Learning to Say No: When FATE is too Late

| Rua M. Williams  University of Florida  Gainesvile, FL, USA  Williams2020@ufl.edu |
| --- |

# ABSTRACT

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

* ACM copyright: ACM holds the copyright on the work. This is the historical approach.
* License: The author(s) retain copyright, but ACM receives an exclusive publication license.
* Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Times New Roman 8-point font. Please do not change or modify the size of this text box.

Each submission will be assigned a DOI string to be included here.

Among growing concerns about the disproportionate dangers AI advances pose to marginalized groups, proposals for a procedural solution to ethics in AI abound. As each framework for enforcing ethics has its exploits exposed, a new cog is added to the orrery. Perhaps it is time to consider that some systems may be inherently violent, even if they are fair. We are not going to program our way into justice. We have to learn to say no to building violent things.

## Author Keywords

Ethics in Artificial Intelligence; Ethics in Machine Learning; Critical Disability Studies; Science and Technology Studies.

## ACM Classification Keywords

# • Social and professional topics~Computing / technology policy • Social and professional topics~Surveillance

# • Social and professional topics~Medical information policy

# INTRODUCTION: When Fate is too Late

In recent years, engineers and ethicists have come together to define tools and frameworks for ensuring the ethical deployment of new technologies. In the field of “applied ethics”—or the application of moral theory to specific practical contexts [5:5]—the ethics of prescriptively applying ethics is rarely explored. I am here to trouble the very notion of finding a universal framework for perfecting ethical technology development. What is considered to be universally true will always be troubled by the particular [18]. Particular troubles shape our understandings of universal moral truths, but that work is never done. The particular is ever expanding, ever shifting, ever lacking in the fixity required for easy universal application.

In Artificial Intelligence (AI) and Machine Learning (ML), the dominant preferred framework for ethical analysis is FAT: an algorithm must be Fair, Accountable, and Transparent in order to be considered ethical [3]. In wrestling with the ways in which FAT can fail, some researchers have attempted to add additional terms to the acronym. For example, FATReCS (Fairness, Account-ability, Transparency, Responsibility, Compliance, Safety) attempts to resolve conflicts with the law (compliance), to acknowledge social responsibility as an additional context for fairness, and to account for the way that recommendation algorithms can shape human perception, thoughts, and beliefs [7]. Other researchers have attempted to map algorithms conforming to FAT guidelines to frameworks of applied ethics in other domains. Heidari et al. [6] explore different frameworks for Equality of Opportunity and propose different models that meet different definitions of fairness as defined by subdomains of parity (Accuracy, Statistical, Equality of Odds, Predictive Value). Passi and Barocas recognize and explore issues of fairness in the way that “problems” are defined and translated into ML features and parameters [13].

These examples make clear that even in practice, FAT is not enough to manage the complexity within and between particular AI applications. However, none of these examples suitably address the burgeoning question of ethics in AI: Should fairness even be the goal?

# Disability: Canaries For Justice

*The following section is adapted from a self-authored unpublished manuscript currently under review at the Canadian Journal of Disability Studies* [19]*.*

An image created by Craig Froehle underwent several viral permutations, as memes are wont to do [4]. In his original, one side of the image labeled ‘equality’ shows three people of varying heights stand on the same wooden crates. Only the tallest two people can see over the fence to view the baseball game in the background. The other side is labeled ‘equity’ and the boxes have been rearranged to give two to the shortest person, one to the middle, and none to the tallest. All three people can now see over the fence. The image was intended to illustrate that equality of opportunity rarely results in equity, or the equality of outcomes.

Future iterations of the image reinforced the concept that sameness and fairness have very different results. The same system can leave different people in different states of oppression. Variations on the image proliferate [4], and eventually, one emerges with the equality illustration on the left, equity in the middle, and finally, liberation on the right, where there is no longer any fence at all, and no boxes. In this version of liberty, there are no barriers, and thus there are no accommodations.

Other iterations replace the equity label with justice. It seems a great number of people felt that equity was either synonymous with justice, or that the image labeled equity was more accurately described as justice. In any case, there is agreement that equality is neither equity nor justice nor liberation. But if equality is none of these, and liberation is also somehow a progression of equity and/or justice, then what is equity, and what is justice?

Martha C. Nussbaum delinks justice from notions of legality— law, order, penalty, retribution. She describes how retributive notions of justice often deflect consequences from the original perpetrators to proxies [12]. In her analysis, retribution is a manifestation of ‘punitive implantation’. The classification of context-void, lawful (normative) consequence sows the seeds of vengeance, crime and punishment— a recursive injustice. Justice then, requires the context-sensitive application of equity in communion with mercy.

If equality is everyone living in the balance of their desire, and equity is everyone having the resources they need to reach the state of equality, then justice is the dismantling of the barriers that seed this state of need in the first place. Barriers implant states of deficit, inequity, injustice. Liberation is a consequence of justice. Equality alone can masquerade as justice in the presumption that everyone desires the same opportunities. Equity alone can masquerade as justice in the delivery of resources without the deconstruction of barriers. Liberation alone masquerades as justice when we are released from bondage into new sites of oppression.

How do we imagine algorithms for fighting the neoliberal perversions of justice? Disability Activists like Patricia Berne, Mia Mingus, Leah Lakshmi Piepzna-Samarasinha, and others describe Disability Justice, Access Intimacy, and Prefigurative Politics as tools of collective resistance, liberation, and love. To disability activists, Justice isn't something that is given or bestowed, nor even taken. It is something that is collectively forged.

Disability Justice centers intersectionality, interdependence, and collective action through a responsibility and accountability to cross-movement solidarities [1]. Disability Justice was born of a need to pull the Disability Rights Movement out of its white-centric reverie. “Disability Justice asserts that ableism helps make racism, christian supremacy, sexism, and queer- and transphobia possible, and that all those systems of oppression are locked up tight” [14:22].

Where Disability Rights are often imagined as “Justice through legally recognized equality of opportunity”, Disability Justice recognizes how the process of law itself can be an access barrier, and how the enforcement of law is inequitable across intersections. Further, Disability Justice contests the naturalization and normalization of what is considered desirable opportunity.

Disability Rights advances access through litigious criteria satisfaction. Access becomes a series of procedures and checkboxes, accommodation and justification— Logistics [10]. When our liberation is imagined as logistical, we become vulnerable to false liberation. Inclusion and Diversity masquerade as Access, Equity, and Justice. Instead, Access Intimacy “moves the work of access out of the realm of only logistics and into the realm of relationships and understanding disabled people as humans, not burdens” [11]. Access Intimacy is a tool of Disability Justice and of Collective Liberation because it calls upon a collective sense of belonging and worth—it is a “hermeneutics of love” [16].

To sow the seeds of a collective, radical, revolutionary, loving justice, we have to dream disabled futures. “Prefigurative politics is a fancy term for the idea of imagining and building the world we want to see now. It’s waking up and acting as if the revolution has happened” [14:149]. Prefigurative Politics compels us to imagine disability not as a problem, but as a site of unique experiential knowledge about how the world can be made more just. Like canaries in the coal mine [2], Disabled experiences of injustice show us where societal injustices converge.

# Examples: Eugenics in the MacHINE

In Keyes et al.’s ‘A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry’ [8], the authors explore how well such a system conforms to the FAT procedure of ensuring ethical AI. In this absurdist illustration, the mulching system is *Fair*: all people across race and gender groups have a roughly equal probability of being selected for mulching. This is achieved by supplying excess training data in statistically marginal identities. The previous version of the system was “unfair” in that cis white men, represented by a majority of the training data, were disproportionately likely to be selected for mulching.

The Mulching System is also Accountable and Transparent according to regulations and guidelines established by the developing company. Here is one of the major flaws in prescriptivist solutions like the FAT(E) Framework. Who defines accountability? Who defines transparency? Even if there were considerable government oversight of AI developers, has governmental policy been historically shown to be fair, accountable, and transparent to the people under its power? Marginalized people the world over have quite a lot to say about how they are regarded by their own governments.

For example, it is presently common practice for judges to determine sentencing based on the results of the COMPASS algorithm that predicts recidivism risk. This algorithm has been shown to be significantly biased against black communities [15,17]. The developers failed to account for the way that data reflects cultural biases [9]. Black communities are overpoliced, and laws are designed specifically to criminalize poverty, blackness, and survival. COMPASS can’t give the risk that someone will re-commit, it can only give a risk that someone will be caught again. However, even this statement is not fully true. COMPASS hasn’t actually been designed to accurately predict an individual’s recidivism risk by any definition. It has only been designed to predict the statistical likelihood of someone with similar *data features* to the subject will be a repeat offender.

Even if COMPASS produced risk calculations that were “fair” and unbiased across all aspects of identity—race, gender, disability, class, sexuality, location of residence, and so on—Is it ever fair or just to decide an individual person’s future based on the statistical aggregate of strangers?

As machine learning moves into the healthcare domain, the implications for AI on disabled people become frankly dystopic. As algorithms are designed to predict which treatments may have the highest chance of success at the cheapest cost, as we build systems which determine dosing, medication alternatives, risk of patient relapse, and so on [20,21]—I question our collective devotion to reckoning with the legacy of eugenics endemic to the work.

Statistics was built for eugenics. The bell-curve, the gaussian, the normal distribution—this symbol is evocative of the culling and shaping of herds. These techniques are inseparable from the mathematics of machine learning. Even if we “overstuffed” our data with disabled features to combat their “outlier status”, there is no medical data that adequately supports disabled life in a society that is so determined to snuff it out. Until my friends can enter a hospital without being handed a DNR as if it were as routine as a sterile sample cup, we are not ready to allow AI to determine medical treatment.

# Position Statement: Learning to Say No

In each of the scenarios presented above, whether they were absurdist, contemporary, or eminent, asking how to make the system fair was the wrong question. We should be asking whether or not the system should be built at all. I am looking for just present, not a future. I am looking for justice today, not later. I am not interested in debiasing data sets and procedures for fair algorithms when we haven’t reckoned with the eugenics inherent in the method from the very start. I don’t mean to say that AI is doomed to wickedness. But I do think we need a lot more people in the room who are worried about it than we have right now.

What would it mean to turn our math around? Rather than looking to big data to find solutions to hegemonically defined problems, what if we used it to find the catalysts of inequitable problems themselves? Some troubling of what constitutes a “problem” is necessary. I do not mean that we should focus our efforts on the detection of “problem bodies”. That’s part and parcel of what we are already doing. A reversal of lenses is required. What would it mean to privilege the outlier? To ask ourselves—what are the conditions in which the outlier is culled? What if we used AI to find the pruning mechanism and dismantle it?

# REFERENCES

1. Patricia Berne, Aurora Levins Morales, and David Langstaff. 2018. Ten Principles of Disability Justice. *Women’s Studies Quarterly* 46, 1 & 2 (2018), 227–230.
2. Jesse Cohen, Carolyn Lazard, and Bonnie Swencionis. 2019. We Are Canaries | Like the “canary in a coalmine,” our autoimmune and other chronic health conditions are a warning about larger imbalances in the world. This is a place to process and explore being a Canary. Retrieved July 1, 2019 from

http://wearecanaries.com/

1. Nicholas Diakopoulos. 2016. Accountability in algorithmic decision making. *Communications of the ACM* 59, 2 (January 2016), 56–62.
2. Craig Froehle. 2016. The Evolution of an Accidental Meme. *Medium*. Retrieved February 17, 2019 from https://medium.com/@CRA1G/the-evolution-of-an-accidental-meme-ddc4e139e0e4
3. Sven Ove Hansson (Ed.). 2017. *The Ethics of Technology: Methods and Approaches*. Rowman & Littlefield International, London, England, UK.
4. Hoda Heidari, Michele Loi, Krishna P. Gummadi, and Andreas Krause. 2019. A Moral Framework for Understanding Fair ML through Economic Models of Equality of Opportunity. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT\* ’19*, 181–190.
5. Toshihiro Kamishima, Pierre-Nicolas Schwab, and Michael D. Ekstrand. 2018. 2nd FATREC workshop. In *Proceedings of the 12th ACM Conference on Recommender Systems - RecSys ’18*, 516–516.
6. Os Keyes, Jevan Hutson, and Meredith Durbin. 2019. A Mulching Proposal: Analysing and Improving an Algorithmic System for Turning the Elderly into High-Nutrient Slurry. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems - CHI EA ’19*, 1–11.
7. Mark MacCarthy. 2017. Standards of Fairness for Disparate Impact Assessment of Big Data Algorithms. *Cumberland Law Review* 48, (2017). Retrieved June 28, 2019 from

https://heinonline.org/HOL/Page?handle=hein.journals/cumlr48&id=75&div=8&collection=journals

1. Mia Mingus. 2011. Changing the Framework: Disability Justice. *Leaving Evidence*. Retrieved February 17, 2019 from

https://leavingevidence.wordpress.com/2011/02/12/changing-the-framework-disability-justice/

1. Mia Mingus. 2011. Access Intimacy: The Missing Link. *Leaving Evidence*. Retrieved February 17, 2019 from https://leavingevidence.wordpress.com/2011/05/05/access-intimacy-the-missing-link/
2. Martha C Nussbaum. 1993. Equity and Mercy. *Philosophy & Public Affairs* 22, 2 (1993), 83–125.
3. Samir Passi and Solon Barocas. 2019. Problem Formulation and Fairness. In *Proceedings of the Conference on Fairness, Accountability, and Transparency - FAT\* ’19*, 39–48.
4. Leah Lakshmi Piepzna-Samarasinha. 2018. *Care Work: Dreaming Disability Justice*. Aresenal Pulp Press, Vancouver, CA.
5. David G. Robinson. 2017. The Challenges of Prediction: Lessons from Criminal Justice. *I/S: a Journal of Law and Policy for the Information Society, Forthcoming* (October 2017).
6. Chela Sandoval. 2000. *Methodology of the Oppressed*. University of Minnesota Press, Minneapolis, MN.
7. Christina Wadsworth, Francesca Vera, and Chris Piech. Achieving Fairness through Adversarial Learning: an Application to Recidivism Prediction. Retrieved July 12, 2018 from

https://www.fatml.org/media/documents/achieving\_fairness\_through\_adversearial\_learning.pdf

1. Traci C. West. 2006. *Disruptive Christian Ethics: When Racism and Women’s Lives Matter*. WJK Press.
2. Rua M. Williams. Fractal Mechanics: Resisting Transinstitutionalization via Prefigurative Politics. *Canadian Journal of Disability Studies*. (under review)
3. Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fiting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, 1–11.
4. Dimitrios Zikos, Ismail Vandeliwala, and Philip Makedon. 2014. A probabilistic algorithm with user feedback loop for decision making during the hospital triage process. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA ’14*, 1–2.