Defining Algorithmic Justice

TWIN CITIES INNOVATION ALLIANCE

Who we are

The Twin Cities Innovation Alliance (TCIA) is a social venture, intended to **spark**, **resource, and guide** entrepreneurs as they grow and scale their businesses across the Twin Cities, operating out of the need for greater diversity, inclusion and equity in technology and entrepreneurship. After 15 years of self-funded initiatives from programing and volunteering in Saint Paul, we launched TCIA with initial seed investment from the Knight Foundation.

Twin Cities Innovation Alliance (TCIA) is a coalition of stakeholders representing a cross sector of public, private and community organizations, corporations and institutions led by visionaries, academics, thought leaders and individuals who are invested in the Twin Cities' continued evolution as a forward-thinking, innovative, 'Smart' global city.

Our mission

Our mission is to build and develop a critical mass of diverse, highly engaged residents, policy makers, and entrepreneurs, made up of minorities and people of color traditionally identified as the end users and consumers of innovation and design, and transforming them into the purveyors and beneficiaries. This will benefit all communities across the nation and our world. We exchange learnings while adapting and evolving our collective work.

What is an algorithm?

Before we can begin to think about algorithmic justice, we must think about what exactly an algorithm is.

At first glance, algorithms can appear difficult to understand. Written by people who-knows-where in the realm of computer science, they are formulas that drive artificial intelligence systems (which is just a fancy way of saying automated technology). Moreover, as the world rapidly becomes more reliant on such technology, algorithms have become an integral force behind how we navigate everyday life. They dictate which advertisements you see on social media, whether your job application is approved for review, which neighborhoods might be more heavily policed, and so much more. Nonetheless, it is not that algorithms are necessarily complex, because in reality, they are not. Algorithms are nothing more than a recipe for making decisions. To understand algorithms, we simply need to demystify them.

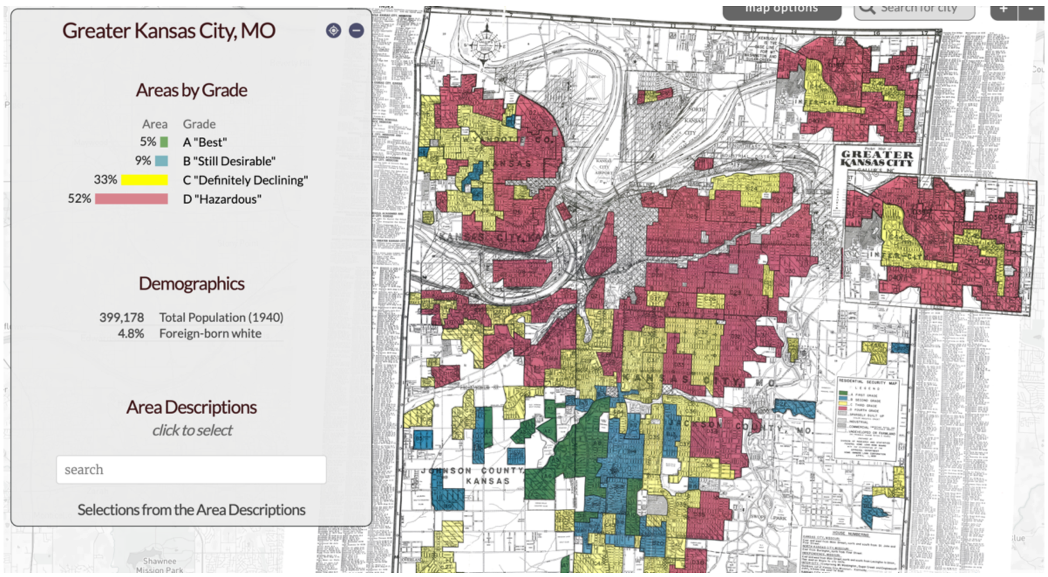
First, it is important to recognize that algorithms do not come from artificial intelligence systems, rather artificial intelligence systems come from algorithms; a computer scientist sits down and uses datasets to craft an algorithm that tells an AI system what to do. And given that humans are notorious for being biased, whether consciously or not, that leaves a lot of room for bias to creep its way into AI systems. Especially if the computer scientist does not scrutinize where the data that they use comes from. Just imagine if a corporation used an algorithm that automatically rejected applications from those who have periods of unemployment. Such an algorithm would seem to discriminate against someone who may have faced an extenuating circumstance like a disability that rendered them unable to work. Or imagine an algorithm that examined previous job applications over a period to determine the best candidate for a tech position. It would only make sense that a history of exclusion of women in the tech industry would make women appear as less ideal candidates. After all, the so-called ideal applications would disproportionately be from men based on historical patterns. Well, these are the exact kinds of biased algorithms that have been developed by corporations like Amazon! Given that humans produce these datasets and write these algorithms, human bias is at the heart of AI systems. Thus, AI systems can never be as objective as we are made to believe. Furthermore, this is the same algorithmic bias that has oppressed minoritized communities since the dawn of U.S. history.

Let's take a deeper look at the history of algorithmic bias!

To understand the continuity of algorithmic bias, consider the history of housing in America. In the 1930s, the federal government created the Home Owners Loan Corporation (HOLC) to determine who could refinance their mortgages. This was supposed to help curb the growing number of foreclosures taking place due to the Great Depression. Simply put, the HOLC was tasked with writing algorithms. To do so, the HOLC sent people all across the United States to appraise neighborhoods. The information that was collected about housing included construction type, average age, repair condition, occupancy, price range, and more. Using this data, the HOLC sorted neighborhoods into the categories of "best, still desirable, definitely declining, and hazardous." Naturally, the best neighborhoods were granted opportunities to refinance while so-called hazardous ones were left unsupported. However, although this was alleged to be objective—after all every neighborhood was appraised the same way—racial bias still informed this process. When HOLC appraisers would compile their data, they considered the presence of black people a detrimental influence, noting the "infiltration of negros" in certain neighborhoods and giving them poor ratings as a result. This racist use of algorithms, known as redlining, explains how minoritized communities were pigeonholed into certain geographic locations and deprived of equal opportunity compared to their white counterparts. It is also important to note that even beyond the blatant consideration of race in these algorithms, the inclusion of datasets such as average income, zip code, etc. was inherently racist as well. These datasets correlate heavily with race, as the historical oppression of the Black community has undoubtedly hindered their economic prosperity. Therefore, even if race is not explicitly considered in an algorithm as it was with redlining, that does not mean an algorithm is race blind. After all, a race blind algorithm does not exist.

Fast forward today, and we can find that such algorithms (and their harmful impacts) continue to drive racial disparity. AI systems used by, say, the Federal Housing Administration (FHA) still used deeply racialized data to determine who gets a home loan. Or consider Facebook who was sued for providing different services to users based on their race, gender, and age. Both of these contemporary practices exclude minoritized demographics from opportunity just like the Home Owners Loan Corporation did—an organization that is now universally considered racist.

So, just how the racialized algorithm used to redline gave way to maps like this…



Modern AI algorithms give way to headlines like this…


The New York Times

Facebook Faces a Reckoning for Redlining

The government says that advertising designed to exclude certain groups violates the Fair Housing Act.

March 29, 2019

These discriminatory practices today are simply better able to hide behind seemingly complex technology and obscured datasets.

The paradox

So, why do we hold the FHA or Facebook to a different standard? Is it because their algorithms are better able to hide behind inaccessible technology that is supposed to be objective? The algorithms of today and those of history are one and the same. As sociologist and Princeton Professor Ruha Benjamin so aptly coined it, today we are simply living in "the new Jim Code." We are made to believe that artificial intelligence takes the human out of decision-making, therefore championing unbiased objectivity. However, such thinking could not be more flawed. That is why we must really consider where the datasets used to create these algorithms come from, and how that might impact certain communities across the globe.

Where does that leave us today?

If bias continues to drive algorithms, then what *has* changed since the rapid expansion of technology? First, is the increased pervasiveness of these algorithms. It is not hard to see that AI is everywhere. AI systems are what power cities all across the United States from surveillance systems to law enforcement to how economic, educational, and other opportunities are dispersed. Next is the magnified impacts of these algorithms. Given that algorithms are everywhere, they are constantly impacting people all the way down to which digital ads show up on your Facebook feed (or Instagram, Snapchat, and Tik Tok for all of the younger folk). Furthermore, this biased decision-making does not only harm individuals, but we must understand how it harms society at large. Faulty algorithms have impacts on the families of those who are discriminated against, the community in which such families reside, and they engrain themselves into the normative structure of daily life. Can you imagine a world where decision making was truly objective and, therefore, just? Finally, is the increasing ability of these algorithms to hide behind seemingly complex, impenetrable AI systems. It has been shown that at their core, algorithms are simple. They are nothing more than recipes. However, as AI systems become infinitely more complex, we lose vital accessibility to their algorithms. Consequently, we tend to lose touch with just how much human bias exists in these systems. Indeed, an algorithm can only ever be as equitable as the society from which it spawns. For if discrimination along the lines of race, gender, and age have been normalized throughout history, why would the writers of an algorithm be equipped to reverse this trend?

In short, artificial intelligence systems can never truly be objective. We cannot create algorithms that don't explicitly consider, say, race or gender then, poof, they are magically race and gender blind. All datasets, be it a zip code where somebody lives, their income, etc. correlate heavily with racialized and gendered histories. This inherently gives algorithms that run AI systems a startling amount of human bias. From here on out, if we constantly remind ourselves that objectivity does not exist, we can begin to think about a future with not just algorithms (because let's be real, they are not going away), but algorithmic justice—where algorithms do not ignore the bias of datasets but, instead, work to recognize and mitigate the harms of their inevitable bias.

Algorithmic justice in the United States

Legislation that regulates artificial intelligence systems justly and effectively is scant to say the least. Currently in the United States, there is no federal legislation aimed at standardizing the regulation of artificial intelligence. As a result, there is a lot of inconsistency throughout the country, leaving AI basically unchecked. For example, given that there is no standard approach to regulating facial recognition in law enforcement—something that can have a drastic effect on the lives of everyday citizens—the City Council in Baltimore banned such technology whereas very similar measures outright failed in the Michigan state legislature. Moreover, there are states that have a relatively expansive range of regulatory bills such as California while others do not have any bills at all like Mississippi (and essentially the entire Southeast for that matter). Once again, it seems paradoxical that one's personal information can be used without their consent in one state yet protected in another. This makes it imperative that stronger federal legislation is passed, as artificial intelligence should not impose on one's life differently depending on their geography. Only with strong federal legislation that is equipped to confront the United States' history of algorithmic discrimination can *all* Americans be protected from the possible harms of artificial intelligence systems.

How does the United States compare abroad?

The state of algorithmic justice abroad is still rather bleak when compared to an ideal standard of algorithmic justice (which we will discuss more in-depth momentarily). States with burgeoning artificial intelligence systems like China have passed a sweeping range of legislation that addresses AI. However, for the most part, legislation in China seems to largely expand the influence of artificial intelligence. It is important to note that China has passed seemingly progressive laws that regulate how businesses use citizens' personal information. A case in point is the Personal Information Protection Law in 2021. Nonetheless, such legislation does little to regulate AI systems at large. For example, China has passed various laws that establish their growing Social Credit System which surveils its citizens and gives them a "social score" based on their behavior. Further legislation has then blacklisted those with low social scores from activities such as flying. So, although there are some AI regulations in China, it is nominal at best. The country lacks proper measure to truly achieve algorithmic justice, placing them on a shockingly similar plane to the United States.

On a more hopeful note, the one-of-a-kind model that best achieves algorithmic justice is the General Data Protection Regulation (GDPR) implemented in the European Union in 2018. This monumental legislation revolutionizes citizens' autonomy over their information, giving them power to control if and how it is used by AI systems. Citizens can demand their personal information to be erased by handlers if they wish, they can opt in and out of AI systems, and penalties are put in place for if they fail to abide by these measures. Moreover, the GDPR takes the important step to standardize AI regulation across all member countries, eliminating much of the inconsistency that currently plagues the United States.

Although the United States is a long way from having strong, federal legislation such as the GDPR, there is local legislation that has a similar, positive effect. In 2018, the California Consumer Privacy Act (CCPA) was passed that, too, protects citizens' personal information from businesses. This is currently the best attempt at algorithmic justice in the United States and is a big step forward. However, local legislation will never be enough. One progressive locality does little to curb the discriminatory nature of uncheck AI systems throughout the United States at large. Thus, good legislation like the GDPR and CCPA should not be seen as the harbinger of algorithmic justice, but rather a baseline for future AI regulation that must follow, both at home and across the globe.

Sources

The 90-year-old financial policy that harms our health. A brief history of redlining. (n.d.). https://a816-dohbesp.nyc.gov/IndicatorPublic/Closerlook/redlining/index.html.

Balraj. (2021, June 22). *Baltimore city Council to ban facial recognition technology*. TechStory. https://techstory.in/baltimore-city-council-to-ban-facial-recognition-technology/.

Burgess, M. (2020, March 24). *What is GDPR? The summary guide to GDPR compliance in the UK*. WIRED UK. https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018.

Calhoun, K. (2020, July 29). *Race after technology*. AAIHS. https://www.aaihs.org/race-after-technology/#:~:text=Benjamin%20defines%20the%20New%20Jim,6).

The Editorial Board. (2019, March 30). *Facebook faces a reckoning for redlining*. The New York Times. https://www.nytimes.com/2019/03/29/opinion/facebook-discrimination-civil-rights.html.

Hickok, M. (2021, February 16). *Why was your job application rejected: Bias in Recruitment algorithms? (part 2)*. Montreal AI Ethics Institute. https://montrealethics.ai/why-was-your-job-application-rejected-bias-in-recruitment-algorithms-part-2/.

Mapping inequality. Digital Scholarship Lab. (n.d.). https://dsl.richmond.edu/panorama/redlining/.

Owens, D. (2020, July 13). *Facebook engages in online segregation and redlining through discriminatory advertising system, lawyers' committee argues*. Lawyers' Committee for Civil Rights Under Law. https://www.lawyerscommittee.org/lawyers-committee-confronts-facebooks-attempts-to-dismiss-digital-redlining-lawsuit-against-its-housing-advertisements/.

Our framework for ideal algorithmic justice

To establish a framework for algorithmic justice, our team has identified five key criteria pulling from good legislation like the GDPR and different conceptualizations of algorithmic justice. Our criteria include concreteness, accountability, community engagement, clarity, and transparency. Although we found that there is currently no policy in the United States that adequately checks off each box (in fact, the majority are nowhere close), we feel as though the presence of these rather idealistic criteria would signal a bill that is most equipped to promote algorithmic justice. Thus, we should not settle for anything less moving forward. A breakdown of each criterion is as followed:

TRANSPARENCY

To ensure algorithmic justice, transparency must be a crucial component of every bill that is passed. It is imperative that the everyday citizen is provided a clear understanding of how the bill will be carried out and how the bill will affect their life. Questions we have been asking to evaluate transparency are:

- Is it clear who the bill aims to benefit, e.g., does it expand corporate influence or protect the consumer?
- Is there clarity of purpose?
- Does the bill analyze and note the minoritized communities it may affect?
- If the bill requires funding, does it state where it will come from?
- Are the impacts (both good and bad) of the bill made accessible to the public?

CONCRETENESS

We believe that a concrete plan leads to better governmental practice, as there is less room for manipulation upon implementation. However, that is not to say that there should not be room for change. There should be enough fluidity so that the government can alter regulations to meet new needs as they arise. Concreteness simply refers to whether the bill contains actionable steps to be effectively rolled out. Questions we have asked to evaluate concreteness are:

- Is there an actionable plan in place to implement the policy?
- Is there a pre-implementation transition period?
- Is the legislation ready for immediate implementation?

ACCOUNTABILITY

To ensure that a regulation is duly enforced, our team looks at whether there are explicit measures in place to achieve accountability. A bill must have a feedback mechanism that allows us to discover if there is a problem. Moreover, should a problem arise, there must be a measure in place so that the feedback can actually influence the algorithm, and the algorithm can be made more just. Such measures can range from fines to technology rollbacks. Questions we asked to evaluate accountability are:

- Is accountability mentioned in the bill?
- Are there measures that actively search for harms of an algorithm?
- Are there repercussions in place for those who fail to abide by regulations, e.g., fines, recalls, etc.?
- Is there some other protocol to ensure accountability?
- Are there institutions clearly outlined that will provide this oversight?

COMMUNITY ENGAGEMENT

Given that the public is the demographic who will be most affected by AI systems, we looked at whether there was ample opportunity for them to offer input before the bill is passed. In some cities like Cleveland, Ohio, we found that the legislature passed many of their policies via "emergency ordinances" which allows them to skip the normal three hearings a bill must go through. This would seem to circumvent vital community engagement. We also looked at whether public input can influence the final legislation. For example, China mandates a 30-day public comment period before legislation is passed, though it is oftentimes questioned how much weight public comments hold. Moreover, we examined whether the legislative body made opportunities for community engagement accessible or, instead, whether the burden was placed on the public to voice their concerns. Questions we asked to evaluate community engagement were:

- Was there an opportunity for community engagement?
- Who initiated opportunities for community engagement?
- Was the window for community engagement a reasonable length?
- Did the bill pass through all the hoops of the democratic process, or was it rushed through such as an emergency ordinance?

CLARITY

With AI systems being a complex area, oftentimes expertise is needed to begin to understand them. As a result, looked at whether bills defined the systems and practices it aimed to regulate. Doing so would allow for greater public oversight, as they would be able to comprehend the systems that affect them daily. For example, Assembly Bill A680A in New York explicitly requires corporations to correspond with consumers at an 8th grade reading level or below. Questions we asked to evaluate clarity were:

- Are the different measures clearly and precisely defined? For example, the GDPR does not explicitly define what "good security measures" are despite mandating them, causing a lot of confusion and leeway.

- Is it clear who the actors are in the legislation, e.g., who will carry it out, who it will affect, etc.?

TWIN CITIES INNOVATION ALLIANCE

There were many existing organizations and bodies of work that influenced our conceptualization of a framework for algorithmic justice. Below are a few!

Approach #1: The Algorithmic Justice League

The AJL was founded in 2016 by a computer scientist named Joy Buolamwini. Their goal, like the Twin Cities Innovation Alliance and the JUST Data Lab, is to raise awareness about the implications of artificial intelligence by intersecting the disciplines of art and research. You can check out a TED Talk from Joy here where she talks about fighting algorithmic bias.

AJL breaks down ethical AI into three main sections, the first one being *agency and control*. This places power in the hands of the public, as the public must be aware of how AI systems function around them and who controls them. Next is *affirmative consent*. This means people are given the opportunity to opt-in and opt-out of an AI system without suffering any penalty. The final section is *centering justice by focusing on impermissible use.* Their definition of justice requires that AI is not used to expand the control that governmental and commercial entities have over the public, e.g., surveillance and policing.

Next, they break down accountable AI into three sections. The first one is *meaningful transparency* which means users of an AI system must provide an explanation of how exactly the system works. The second component is *continuous oversight*. Under this standard, AI systems are held accountable by an accredited third party. Lastly, an accountable system must *redress harms.* In other words, there must be an actionable plan set out for the public to easily “contest and correct” harmful decisions from an AI system.

For more information, check out the AJL website here.

Approach #2: The Berkman Klein Center at Harvard University

The Berkman Klein Center is a part of Harvard University. They focus their work on researching and understanding cyberspace. With that, they have multiple bodies of work centering on how to achieve fairness in AI systems.

First, the Berkman Klein Center finds that current approaches to AI regulation do not require an explicit analysis of the harms an algorithm can have. This is in stark contrast to other common uses of personal information like "*the ethical framework for the protection of human subjects participating in research.*" This rigorous framework mandates that the possible harms that a research subject may face are discovered and taken care of before the subject ever participates.

To achieve this more just end, the Berkman Klein Center points out four things that should be considered in an analysis of algorithmic harm:

1. One must identify the major choices in algorithmic design, implementation and application that have the potential to affect someone's well-being.
2. One must assess the effects of these decisions on the wellbeing of individuals.
3. One must measure well-being broadly, to include lifetime wealth, health, longevity, subjective life satisfaction, and the ability to make ethically relevant choices.
4. One must recognize algorithmic unfairness as choices in algorithmic design, implementation, and application that have disproportionate effects on members of different groups.

Their work heavily considers the *breadth of algorithmic harm*. They note that harms extend beyond an individual affected by a decision and impacts family members, communities, and society at large. For example, they note that children with incarcerated family members, which is a plausible harm of AI decision making, experience mental illnesses at higher rates.

For more information on the Berkman Klein Center's work, look here.

BERKMAN KLEIN CENTER
FOR INTERNET & SOCIETY
AT HARVARD UNIVERSITY

Approach #3: The Santa Fe Institute

Researchers for the Santa Fe Institute and New Mexico University have banded together to fight for algorithmic justice. The backgrounds of these researchers range from computer science, political science, mathematics, and law, which exhibits the importance of having an AI system created and audited by a cross perspective (people from different disciplines) against its outcomes.

The first focus area of this team is the intersection of artificial intelligence and housing. They argue that the most important way to secure justice is to *interrogate the historical and geographic data* that is used to make decisions, e.g., approving or denying a loan or accepting a rental application. Only by doing this can the pervasive biases in data collection be illuminated. It is of the Santa Fe Institute's opinion that the best recourse for AI systems in this arena is to subject them to *a test of disparate impact* before they are accepted.

The next focus area of this team is the criminal justice system. Their procedures to achieve racial justice center around issues of transparency. Therefore, when they testified before the New Mexico Legislature, they came up with *four key questions* to ensure justice is held at the forefront of AI decision making.

1. How does the algorithm work? Can everyone (defendants, prosecutors, judges) understand how a score was obtained?
2. Can we validate its performance independently? How well does it work on our local population in New Mexico?
3. When should a human be in the loop? Should an algorithm ever be used for detention before a trial?
4. What does the data really mean? Does a single zero or one capture the full story behind a failure to appear or rearrest?

For more information on the Santa Fe Institute's work, explore their site here.