 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 a 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 to deny AT's role in increasing quality of life for some people with disabilities, if incremental. At the same time, "fairness", like "equality" [1], is not an uncontested concept. Ethicists have troubled the notion that it can produce justice in and of itself [14]. A recent paper by Anna Lauren Hoffmann, for example [12], pointing to

the way that fairness is modelled on U.S. anti-discrimination law, surfaces the gaps and injustices a fairness framing remains silent on, 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 aimed at repairing past harm and 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 when concerning AI ethics and disability.

To highlight the necessity of this reframing, we present two case studies of AI/AT in which the application of a principle of fairness, while an improvement over inaction, does not prevent the harms that the technology opens space for.

Fairness: Defined and Contested

To begin, we briefly define and critique fairness according to the scoped scholarship on AI, ethics and people with disabilities that we use to form our arguments. Shari Trewin and Anna Laura Hoffmann summarise fairness as it has been articulated by statisticians. These articulations (as applied to disability) largely evaluate whether similar cases, separated only by the presence of disability, produce the same outcome. In the event there is a disparity, the system is considered unfair. Approaches to remedying this assume that a lack of fairness is a failure of *implementation*, and largely centre increasing the representativeness of the data underlying the algorithm, and improving how the algorithm integrates “outliers”[30, 12]. An initial reading of fairness may make it seem a reasonable goal to centre in algorithmic systems; after all, what is the problem with addressing disparities?

Hoffmann delineates four primary limitations of an approach to the consequences of AI that centres fairness. First, fairness is not fair; it aims to increase the status of disadvantaged people without explicitly addressing how the privileges of more powerful people will change. Second, fairness relies on traits being well defined so a system can know what to do with them. Third, fairness historically aims to improve one contested identity at a time when in reality, many people are multiply marginalized, and it is not just aspects of their identity but interactions among multiple facets of their identities and oppressive structures which produce systematic discrimination in different ways in different situations. Finally, fairness frames marginalization as occurring in relation to specific things, namely assumed desires like employment. However, much injustice is produced by the development of standards (both formally and informally) which are then applied across domains. In other words, someone is not marginalized when applying for jobs and then not. Instead, oppression is threaded through what they do according to predetermined norms and disciplinary institutions which enforce them. In summary, fairness is premised on understanding how oppression manifests in an individual and aims to promote equality through the remediation of technologies. It does not question the structures (including those which rely on AI to surveil and make decisions) that allow people with disabilities to be disadvantaged in the first place.

For the rest of this paper, we will present two case studies—one on the use of AI to diagnose neurodiversity including autism and the second about computer vision which provides information for blind people. After introducing a case, we will overview some concerns that might be raised through a fairness lens and then some concerns which might be raised with a justice lens. 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 which can (using facial recognition) automatically diagnose certain neurodiverse states – including autism [28, 24]. Using already-recognized and diagnosed autistic children, researchers rely on examining facial expressions, degrees of emotiveness, and repetitive behaviours to provide diagnostic tools, arguing that doing so would reduce the delay of diagnosis in a child's life [23, 11].

Concerns Raised through a Fairness Lens

With diagnosis, researchers are confronted with biases in the pre-existing framework of autism – particularly the widely-studied gender bias in symptoms [25], and the consequential discrepancies in diagnostic rates [10] – and less-studied but firmly established biases around race and ethnicity, class and geography [4, 17]. Dependence on diagnostic tools which are based on the experiences of *those already diagnosed* thus risk 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 [7].

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, reinforcing the power that psychiatrists hold. Examinations of medicalization – the process by which this notion of formal gatekeeping becomes legitimized [6] – have already identified it within autism diagnostics [32], simultaneously finding little validity to the diagnostic systems computer vision researchers are using as their baseline [29]. By adding technical and scientific authority to medical authority, people subject to medical contexts are not only not granted power, but are even further disempowered, with even less legitimacy given to the patient's voice. Once again, fairness is not a solution; the issue is not one of discrimination against the patient for being autistic but for being a *patient*. Just outcomes in this area, in other words, require not a consideration of fairness but of power, and of the wider social context into which technical systems are placed.

Finally, and more cut-and-dried, there is the question of what the consequences and implications of an autism diagnosis are. AI systems in this domain are built on the premise that an early diagnosis is a good outcome, that diagnosis leads to possibilities of treatment, support and consideration. Notwithstanding the already-discussed biases in who can access diagnosis (and how diagnostic tests are constructed), there are serious questions about whether an earlier diagnosis is a better one [26]. 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 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” [33, 19]. As Mitzi Waltz puts it, “autism = death”. Morally and ethically, computer vision systems to 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 is not one that can be solved simply through examining the immediate algorithmic inputs and outputs of the computer vision system. Instead, we need a model that considers holistic, societal implications, and the way that 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 – is that of using computer vision (AI that “sees”) to assist vision-impaired people [34, 15, 20, 37]. Projects presented at ASSETS alone include haptic/vision-based systems for detecting and representing the emotions behind facial expressions [5], augmented touch for communicating visual information [9], facial recognition for communicating conversational partners’ identity [2], and object and scene recognition [21].

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 [13]. 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 centring of such systems in AT design risks further harm to people already marginalized within both society widely, and the disability community. And improving algorithms to recognize more genders, races and objects still predispose futures where surveillance technologies may be justified for their utility for blind people while ignoring their ongoing documented misuse [12, 13].

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 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 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 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 far more accurate than they are [16]. The result is that a computer vision system for accessibility, while rendering things more accessible, does so by shifting the centre of analysis and judgment away from the user and towards 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, including that cited to inform AI for accessibility, acknowledge and even praise 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 which train in or privilege nonvisual sensemaking [35]. Next, surveillance technologies are controversial, and disability studies scholars have critiqued the ubiquity and inaccuracy of technology saviour narratives which hail automation for increasing the quality of life of people with disabilities [31]. Here we risk glorifying surveillance without questioning its misuse. How could technology to assist a blind person be kept from integration into policing technologies; who's to say blind people aren't among the users of policing technologies? Instead, until significant work is done to correct and nuance stories about disability, those who question using surveillance technologies even when used for the purposes of assisting disabled people may be shamed [22].

These are not issues that notions of "fairness" can surface, articulate and tackle, because the issue is not only that disabled sub-populations may be treated unequally between each other or compared to normative society, but that the technologies' model of liberation is liberation without challenging 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 already being shifted away from in critical data studies – is highly dangerous for conversations around disability and AI to centre. 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 a centring of our work not on questions of fairness, but on questions of *justice* [12].

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, and continues to do so. Just as Mankoff *et al.* urged the integration of disability studies into assistive technology [18], 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 [3, 35].

Similarly, though disability alone 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 centre multiply marginalized disabled people, or we risk only helping the most privileged. Additionally, we need to carve out space in AI ethics programs which 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 give rise to disability are very, very old. They will not be unravelled 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, Anna Lauren Hoffmann and Adam Hyland for their feedback, help and support.

References

1. Elizabeth S. Anderson. 1999. What is the Point of Equality?. *Ethics* 109, 2 (January 1999), 287-337.
2. Douglas Astler, Harrison Chau, Kailin Hsu, Alvin Hua, Andrew Kannan, Lydia Lei, Melissa Nathanson, Esmaeel Paryavi, Michelle Rosen, Hayato Unno, Carol Wang, Khadija Zaidi, Xuemin Zhang, and Cha-Min Tang. 2011. Increased accessibility to nonverbal communication through facial and expression recognition technologies for blind/visually impaired subjects. In *Proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility* (ASSETS '11). ACM, New York, NY, USA, 259-260. DOI: <https://doi.org/10.1145/2049536.2049596>
3. Olivia Banner. 2019. Technopsycence and Afro-Surrealism's Cripistemologies. *Catalyst: Feminism, Theory, Technoscience*. 5, 1 (2019). DOI: <https://doi.org/10.28968/cftt.v5i1.29612>
4. Sander Begeer, Saloua El Bouk, Wafaa Boussaid, Mark Meerum Terwogt, and Hans M. Koot. 2009. Underdiagnosis and referral bias of autism in ethnic minorities. *J Autism Dev Disord*. 39, 1 (January 2009), 142-148. DOI: <https://doi.org/10.1007/s10803-008-0611-5>
5. Hendrik Buimer, Thea Van der Geest, Abdellatif Nemri, Renske Schellens, Richard Van Wezel, and Yan Zhao. 2017. Making Facial Expressions of Emotions Accessible for Visually Impaired Persons. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17). ACM, New York, NY, USA, 331-332. DOI: <https://doi.org/10.1145/3132525.3134823>
6. Adele E. Clark, Janet K. Shim, Laura Mamo, Jennifer Ruth Fosket, and Jennifer R. Fishman. 2010. Biomedicalization: Technoscientific transformations of health, illness, and US biomedicine. In *Biomedicalization: Technoscience, health, and illness in the US* (2010), 47-87.
7. Peter Conrad. 1992. Medicalization and social control. *Annual Review of Sociology* 18, 1 (August 1992), 209-232.
8. Terrance DeVries, Ishan Misra, Changan Wang, and Laurens van der Maaten. 2019. Does Object Recognition Work for Everyone?. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 52-59.
9. Leah Findlater, Lee Stearns, Ruofei Du, Uran Oh, David Ross, Rama Chellappa, and Jon Froehlich. 2015. Supporting Everyday Activities for Persons with Visual Impairments Through Computer Vision-Augmented Touch. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (ASSETS '15). ACM, New York, NY, USA, 383-384. DOI: <https://doi.org/10.1145/2700648.2811381>
10. Ellen Giarelli, Lisa D. Wiggins, Catherine E. Rice, Susan E. Levy, Russell S. Kirby, Jennifer Pinto-Martin, and David Mandell. 2010. Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disabil. Health J*. 3, 2 (April 2010), 107-116. DOI: <https://doi.org/10.1016/j.dhjo.2009.07.001>
11. Jordan Hashemi, Mariano Tepper, Thiago Vallin Spina, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, Helen Egger, Geraldine Dawson, and Guillermo Sapiro. 2014. Computer vision tools for low-cost and noninvasive measurement of autism-related behaviors in infants. *Autism research and treatment* (2014). DOI: <https://doi.org/10.1155/2014/935686>
12. Anna Lauren Hoffmann. (2019) Where fairness fails: Data, algorithms, and the limits of antidiscrimination discourse. *Information, Communication, & Society*, 22, 7 (2019), 900-915.
13. Os Keyes. 2018. The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition. In *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 88 (November 2018), 22 pages. DOI: <https://doi.org/10.1145/3274357>
14. Os Keyes, Jevan Hutson, and Meredith Durbin. 2019. A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI EA '19). ACM, New York, NY, USA, Paper alt06, 11 pages. DOI: <https://doi.org/10.1145/3290607.3310433>
15. Marco Leo, G. Medioni, M. Trivedi, Takeo Kanade, and Giovanni Maria Farinella. 2017. Computer vision for assistive technologies. *Computer Vision and Image Understanding* 154 (January 2017), 1-15.
16. Haley MacLeod, Cynthia L. Bennett, Meredith Ringel Morris, and Edward Cutrell. 2017. Understanding Blind People's Experiences with Computer-Generated Captions of Social Media Images. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (CHI '17). ACM, New York, NY, USA, 5988-5999. DOI: <https://doi.org/10.1145/3025453.3025814>
17. David S. Mandell, Maytali M. Novak, and Cynthia D. Zubritsky. 2005. Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics* 116, 6 (December 2005) 1480.

18. Jennifer Mankoff, Gillian R. Hayes, and Devva Kasnitz. 2010. Disability studies as a source of critical inquiry for the field of assistive technology. In *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility (ASSETS '10)*. ACM, New York, NY, USA, 3-10. DOI: <https://doi.org/10.1145/1878803.1878807>
19. Bernard Metz, James A. Mulick, and Eric M. Butter. Autism: A late-20th-century fad magnet. *Controversial therapies for developmental disabilities: Fad, fashion and science in professional practice*. 237-263, 2005.
20. Microsoft. 2017. Seeing AI: Talking Camera App for Those with a Visual Impairment (July 2017). Retrieved 30 June, 2019 from <https://www.microsoft.com/en-us/ai/seeing-ai>
21. Davide Mulfari. 2018. A TensorFlow-based Assistive Technology System for Users with Visual Impairments. In *Proceedings of the Internet of Accessible Things (W4A '18)*. ACM, New York, NY, USA, Article 11, 2 pages. DOI: <https://doi.org/10.1145/3192714.3196314>
22. Halley Profita, Reem Albaghli, Leah Findlater, Paul Jaeger, and Shaun K. Kane. 2016. The AT Effect: How Disability Affects the Perceived Social Acceptability of Head-Mounted Display Use. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 4884-4895. DOI: <https://doi.org/10.1145/2858036.2858130>
23. Shyam Rajagopalan, Abhinav Dhall, and Roland Goecke. 2013. Self-stimulatory behaviours in the wild for autism diagnosis. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 755-761.
24. James M. Rehg. Behavior Imaging: Using Computer Vision to Study Autism. In *Proceedings of the 2011 IAPR Conference on Machine Vision Applications*, 14-21.
25. Tessa Taylor Rivet and Johnny L. Matson. 2011. Review of gender differences in core symptomatology in autism spectrum disorders. *Research in Autism Spectrum Disorders* 5, 3 (July 2011), 957-976.
26. Ginny Russell. Early Diagnosis of Autism. *Re-thinking Autism: Diagnosis, Identity and Equality* 252, 2016.
27. Huiying Shen, Owen Edwards, Joshua Miele, and James M. Coughlan. 2013. CamIO: a 3D computer vision system enabling audio/haptic interaction with physical objects by blind users. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. ACM, New York, NY, USA, Article 41, 2 pages. DOI: <http://dx.doi.org/10.1145/2513383.2513423>
28. Jerome Thevenot, Miguel Bordallo López, and Abdenour Hadid. 2017. A survey on computer vision for assistive medical diagnosis from faces. *IEEE journal of biomedical and health informatics* 22, 5 (October 2017), 1497-1511.
29. Sami Timimi, Brian McCabe, K. Runswick-Cole, R. Mallett, and S. Timimi. Autism screening and diagnostic tools. *Re-thinking Autism: Diagnosis, Identity and Equality* 159-183, 2016.
30. Shari Trewin. 2018. AI fairness for people with disabilities: point of view. arXiv preprint arXiv: 1811.10670
31. Bonnie Tucker. 2017. Technocapitalist disability rhetoric: When technology is confused with social justice. *Enculturation* 24 (April 2017).
32. Kylie Valentine. 2010. A consideration of medicalisation: Choice, engagement and other responsibilities of parents of children with autism spectrum disorder. *Social Science & Medicine* 71, 5 (September 2010) 950-957.
33. Mitzi Waltz. 2008. Autism= death: The social and medical impact of a catastrophic medical model of autistic spectrum disorders. *Popular narrative media* 1, 1 (2008), 13-24.
34. Linda Wang, and Alexander Wong. 2019. Implications of Computer Vision Driven Assistive Technologies Towards Individuals with Visual Impairment. arXiv preprint arXiv:1905.07844
35. Rua M. Williams and Juan E. Gilbert. 2019. Cyborg Perspectives on Computing Research Reform. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems (CHI EA '19)*. ACM, New York, NY, USA, Paper alt13, 11 pages. DOI: <https://doi.org/10.1145/3290607.3310421>
36. World Institute on Disability. 2019. AI and Accessibility (July 2019). Retrieved 23 June, 2019 from <https://wid.org/2019/06/12/ai-and-accessibility/>
37. Yuhang Zhao, Cynthia L. Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. Enabling People with Visual Impairments to Navigate Virtual Reality with a Haptic and Auditory Cane Simulation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Paper 116, 14 pages. DOI: <https://doi.org/10.1145/3173574.3173690>

About the Authors:



Cynthia Bennett is a Ph.D. candidate at the University of Washington. Her research studies Human-Centered Design methods with a disability studies lens. Her work is supported by the National Science Foundation Graduate Research Fellowship and a Microsoft Research Dissertation Grant.



Os Keyes is a Ph.D. at the University of Washington, where they study gender, infrastructure and power. They have published at venues including CHI, CSCW, ASSETS, and written for (amongst others) *Vice*, *Logic* and *Scientific American*. They are the inaugural recipient of an Ada Lovelace Fellowship.