May 2020

Automating autism: Disability, discourse, and Artificial Intelligence

Os Keyes
University of Washington, okeyes@uw.edu

Automating autism: Disability, discourse, and Artificial Intelligence

Automating autism: Disability, discourse, and Artificial Intelligence

Os Keyes
Department of Human Centered Design & Engineering, University of Washington, Seattle, United States of America

As Artificial Intelligence (AI) systems shift to interact with new domains and populations, so does AI ethics: a relatively nascent subdiscipline that frequently concerns itself with questions of "fairness" and "accountability." This fairness-centred approach has been criticized for (amongst other things) lacking the ability to address discursive, rather than distributional, injustices. In this paper I simultaneously validate these concerns, and work to correct the relative silence of both conventional and critical AI ethicists around disability, by exploring the narratives deployed by AI researchers in discussing and designing systems around autism. Demonstrating that these narratives frequently perpetuate a dangerously dehumanizing model of autistic people, I explore the material consequences this might have. More importantly, I highlight the ways in which discursive harms—particularly discursive harms around *dehumanization*—are not simply inadequately handled by conventional AI ethics approaches, but actively invisible to them. I urge AI ethicists to critically and immediately begin grappling with the likely consequences of an approach to ethics which focuses on personhood and agency, in a world in which many populations are treated as having neither. I suggest that this issue requires a substantial revisiting of the underlying premises of AI ethics, and point to some possible directions in which researchers and practitioners might look for inspiration.

Keywords: AI ethics, autism, feminist epistemology, critical disability studies, personhood

Conventional approaches to AI ethics frequently concern themselves with notions of "fairness," "accountability," and "transparency"—notions which have been troubled repeatedly by scholars as failing to attend to (amongst other issues) the *discursive* consequences of AI systems and development (Hoffmann, 2019). Further, AI ethics investigations of these principles—particularly fairness—are frequently silent on questions of disability, and the lives of disabled people, usually attending to questions of race and gender in isolation. Beginning from Elizabeth Ellcessor's argument that adopting a disabilities studies framework allows us to "find new questions" (Ellcessor, 2016, p. 4), I seek to demonstrate not only the harms to disabled people that occur when ethical frameworks fail to explicitly consider us, but ways in which incorporating disability into our evaluations of AI systems reveals more general and fundamental problems in the ethical frameworks being adopted and normalised.

In order to demonstrate this, I use this paper to inquire into the discursive frameworks perpetuated by AI developers' work, and the consequences of these frameworks, in the case of AI research around *autism*. Depending on who one asks, autism can be conceptualised in many different ways. Under a "medical" model of disability, which treats diagnostic criterion as representing the core truths of autistic minds and lives, autistic people (autists) are individuals suffering from a *disorder*, one characterized by stunted or absent social skills and emotional awareness. In contrast, figures aligned with the "neurodiversity" movement seek to portray autism and autists in a more positive light (Dyck & Russell, 2020). Floating between these two points are an array of other perspectives that seek to contextualise and historicise the construction of "autism" as a concept and the already-political nature of diagnostic procedures (Eyal, 2010; Timimi et al, 2019), trace the (often racialized and gendered) variations in autistic experiences of the world (Brown, Ashkenazy & Onaiwu, 2017), and articulate and critique the violence that both positive and negative stereotypes of autistic lives and minds produce (McGuire, 2016).

Discourses surrounding autism are a particularly apt site of inquiry into AI ethics' consideration of disability, for two reasons. First; a large body of work has demonstrated the particular power and relevance of popular conceptions of autism to autistic lives. Narratives of autism—largely authored by non-autistic family members, academics and other self-appointed experts—play an outsize role in defining the shape of autism as a concept, and the conventional approach that society takes in evaluating, interpreting and *governing* autistic lives and futures; as Ann McGuire writes, drawing on Ian Hacking, "contemporary stories of autism are…functioning to constitute what autism is and can be" (McGuire, 2016, p.11).

Second; these narratives—drawing on the medical model, and consequently the idea of autists as asocial and overly rational—resonate strongly with ideas of algorithms, automation and machinery. As demonstrated by Jordynn Jack, M. Remi Yergeau and other rhetoricians, there is a longstanding symbolic link between autists and machines. This not only takes the form of representations of autists as "computer geeks"—of, as Maji Nadesan puts it, "the idea that people with autism are technologically gifted and are particularly adept with computer technology" (Nadesan, 2013, p. 4)—but through representing autists as *computers*; as robotic, machine-like "others" (Waltz, 2008). Indeed, even academic scholarship and popular media inquiries *into* technology frequently deploy these popular frames of autism as a way of conceptualizing technical systems; Pinchevski and Peters describe purported "elective affinities between autism and new media" (Pinchevski & Peters, 2015), while a recent *Forbes* editorial is bold enough to parallel autistic people and self-driving cars (Razdan, 2020).

My inquiry focuses on discourses of autism in two AI spaces; researchers working on algorithms to automatically identify and diagnose autism, and the company *Daivergent*, which employs autistic people in labelling machine learning data. Through a critical analysis of papers, press releases and media coverage, I conclude that in both cases autists are portrayed as asocial, fundamentally lacking in the ability to know and understand, and consequently, lacking in agency and personhood.

Perpetuating and reinforcing these discourses has profound material consequences for autistic people inside and outside the spaces impacted by these technologies and the organisations developing and deploying them. More widely, however, the examination of disability and disempowerment highlights limitations in conventional AI ethics—limitations which go far further than "fairness" and include transparency, accountability and other principles which demand harmed populations be in communication with developers, even as AI systems and researchers actively undermine the legitimacy of some communicators. This calls not just for an avoidance of discourses which treat autistic people as inhuman, but a wider revisitation of the premise of communication and capacity that AI ethics is often dependent on.

Background

Conventional views of AI ethics

With the increasing development and deployment of Artificial Intelligence (AI), attention has turned to the question of "AI ethics:" the articulation of various approaches to the appropriate and "good" use of AI. It is important to avoid treating this as entirely novel: within the field of surveillance studies, for example, scholars have been investigating the rise of automated, algorithmic decision-making for decades (Gandy, 1993; Lyon, 1994; Norris & Armstrong, 1999). But the urgency of it—the widespread feeling that AI is "a significant emerging and future-shaping technological field that is developing at an accelerating rate" (Goode, 2018), and a corresponding rise in public, governmental, scholarly and corporate interest—has led to a particular flourishing of both applied and theoretical scholarship. The result has been myriad sets of principles, guidelines and policies around "good" AI, what it constitutes, and what is necessary to produce it (Whittlestone et al., 2019; Jobin et al., 2019).

As would be expected, the rapidly expanding nature of the field and the wide range of stakeholders means that these principles are yet to "stabilize:" theorists and practitioners frequently disagree over precisely what constitutes an ethical approach. But some components appear fairly consistently and frequently—in particular, notions of *fairness*,

accountability, and *transparency* (Stark & Hoffmann, 2019; Hagendorff, 2019; Floridi et al., 2018). Although each of these principles have been conceptualized and articulated in many different ways, (van Nuenen et al, 2020; Wieringa, 2020), a broad-strokes summary would be that *fairness* requires an avoidance of discrimination in making algorithmic decisions, *transparency* the disclosure of the rationale behind any such decision, and *accountability* a mechanism of addressing any harmful consequences or algorithmic failures.

Fairness has been a particularly frequent topic of discussion. Approaching fairness as a technical problem—does a system produce uneven outcomes for different demographic groups?—both academic and industry researchers have begun focusing on technical tools to identify and correct discriminatory systems, seeking to fix one algorithm with another (Bellamy et al., 2019; Spiecher et al., 2018). Interdisciplinary researchers have similarly attended to fairness, treating questions of bias as a primary component of an algorithmic system's moral valence (Buolamwini & Gebru, 2018; Chouldechova, 2017).

AI meets Disability

Although there is value in much of this scholarship, there are two gaps in it that are—for my work—particularly relevant. The first is in how fairness specifically (and ethical duties in general) are discussed; while gender and race are frequently deployed as protected characteristics to be scrutinized in evaluating algorithmic systems, disability is not. Instead, it is often left (at best) unmarked. Although a small number of works that substantively discuss the ways that algorithmic systems could discriminate against disabled people, a 2019 review of 1,659 AI ethics article abstracts found *eleven* containing disability-related keywords (Lillywhite & Wolbring, 2019). This is particularly concerning given the increasing interest in explicitly applying algorithmic systems to questions of disability.

There are signs this is beginning to change. A call by Meredith Ringel Morris on "AI and Accessibility" (Morris, 2019), in parallel with a dedicated workshop at the ACM SIGACCESS Conference on Computers and Accessibility (ASSETS) 2019 (Trewin, 2018), provided a rare centring of disability in discussions of AI ethics. Taken in concert with the World Institute on Disability's statement on AI (World Institute on Disability, 2019), and a report by AI Now directly addressing disability as a vital area of consideration (Whittaker et al., 2019), we can see an increasing (and much-needed) attentiveness to disability from AI ethicists. As would be expected, given the already-discussed trend towards fairness as a general value in AI ethics, this attentiveness frequently centres notions of bias and discrimination as their core concern. The ASSETS workshop, for example, was specifically titled "AI Fairness for People with Disabilities;" the World

Institute on Disability's statement repeatedly evoked questions of dataset bias and disproportionate outcomes as their primary concern.

The problem, however, is that this framing of ethics is anything but uncontested, as suggested by a paper *at* that workshop specifically contesting it (Bennett & Keyes, 2019). There are frequent critiques raised about fairness as a sole or primary ethical value for artificial intelligence, both generous and pointed. Some researchers are concerned by the *immediacy* of fairness: the way that fairness-based approaches to ethics typically evaluate the immediate outputs of an algorithm, while leaving the longer-term consequences unexamined (Selbst et al., 2019). Others point to the manifold definitions of fairness, and the vastly different material outcomes produced by each one (Hutchinson & Mitchell, 2019). Less optimistically, some critics highlight the treatment of "fairness" as a value that can, at least theoretically, be modelled, as anything but an accident. Instead they contend that the focus on computable ethical principles that do not address more structural and longitudinal outcomes is *precisely the point*, constituting "ethics-washing" that allows organisations to continue with "business as usual" (Wagner, 2018; Sloane, 2019).

Discrimination and Discourse

One particular issue—the second of the two gaps mentioned above—is the question of *discursive*, rather than *directly material* harms. As Hoffmann notes in her work on "Where Fairness Fails" (Hoffmann, 2019), a fairness-oriented frame, with its focus on materiality, "fails to appropriately attend to the legitimising, discursive or dignitary dimensions of data...algorithms do not merely shape distributive outcomes, but they are also intimately bound up in the production of particular kinds of meaning, reinforcing certain discursive frames over others" (Hoffmann, 2019, p.908). In other words, what algorithms do is not just a question of material goods and (direct) material harms, but a question of the discourses and narratives they depend on, perpetuate and legitimise. A large body of work has looked at the discourses surrounding artificial intelligence and AI ethics; Hoffmann's paper, along with the work of Peña Gangadharan & Niklas, Dencik *et al.* and several others (Peña Gangadharan & Niklas, 2019; Dencik *et al.*, 2018) asks us to look at the *consequences* of those discourses.

By "discourse," critical scholarship does not mean something as simple as statements, or rhetoric: rather, it means how statements fit into knowledge; how they shape and signify what can be known, through what methods, and through what actors (McHoul & Grace, 2015). To illustrate, we can examine Bivens & Hoque's "Programming sex, gender and sexuality" (Bivens & Hoque, 2018), which Hoffmann highlights as an example of critical discourse analyses in technological domains. Bivens & Hoque investigate the discourses deployed in and around "Bumble," a

dating app billed as embodying feminist values. Exploring public relations statements by the company that designs it and its figureheads, media coverage and aspects of the app's design, the researchers articulate how the "feminist" figure the app is designed for is specifically a middle-class, white, cisgender and heterosexual woman, with sometimes-dangerous consequences (including the possibility of assault) for those who fall outside that mold.

Bivens & Hoque's point is not just that the app is exclusionary to a vast range of people, but that this exclusion *generates meaning*: within the world of Bumble, to be feminist is to be a white, cisgender and heterosexual woman; to be male is to be a threat; to be a lesbian is to be non-existent. These frames, and the way that they resonate with wider cultural narratives, delegitimises particular populations. Bumble is not simply an app but a tool for meaning-making and knowledge generation— one that cannot, as designed, be positively applied to those outside a narrow norm.

Similarly, both the technologies and cultural imaginaries entangled with "AI" serve as a source of meaning and knowledge. As a consequence, we should attend not just to whether particular populations are excluded, or not, but the terms under which that happens: the justifications used, the framings they are subject to, and how this might reinforce or undermine damaging cultural frames regardless of what "the software" is *intended* to do. If applications of AI ethics to disability do not (or *cannot*) investigate this, then the model of ethics we are using may allow vast harms to go unnoticed by those with the structural power to address them. Autism is a particularly pertinent case study given that—as discussed earlier—it is strongly subject to and shaped by cultural narratives and assumptions. AI interventions around it, and any ethical silence on discursive harms, thus involve pressing on an already-skewed scale.

AI Interventions in Autism

To demonstrate the importance of inquiring into discursive framings within AI, and the harm such framings can cause, I undertake a critical discourse analysis of AI research publications and popular coverage that concern themselves with autism as a phenomenon, and autistic lives as a site of utility or intervention. This analysis concerns itself with how "dominant discourses (indirectly) influence…socially shared knowledge, attitudes and ideologies…[and] facilitate the formation of specific social representations" (van Dijk, 1993, pp. 258–9). In the case of autism, I centre questions of what social representations of autism (and autists) are (re)produced in the corpus, and whose voices are included or excluded from the process of shaping those representations. Such an approach has been undertaken in

other research on disability and technology, including Julie Elman's work on wearable technologies (Elman, 2018) and Spiel et al.'s inquiries into the experiences of autistic children using co-designed technologies (Spiel et al., 2017).

Sites of Analysis

The analysis focuses on two different projects—more accurately, one project, and one research subfield—both of which concern themselves with autism as a phenomenon, and autistic lives as a site of utility or intervention. My particular sites of inquiry—sites where autism and AI take form together—are (respectively) *Artificial Intelligence for Autism Diagnosis* (or *AIAD*) and *Daivergent.*

AI for Autism Diagnostics (AIAD) originates in the perception that current autism diagnostics are "expensive, subjective and time-consuming" (Jiang & Zao, 2017). By replacing existing mechanisms (which are centered on conversations between doctors and patients, and/or their friends and family), researchers hope to provide "efficient objective measures that can help in diagnosing this disease [sic] as early as possible with less effort" (Thapaliya et al., 2018).

Such replacements come in a range of different forms. Many papers use computer vision—machine learning systems that "see"—but this is sometimes to examine behavioural or social responses (Hashemi et al., 2018), sometimes to specifically evaluate eyeball movement (Jiang & Zao, 2017), or similarly, gait (Hasan et al., 2018), head movement (Bovery et al., 2019), or general upper-body form (Wedyan & Al-Jumaily, 2016). Outside of computer vision-based approaches, researchers have also looked at the analysis of conversational or voice recordings for auditory cues (Santos et al., 2013), the administration of screeners through a mobile app (featuring a predictive model that interprets the data) (Omar et al., 2019), or tracking the participant's attempts to assess emotions while observing videos (Uluyagmur-Ozturk et al., 2016). Many of the methodological approaches also involve a robotic component, with robots serving as the "unbiased" administrators or recorders of audio, video or written tests (Prescan et al., 2018; Petric et al., 2017).

To analyse work in AIAD, I constructed a corpus of 82 papers that investigated the use of machine learning systems for autism diagnosis. Drawing influence from Waidzunas & Epstein's investigation of the history of the plethysmograph (Waidzunas & Epstein, 2015), I followed the citation networks of papers that featured the terms ("autism" OR "autistic") AND "machine learning," incorporating into the corpus any papers that both cited a work in the initial "seed" dataset, and concerned themselves with autism diagnostic or screening tools. In and of themselves, these are narrow keywords; one might ask whether incorporating (for example)

"artificial intelligence," "neural network" or more precise machine learning terminology would produce different results for the initial seed papers. The answer is undoubtedly yes, but the reliance on citational networks rather than keywords alone goes some way towards mitigating this limitation. The resulting corpus is, while not comprehensive, fairly *cohesive*, with papers regularly citing not simply one other work within the corpus but many.

Daivergent (the name of which plays on AI and the idea of autistic people as deviant or other) originates with a very different perceived problem: the question of autistic people's unemployment. The company was founded by two data scientists, Bryon Dai and Rahul Mahida, both of whom have autistic relatives—a brother and a cousin, respectively—and funded by the venture capitalist Brian Jacobs, whose son is autistic (Galer, 2019; Levy, 2019). Concerned about their relatives' future after child-oriented disability services stopped being applicable, Dai and Mahida began Daivergent in 2017 to provide a bridge between autistic people and the technology industry.[1]

This bridge consists of, in parallel, offering autistic people jobs in classifying and "hand-coding" the input data for AI, and training in workplace norms and practices. To the founders, pairing autistic people with hand-coding takes advantage of what they see as the nature of autism: a "unique aptitude" for "intensively focused, complex, repetitive processes" (Galer, 2019). While most people get bored of such work, autists are seen as individuals who "can do it for the day, can do it for the week, can do it month after month" (Kadet, 2019). In exchange, they receive salaries of $15-20 an hour, and the opportunity to "gain a meaningful life" (Kung, 2019), with the founders pointing to ex-employees who have gone on to work as a clerk, in a payroll role, or "even in other places such as game design" (Galer, 2019).

Daivergent is hardly the only company seeking to market itself as rendering autists "productive" in the technology sector—other examples include Auticon, which describes itself as "[providing] a neurodiverse and agile workforce to improve our client's IT projects" (Auticon, 2020), and Aspiritech, "a world-class QA testing company that empowers individuals on the autism spectrum to fulfill their potential through meaningful employment combined with social opportunity" (Aspiritech, 2020). But Daivergent is (so far as I can determine) singular in positioning autists as a specialised workforce within *Artificial Intelligence*; as unique assets in developing AI systems and the datasets they depend on. Interestingly, and in juxtaposition with (or perhaps reflecting?) the needs around

---

[1] The issue of the employment and segmentation of disabled people has been widely problematized and historicised; see (for example) Parker, Owen & Gould, 2012; Friedner, 2015; and Simon, 1994

automated diagnostic systems, Daivergent's registration process does not require medical documentation—although they are silent on their expectations, it seems likely that self-diagnosed autists make up some percentage of their users.

Corpus contributions pertaining to Daivergent consisted of all media and marketing coverage of the company—both traditional venues (such as *The Wall Street Journal*) and non-traditional (the marketing blog of Amazon, whose software Daivergent uses)—that could be discovered through LexisNexis, along with the contents of Daivergent's website and marketing materials. This corpus spanned approximately two years, from the founding of the company in December 2017 until December 2019.

On the surface, these are very different: academic researchers versus a private company, AI usage versus AI design, attempts to label autism versus attempts to label *with* autism. But that is precisely the point: they are highly distinct environments, meaning that the strong alignment between the narrative representations of autism they deploy demonstrates that these representations are likely to appear far more frequently in AI than in these two cases alone. Similarly, while there are obvious differences in the types of source documents (publications versus news coverage), both constitute the most available material in which the actors represent themselves to their community and to the wider world. While they have different audiences, they are ultimately the same *kinds* of audience within the worlds that AI researchers and startup founders, respectively, occupy.

Analysing Discourses of Autism

After obtaining the source texts, I analysed and coded them following an approach based on Critical Discourse Analysis (CDA), which (as discussed above) focuses on the replication of dominant discursive frames, and the ways in which those frames constrain individuals and communities subject to them. In this case, my approach is one of "sociodiagnostic critique:" I seek not simply to analyse the texts for internal contradictions, but to situate them in the context of wider discourses and society, and my own background knowledge (Reisgl & Wodak, 2005).

To a certain degree, CDA is methodologically agnostic; there are few consistent approaches in how data should be collected and analysed (Meyer, 2001). My approach consisted of collecting the source texts, and then generating a set of themes through an inductive coding of the texts— looking particularly at how these texts described or discussed autistic people or autism, and how autistic people were positioned in relation to the works.

This approach generated a range of common themes. This included (1) framings of autism in a highly medicalized fashion—as something embodied, and something *wrong,* worth addressing and correcting—(2) framings of autism as an epidemic, and a concern of particular urgency, (3) positioning both AIAD and Daivergent as attempts to intervene in and normalise autistic existences (4) a strong neoliberal motivation behind those interventions, with the importance of addressing autism framed in terms of financial productivity, and (5) a general lack of explicit inquiry into race and gender, accompanied by *implicit* framings of autism as a fundamentally white and male phenomenon.

Each of these themes are important and deserving of investigation. Investigating all of them in a single paper, however, would produce rather shallow analyses. I highlight them here largely for methodological transparency, and to surface them as areas of future research for other scholarly works (and/or scholars). Instead, I would like to focus in depth on one particular theme that arose—the theme of autistic communication, and of knowledge. How do AIAD and Daivergent materials construct the ability of autists to know, and to communicate that knowledge?

Autism Discourses in AI

The terms "to communicate" or "to know" have a range of possible meanings and interactions. My understanding and use of those terms in this paper draw from feminist philosophers of knowledge, specifically the work of feminist epistemologists who (from the 1970s onwards) have consistently attended to questions of knowledge and communication. This consists not just of examining what constitutes knowledge, but "attention to what kind of subject one must be in order to be (seen as) a knowing subject" (Tuana, 2017, p. 126); attending to *who* can know. Under this frame (and many other approaches to questions of knowledge),[2] knowledge and communication are deeply bound up in each other. Knowledge must be communicated, recognised, and treated as credible in how it is evaluated (McConkey, 2004).

The social and reciprocal nature of knowledge and its construction is well-established in Science and Technology Studies; as Helen Longino summarises, "scientific inquiry is a collaborative human activity [and is] socially organized in certain ways that affect both goals and criteria of success" (Longino, 1990, p. 17). This is neatly demonstrated in the work of historian Ann Johnson, whose explorations of the social circumstances of engineering design processes treat technology as the result of socially-

---

[2] Wittgenstein, for example, noted that "knowledge is in the end based on acknowledgement" (Wittgenstein, 1969), and it is difficult to place him within feminist epistemology due to (amongst other things) his renowned misogyny.

mediated knowledge, and (in some respects) as an instantiation of such knowledge (Johnson, 2005; Johnson, 2009). But this relationship between knowledge and social recognition is not just a matter of professionalized research, or an abstract, theoretical concern; knowledge-making is deeply important to day-to-day activities and individuals' status in society. As Genevive Lloyd notes in her foundational work on feminist epistemology, in a society that conceives of itself around notions of rationality, the ability to know (and be seen to know) is deeply tied up with one's humanity (Lloyd, 2002). Put simply, someone who is not a recognisable knower is not a person, and vice versa. Both the ability to communicate and the ability to know thus have deep implications for personhood (Congdon, 2018).

Scholarship on epistemic personhood has frequently identified disability a factor in its extension—so, too, race and gender. Given this, and the long history of how race, gender and disability are interwoven in both discursive and material structures of oppression (Samuels, 2014), my analysis aims to consider not only the positioning of autists writ large but the role that gender and race play in the narratives of autism on display, and the consequences of those narratives.

Communication and Knowledge of Others
Questions of sociality are deeply implicated in both societal framings of autism, and framings of knowledge—and so it is no surprise to see it appear in the ideas of autism deployed by AI researchers and developers.

By *sociality* I mean the ability to appropriately and properly interact with others—something that has implications around both communication and knowledge. An absence of sociality is "often deemed to be a major feature of those diagnosed as being on the autism spectrum" (Milton, 2012): it is a core component of narratives within research (Verhoeff, 2012), current and defunct diagnostic criteria (O'Reilly et al., 2019), and public perception (Billawala & Wolbring, 2014). When it comes to communication, autists are sometimes framed as literally lacking the ability to communicate with others—many of us are nonverbal. For those of us who are verbal, our particular tropes are treated as inappropriate or invalid. These include echolalia (repeating the words of another), which is seen as containing no value (Roberts, 1989), and overly direct styles of communication, frequently treated as rude or disruptive. Verbal and non-verbal autists often engage in stimming—repetitive motions to ground one's sensory presence in an environment—which is often seen as disruptive to the *status quo* and a deviation from appropriate communication (Nolan & McBride, 2015).

There are alternative interpretations of each of these: stimming as a coping mechanism (Kapp et al., 2019), sociality as being constitutable in

autistic ways (Heasman & Gillespie, 2019), echolalia *as* a form of communication (de Jaegher, 2013)—but that is not the point. Autistic modes of communication are treated as less-valid, and less intelligible. Moreover, the explanation for unconventional communication is often one of *knowledge*: rather than simply being ignorant, autistic people are seen as not being able to understand what is appropriate communication, and incapable of understanding others.

In both AIAD literature and the materials released to frame Daivergent, we see the same stereotypes and tropes replicated. Both diagnostic AI researchers and Daivergent figureheads are unified in pointing to abnormal social behaviour and communication as an autistic attribute: "serious shortcomings in their social skills" (Irani et al., 2018). More specifically, "deficiency" in making eye contact (Uluyagmur-Ozturk et al., 2016), "serious problems with being creative" (Lund, 2009), "difficulties" in recognising the emotions of others, and "delay or perversion in language" (Altay & Ulas, 2018). It "makes ordinary social interactions particularly challenging" (Levy, 2019), explaining the high unemployment rate: as one set of researchers mournfully inform us, "about 50% of people with autism can never ... make useful speech" (Altay & Ulas, 2019).

Unsurprisingly, then, the interventions themselves build on and replicate these assumptions. In the case of AIAD, one way of framing the computer vision-oriented diagnostic tools—intended to replace, again, subjective interviewing—is to understand it as dependent on the presumption that diagnosis cannot rely on purposeful autistic communication. In the case of Daivergent, we see a repeated emphasis on the fact that the company provides not only jobs, but social skills opportunities: it emphasizes that "Daivergent stands out for the training it provides…not just technical skills but social and communication skills-training" (Prafder, 2019), and offers employees the ability to "Join any of our 15 shared interest groups to meet like-minded individuals that share your passions" (Daivergent, 2019a). Positioning itself as a provider of "unique social and communication training opportunities...alongside shared interest groups that help connect the autism community to one another," Daivergent operates from the implicit assumption that autistic community—of which there is much, including organisations serving/led by non-verbal autistic people (Yergeau, 2018)—does not exist; that autistic communication must be guided and shaped to be legitimate or capable of being recognised and understood (Daivergent, 2019b; Demo, 2017).

Knowledge of Self

So if autism is defined by a lack of communicative competence, and a lack of understanding others—where does that come from, and what are the implications? The dominant explanation within popular and academic ideas of autism is Baron-Cohen's model of "Theory of Mind:" the idea that

autistic people simply lack empathy and an understanding of others (Dinishak, 2016).[3] From this comes the associations autism has with a lack of empathy, bluntness, and difficulties communicating.

But there are other implications that stem from this as well; implications about autistic *knowledge of self*. "Empathy," in much theory and philosophy, is not something that just appears *de novo*: it is something learned, and premised on our own experiences. The analogy of a "simulation" is used; we model our idea of others on our own senses of self, and simulate how *we* would interpret the situation were we in "their shoes." A consequence of this is that normative theories of autistic minds do not just imply a lack of understanding of others, but that this stems from *a lack of understanding of self*. Jeanette Kennett, for example, uses "the highest-functioning autistic people" as an intellectual foil, positioning them as "[having) *some* capacity for introspection about their condition" (emphasis mine) and thus implying that the default state for autists is total ignorance of self (Kennett, 2002). Autists are framed as unreliable narrators of their internal state, incapable of knowing and representing their needs or desires, much less communicating them. Indeed, as noted by M. Remi Yergeau,[4] "clinical constructions of autism frequently position expertise and self-knowledge as antithetical to autism itself" (Yergeau, 2018, p.140).

In the absence of such knowledge, autistic people cannot be credible sources of information—not even information about ourselves. With Daivergent, it is notable that (with one exception, discussed later) no autistic people speak in their materials, press coverage, interviews or marketing reports. Instead, the idea of autism and the needs of autists are communicated by non-autistic people, pointing to the existence of autistic family members as a source of their expertise. Dai, for example, is depicted as having "first-hand experience with the challenge" by dint of having a *brother* who directly experiences autism (Galer, 2019), while Mahida states that "We both [have) family members with autism. We know the type of things they enjoy doing," generalising those "things" to autists as a whole, and speaking for autists as a population when he states emphatically that "They want to work in tech. They want to work doing things for AI" (Kung, 2019).

Within the diagnostic AI research, the bulk of users and perspectives center familial voices rather than autistic ones. In Thabtah's study, the app was designed for use by "a variety of stakeholders including parents,

---

[3] Simon Baron-Cohen is, interestingly, the cousin of *Sacha* Baron Cohen. The two could not be more different; one creates deeply bigoted and distressingly popular media content that describes and normalises the worst excesses of society's violence towards marginalized persons. The other wrote Borat.
[4] From whose book, *Authoring autism*, the title of this paper is lovingly purloined.

caregivers and more importantly health professionals", but never self-diagnosis (Thabtah, 2019); Irani et al.*'s* project adapted to feedback solicited from "the parents" (Irani et al., 2018); in Tariq et al.'s study, participation was determined by the parents—referred to throughout as the "participants"—despite the data covering autistic people up to the age of 17 (Tariq et al., 2018). Under the discourse of autism used, autistic people cannot consent or give feedback, not simply because they cannot communicate but because they have nothing *to* communicate.

Knowledge, Agency and Personhood

If autists are entities lacking in the ability to communicate and be social, and further, lacking the ability to have knowledge of self (much less knowledge of others): do autists have agency? Personhood? Are autists, really, human?

I raise this question because the answer that dominant frames of autism provide is "no." Indirect inhumanity is communicated through representations of autists as alien (Reddington & Price, 2016), robotic (Belek, 2017), or (in much of ethics, and in "autism advocacy") analogous to psychopathy: an interesting thought experiment in whether one can be a moral agent while quite so neurologically deviant (McGeer, 2008; Krahn & Fenton, 2009; Saunders, 2018). More directly, autism is treated as oppositional to the traits that "make" a person a person (Duffy & Dorner, 2011). Yergeau, while critiquing such ideas, summarises them with the statement that "humans are human because they possess a theory of mind, and autistics are inhuman because they do not" (Yergeau, 2013).

Portrayals of inhumanity in AIAD research and Daivergent's materials are largely indirect. But one telling illustration of the wider industry context and approaches comes from media coverage of the company, discussing broader efforts to employ autistic people and other people with "intellectual disabilities" (IDD):

> "At Salesforce.com, a customer relationship management (CRM) software company headquartered in San Francisco, 46 IDD workers are currently core to the firm's operations, says Benny Ebert-Zavos, manager of real estate communications for the organization. 'We hire them to organize and maintain conference rooms, assist with event setup, support our reusable dish program, stock pantries, upkeep our social lounges, stock office supplies and brew coffee,' he says. 'These folks are the key to making sure that when people come in, they can focus on work.'" (Prafder, 2019)

Notable is the distinction between "these folks" and people; the distinction between their labour and "work."

Similarly, we can see instances of how the treatment of autists as inhuman correspondingly portray autists as "divorced from concepts of

agency and autonomy" (Quirici, 2015). This is hardly surprising: Korsgaard, in reviewing the nature of agency, describes conventional conceptions of agency as consisting of the ideas "that the capacity for agency consists in or depends on the existence of certain normative relations, and that the realization of that capacity—success in action— depends on conformity to the norms in question" (Korsgaard, 2014). In the case of autists, we can see the construction of autistic relating as abnormal, and of autists as incapable *of* adhering to normative relations, as precluding the possibility of autistic agency—at least, absent confirmation from normative actors.

These ideas of autists as non-agentic feature heavily, albeit indirectly, in Daivergent's literature—as a positive. Consider the rationales provided for hiring autistic people in particular; their dedication to engaging in the same tasks "month after month" (Kadet, 2019), a status that resonates more strongly with metaphors of machines than of people. But that is not all: companies should hire autistic people because they have "perseverance" and a "sense of loyalty;" because they are not going to *leave* (Levy, 2019). To be autistic, after all, is to neither know what you want nor how to communicate it, and so the concept of autonomous autists is as alien to a normative view of autism as autists allegedly are to themselves.

Non-agentism and inhumanity also feature, albeit more implicitly, in much of the AIAD literature. As well as discussing communication, the literature also discusses violence and risk. An autistic person has "a very high risk for wandering; he can become very dangerous for himself, his family and the society as he can harm others as well as himself in an aggression" (Omar et al., 2019). Autism impacts "self-control and [the] person's ability to learn" (Pahwa et al., 2016). Stimming and other "stereotypical motor movements can lead to self-injurious behaviour under certain environmental conditions" (Albinali et al., 2009). In all of these framings, autists appear as figures who are—as a consequence of this dearth of outer awareness and communication—fundamentally *lacking*; lacking control over self, lacking the ability to engage in inference, lacking, in other words, in agency and the ability to choose. An autist is not a person—an autist is a machine, one whose misfiring outputs betray faults in their wiring.

Discussion

In my analysis above, I examined the discursive framing of autistic communication and self that is deployed by Artificial Intelligence for Autism Diagnosis (AIAD) research, and the autist-employing startup "Daivergent." In doing so I suggested that in both cases, work follows a normative approach in describing autists as lacking in communication,

sociality and sense of others—and further, lacking in sense of self. This approach results in turn in a perception of autists as unpersons; as inhuman, and as lacking in agency and autonomy. Below, I discuss the material and conceptual implications of AI research perpetuating and internalising this logic.

The consequences of conventional frames

Discourses are an important part of the "output" of sociotechnical systems and narratives—and a part inextricably wound up in how those discourses help shape the structure of society. Simply because something is discursive does not mean it does not have material consequences. So what are the implications of AI development perpetuating this conventional framing of autism? What consequences does this have for autistic people? Answering this requires us to take a step back and look at the existing consequences of these narratives. If we look at how autists are already treated in other sites, as a result of the conventional framing of autism, we see some profound and disturbing phenomena. Some of it is interactional, and day to day: the treatment of autistic sociality and communication as invalid and less-than creates heightened feelings of stigma and "negative difference and feeling lesser," leading to the "exhausting" work of hiding one's otherness, simulating normativity, for the fear of ostracisation should one be detected (Hodge et al., 2019).

Other material consequences of conventional framings—and of their reinforcement—are far more tangible, and far more clearly violent. As a result of autistic communication being seen as an oxymoron, approaches to repairing communication failures between autists and non-autists are ones of *normalisation*: "fixing" autistic people, rather than attempting to meet autistic people in the middle. Such "repair" is frequently violent, featuring—in the case of Applied Behavioural Analysis (ABA), the standard "treatment" for autists—training centered on "aversives:" responding to autists stimming, communicating non-normatively or "acting out" through withdrawing access to food, social interaction or touch. Children may be subject to aversives "in the forms of time-outs (often in closets, cells or segregated rooms), Tabasco sauce on one's tongue, spray bottles filled with vinegar, forced proximity to a cold or hot surface, physical restraint, screams directed at the child, and so on"—all for something as simple as refusing to touch one's nose (Yergeau, 2018, p.97). The most extreme form of this (or the logical conclusion of it, depending on one's level of cynicism) can be seen in the form of the Judge Rothenberg Center, located in Massachusetts, which uses "aversives" such as straightjacketing, electrocution to the point of third-degree burns, and the inhalation of ammonia (Adams & Erevelles, 2017).

An immediate reaction to this is one of horror; what monstrosity! What inhumanity! But "inhumanity" is the point; of *course* these are the

therapies, of *course* the center has not been shut down: those subject to these tortures are not *people*. They cannot consent, in the sense that they cannot say "no;" what they say, even about the treatment of their own body, cannot be taken seriously at all. As Adams & Erevelles point out, it is not that (in the ensuing lawsuits) autistic people did not testify as to their experience and assaults—it is that their voices were not taken as communicating valid knowledge compared to the (professional, well-adjusted, and credible) doctors (Adams & Erevelles, 2017).

As this paper demonstrates, the attitude taken by Daivergent and AIAD researchers neatly fits into the conventional understanding of autism. Correspondingly, it reinforces them—in a small way, to be sure, but in a way that is still worthy of ethical and political notice. AI does not exist outside of the world, and so narratives within AI that further cement violent discourses have an impact on the broader domains—employment, and healthcare, in this case—in which autistic people are subject to this ongoing violence.

More directly, one might ask what the construction of autism taken by these researchers and founders means for those *directly* interacting with them. What about AIAD patients, or Daivergent employees? If autists cannot validly know, whose perspective is foregrounded in the event that an autist disputes the outcome of a diagnostic algorithm? Whose perspective is foregrounded in the event that an autist disputes the morality of this algorithmic work in general? When autistic employment is oriented around assumptions of roboticism and machinic lack of self, what happens when autistic employees *display* autonomy? It is hard to imagine an AI company that sees autists as asocial or non-agentic as taking seriously, for example, attempts to unionise: a union of autists would be a contradiction in terms.

The consequences of these framings are hardly evenly distributed; although race and gender are seldom, if ever, *substantively* raised in the sources (hence the lack of discussion in my findings), this is distinct from stating that the framings are not gendered and racialized. Mainstream conceptions of autism have always been both; as a few (but by no means exhaustive) set of examples, I would point to the way that Baron-Cohen's "Theory of Mind" positions autism as constituting an "extreme male brain," the broadening of autism diagnoses and criterion's origins, in part, in a desire by white parents to differentiate their children from the (predominantly black) people diagnosed with "childhood schizophrenia" (Eyal *et al.*, 2010), and the portrayal of Black autists as violent, and in need of controlling (Erevelles, 2014). So, too, is epistemic personhood gendered and racialized; there is a reason much of the work on knowledge and injustice originates in feminist and Black philosophy (Mills, 2007; Ortega, 2006; Code, 2014). As a result, it would be unsurprising—

indeed, it seems somewhat inevitable—to see the narratives of dehumanization I highlight, and their consequences, deployed particularly powerfully against gender and/or racial minorities.

The consequences of conventional AI ethics

Beyond the material consequences, however, these questions of autonomy, humanity and their discursive construction raise wider questions about the nature of AI ethics itself. They raise questions not simply about whether AI systems can be developed in a way that is just to autistic people, but whether AI ethics can reliably conceive of exclusionary injustices as a problem.

Understanding what I mean by that requires that we take a step back. When I say "AI ethics" I mean something very particular; the normative, default set of values which (as discussed above) seem widely agreed upon as, if not an entire system for achieving justice, then at least a good starting point. Fairness, true, but also accountability and transparency. These principles are fundamentally premised on a recognition of injured parties as people, and as social creatures, and as viable knowers. Accountability requires being answerable to people and in conversation with them; transparency requires a relationship of shared exchange and reasoning; fairness requires a form of recognition and acknowledgement of a population as distinct and worthy of equality as "people." In other words, *AI practitioners are discursively framing certain populations as non-human and non-agentic in parallel with ethical frameworks that depend on humanity and agency for addressing harms.*

In this case study we have seen how discourses of autists as asocial and non-agentic produce material harms—but it seems to me we should also ask what flaws they highlight in AI ethics frameworks for addressing those selfsame harms. If autistic people are being constructed by AI practitioners as incapable of agency and full humanity, in an ethical framework that treats agency and full humanity as mutual dependencies of each other, and both as a necessary prerequisite for participating in the frameworks to address injustices, then we have an impasse. If our approach to ethics is simultaneously (1) framed around notions of communication, credibility and recognition and (2) framing autistic people as lacking in those things, there is no viable way for autists to participate in processes that are frequently treated as the panacea to any injustice this domain generates. Autists will be subject to both discursive and material violence, and the discursive violence will strip us of the ability to viably dispute either.

Now, one immediate solution to this might appear to be to move the goalposts—to declare that discursively framing autists as less-than-human is wrong. This would certainly help, although the issue is far more

widespread than one of discourses within AI ethics. Maintaining rigid boundaries around who counts as a person, and as a knower, is nearly ubiquitous in normative philosophy more generally. Liberal philosophy often treats people as "fully rational, mutually-independent decisionmakers" (Schwartzman, 2006); individuals who do not meet these conditions are denied access to decision-making processes and modes of political or ethical engagement. Such an approach is frequently criticised, for the same reasons that I raise concerns with AI ethics. In particular, feminist philosophers and philosophers of disability have taken issue with the way that the idea of rationality and independent agency as a basis for personhood risks silencing, harming and literally dehumanising those who (for whatever reason) cannot make a claim to such status. Moreover, because the resulting ethical frameworks *assume* such status, they are frequently "strangely silent about the predicaments of outsiders" (O'Neill, 2000, p. 4)—as Lauren Davy notes in reviewing the work of John Rawls, disability is "relegated ... to a footnote ... a problem to be worked out later when all other matters of justice are settled" (Davy, 2019, p. 105).

Critiques of these approaches, and the uneven distribution of what counts as rationality and interdependence, provide a set of ideas to ameliorate the harms that result—ideas that we may be able to use in AI ethics. The work of José Medina and Miranda Fricker, in particular, highlights the need for us to engage in work that includes not only openness and self-criticality in how we interpret people and perspectives on an individual basis, but the construction of forms of "hermeneutical resistance:" ways of knowing and communicating that actively push back against monolithic ideas of personhood and knowledge (Medina, 2013; Medina, 2017; Fricker, 2007). In the case of AI ethics, this might look like actively pushing back against proposals for monolithic conceptions of justice, or mechanisms for achieving it, while developing more polyphonic and adaptive approaches.

More broadly, we might consider different ways of conceptualizing personhood altogether. A feminist "care ethics" approach to personhood and knowledge that treats not just disabled people but all people as dependent on communities, infrastructure and relationships: that treats personhood as relational and wrapped up in our relationships of care to each other (Davy, 2019). To go further we might examine the "posthumanist ethics" of Karen Barad and others, which (as adroitly explained by Natasha Mauthner) "seeks to conceptualize ontological, epistemological, and ethical agency without recourse to the human subject" (Mauthner, 2019, pp. 680–681). Paying critical attention to how we conceptualise personhood and knowledge, and what we make those concepts depend on (and dependents of) would not simply allow for more disability-aware approaches to AI ethics, but additionally address concerns around race, gender and sexuality; Medina's scholarship,

discussed above, is rooted in part in the recognition that knowledge—and humanity—are additionally (some would say, fundamentally) racialised and gendered.

Still, I am cautious and cognizant that these suggestions are ultimately efforts to *ameliorate* dehumanization in the structure of mechanisms for correcting injustice. As demonstrated by the work of agonistic theorists in political philosophy, there is no singular approach that will "solve" the question of otherness and silencing (Honig, 2016; Mouffe, 2000). Regardless of where we draw the line with regards to personhood, knowledge and access to justice, we are drawing a line—marking some as legitimate and some as not. Pragmatically, my intention with this paper is to highlight a substantial loophole in conventional framings of AI ethics— the loophole of *personhood*—and the need to address it. I do not believe we can escape silencing and perpetuating injustices altogether. But what we can do is confront it in how we theorise about justice, and mechanisms to achieve it. We can understand harm as an inescapable consequence of efforts to reduce it, and understand those efforts as ultimately contingent, and open to challenge. Most broadly, then, my demand is not simply for a consideration of discursive harm, and disabled voices, in the development of AI ethics, but a more wide-ranging demand that we avoid the fatal mistake of treating any mechanism, or set of principles, as settled.

Conclusion

I have analysed the approach that both public and private-sector entities seeking to "help" autistic people with Artificial Intelligence take to defining and framing autism, and autists. Doing so has revealed that these entities, their research and their public materials push a vision of autists as asocial, unknowing and somewhat non-human creatures, lacking in agency and autonomy.

This validates concerns that AI ethics is failing to attend to the consequences of AI discourses, and failing to attend to disability. But it also raises tensions and issues with the very premise of conventional AI ethics, which is often dependent on practices of communication and recognition, and so (tacitly) requires the treatment of individuals as "people" before their concerns can be heard. If autistic communication is invalid—if autists are invalid—and resolving harms is a communicative practice conducted between valid people, then harms to autistic people simply cannot be resolved under such a frame.

One approach would be to make sure that we simply treat autistic people as people—but keeping a dependency on and bar to personhood still leaves some individuals dehumanised and unable to access our

frameworks for addressing harms, even if those individuals are not autistic. Instead, I advocate that critical attention be paid not only to the immediate barriers to accessing justice, but the status we give to "personhood" in the first place. This includes, as discussed, a greater attentiveness to the conditions under which we evaluate knowledge and communication, and efforts to reshape our idea of the ethical agent (and so victim of unethical behaviour) to be relational, rather than hyper-individualised. I encourage researchers and practitioners concerned about disability justice specifically, or weaknesses in our ethical frameworks more generally, to consider these possibilities.

My intention here is not to demand some particular universal ethic to replace the current one; I am unsure whether *any* universal approach can resolve these issues, rather than replicate them in new forms. Instead, my goal is simply to encourage an urgent recognition of how violence too-often depends on our willingness to treat the terms of "humanity" and "personhood" uncritically, and accept them as a prerequisite for ethical attention. To this end, I want to underscore how vital it is that we retain and reinforce that critical lens—that we avoid treating any term of art or scheme of justice as unquestionable and "settled law." Who can play the game is a vital question, but so, too, is whether the dice are loaded. Working towards justice requires us to continually ask the second, as well as the first.

References

Adams, D. & Erevelles, N. (2017). Unexpected spaces of confinement: Aversive technologies, intellectual disability, and "bare life." *Punishment & Society, 19*(3), 348–365.

Albinali, F., Goodwin, M. S., & Intille, S. S. (2009, September). Recognizing stereotypical motor movements in the laboratory and classroom: a case study with children on the autism spectrum. In *Proceedings of the 11th international conference on Ubiquitous computing* (pp. 71–80).

Aspiritech (2020) *About.* Aspiritech https://web.archive.org/web/20200704214524/https://www.aspiritech.org/about

Auticon (2020) *About Auticon.* Auticon https://web.archive.org/web/20200420221854/https://auticon.us/company/

Altay, O., & Ulas, M. (2018, March). Prediction of the autism spectrum disorder diagnosis with linear discriminant analysis classifier and K-

nearest neighbor in children. In *2018 6th International Symposium on Digital Forensic and Security (ISDFS)* (pp. 1–4). IEEE.

Belek, B. (2017). I feel, therefore I matter: Emotional rhetoric and autism self-advocacy. *Anthropology Now*, 9(2), 57–69.

Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Nagar, S. (2019). AI fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, *63*(4/5), 4–1.

Bennett, C. L., & Keyes, O. (2019). What Is the point of fairness? Disability, AI and the complexity of justice. In *ASSETS 2019 Workshop—AI Fairness for People with Disabilities*.

Billawala, A., & Wolbring, G. (2014). Analyzing the discourse surrounding autism in the New York Times using an ableism lens. *Disability Studies Quarterly*, *34*(1).

Bivens, R., & Hoque, A. S. (2018). Programming sex, gender, and sexuality: Infrastructural failures in the" feminist" dating app Bumble. *Canadian Journal of Communication*, *43*(3), 441–459.

Bovery, M. D. M. J., Dawson, G., Hashemi, J., & Sapiro, G. (2019). A scalable off-the-shelf framework for measuring patterns of attention in young children and its application in autism spectrum disorder. *IEEE Transactions on Affective Computing*.

Brown, L. X., Ashkenazy, E., & Onaiwu, M. G. (Eds.). (2017). *All the weight of our dreams: On living racialized autism*. DragonBee Press.

Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency* (pp. 77–91).

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big data*, *5*(2), 153–163.

Code, L. (2014). Ignorance, injustice and the politics of knowledge: Feminist epistemology now. *Australian Feminist Studies*, *29*(80), 148–160.

Congdon, M. (2018). "Knower" as an ethical concept: From epistemic agency to mutual recognition. *Feminist Philosophy Quarterly*, *4*(4).

Daivergent (2019a) *Interested candidates*. Daivergent. https://web.archive.org/web/20191114170952/https:/daivergent.com/work-readiness/candidates

Daivergent (2019b) *Interested organizations*. Daivergent. https://web.archive.org/web/20191114171242/https:/daivergent.com/work-readiness/organizations

Davy, L. (2019). Between an ethic of care and an ethic of autonomy: Negotiating relational autonomy, disability, and dependency. *Angelaki*, *24*(3), 101–114.

De Jaegher, H. (2013). Embodiment and sense-making in autism. *Frontiers in Integrative Neuroscience*, *7*, 15.

Demo, A. T. (2017). Hacking agency: Apps, autism, and neurodiversity. *Quarterly Journal of Speech*, *103*(3), 277–300.

Dinishak, J. (2016). The deficit view and its critics. *Disability Studies Quarterly*, *36*(4).

Duffy, J., & Dorner, R. (2011). The pathos of "mindblindness:" Autism, science, and sadness in "theory of mind" narratives. *Journal of Literary & Cultural Disability Studies*, *5*(2), 201–215.

Dyck, E., & Russell, G. (2020). Challenging psychiatric classification: Healthy autistic diversity and the neurodiversity movement. In *Healthy minds in the twentieth century* (pp. 167–187). Palgrave Macmillan.

Erevelles, N. (2014). Crippin' Jim Crow: Disability, dis-location, and the school-to-prison pipeline. In *Disability incarcerated* (pp. 81–99). Palgrave Macmillan.

Eyal, G., Hart, B., Onculer, E., Oren, N., & Rossi, N. (2010). *The autism matrix*. Polity.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines*, *28*(4), 689–707.

Friedner, M. (2015). Deaf bodies and corporate bodies: New regimes of value in Bangalore's business process outsourcing sector. *Journal of the Royal Anthropological Institute*, *21*(2), 313–329.

Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing.* Oxford University Press.

Galer, S. (2019, March 28). Mother's wish spurs childhood friends to pioneer autism talent pool startup. *Forbes.* https://www.forbes.com/sites/sap/2019/03/28/mothers-wish-spurs-childhood-friends-to-pioneer-autism-talent-pool-startup

Gandy, O. H. (1993). *The panoptic sort: A political economy of personal information.* Westview Press.

Goode, L. (2018). Life, but not as we know it: AI and the popular imagination. *Culture Unbound, 10*(2), 185–207.

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. *Minds and Machines, 30*, 99–120.

Hasan, C. Z. C., Jailani, R., & Tahir, N. M. (2018, October). ANN and SVM Classifiers in Identifying Autism Spectrum Disorder Gait Based on Three-Dimensional Ground Reaction Forces. In *TENCON 2018-2018 IEEE Region 10 Conference* (pp. 2436–2440). IEEE.

Hashemi, J., Dawson, G., Carpenter, K. L., Campbell, K., Qiu, Q., Espinosa, S., ... & Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing.*

Heasman, B., & Gillespie, A. (2019). Neurodivergent intersubjectivity: Distinctive features of how autistic people create shared understanding. *Autism, 23*(4), 910–921.

Hodge, N., Rice, E. J., & Reidy, L. (2019). 'They're told all the time they're different': how educators understand development of sense of self for autistic pupils. *Disability & Society, 34*(9–10), 1353–1378.

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

Honig, B. (1993). *Political theory and the displacement of politics.* Cornell University Press.

Hutchinson, B., & Mitchell, M. (2019, January). 50 years of test (un)fairness: Lessons for machine learning. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 49–58).

Irani, A., Moradi, H., & Vahid, L. K. (2018, November). Autism screening using a video game based on emotions. In *2018 2nd National and 1st International Digital Games Research Conference: Trends, Technologies, and Applications (DGRC)* (pp. 40–45). IEEE.

Jaiswal, S., Valstar, M. F., Gillott, A., & Daley, D. (2017, May). Automatic detection of ADHD and ASD from expressive behaviour in RGBD data. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)* (pp. 762–769). IEEE.

Jiang, M., & Zhao, Q. (2017). Learning visual attention to identify people with autism spectrum disorder. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3267–3276).

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399.

Johnson, A. (2005). Revisiting technology as knowledge. *Perspectives on Science, 13*(4), 554–573.

Johnson, A. (2009). *Hitting the brakes: Engineering design and the production of knowledge*. Duke University Press.

Kadet, A. (2019). *Startup touts unique talent pool: Workers with autism*. Wall Street Journal. https://www.wsj.com/articles/startup-touts-unique-talent-pool-workers-with-autism-1536069600

Kapp, S. K., Steward, R., Crane, L., Elliott, D., Elphick, C., Pellicano, E., & Russell, G. (2019). 'People should be allowed to do what they like': Autistic adults' views and experiences of stimming. *Autism*, *23*(7), 1782–1792.

Kennett, J. (2002). Autism, empathy and moral agency. *The Philosophical Quarterly, 52*(208), 340–357.

Korsgaard, C. M. (2014). The normative constitution of agency. In *Rational and social agency: The philosophy of Michael Bratman*, 190–214.

Krahn, T., & Fenton, A. (2009). Autism, empathy and questions of moral agency. *Journal for the Theory of Social Behaviour*, *39*(2), 145–166.

Kung, M. (2019). *How Daivergent is providing autistic people with a stepping-stone to a career in tech*. AWS Startups Blog. https://aws.amazon.com/blogs/startups/how-daivergent-provides-career-help-to-autistic-people/

Levy, A. (2019). Why this investor left the pursuit of billion-dollar exits to help employers hire people with autism. *CNBC*. https://www.cnbc.com/2019/08/25/brian-jacobs-started-moai-capital-to-invest-in-autism-employment.html

Lillywhite, A., & Wolbring, G. (2019). Coverage of ethics within the artificial intelligence and machine learning academic literature: The case of disabled people. *Assistive Technology*, 1–7.

Lloyd, G. (2002). *The man of reason: "Male" and "female" in Western philosophy*. Routledge.

Lund, H. H. "Modular playware as a playful diagnosis tool for autistic children." *2009 IEEE International Conference on Rehabilitation Robotics*. IEEE, 2009.

Longino, H. E. (1990). *Science as social knowledge: Values and objectivity in scientific inquiry*. Princeton University Press.

Lyon, D. (2013). *The electronic eye: The rise of surveillance society.* John Wiley & Sons.

Mauthner, N. S. (2019). Toward a posthumanist ethics of qualitative research in a Big Data era. *American Behavioral Scientist*, *63*(6), 669–698.

McConkey, J. (2004). Knowledge and acknowledgement: 'epistemic injustice' as a problem of recognition. *Politics*, *24*(3), 198–205.

McGeer, V. (2008). Varieties of moral agency: Lessons from autism (and psychopathy). *Moral psychology*, *3*, 227–57.

McGuire, A. (2016). *War on autism: On the cultural logic of normative violence*. University of Michigan Press.

McHoul, A., & Grace, W. (1997). *A Foucault primer: Discourse, power, and the subject*. NYU Press.

Medina, J. (2013). *The epistemology of resistance: Gender and racial oppression, epistemic injustice, and the social imagination*. Oxford University Press.

Medina, J. (2017). Varieties of hermeneutical injustice. In I.J. Kidd, J Medina, G. Pohlhaus, Jr. (Eds.), *The Routledge handbook of epistemic injustice* (pp. 41-52). Taylor and Francis.

Meyer, M. (2001). Between theory, method, and politics: positioning of the approaches to CDA. *Methods of critical discourse analysis*, *1431*.

Mills, C. (2007). White ignorance. In S. Sullivan & N. Tuana (Eds.), *Race and epistemologies of ignorance*, 26–31. SUNY Press

Milton, D. (2012). On the ontological status of autism: The 'double empathy problem'." *Disability & Society, 27*(6), 883–887.

Morris, M. R. (2019). AI and accessibility: A discussion of ethical considerations. *arXiv preprint arXiv:1908.08939*.

Mouffe, C. (2000). For an agonistic model of democracy. In *The democratic paradox*. Verso.

Nadesan, M. H. (2013). *Constructing autism: Unravelling the 'truth' and understanding the social*. Routledge.

Nolan, J., & McBride, M. (2015). Embodied semiosis: Autistic 'stimming' as sensory praxis. In P. P. Trifonas (Ed.), *International handbook of semiotics* (pp. 1069–1078). Springer, Dordrecht.

Norris, C., & Armstrong, G. (1999). *The maximum surveillance society: The rise of CCTV* (Vol. 2). Oxford.

Omar, K. S., Mondal, P., Khan, N. S., Rizvi, M. R. K., & Islam, M. N. (2019, February). A machine learning approach to predict autism spectrum disorder. In *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1–6). IEEE.

Ortega, M. (2006). Being lovingly, knowingly ignorant: White feminism and women of color. *Hypatia*, *21*(3), 56–74.

O'neill, O. (2000). *Bounds of justice*. Cambridge University Press.

O'Reilly, M., Lester, J. N., & Kiyimba, N. (2020). Autism in the twentieth century: An evolution of a controversial condition. In S. Taylor & A. Brumby (Eds.), *Healthy minds in the twentieth century* (pp. 137-165). Palgrave Macmillan.

Pahwa, A., Aggarwal, G., & Sharma, A. (2016, April). A machine learning approach for identification & diagnosing features of Neurodevelopmental disorders using speech and spoken sentences. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 377–382). IEEE.

Parker Harris, S., Owen, R, & Gould, R. (2012). Parity of participation in liberal welfare states: Human rights, neoliberalism, disability and employment. *Disability & Society, 27*(6), 823–836.

Peña Gangadharan, S., & Niklas, J. (2019). Decentering technology in discourse on discrimination. *Information, Communication & Society*, *22*(7), 882–899.

Petric, F., Miklić, D., Cepanec, M., Cvitanović, P., & Kovačić, Z. (2017, August). Functional imitation task in the context of robot-assisted autism spectrum disorder diagnostics: Preliminary investigations. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1471–1478). IEEE.

Prafder, E. (2019). There are some jobs where candidates on the autism spectrum excel. *New York Post*. https://nypost:com/2019/04/07/there-are-some-jobs-where-candidates-on-the-autism-spectrum-excel/

Presecan, M., Petric, F., & Kovacic, Z. (2018, September). Object Classification for Child Behavior Observation in the Context of Autism Diagnostics Using a Deep Learning-based Approach. In *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (pp. 1–6). IEEE.

Quirici, M. (2015). Geniuses without imagination. *Journal of Literary & Cultural Disability Studies*, *9*(1), 71–89.

Reddington, S., & Price, D. (2016). Cyborg and autism: Exploring new social articulations via posthuman connections. *International Journal of Qualitative Studies in Education*, *29*(7), 882–892.

Reisgl, M. & Wodak, R. (2005). *Discourse and discrimination: Rhetorics of racism and antisemitism*. Routledge.

Roberts, J. (1989). Echolalia and comprehension in autistic children. *Journal of Autism and Developmental Disorders, 19*(2), 271–281.

Samuels, E. (2014). *Fantasies of identification: Disability, gender, race* (Vol. 10). NYU Press.

Santos, J. F., Brosh, N., Falk, T. H., Zwaigenbaum, L., Bryson, S. E., Roberts, W., ... & Brian, J. A. (2013, May). Very early detection of autism spectrum disorders based on acoustic analysis of pre-verbal

vocalizations of 18-month old toddlers. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 7567–7571). IEEE.

Saunders, P. (2018). Neurodivergent rhetorics: Examining competing discourses of autism advocacy in the public sphere. *Journal of Literary & Cultural Disability Studies*, *12*(1), 1–17.

Schwartzman, L. H. (2006). *Challenging liberalism: Feminism as political critique*. Penn State Press.

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019, January). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68).

Simon, B. L. (1994). *The empowerment tradition in American social work: A history*. Columbia University Press.

Sloane, M. (2019). Inequality is the name of the game: Thoughts on the emerging field of technology, ethics and social justice. In *Weizenbaum Conference* (p. 9). DEU.

Spiel, K., Frauenberger, C., & Fitzpatrick, G. (2017). Experiences of autistic children with technologies. *International Journal of Child-Computer Interaction*, *11*, 50–61.

Stark, L. & Hoffmann, A. L. (2019). Data is the new what?: Popular metaphors & professional ethics in emerging data culture. *Journal of Cultural Analytics, 1*(1)*.

Tariq, Q., Daniels, J., Schwartz, J. N., Washington, P., Kalantarian, H., & Wall, D. P. (2018). Mobile detection of autism through machine learning on home video: A development and prospective validation study. *PLoS medicine*, *15*(11), e1002705.

Thabtah, F. (2019). An accessible and efficient autism screening method for behavioural data and predictive analyses. *Health Informatics Journal*, *25*(4), 1739–1755.

Thapaliya, S., Jayarathna, S., & Jaime, M. (2018, December). Evaluating the EEG and eye movements for autism spectrum disorder. In *2018 IEEE international conference on big data (Big Data)* (pp. 2328–2336). IEEE.

Timimi, S., Milton, D., Bovell, V., Kapp, S., & Russell, G. (2019). Deconstructing diagnosis: Four commentaries on a diagnostic tool to assess individuals for autism spectrum disorders. *Autonomy, 1*(6).

Trewin, S. (2018). AI fairness for people with disabilities: Point of view. *arXiv preprint arXiv:1811.10670*.

Tuana N (2017) Feminist epistemology: The subject of knowledge In *The Routledge Handbook of Epistemic Injustice*. Routledge, pp. 125–138.

Uluyagmur-Ozturk, M., Arman, A. R., Yilmaz, S. S., Findik, O. T. P., Genc, H. A., Carkaxhiu-Bulut, G., ... & Cataltepe, Z. (2016, December). ADHD and ASD classification based on emotion recognition data. In *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 810–813). IEEE.

Verhoeff, B. (2012). What is this thing called autism? A critical analysis of the tenacious search for autism's essence. *BioSocieties*, *7*(4), 410–432.

Wagner, B. (2018). Ethics as an escape from regulation: From ethics-washing to ethics-shopping. *Being profiling. Cogitas ergo sum*, 1–7.

Waidzunas, T., & Epstein, S. (2015). 'For men arousal is orientation': Bodily truthing, technosexual scripts, and the materialization of sexualities through the phallometric test. *Social studies of science*, *45*(2), 187–213.

Waltz, M. (2008). Autism=death: The social and medical impact of a catastrophic medical model of autistic spectrum disorders. *Popular Narrative Media*, *1*(1), 13–23.

Wedyan, M., & Al-Jumaily, A. (2016, December). Upper limb motor coordination based early diagnosis in high risk subjects for Autism. In *2016 IEEE symposium series on computational intelligence (SSCI)* (pp. 1–8). IEEE.

Whittaker, M., Alper, M., Bennett, C. L., Hendren, S., Kaziunas, L., Mills, M., ... & West, S. M. (2019). Disability, bias, and AI. *AI Now Institute*.

Whittlestone, J., Nyrup, R., Alexandrova, A., & Cave, S. (2019, January). The role and limits of principles in AI ethics: towards a focus on

tensions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 195–200).

Wieringa, M. (2020, January). What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1–18).

Wittgenstein L. (1969) *On certainty*. Blackwell, Oxford.

World Institute on Disability (2019) *AI and accessibility*. World Institute on Disability. https://wid.org/2019/06/12/ai-and-accessibility/

Yergeau M. R. (2013). Clinically significant disturbance: On theorists who theorize theory of mind. *Disability Studies Quarterly, 33*(4).

Yergeau M. R. (2018). *Authoring autism: On rhetoric and neurological queerness*. Duke University Press.