Disability Rights Education & Defense Fund

DREDF

Addressing Disability & Ableist Bias in Autonomous Vehicles: Ensuring Safety, Equity & Accessibility in Detection, Collision Algorithms & Data Collection

Brief by Ian Moura for the Disability Rights Education & Defense Fund
November 7, 2022, C. Tyson, ed

Ian Moura is a PhD student in Social Policy at the Heller School for Social Policy and Management at Brandeis University. Moura was a 2022 Marilyn Golden Policy Intern at the Disability Rights Education & Defense Fund (DREDF). Marilyn Golden (1954 – 2021) was a senior policy analyst at DREDF and a long-time disability rights advocate. Golden played a key role in the development, passage, and implementation of the Americans with Disabilities Act of 1990 (ADA), including implementation of the transportation provisions. Golden advocated for access and equity across transportation modes, including transportation network company services and autonomous vehicles (AVs). Golden's advocacy molded and shaped accessibility in the United States. She spent more than three decades teaching others the value of disability civil rights and disabled lives via, and beyond the law. Moura's brief on ableist algorithm bias in autonomous vehicles and policy solutions continues Golden's legacy. Moura can be reached at ianmoura@brandeis.edu. Learn more about DREDF's past and ongoing transportation and AV advocacy at https://dredf.org/public-policy/transportation/. Contact info@dredf.org with questions.

Executive Summary

Autonomous vehicles (AVs) have the potential to transform access to transportation and related infrastructure for people with disabilities. AVs and new technologies also come with significant risks of embedding and perpetuating bias and discrimination that permeate society. This brief considers bias within pedestrian detection and collision behavior algorithms, lack of disability representation in the datasets used to train and test AVs, and the need for ethics frameworks that give full recognition to disabled people's humanity and fundamental rights. Policy measures to mitigate the identified risks are proposed.

Previous Work on AVs and Disability

AVs have been a point of interest within both academic and industry research for well over a decade. Several studies have explored the attitudes of people with disabilities towards AVs or accessible vehicle design or standards, others have looked at legal and policy implications of AV adoption related to disability protections and services. A small number of recent works, while not specifically focused on AVs, consider the implications of algorithms and artificial intelligence for disabled people. Few papers that consider the broader ethical frameworks of AVs consider the needs of disabled people or the influence ableism has had on many ethical standards.

Algorithmic Bias

Algorithmic bias can be understood as decision-making by an algorithmic tool or system that is prejudicial towards a particular person or group, especially in ways considered unfair. In defining what constitutes "normal" human appearance and movement for the purposes of identification by an AV's algorithms, the designers, computer programmers, and engineers responsible for developing AVs may translate society's ableist biases into software. The decision to treat disabled people as "edge cases," which do not merit equal consideration during the design and development of algorithms, all but guarantees unequal treatment when these systems are implemented, and encodes an existing social reality where disabled people are seen as less fundamentally human and less representative of the range of human experiences.

Bias in Pedestrian Detection Algorithms

AVs rely on a number of different types of sensors in order to navigate their environment and perceive obstacles within it. These sensors incorporate a range of technologies, such as LiDAR, RADAR, and cameras, and may or may not incorporate algorithms as part of their operation. While it is important that all sensors used by AVs are properly calibrated to detect disabled people in the vehicle's environment, the aspects of sensor technology that integrate algorithms are of particular concern in developing AVs that are safe and equitable for people with disabilities.

Recent research has identified potential issues of bias within both machine vision in general and the specific pedestrian detection algorithms used by AVs. Facial recognition technology's accuracy is decreased when used to detect darker-skinned subjects, and there are concerns about the accuracy and ethics of automatic gender recognition algorithms, particularly for transgender and gender-nonconforming individuals.

Anecdotal evidence suggests similar issues with such algorithm's ability to detect disabled people in or around roadways, particularly if those individuals do not present or move as the algorithm has been trained to expect them to. For example, when a researcher tested a model with visual captures of a friend who propels herself backward in her wheelchair using her feet and legs, the system not only failed to recognize her as a person, but indicated that the vehicle should proceed through the intersection, colliding with her.

Bias in Collision Behavior Algorithms

Standards for AV behavior when faced with an unavoidable collision may base recommendations off of public perceptions of the "best" choice in a hypothetical crash scenario. When the expectation is that the majority opinion determines fair conduct, particular care needs to be taken to ensure that the needs and preferences of minority groups, and in particular, of multiply marginalized people are adequately considered and that their rights are fully protected.

Discussion of algorithms that might deprioritize the safety and well-being of people with disabilities needs to be balanced with consideration of current safety issues, such as the need for collision testing with crash dummies that accurately represent people with disabilities, and AV design and safety testing that explicitly considers scenarios which disproportionately

present risk to disabled people, both due to the nature of vehicles and as a result of their interaction with an often inaccessible built environment.

Bias in Data Collection for Algorithms

It is critical that both developers and policymakers recognize that these data sets and algorithms frequently contain the same biases and prejudices that permeate society. Collision prediction algorithms that consider individual health information or estimate an individual's likelihood of survival may disproportionately harm people with disabilities if taking in to account characteristics of occupants and non-occupants.

A lack of disability representation within datasets creates significant risk for disabled people as AVs become a reality. While there is a general lack of representation of disabled people in datasets used to train and test algorithms, disabled people who have additional marginalized identities, such as disabled women and non-binary people, or disabled people of color, are particularly underrepresented. A lack of disability data is not unique to AVs, or even to algorithms more broadly. Disability is often treated as an afterthought or an exception, when it is considered at all.

Proposed Policy Measures

Recommendations for improving AV safety for disabled people include: improving datasets, preventing and remediating algorithmic bias, establishing standards for external oversight and regulation ensuring the burden of proof does not fall on those most impacted, and developing disability inclusive ethics.

Measures to Improve Datasets

In order to identify disability-related discrepancies in algorithmic performance, disabled people need to be accurately identified within datasets. This means that examples of disabled people must demonstrate a variety of disability types, and must include people of color and from a diverse array of ethnic backgrounds, people with a variety of gender identities and presentations, and people of a wide range of ages. Part of the work of increasing and improving disability data may well be creating more inclusive processes for the determination of outlying data.

Increasing representation of disabled people within datasets must come with acknowledgment of recent critiques of the reliance on "more data" as the solution to algorithmic bias. Disabled people are already subject to excess surveillance. Disability is frequently stigmatized, and gathering greater data about disabled people in general, and about the nature and characteristics of their disabilities in particular, can make people vulnerable to discrimination. Standards and procedures should be established to ensure that synthetic data are not used as a substitute for increasing the quality and inclusion of disability data, that data are appropriately protected and governed, and that integration with existing datasets is done responsibly so as to avoid exposing people with disabilities to unnecessary risks. Protecting people's right to know what information about them may be considered in an algorithm's predictions or classifications should be prioritized. Protections must also be developed to

ensure those choosing to share less of their data are not harmed or at fault for any algorithmic recommendations.

Disability data must also encompass non-individual information about accessibility features within the places where AVs operate, such as the location of curb-cuts. Consideration should be given not just to data quantity and quality, but also to identifying specific kinds of data which can help to ensure that AVs adequately meet the needs of disabled occupants.

Measures to Prevent and Remediate Algorithmic Bias

Addressing conscious and unconscious biases among developers is a critical part of preventing algorithmic bias. AV developers should take care to include not just disabled researchers, industry experts, and policymakers, but also disabled people who do not have these kinds of credentials or qualifications. AV developers should also ensure that they are seeking collaboration with people and organizations who represent a variety of disabilities, and that their collaborators include disabled people of color, disabled women and non-binary individuals, and others who can speak to the variety of disability experiences and perspectives and the impact of intersecting identities.

Measures to Establish Standards for External Oversight and Regulation

Like datasets, algorithms themselves should be subject to outside oversight and auditing. Ideally, audits should be conducted both by external experts, who were not involved in development of the algorithms under review, as well as disabled people. Involving the disability community in audits is important both because of the direct impact of AVs on people with disabilities and because of the relatively small number of disabled people who are involved in designing and developing AVs.

Beyond specific external review, standards should be developed regarding the kinds of algorithms that are used in AVs. When possible, the algorithms in AVs should be transparent and interpretable, and algorithms themselves should be publicly accessible.

The nature of AVs also means that some proposed forms of auditing and oversight that have been developed in the context of other kinds of algorithmic tools cannot be relied on as a means of regulating AVs. Preventing algorithmic bias and other safety risks associated with AVs should be prioritized over attempting to rectify them after they occur.

Measures to Develop Disability Inclusive Ethics

If ethical standards are to be based on what "society" believes, then researchers, policymakers, and industry experts need to be explicit about who is fully included in society and whose opinion is elevated and enshrined into ethical codes. Rectifying this situation, both with regard to AVs and more broadly, means questioning assumptions about who is valuable to society, and in what ways. AV developers would do well to draw on work that has emerged from the disability rights movement and critical disability studies, which could inform a more inclusive ethical vision for AV design, development, and implementation.

Table of Contents

Introduction

Autonomous vehicles (AVs) have the potential to transform transportation and related infrastructure, and recent scholarship has frequently proposed their adoption as part of improving roadway safety and public transit feasibility.[1] The design, development, adoption, and regulation of AVs is of particular concern to people with disabilities, both in terms of greater autonomy and access to transit and public spaces, and in terms of potential safety risks.[2]

While a number of recent resources and reports have considered AV design accessibility,[3] there has been little research that focuses on issues related to ethics, data collection and use, or algorithmic bias in the context of AVs from a disability perspective. This report seeks to rectify some of these omissions by exploring the implications of AVs on people with disabilities, specifically focusing on bias within pedestrian detection and collision behavior algorithms, lack of disability representation in the datasets used to train and test AVs, and the need for ethics frameworks that give full recognition to disabled people's humanity and fundamental rights. Policy measures are proposed to mitigate the identified risks.

I. Previous Work on AVs and Disability

AVs have been a point of interest within both academic and industry research for well over a decade. However, scholarly works in particular have tended to focus on only a small number of issues at a time, failing to adequately address the ways in which the impacts of AVs are frequently interrelated.[4] The design, development, and implementation of AVs is a fundamentally interdisciplinary area.[5] An understanding of AVs necessitates engagement on not just automotive regulation or machine vision techniques, but also broader participation in conversations about topics such as data, privacy, and algorithmic bias.

In addition, related research and policy work must include specific considerations of disability and disabled people. Though technology is often touted as a solution to providing disabled people greater freedom and integration, it also comes with significant risks of embedding and perpetuating the bias and discrimination that permeate society.[6]

Of the existing scholarly articles on AVs that center disability, most are limited in scope or focus. Several studies have explored the attitudes of people with disabilities towards AVs.[7] Others have looked at legal and policy implications of AV adoption related to disability protections and services.[8] A small number of recent works, while not specifically focused on AVs, consider the implications of algorithms and artificial intelligence for disabled people.[9] However, there has been little work considering the algorithms specific to AVs, such as those used for pedestrian detection and collision behavior, and how they may impact people with disabilities. Similarly, few papers that consider the broader ethical frameworks through which AVs are considered include consideration of the needs of disabled people or the influence ableism has had on many ethical standards.

II. Algorithmic Bias

At the most basic level, an algorithm is a set of instructions which can be followed step by step to arrive at a particular outcome or decision; in the context of AVs, as with other emerging digital technology, these instructions are fed into a computer or computational system, for arrival at a particular outcome which may or may not be mediated by human oversight.[10] Though algorithms are often presented as a way to quantify fairness there are numerous ways in which to define "fairness" for the purpose of an algorithm, and many of these definitions are incompatible with one another or involve significant tradeoffs.[11] Adding to the complexity, fairness definitions are both subjective and contextually bound.[12]

As algorithmic tools become more widespread, there is increasing awareness of and concern about algorithmic bias arising from these technologies. Algorithmic bias can be understood as decision-making by an algorithmic tool or system that is prejudicial towards a particular person or group, especially in ways considered unfair.[13] Consequently, algorithmic bias is closely linked to the concepts of fairness and equity. Fairness is not only a subjective concept; it is also frequently linked to the distribution of power within society, and to which particular members have the ability to regulate outcomes.[14] There is increasing recognition of the need to re-center an understanding of algorithms within algorithmic fairness.[15]

Concerns about algorithmic bias in AVs primarily relate to programmed responses to unavoidable collisions, and the machine vision algorithms that allow an AV to detect people and other obstacles in its environment.

Bias in Pedestrian Detection Algorithms

AVs rely on different sensors in order to navigate their environment and perceive obstacles within it. These sensors incorporate a range of technologies, such as LiDAR, RADAR, and cameras, and may or may not incorporate algorithms as part of their operation.[16] While it is important that all AV sensors are properly calibrated to detect disabled people in the vehicle's environment, the aspects of sensor technology that integrate algorithms are of particular concern. Research has identified potential bias within both machine vision in general and the specific pedestrian detection algorithms used by AVs. For example, facial recognition technology's accuracy is decreased when used to detect darker-skinned subjects, and on subjects who are both darker-skinned and female.[17] Scholars have also highlighted concerns regarding automatic gender recognition algorithms, particularly for transgender and gender-nonconforming individuals.[18] Analyses of pedestrian detection algorithms have raised similar concerns, finding that standard models are more precise for pedestrians with lighter skin tone than those with darker skin tone, even in situations that do not present particular challenges to detection based on occlusion or time of day.[19] Anecdotal evidence suggests similar issues with such algorithms' ability to detect disabled people in or around roadways, particularly if those individuals do not present or move as the algorithm has been trained to expect them to. For example, when a researcher tested a model with visual captures of a friend who propels herself backward in her wheelchair using her feet and legs, the system not only failed to recognize her as a person, but indicated that the vehicle should proceed through the intersection, colliding with her.[20]

In each of the cases, the failure of algorithms to accurately recognize or classify a particular individual is not the result of certain people being more difficult to identify or categorize. Instead, these failures are the product of specific decisions made during the design and development of the algorithms themselves. Like any algorithm, pedestrian detection systems are designed to accomplish specific goals, assessed by particular metrics reflecting human choices and prioritizations.[21] Not only that, but they represent the judgments and choices of a particular group of people, who have the expertise and corresponding power to set standards for technical systems.[22] In defining what constitutes "normal" human appearance and movement for the purposes of identification by an AV's algorithms, the designers, computer programmers, and engineers responsible for developing AVs may translate society's ableist biases into software. The decision to treat disabled people as "edge cases," which do not merit equal consideration during the design and development of algorithms, all but guarantees unequal treatment when these systems are implemented, and encodes an existing social reality where disabled people are seen as less fundamentally human and less representative of the range of human experiences.[23]

It is important to emphasize that the process through which such biases are incorporated into algorithmic systems is not necessarily intentional. Algorithmic bias can, and often does, arise from issues related to training data. Additionally, elements of the architecture of certain kinds of algorithms, including those commonly used for machine vision, make them susceptible to replicating societal disparities and biases as represented in datasets, often without the awareness of designers and developers. For example, machine learning algorithms, commonly used for functions including pedestrian detection, include supervised, semi-supervised, and unsupervised models, each of which comes with specific advantages and challenges in terms of preventing bias. In a supervised model, the training data include expected results, and

performance is assessed in part by comparing the model's output – its predictions or classifications – with what developers expect, and then generalizing that performance to work with novel data and real-world situations.[24] While the use of labeled training data as a comparison may help developers ensure fairness this is predicated on inclusion of adequate examples. Consequently, an algorithm that performs well on training data may or may not perform well when actually implemented.[25] In contrast to supervised models, unsupervised machine learning relies on novel input data alone, and identifies patterns and groups together repeated or similar elements of these data without a reference or training dataset for comparison. Semi-supervised machine learning relies on a limited amount of labeled data to support categorization of a much larger unlabeled dataset.[26] While unsupervised and semi-supervised machine learning most obviously run the risk of arriving at biased groupings or finding patterns that operate in ways unintended by their developers, even supervised machine learning runs the risk of behaving unpredictably when they encounter a situation outside those represented in training data.[27]

Lack of Transparency in Pedestrian Detection Algorithms

Adding to the challenge is the fundamental opacity of the algorithms used for machine vision, and by extension, for pedestrian detection in AVs. There has been a push for greater use of transparent, interpretable algorithms rather than those that are "opaque," particularly in high-stakes situations, such as those where human lives are potentially at risk.[28] Algorithms may be "opaque" for several different reasons; for example, some algorithms are proprietary. Although their architecture is comprehensible and interpretable, intellectual property protections render them uninterpretable to anyone outside the company or individuals who own the rights to them.[29] Other algorithms are fundamentally opaque by nature of their construction, even to the people who construct them. This is particularly true for deep convolutional neural networks (CNNs), a machine learning technique that involves passing an input, such as an image, through multiple layers of an interconnected network, and which rely on hidden, and often non-intuitive, correlations to create output.[30] The inscrutability of CNNs results from a fundamental misalignment between the way these algorithms operate and the capacity of human beings to understand their behavior.[31] Consequently, there is an architectural component to bias in pedestrian detection systems. While such algorithms are commonly referred to as machine vision, it is a misnomer to assume that AVs are really seeing anything in the way humans ordinarily understand the word. Though CNNs can identify specific features based on their training and exposure to previous data, they lack representational capacity; they detect and categorize parts of images, but they do not imbue any human meaning into those images, and the information within an image that a CNN identifies as salient contains no meaningful information for a human who views it.[32] Attempts to address bias in such systems must be accompanied by an awareness that achieving AV performance that mimics that of a human driver does not mean that an AV's underlying algorithms are in any way actually behaving like a human driver.

Bias in Collision Behavior Algorithms

There is additional potential for bias in how AVs behave in the event of an unavoidable collision. The algorithms used to dictate collision behavior are typically rule-based, explicitly instructing

AVs in what to do in the event of a crash. Scholarly literature on collision algorithms frequently contextualizes them in terms of societal acceptance of AVs.[33] Consequently, much of the research that seeks to establish standards for AV behavior when faced with an unavoidable collision bases recommendations off of public perceptions of the "best" choice in a hypothetical crash scenario.

Perhaps the best-known example of these collision scenarios is the Trolley Problem. The Trolley Problem is a popular thought experiment in moral philosophy that describes a situation in which the operator of a trolley car must decide whether to change tracks, colliding with one person, or remain on the original track and collide with five people.[34] When adapted to specifically relate to AVs, variations of the Trolley Problem typically involve decisions about steering a vehicle to influence who or what it collides with, generally with the assumption that avoiding a collision entirely is impossible.

A prominent example of research based on the Trolley Problem is the Moral Machine, a web-based version of an autonomous vehicle focused scenario. Over 2 million people participated in the experiment by answering questions about AV-related moral dilemmas during an unavoidable collision, and a subgroup of responses from over 400,000 participants who completed an optional demographic survey was analyzed to determine preferences for AV behavior.[35] Researchers have conducted similar experiments, describing or illustrating scenarios in which participants must determine the AV's behavior regarding an impending collision.[36] Notably, inconsistencies are reported in how people say they want an AV to behave in the abstract, and the behavior they want in a car that they would ride in or purchase.[37]

Much of the academic literature on the Trolley Problem specifically, and on algorithmic responses to an unavoidable collision, approaches the issue from one of two standpoints. The first seeks to establish a set of universal ethics for how an AV should behave in a variety of hypothetical crash scenarios. In these experiments, fairness is understood to result from consensus or majority opinion, with the assumption that if most people think a certain choice or action is moral, that is an appropriate proxy for a more formally defined ethical code. When the standard for determining moral or ethical behavior is that decisions be left up to society, it is important to consider who is fully recognized and included as part of society.[38] Particular care needs to be taken to ensure that the needs and preferences of minority groups, and in particular, of multiply marginalized people, are adequately considered and that their rights are fully protected.

The second common approach in academic literature on collision behavior algorithms is to relate possible responses an AV might take in the event of a crash to established philosophical positions. For example, some scholars have described "Rawlsian" algorithms that might be used to outline collision behavior, or connect particular approaches to schools of thought such as deontological ethics or utilitarianism.[39,40] While connecting the concrete approaches to solve social problems to theoretical ideas is a worthwhile part of exploring solutions, consideration should be given to the implications for people with disabilities. There is potential for ethical systems to encode the biases that are entwined with particular theories, especially if they become the basis for algorithmic decision-making.

There are significant issues with the Trolley Problem as a framework for determining ethical decision-making in the event on an unavoidable collision. Perhaps most concerningly, the framing of these scenarios as a choice between braking in a straight line and braking while swerving obscures the fact that these are not equivalent actions. Simultaneously applying force to a car's brakes and attempting to turn the wheel increases the risk of skidding and of losing

control of the vehicle, lowering the ability of a driver – whether autonomous or human – to dictate the direction of travel.[41] Additionally, by failing to acknowledge existing automation such as lane assist and emergency-braking technology, scenarios like the Trolley Problem position vehicle behavior in a collision scenario as a novel problem, unrelated to existing dilemmas around shared roadways and vehicle operation. Research on both public attitudes about the collision behavior of AVs and on collision algorithms themselves often feeds into somewhat sensationalist narratives about AVs, while obscuring more mundane and pressing concerns.[42]

Establishing standards for data use and governance in the context of AVs, for example, tend to be less of a focus of consideration, particularly outside academic circles, than Trolley Problem scenarios. This discrepancy is particularly noteworthy considering that at present, AVs do not have sufficient perceptual capabilities to make the kinds of distinctions between people to carry out collision decision-making on which the Trolley Problem hinges.[43] While it is worthwhile to contemplate what may happen when such advances become part of AV technology, such discussion should not come at the expense of addressing issues that are already occurring or that are possible based on existing functionality. It should go without saying that AVs must not discriminate against disabled people in their decisions in the event of an unavoidable collision.

Addressing algorithms that might theoretically deprioritize the safety and well-being of people with disabilities must be balanced with robust consideration of current safety issues. There is a need for collision testing with crash dummies that accurately represent people with disabilities. In addition, AV design and testing must consider scenarios which disproportionately present risk to disabled people, including the AVs and disabled pedestrian's interaction with often inaccessible built environments. People with disabilities, including wheelchair users, often must travel in the road when a sidewalk has no curb cuts or is damaged, or during snowstorms when clearing streets is prioritized over sidewalks.

Bias in Algorithmic Data

Addressing the issues related to algorithmic bias and disability requires a corresponding consideration of the data on which algorithms rely and operate. The data used to develop and train algorithms for use in AVs encompass a broad array of information, including large databases of images, estimations and results of various crash scenarios, and measurements and specifications representing built environments that vehicles are expected to navigate. It is critical that both developers and policymakers recognize that these data are not an objective reflection of the world; rather, they result from a series of choices and decisions, and frequently contain the same biases and prejudices that permeate society.[44]

Pedestrian detection algorithms generally rely on neural networks, which pass images through multiple layers in order to categorize them. Because these algorithms are expected to construct inferences from data, it matters a great deal what those data are. Compared to fields in which datasets are intentionally and systematically constructed, the data used in such algorithms are frequently scraped from publicly available sources, such as search engine results and social media sites. In the case of images, training data are labeled manually, often through crowdsourcing platforms like Amazon Mechanical Turk.[45] Labeling is an interpretive process, subject to cultural, contextual, and personal biases, and labels are only approximations of the actual objects they are assigned to, rather than direct representations.[46] Additionally, the process of compiling and labeling image data is often poorly documented

making it difficult to replicate or even understand how a particular dataset was created, or identify the ways in which it may contribute to algorithmic bias.[47] A recent study investigating models trained on ImageNet, a popular dataset created from internet images, found that the models replicated human biases that have been documented within social psychology, including several related to race and gender, and suggest that reliance on images pulled from public internet sources can result in models that recreate human biases, based on stereotypical ways that individuals from certain sociodemographic groups are portrayed online.[48]

A primary challenge for the data used to develop and train pedestrian detection algorithms is gathering a wide enough range of images to fall into multiple categories. An AV needs, potentially, basic information about the crash situation, such as the vehicle's speed, location, and immediate environment; additional data about individuals involved in the collision; and information related to set priorities for how to respond to different types of collisions.[49] The data's meaning and source must be interpreted and designed by the people who specify the architecture of the collision behavior algorithm. In addition to the encoding of ethics in dictating how AVs should respond to crash scenarios, what passenger or pedestrian information to include has the potential to create algorithms that treat disabled people unfairly. For example, should AVs take individual characteristics of occupants and non-occupants into account when determining how to respond to an unavoidable collision, individual health information or estimates of an individual's likelihood of survival may disproportionately harm people with disabilities.

As both pedestrian detection algorithms and collision behavior algorithms demonstrate, a lack of disability representation within datasets creates significant risk for disabled people as AVs become a reality. The lack of representation of disabled people within algorithmic data used by AVs has multiple causes, including limited systematic approaches to dataset development, as well as general limitations on the amount, quality, and representativeness of data on people with disabilities currently available.[50] Disability is complex and multifaceted, and disabled people are a diverse, heterogeneous group. Disabled people who have additional marginalized identities, such as disabled women and nonbinary people, or disabled people of color, are particularly underrepresented.[51]

It should be noted that a lack of disability data is not unique to AVs, or even to algorithms more broadly. Disability is often treated as an afterthought or an exception, when it is considered at all.[52] In the case of the researcher who discovered that pedestrian detection algorithms failed to identify a wheelchair user who moved in a "non-standard" way, when the researcher raised concerns about the algorithm's performance, she was told that the model would improve with greater exposure to images of people using wheelchairs, suggesting that incorporation of disabled people into the original training data was not a priority for developers.[53] Apart from the obvious safety risks this creates, a lack of data from and about disabled people also makes it difficult, or even impossible, to identify disparities in how algorithms treat disabled people compared to nondisabled people.

Disability Data Collection Potential for Increased Discrimination

The large amounts of data needed to train and test algorithms for use in AVs are assembled from smaller collections of data, including records from service providers, companies, and government entities [54]. Disabled people, and particularly disabled people of color and disabled people who are socioeconomically disadvantaged, are often subject to what Virginia Eubanks

has termed the "digital poorhouse:" a growing web of algorithmic surveillance that replicates many of the functions of the physical poorhouses of the past.[55] Disability is frequently stigmatized, and gathering greater data about disabled people in general, and about the nature and characteristics of their disabilities in particular, can make people vulnerable to discrimination.[56] Additionally, beyond simply being at greater risk from data misuse and compromised privacy, disabled people are often vulnerable to reidentification even when data are supposedly anonymized.[57] Attempts to increase data about disability can also involve labeling people as disabled based on guesswork and assumptions, and even constructing synthetic disability data, similar to efforts to create more racially inclusive datasets that have relied on simulated images of darker-skinner individuals.[58] Disabled people may also be at risk of harm stemming from the dynamics of data production and use, which create stratified classes of people based on who creates data (consciously or not), who collects data, and who analyses data.[59]

III. Proposed Policy Measures

Recommendations for improving AV safety for disabled people, specifically in relation to their use of algorithms and data, target four aspects of the issue: improving datasets, preventing and remediating algorithmic bias, establishing standards for external oversight and regulation ensuring the burden of proof does not fall on those most impacted on, and developing disability inclusive ethics.[60]

Measures to Improve Datasets

In order to identify disability-related discrepancies in algorithmic performance, disabled people need to be accurately identified within datasets. Several researchers have discussed the importance of reliable identification of sociodemographic group membership as a step towards improving data and algorithmic equity.[61] Furthermore, because of the mounting evidence that individuals with intersecting identities are particularly likely to experience biased treatment from algorithms, additional effort must be dedicated to gathering disability data that adequately represents the complexity and heterogeneity of the disability community. This means that examples of disabled people must demonstrate a variety of disability types, and must include people of color and from a diverse array of ethnic backgrounds, people with a variety of gender identities and presentations, and people of a wide range of ages. Additionally, extra consideration must be given before omitting or removing "outliers" from the dataset, as many of the standards against which data are judged to determine their accuracy routinely exclude disabled people.[62]

Increasing representation of disabled people within datasets should come with acknowledgment of recent critiques of the reliance on "more data" as the solution to algorithmic bias. Disabled people are already often subject to excess surveillance; consequently, greater collection and inclusion of disability data must come with consideration of the risks to privacy that accompany this process.

Ideally, part of the work to collect more and better disability data will include establishing standards and procedures to ensure that synthetic data are not used as a substitute for increasing the quality and inclusion of disability data, that data are appropriately protected and governed, and that integration with existing datasets is done responsibly so as to avoid

exposing people with disabilities to unnecessary risks. Protecting people's right to know what information about them may be considered in an algorithm's predictions or classifications should be prioritized. Protections must also be developed to ensure those choosing to share less of their data are not harmed or at fault for any algorithmic recommendations.

Measures to Improve Datasets: Disability Data Including Accessibility of Infrastructure

Disability data in this context are not limited to increased information on and identification of disabled individuals within large datasets. Disability data must also encompass information about accessibility features within the places where AVs operate, such as the location of curb-cuts and elevators, or less crowded entrances. A recent study found significant gaps in municipal data on features identified by disabled people as promoting safe pedestrian travel.[63] As part of efforts to improve disability data, therefore, consideration should be given not just to data quantity and quality, but also to identifying specific kinds of data which can help to ensure that AVs adequately meet the needs of disabled occupants.

In addition to improving disability data overall, the specific datasets used to train and test autonomous vehicle algorithms require increased transparency and accountability. Datasets need to be more widely, and publicly, available, and the processes through which they were created should be clearly detailed in a way that would theoretically allow others to replicate their construction. Datasets should be subject to an external auditing process, and the results of audits should be available to the public. At the same time, audits should be conducted in a way that does not compromise individuals' privacy or artificially collapse complex and intersecting identities for the sake of creating easy-to-measure benchmarks.[64]

Measures to Prevent and Remediate Algorithmic Bias

Specific steps must also be taken to mitigate and ideally, prevent, biased treatment of disabled people by the algorithms on which AVs rely. Scholars and industry experts have proposed developing standards for algorithmic accountability, highlighting different elements of algorithmic systems that must be considered in order to prevent bias.[65] Others have suggested specific questions that should be asked of a model in order to understand its potential to adversely impact human lives, and to determine whether it may warrant additional examination to prevent biased outcomes.[66] In the context of AVs, two specific measures have significant promise for reducing the risk of algorithmic bias for people with disabilities. First, increasing the consideration and incorporation of contextual information into AV development. Second, integrating participatory methods that engage the disability community into the standard process for AV design and implementation.

Contextual information can mean a number of different things. For autonomous vehicle development, relevant context includes data about the environment in which AVs are expected to operate and specific information regarding the needs and preferences of people with a range of disabilities. It should also draw on fields like Science and Technology Studies and Critical Disability Studies to capture the ways in which societal factors, including ableism, impact the way technology operates in the world. Specific trainings could provide an overview from fields that investigate the ways technology is intertwined with both its human users and society as a whole and seek to increase developers' familiarity with the disability community and the disability rights movement. Because human biases easily become incorporated into algorithms,

even when those who develop them have every intention of creating technology that is fair and equitable, addressing conscious and unconscious biases among developers is a critical part of preventing algorithmic bias.

Disabled people must also be fully included in the development process. One way to do this is through the use of participatory methods for AV research and design. Participatory methods have been proposed within the context of other situations in which algorithms are used, often as work to acknowledge and rectify the tendency for algorithms to reinforce existing power structures.[67] Recent research into preventing algorithmic bias has noted that while algorithmic bias can emerge at any stage of development, critical issues most commonly emerge during early steps.[68] Therefore, participatory methods should be incorporated from the beginning of AV development, rather than brought in solely during later stages, such as once AVs are being adopted and implemented within a city or region. Additionally, AV developers should take care to include not just disabled researchers, industry experts, and policymakers, but also disabled people who do not have these kinds of credentials or qualifications. In order to foster real inclusion and collaboration, lived experience of disability needs to be treated as an equally valuable form of expertise. AV developers should also ensure that they are seeking collaboration with people and organizations who represent a variety of disabilities, and that their collaborators include disabled people of color, disabled women and non-binary individuals, and others who can speak to the variety of disability experiences and perspectives and the impact of intersecting identities.

Measures to Establish Standards for External Oversight and Regulation

Like datasets, algorithms themselves should be subject to outside oversight and auditing. Ideally, audits should be conducted both by external experts, who were not involved in development of the algorithms under review, as well as disabled people. While understanding the technical aspects of algorithms may require specific training and background knowledge, involving the disability community in audits is important both because of the direct impact of AVs on people with disabilities and because of the relatively small number of disabled people who are involved in designing and developing AVs. Establishing independent audit panels can support greater connection and collaboration between technical experts and people with disabilities. Disabled people and other marginalized communities should be included at the highest levels in oversight and auditing. However, the burden of proof for existing or potential algorithmic bias must not fall on those most impacted.

Beyond specific external review, standards should be developed regarding the kinds of algorithms that are used in AVs. When possible, the algorithms in AVs should be transparent and interpretable, and algorithms themselves should be publicly accessible. This is in line with recommendations from computer science experts regarding the use of algorithmic decision-making in high stakes situations.[69] However, pedestrian detection systems commonly rely on computational architecture that is fundamentally uninterpretable. In cases where no equivalent interpretable model exists, developers must be cautious about the use of explanatory models built alongside the algorithm, which can distort and misrepresent the algorithm's actual operation. When uninterpretable models such as CNNs are required for AV functions, additional scrutiny of training and test data should be standard. Particular attention must be paid to the representation of marginalized groups within the data and the degree to which the data adequately relate to the conditions under which AVs will be expected to operate.

The nature of AVs also means that some proposed forms of auditing and oversight cannot be relied on as a means of regulating AVs. For example, some recent work has proposed the use of features that allow contestation of algorithmic decisions as part of increasing transparency and addressing algorithmic bias.[70] Others have considered safety measures informed by due process.[71] While these suggestions are valuable in addressing certain situations where algorithmic bias arises, the instantaneous nature of algorithmic use within AVs, where decisions such as how to behave in an unavoidable impending collision require immediate action, renders these strategies less useful. This underscores the importance of focusing on preventing algorithmic bias and other safety risks associated with AVs, rather than attempting to rectify them after they occur.

Measures to Develop Disability Inclusive Ethics

Finally, technological development is intertwined with society as a whole. Seemingly neutral technologies incorporate biases that are present in the societies from which they emerge. Therefore, an important overarching aspect of the work to ensure that AVs are safe and equitable for people with disabilities involves working towards full and equal partnerships with disabled people during the design, development, and implementation of AVs. Beyond simply establishing a standard for collaborative practice, ideas about ethical behavior by AVs needs to be based upon a sense of ethics that recognizes the full humanity of disabled people and protects their rights within society.

Many of the existing analyses of ethical implications of AVs focus on broad societal acceptance of AVs and refer to public opinion or social norms regarding choices that many ultimately amount to life-or-death decisions. These studies, while valuable in providing insight into how many people understand and think about AVs, suffer from a lack of recognition of the pervasiveness of ableism within society. If ethical standards are to be based on what "society" believes, then researchers, policymakers, and industry experts need to be explicit about who is fully included in society and whose opinion is elevated and enshrined into ethical codes. Simply having ethics codes for AVs, or for any technology, is insufficient. Ethics codes often do little to support vulnerable groups and rarely create real accountability to such communities, often due to a lack of authentic engagement with them.[72] Similarly, although many of the prescribed responses to algorithmic bias, such as those that emphasize fairness, accountability, and transparency, offer up solutions, they ultimately do little to shift established balances of power unless accompanied by consideration of who gets to decide what is fair, accountable, or transparent.[73] Rectifying this situation, both with regard to AVs and more broadly, means questioning assumptions about who is valuable to society, and in what ways. AV developers, policymakers and regulators would do well to draw on work that has emerged from the disability rights movement and critical disability studies, which could inform a more inclusive ethical vision for AV design, development, marketing, and implementation.

References

Andrus, McKane, Elena Spitzer, Jeffrey Brown, and Alice Xiang. "'What We Can't Measure, We Can't Understand': Challenges to Demographic Data Procurement in the Pursuit of Fairness." *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 2020, 249–60.

Araujo, Theo, Natali Helberger, Sanne Kruikemeier, and Claes H. de Vreese. "In AI We Trust? Perceptions about Automated Decision-Making by Artificial Intelligence." *AI & Society* 35, no. 3 (September 2020): 611–23. https://doi.org/10.1007/s00146-019-00931-w.

Awad, Edmond, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. "The Moral Machine Experiment." *Nature (London)* 563, no. 7729 (2018): 59–64. https://doi.org/10.1038/s41586-018-0637-6.

Aysolmaz, Banu, Nancy Dau, and Deniz Iren. "Preventing Algorithmic Bias in the Development of Algorithmic Decision-Making Systems: A Delphi Study." *Proceedings of the 53rd Hawaii International Conference on Systems Sciences*, 2020, 5267–76.

Barabas, Chelsea, Colin Doyle, JB Rubinovitz, and Karthik Dinakar. "Studying Up: Reorienting the Study of Algorithmic Fairness around Issues of Power." *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020, 167–76.

Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review* 104, no. 3 (2016): 671–732.

Bennett, Roger, Rohini Vijaygopal, and Rita Kottasz. "Attitudes towards Autonomous Vehicles among People with Physical Disabilities." *Transportation Research. Part A, Policy and Practice* 127 (2019): 1–17. https://doi.org/10.1016/j.tra.2019.07.002.

Bergmann, Lasse T., Larissa Schlicht, Carmen Meixner, Peter König, Gordon Pipa, Susanne Boshammer, and Achim Stephan. "Autonomous Vehicles Require Socio-Political Acceptance-An Empirical and Philosophical Perspective on the Problem of Moral Decision Making." *Frontiers in Behavioral Neuroscience* 12 (2018): 31–31. https://doi.org/10.3389/fnbeh.2018.00031.

Bigman, Yochanan E., and Kurt Gray. "Life and Death Decisions of Autonomous Vehicles." *Nature (London)* 579, no. 7797 (2020): E1–2. https://doi.org/10.1038/s41586-020-1987-4.

Binns, Reuben, and Reuben Kirkham. "How Could Equality and Data Protection Law Shape AI Fairness for People with Disabilities?" *ACM Transactions on Accessible Computing* 14, no. 3 (2021): 1–32. https://doi.org/10.1145/3473673.

Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan. "The Social Dilemma of Autonomous Vehicles." *Science (American Association for the Advancement of Science)* 352, no. 6293 (2016): 1573–76. https://doi.org/10.1126/science.aaf2654.

boyd, danah, and Kate Crawford. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15, no. 5 (2012): 662–79. https://doi.org/10.1080/1369118X.2012.678878.

Bradshaw-Martin, Heather, and Catherine Easton. "Autonomous or 'Driverless' Cars and Disability: A Legal and Ethical Analysis." *European Journal of Current Legal Issues* 20, no. 3 (December 11, 2014). https://webjcli.org/index.php/webjcli/article/view/344.

Brewer, Robin, and Vaishnav Kameswaran. "Understanding the Power of Control in Autonomous Vehicles for People with Vision Impairment," 185–97. ASSETS '18. ACM, 2018. https://doi.org/10.1145/3234695.3236347.

Brinkley, Julian, Jr Huff, Briana Posadas, Julia Woodward, Shaundra Daily, and Juan Gilbert. "Exploring the Needs, Preferences, and Concerns of Persons with Visual Impairments Regarding Autonomous Vehicles." *ACM Transactions on Accessible Computing* 13, no. 1 (2020): 1–34. https://doi.org/10.1145/3372280.

Buolamwini, Joy, and Timnit Gebru. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." In *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 77–91. PMLR, 2018. https://proceedings.mlr.press/v81/buolamwini18a.html.

Campbell, Sean, Niall O'Mahony, Lenka Krpalcova, Daniel Riordan, Joseph Walsh, Aidan Murphy, and Conor Ryan. "Sensor Technology in Autonomous Vehicles: A Review," 1-. Piscataway: The Institute of Electrical and Electronics Engineers, Inc. IEEE, 2018. https://search.proquest.com/docview/2159992893?pq-origsite=primo&accountid=9703.

Carabantes, Manuel. "Black-Box Artificial Intelligence: An Epistemological and Critical Analysis." *AI & SOCIETY* 35, no. 2 (June 2020): 309–17. https://doi.org/10.1007/s00146-019-00888-w.

Citron, Danielle Keats, and Frank A. Pasquale. "The Scored Society: Due Process for Automated Predictions." *Washington Law Review* 89, no. 1 (2014): 1-.

Claypool, Henry, Amitai Bin-Nun, and Jeffrey Gerlach. "Self-Driving Cars: The Impact on People with Disabilities." Newton, MA: Ruderman Family Foundation, 2017.

Corbett-Davies, Sam, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. "Algorithmic Decision Making and the Cost of Fairness." In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 797–806. KDD '17. New York, NY, USA: Association for Computing Machinery, 2017. https://doi.org/10.1145/3097983.3098095.

Cunneen, Martin, Martin Mullins, Finbarr Murphy, Darren Shannon, Irini Furxhi, and Cian Ryan. "Autonomous Vehicles and Avoiding the Trolley (Dilemma): Vehicle Perception, Classification, and the Challenges of Framing Decision Ethics." *Cybernetics and Systems* 51, no. 1 (2020): 59–80. https://doi.org/10.1080/01969722.2019.1660541.

Davnall, Rebecca. "The Car's Choice: Illusions of Agency in the Self-Driving Car Trolley Problem." *Artificial Intelligence*, June 24, 2020, 189–202. https://doi.org/10.30965/9783957437488_013.

Deitz, Shiloh, Amy Lobben, and Arielle Alferez. "Squeaky Wheels: Missing Data, Disability, and Power in the Smart City." *Big Data & Society* 8, no. 2 (2021): 205395172110477-. https://doi.org/10.1177/20539517211047735.

Deka, Devajyoti, and Charles T. Brown. "Self-Perception and General Perception of the Safety Impact of Autonomous Vehicles on Pedestrians, Bicyclists, and People with Ambulatory Disability." *Journal of Transportation Technologies* 11, no. 3 (May 18, 2021): 357–77. https://doi.org/10.4236/jtts.2021.113023.

Dicianno, Brad E., Sivashankar Sivakanthan, S. Andrea Sundaram, Shantanu Satpute, Hailee Kulich, Elizabeth Powers, Nikitha Deepak, Rebecca Russell, Rosemarie Cooper, and Rory A. Cooper. "Systematic Review: Automated Vehicles and Services for People with Disabilities." *Neuroscience Letters* 761 (2021): 136103–136103. https://doi.org/10.1016/j.neulet.2021.136103.

Dwork, Cynthia, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. "Fairness through Awareness," 214–26. ITCS '12. ACM, 2012. https://doi.org/10.1145/2090236.2090255.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. First edition. New York, NY: St Martin's Press, 2018.

Frank, Darius-Aurel, Polymeros Chrysochou, Panagiotis Mitkidis, and Dan Ariely. "Human Decision-Making Biases in the Moral Dilemmas of Autonomous Vehicles." *Scientific Reports* 9, no. 1 (2019): 13080–19. https://doi.org/10.1038/s41598-019-49411-7.

Grgic-Hlaca, Nina, Elissa M. Redmiles, Krishna P. Gummadi, and Adrian Weller. "Human Perceptions of Fairness in Algorithmic Decision Making: A Case Study of Criminal Risk Prediction." In *Proceedings of the 2018 World Wide Web Conference*, 903–12. WWW '18. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018. https://doi.org/10.1145/3178876.3186138.

Gurney, Jeffrey K. "Crashing into the Unknown: An Examination of Crash-Optimization Algorithms through the Two Lanes of Ethics and Law." *Albany Law Review* 79, no. 1 (2015): 183-.

Henin, Clément, and Daniel Le Métayer. "A Framework to Contest and Justify Algorithmic Decisions." *AI and Ethics* 1, no. 4 (November 1, 2021): 463–76. https://doi.org/10.1007/s43681-021-00054-3.

Hoffmann, Anna Lauren. "Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse." *Information, Communication & Society* 22, no. 7 (June 7, 2019): 900–915. https://doi.org/10.1080/1369118X.2019.1573912.

Kasinidou, Maria, Styliani Kleanthous, Pınar Barlas, and Jahna Otterbacher. "I Agree with the Decision, but They Didn't Deserve This: Future Developers' Perception of Fairness in Algorithmic Decisions." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 690–700. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445931.

Kasy, Maximilian, and Rediet Abebe. "Fairness, Equality, and Power in Algorithmic Decision-Making." In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 576–86. FAccT '21. New York, NY, USA: Association for Computing Machinery, 2021. https://doi.org/10.1145/3442188.3445919.

Katell, Michael, Meg Young, Dharma Dailey, Bernease Herman, Vivian Guetler, Aaron Tam, Corinne Bintz, Daniella Raz, and P. M. Krafft. "Toward Situated Interventions for Algorithmic

Equity: Lessons from the Field." In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 45–55. Barcelona Spain: ACM, 2020. https://doi.org/10.1145/3351095.3372874.

Keyes, Os. "The Misgendering Machines: Trans/HCI Implications of Automatic Gender Recognition." *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW (2018): 1–22. https://doi.org/10.1145/3274357.

Koopman, Philip, and Michael Wagner. "Autonomous Vehicle Safety: An Interdisciplinary Challenge." *IEEE Intelligent Transportation Systems Magazine* 9, no. 1 (2017): 90–96. https://doi.org/10.1109/MITS.2016.2583491.

Kuzio, Jacqueline. "Autonomous Vehicles and Paratransit: Examining the Protective Framework of the Americans with Disabilities Act." *Case Studies on Transport Policy* 9, no. 3 (2021): 1130–40. https://doi.org/10.1016/j.cstp.2021.06.001.

Leben, Derek. "A Rawlsian Algorithm for Autonomous Vehicles." *Ethics and Information Technology* 19, no. 2 (2017): 107–15. https://doi.org/10.1007/s10676-017-9419-3.

Milakis, Dimitris, Bart van Arem, and Bert van Wee. "Policy and Society Related Implications of Automated Driving: A Review of Literature and Directions for Future Research." *Journal of Intelligent Transportation Systems* 21, no. 4 (2017): 324–48. https://doi.org/10.1080/15472450.2017.1291351.

Millan-Blanquel, L., S. M. Veres, and R. C. Purshouse. "Ethical Considerations for a Decision Making System for Autonomous Vehicles during an Inevitable Collision." IEEE, 2020. https://doi.org/10.1109/med48518.2020.9183263.

Mitchell, Shira, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. "Algorithmic Fairness: Choices, Assumptions, and Definitions." *Annual Review of Statistics and Its Application* 8, no. 1 (March 7, 2021): 141–63. https://doi.org/10.1146/annurev-statistics-042720-125902.

Nakamura, Karen. "My Algorithms Have Determined You're Not Human: AI-ML, Reverse Turing-Tests, and the Disability Experience," 1–2. ASSETS '19. ACM, 2019. https://doi.org/10.1145/3308561.3353812.

Ntoutsi, Eirini, Pavlos Fafalios, Ujwal Gadiraju, Vasileios Iosifidis, Wolfgang Nejdl, Maria-Esther Vidal, Salvatore Ruggieri, et al. "Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey." *WIREs Data Mining and Knowledge Discovery* 10, no. 3 (2020): e1356. https://doi.org/10.1002/widm.1356.

Offert, Fabian, and Peter Bell. "Perceptual Bias and Technical Metapictures: Critical Machine Vision as a Humanities Challenge." *AI & Society* 36, no. 4 (2020): 1133–44. https://doi.org/10.1007/s00146-020-01058-z.

Packin, Nizan Geslevich. "Disability Discrimination Using Artificial Intelligence Systems and Social Scoring: Can We Disable Digital Bias?" *Journal of International and Comparative Law* 8, no. 2 (2021): 487–511.

Papa, Enrica, and António Ferreira. "Sustainable Accessibility and the Implementation of Automated Vehicles: Identifying Critical Decisions." *Urban Science* 2, no. 1 (2018): 5-. https://doi.org/10.3390/urbansci2010005.

Paullada, Amandalynne, Inioluwa Deborah Raji, Emily M. Bender, Emily Denton, and Alex Hanna. "Data and Its (Dis)Contents: A Survey of Dataset Development and Use in Machine Learning Research." *Patterns (New York, N.Y.)* 2, no. 11 (2021): 100336–100336. https://doi.org/10.1016/j.patter.2021.100336.

Raji, Inioluwa, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. "Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing," 145–51. AIES '20. ACM, 2020. https://doi.org/10.1145/3375627.3375820.

Robinson, Jonathan, Joseph Smyth, Roger Woodman, and Valentina Donzella. "Ethical Considerations and Moral Implications of Autonomous Vehicles and Unavoidable Collisions." *Theoretical Issues in Ergonomics Science* ahead-of-print, no. ahead-of-print (2021): 1–18. https://doi.org/10.1080/1463922X.2021.1978013.

Rudin, Cynthia. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1, no. 5 (May 2019): 206–15. https://doi.org/10.1038/s42256-019-0048-x.

Saxena, Nripsuta Ani, Karen Huang, Evan DeFilippis, Goran Radanovic, David C. Parkes, and Yang Liu. "How Do Fairness Definitions Fare? Testing Public Attitudes towards Three Algorithmic Definitions of Fairness in Loan Allocations." *Artificial Intelligence* 283 (2020): 103238–15. https://doi.org/10.1016/j.artint.2020.103238.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. "Fairness and Abstraction in Sociotechnical Systems." In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 59–68. FAT* '19. New York, NY, USA: Association for Computing Machinery, 2019. https://doi.org/10.1145/3287560.3287598.

Shaw, David, Bernard Favrat, and Bernice Elger. "Automated Vehicles, Big Data and Public Health." *Medicine, Health Care, and Philosophy* 23, no. 1 (2020): 35–42. https://doi.org/10.1007/s11019-019-09903-9.

Steed, Ryan, and Aylin Caliskan. "Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases." Ithaca: Cornell University Library, arXiv.org, 2021. https://doi.org/10.1145/3442188.3445932.

Treviranus, Jutta. "Sidewalk Toronto and Why Smarter Is Not Better." *Medium* (blog), October 31, 2018. https://medium.datadriveninvestor.com/sidewalk-toronto-and-why-smarter-is-not-better-b233058d01c8.

Trewin, Shari, Sara Basson, Michael Muller, Stacy Branham, Jutta Treviranus, Daniel Gruen, Daniel Hebert, Natalia Lyckowski, and Erich Manser. "Considerations for AI Fairness for People with Disabilities." *AI Matters* 5, no. 3 (December 6, 2019): 40–63. https://doi.org/10.1145/3362077.3362086.

Wang, Ruotong, F. Maxwell Harper, and Haiyi Zhu. "Factors Influencing Perceived Fairness in Algorithmic Decision-Making: Algorithm Outcomes, Development Procedures, and Individual Differences." In *Proceedings of the 2020 CHI Conference on Human Factors in Computing*

Systems, 1–14. CHI '20. New York, NY, USA: Association for Computing Machinery, 2020. https://doi.org/10.1145/3313831.3376813.

Washington, Anne, and Rachel Kuo. "Whose Side Are Ethics Codes on?: Power, Responsibility and the Social Good," 230–40. FAT '20. ACM, 2020. https://doi.org/10.1145/3351095.3372844.

Whittaker, Meredith, Meryl Alper, Olin College, Liz Kaziunas, and Meredith Ringel Morris. "Disability, Bias, and AI." AI Now Institute at NYU, 2019.

Wieringa, Maranke. "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*, 1–18. Barcelona Spain: ACM, 2020. https://doi.org/10.1145/3351095.3372833.

Williams, Betsy Anne, Catherine F. Brooks, and Yotam Shmargad. "How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications." *Journal of Information Policy (University Park, Pa.)* 8, no. 1 (2018): 78–115. https://doi.org/10.5325/jinfopoli.8.1.0078.

Williams, Rua M., Simone Smarr, Diandra Prioleau, and Juan E. Gilbert. "Oh No, Not Another Trolley! On the Need for a Co-Liberative Consciousness in CS Pedagogy." *IEEE Transactions on Technology and Society* 3, no. 1 (2022): 67–74. https://doi.org/10.1109/TTS.2021.3084913.

Wilson, Benjamin, Judy Hoffman, and Jamie Morgenstern. "Predictive Inequity in Object Detection," 2019. https://arxiv.org/abs/1902.11097.

---

[1] Milakis, van Arem, and van Wee, "Policy and Society Related Implications of Automated Driving."

[2] Bennett, Vijaygopal, and Kottasz, "Attitudes towards Autonomous Vehicles among People with Physical Disabilities"; Bradshaw-Martin and Easton, "Autonomous or 'Driverless' Cars and Disability"; Brewer and Kameswaran, "Understanding the Power of Control in Autonomous Vehicles for People with Vision Impairment"; Brinkley et al., "Exploring the Needs, Preferences, and Concerns of Persons with Visual Impairments Regarding Autonomous Vehicles"; Claypool, Bin-Nun, and Gerlach, "Self-Driving Cars: The Impact on People with Disabilities"; Deka and Brown, "Self-Perception and General Perception of the Safety Impact of Autonomous Vehicles on Pedestrians, Bicyclists, and People with Ambulatory Disability"; Dicianno et al., "Systematic Review"; Kuzio, "Autonomous Vehicles and Paratransit."

[3] See Identifying Automated Driving Systems-Dedicated Vehicles Passenger Issues for Persons with Disabilities (SAE International, November 2019), https://www.sae.org/standards/content/j3171_201911; AVs and Increased Accessibility Workshop Series (Alliance of Automobile Manufacturers, now the Alliance for Automotive Innovation, May/July/September 2019), https://www.autosinnovate.org/avaccessibility; Driverless Cars and Accessibility: Designing the Future of Transportation for People with Disabilities (ITS America, April 2019), https://itsa.org/advocacy-material/driverless-cars-and-accessibility/ ; and the USDOT Inclusive Design Challenge Resources page at https://www.transportation.gov/inclusive-design-challenge/resources. In addition to vehicle accessibility needs, advocates have noted the potential for algorithmic bias in pedestrian detection, collisions, and data, and implications for disabled travelers. This brief serves to address those concerns in depth and propose solutions. See the National Council on Disability's Self Driving Cars Mapping Access to a Technology Revolution report (November 2015), https://ncd.gov/publications/2015/self-driving-cars-mapping-access-

technology-revolution; and the Consortium for Constituents with Disabilities Transportation Taskforce AV Principles (May 2022), https://www.c-c-d.org/fichiers/CCD-Transpo-TF-AV-Principles-May-2022.pdf.

[4] Papa and Ferreira, "Sustainable Accessibility and the Implementation of Automated Vehicles."

[5] Koopman and Wagner, "Autonomous Vehicle Safety."

[6] Nakamura, "My Algorithms Have Determined You're Not Human."

[7] Bennett, Vijaygopal, and Kottasz, "Attitudes towards Autonomous Vehicles among People with Physical Disabilities"; Brewer and Kameswaran, "Understanding the Power of Control in Autonomous Vehicles for People with Vision Impairment"; Brinkley et al., "Exploring the Needs, Preferences, and Concerns of Persons with Visual Impairments Regarding Autonomous Vehicles"; Deka and Brown, "Self-Perception and General Perception of the Safety Impact of Autonomous Vehicles on Pedestrians, Bicyclists, and People with Ambulatory Disability."

[8] Bradshaw-Martin and Easton, "Autonomous or 'Driverless' Cars and Disability"; Kuzio, "Autonomous Vehicles and Paratransit."

[9] Binns and Kirkham, "How Could Equality and Data Protection Law Shape AI Fairness for People with Disabilities?"; Trewin et al., "Considerations for AI Fairness for People with Disabilities"; Whittaker et al., "Disability, Bias, and AI."

[10] Wieringa, "What to Account for When Accounting for Algorithms."

[11] Corbett-Davies et al., "Algorithmic Decision Making and the Cost of Fairness"; Mitchell et al., "Algorithmic Fairness"; Saxena et al., "How Do Fairness Definitions Fare?"

[12] Araujo et al., "In AI We Trust?"; Grgic-Hlaca et al., "Human Perceptions of Fairness in Algorithmic Decision Making"; Kasinidou et al., "I Agree with the Decision, but They Didn't Deserve This"; Wang, Harper, and Zhu, "Factors Influencing Perceived Fairness in Algorithmic Decision-Making."

[13] Ntoutsi et al., "Bias in Data-Driven Artificial Intelligence Systems—An Introductory Survey."

[14] Barabas et al., "Studying Up: Reorienting the Study of Algorithmic Fairness around Issues of Power"; Kasy and Abebe, "Fairness, Equality, and Power in Algorithmic Decision-Making."

[15] Barabas et al., "Studying Up: Reorienting the Study of Algorithmic Fairness around Issues of Power"; Kasy and Abebe, "Fairness, Equality, and Power in Algorithmic Decision-Making"; Selbst et al., "Fairness and Abstraction in Sociotechnical Systems."

[16] Campbell et al., "Sensor Technology in Autonomous Vehicles"; Cunneen et al., "Autonomous Vehicles and Avoiding the Trolley (Dilemma)."

[17] Buolamwini and Gebru, "Gender Shades."

[18] Keyes, "The Misgendering Machines."

[19] Wilson, Hoffman, and Morgenstern, "Predictive Inequity in Object Detection."

[20] Treviranus, "Sidewalk Toronto and Why Smarter Is Not Better."

[21] Selbst et al., "Fairness and Abstraction in Sociotechnical Systems."

[22] Barabas et al., "Studying Up: Reorienting the Study of Algorithmic Fairness around Issues of Power"; Kasy and Abebe, "Fairness, Equality, and Power in Algorithmic Decision-Making"; Selbst et al., "Fairness and Abstraction in Sociotechnical Systems."

[23] Nakamura, "My Algorithms Have Determined You're Not Human."

22

[24] Packin, "Disability Discrimination Using Artificial Intelligence Systems and Social Scoring."

[25] Koopman and Wagner, "Autonomous Vehicle Safety."

[26] Packin, "Disability Discrimination Using Artificial Intelligence Systems and Social Scoring."

[27] Koopman and Wagner, "Autonomous Vehicle Safety."

[28] Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead."

[29] Carabantes, "Black-Box Artificial Intelligence."

[30] Offert and Bell, "Perceptual Bias and Technical Metapictures."

[31] Carabantes, "Black-Box Artificial Intelligence"; Offert and Bell, "Perceptual Bias and Technical Metapictures."

[32] Offert and Bell, "Perceptual Bias and Technical Metapictures."

[33] Gurney, "Crashing into the Unknown"; Milakis, van Arem, and van Wee, "Policy and Society Related Implications of Automated Driving"; Robinson et al., "Ethical Considerations and Moral Implications of Autonomous Vehicles and Unavoidable Collisions."

[34] Williams et al., "Oh No, Not Another Trolley! On the Need for a Co-Liberative Consciousness in CS Pedagogy."

[35] Awad et al., "The Moral Machine Experiment."

[36] Bergmann et al., "Autonomous Vehicles Require Socio-Political Acceptance-An Empirical and Philosophical Perspective on the Problem of Moral Decision Making"; Bigman and Gray, "Life and Death Decisions of Autonomous Vehicles"; Bonnefon, Shariff, and Rahwan, "The Social Dilemma of Autonomous Vehicles"; Frank et al., "Human Decision-Making Biases in the Moral Dilemmas of Autonomous Vehicles."

[37] Bonnefon, Shariff, and Rahwan, "The Social Dilemma of Autonomous Vehicles"; Robinson et al., "Ethical Considerations and Moral Implications of Autonomous Vehicles and Unavoidable Collisions."

[38] Williams et al., "Oh No, Not Another Trolley! On the Need for a Co-Liberative Consciousness in CS Pedagogy."

[39] Leben, "A Rawlsian Algorithm for Autonomous Vehicles"; Millan-Blanquel, Veres, and Purshouse, "Ethical Considerations for a Decision Making System for Autonomous Vehicles during an Inevitable Collision"; Robinson et al., "Ethical Considerations and Moral Implications of Autonomous Vehicles and Unavoidable Collisions."

[40] Philosopher John Rawls suggested we should imagine we sit behind a veil of ignorance that keeps us from knowing who we are. By being ignorant of our circumstances, we can more objectively consider how societies should operate, and agree on principles of social and political justice. Deontology is an ethical theory that uses rules to distinguish right from wrong. Utilitarianism is a theory that determines right from wrong by focusing on outcomes. Source: University of Texas McCombs School of Business: Ethics Unwrapped Glossary (2002) https://ethicsunwrapped.utexas.edu/glossary

[41] Davnall, "The Car's Choice."

[42] Cunneen et al., "Autonomous Vehicles and Avoiding the Trolley (Dilemma)."

[43] Cunneen et al.; Davnall, "The Car's Choice."

[44] Hoffmann, "Where Fairness Fails"; Paullada et al., "Data and Its (Dis)Contents."

[45] Paullada et al., "Data and Its (Dis)Contents."

46 Barocas and Selbst, "Big Data's Disparate Impact"; boyd and Crawford, "Critical Questions for Big Data"; Hoffmann, "Where Fairness Fails"; Paullada et al., "Data and Its (Dis)Contents."

47 Paullada et al., "Data and Its (Dis)Contents."

48 Steed and Caliskan, "Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases."

49 Shaw, Favrat, and Elger, "Automated Vehicles, Big Data and Public Health."

50 Packin, "Disability Discrimination Using Artificial Intelligence Systems and Social Scoring"; Paullada et al., "Data and Its (Dis)Contents."

51 Packin, "Disability Discrimination Using Artificial Intelligence Systems and Social Scoring."

52 Nakamura, "My Algorithms Have Determined You're Not Human."

53 Treviranus, "Sidewalk Toronto and Why Smarter Is Not Better."

54 Williams, Brooks, and Shmargad, "How Algorithms Discriminate Based on Data They Lack."

55 Eubanks, *Automating Inequality*.

56 Whittaker et al., "Disability, Bias, and AI."

57 Trewin et al., "Considerations for AI Fairness for People with Disabilities."

58 Whittaker et al., "Disability, Bias, and AI."

59 boyd and Crawford, "Critical Questions for Big Data."

60 These proposed policy measures are inline with the 2022 White House Office of Science and Technology Policy Blueprint for an AI Bill of Rights. The Blueprint recommendations include consultation with diverse communities, proactive and continuous measures to protect against algorithmic discrimination, ensuring accessibility for people with disabilities, and protection from abusive data practices.

61 Andrus et al., "What We Can't Measure, We Can't Understand"; Dwork et al., "Fairness through Awareness"; Williams, Brooks, and Shmargad, "How Algorithms Discriminate Based on Data They Lack."

62 Trewin et al., "Considerations for AI Fairness for People with Disabilities."

63 Deitz, Lobben, and Alferez, "Squeaky Wheels."

64 Raji et al., "Saving Face."

65 Wieringa, "What to Account for When Accounting for Algorithms."

66 Trewin et al., "Considerations for AI Fairness for People with Disabilities."

67 Barabas et al., "Studying Up: Reorienting the Study of Algorithmic Fairness around Issues of Power"; Katell et al., "Toward Situated Interventions for Algorithmic Equity."

68 Aysolmaz, Dau, and Iren, "Preventing Algorithmic Bias in the Development of Algorithmic Decision-Making Systems: A Delphi Study."

69 Rudin, "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead."

70 Henin and Le Métayer, "A Framework to Contest and Justify Algorithmic Decisions."

[71] Citron and Pasquale, "The Scored Society."

[72] Washington and Kuo, "Whose Side Are Ethics Codes On?"

[73] Williams et al., "Oh No, Not Another Trolley! On the Need for a Co-Liberative Consciousness in CS Pedagogy."