Who is Responsible for Algorithmic Bias?

And what can we do about it?

What is an algorithm?

An **algorithm** is a series of steps used to solve a computational problem.

Algorithmic bias is a series of systematic and repeatable errors in a computer system which generate unfair outcomes such as privileging one group over another, highlighting certain results, etc.

Where do we see algorithmic bias?

As processes become increasingly more automated, jobs that were once done by humans are now done by computers. Because of this, we see algorithmic bias cropping up more and more. When humans did these jobs, they were often governed not just by processes of fairness and transparency, but also by federal, state, and local laws.

Let's look at some examples of algorithmic bias in practice.

In 2014, Amazon implemented an algorithm that would screen resumes for its engineering roles. Over time, the algorithm learned that Amazon hires far more men than women for its engineering roles and began to favor men's resumes over women's resumes.

The algorithm mistakenly assumed that Amazon hired more men than women because men were better candidates for the engineering jobs, rather than because there are fewer women in engineering, overall, and therefore fewer women hired to engineering roles. The algorithm went so far as to downgrade resumes that contained the word "women's" and blacklist candidates out of women's colleges, before it was ultimately decommissioned.

In 2020, as the world began to lock down due to the COVID-19 pandemic, the number of users on video conferencing platforms like Zoom skyrocketed. Zoom users very quickly noticed issues in its facial recognition algorithms which were used to generate virtual backgrounds while keeping the user's face visible. The algorithm, when in use, would remove the heads of Black users, thinking that their faces were part of the background. The same was not the case for users of other skin tones.

A similar issue was also seen in Twitter's automated cropping algorithm. In 2020, users noticed that Twitter's cropping algorithm was removing Black faces from its image previews in Tweets. In response, Twitter conducted a study on this "cropping bias" and found that, between men and women there was an 8% demographic parity in favor of women, between Black and white individuals, there was a 4% demographic parity in favor of white individuals. Between Black and white women, there was a 7% demographic parity in favor of white women, and between Black and white men, there was a 2% demographic parity in favor of white men. As a result, Twitter changed its cropping algorithm to display standard aspect ratio images in full, i.e. without cropping.

Why does algorithmic bias exist?
Computers, themselves, are not smart.

Machines are built by people with bias, trained on data collected by humans with bias, and often tested with humans with bias. Because of this, algorithms risk replicating and amplifying human biases.

But what specifically causes algorithmic bias to occur?

Data set was bad or not comprehensive.
In 2015, users of Google Photos discovered that the automated tagging function of the application was labeling Black people as gorillas. This feature was implemented as a way of being able to search through photos more easily, identifying specific faces, objects, etc. When this was discovered, Google immediately put in a bug fix that removed the gorilla label entirely such that nothing within the photo library could be labeled “gorilla” and stated that they would be working on “longer-term fixes.”

As of 2018, Google Photos still could not identify any gorilla images. Wired Magazine tested 40,000 images of animals in Google Photos and the application could identify many animals including pandas and poodles, but returned “no results” for search terms “gorilla”, “chimp”, “chimpanzee”, or “monkey”. The application could, however, identify “baboon”, “gibbon”, “marmoset”, and “orangutan”.

Algorithm wasn’t flexible enough.
In 2018, Nijeer Parks was falsely accused of shoplifting and arrested based on facial recognition software used by local police departments. This is not an isolated case, more and more, facial recognition is being used by law enforcement, courts, and other entities to identify crime suspects and make arrests.

Facial recognition software is used by law enforcement around the world in order to identify potential suspects based on security camera footage and pictures. A study by Georgetown Law’s Center on Privacy and Technology found that over 117 million American adults (nearly 1 in 2) are in facial recognition networks used by police departments.

The databases used to populate these systems mainly come from drivers license photos and mugshot data. Historically, the data used to investigate crimes was mainly made up of fingerprint and DNA data, which could only be obtained through previous criminal arrests or investigations. Now, these databases are primarily made up of law-abiding citizens.

This has many consequences.

Algorithm wasn’t flexible enough. (Cont.)

Black Americans are more likely to be singled out, especially when mug shot data is used. The overrepresentation of Black faces in mug shot data means that Black faces are consequently more likely to be falsely matched and have more opportunity to be falsely matched, producing biased results.

Facial recognition tends to perform more poorly on darker faces generally. The research on this subject is limited, however, there could be a few factors at play. First, the dataset may not be diverse enough. Second, it could be that darker skin has less contrast and generally machine learning models are looking for very minute patterns. Third, security camera footage tends to be top down and drivers license photographs vary state by state which can cause discrepancies in results even when a person is present in the database.

In general, people have dopplegangers, people who look vaguely like them, and inevitably, looking at a billion people, statistically, there will be people who look very similar even if they are not related.

The team didn't think through all the use cases.
In 2019, researchers found that a health care risk-prediction algorithm used by hospitals across the United States (on more than 200 million patients) was giving racially biased results. The algorithm was used to help hospitals and insurance companies direct resources to patients that would benefit from "high-risk care management," such as specialized nursing, extra primary-care visits, etc. in order to preempt more serious medical care and hospitalization, thereby reducing cost.

The algorithm was found to use healthcare spending as a proxy for medical need, resulting in Black patients receiving lower risk scores than their corresponding White counterparts, while not accounting for the implicit racial bias in the healthcare system, which causes Black patients to have more distrust in the healthcare system overall, thereby making them less likely to request extra care and less likely to spend money on healthcare.

Humans interacted with an algorithm enough to manipulate it.
In 2016, Microsoft released their artificial intelligence chatbot Tay onto Twitter. Tay was designed to mimic the language patterns of a 19-year-old girl and learn through interacting with Twitter users. When it was released on March 23, 2016, Tay began replying to Twitter users and captioning photos to be memes.

Twitter users quickly exploited Tay, tweeting politically incorrect and inflammatory messages to the bot. Tay had been programmed with some topic blacklisting, including the murder of Eric Garner, for which the bot would generate default, safe, canned answers. However, the bot, as a result of these tweets, began to tweet racist and sexually-charged messages to other Twitter users.

Tay’s behavior was unexpected but understandable. Tay learned only by mimicking other users and had no concept of inappropriate behavior. The bot was only able to copy the deliberately offensive behavior that it was seeing from users with whom it was interacting.

Culture issues in technology

“Move Fast, Break Things”: this was Facebook’s old slogan. And while the slogan itself is now defunct, it remains a good description of the way much of the technology industry still operates. And it is detrimental to people, and detrimental to technology.

It's all of the above.

Who is responsible for algorithmic bias?

- Is it developers?
- Is it companies?
- Is it institutions?

It's all three.

Individual developers must do more to push internally, encourage their teams to fully test out their products before release, and encourage data cleanliness practices.

Companies have economic incentives to keep their platforms running even when there are known bugs and issues, even at the expense of their users.

Institutions must encourage students to understand the ethics and pitfalls of algorithms and bias, as well as institute ethical best practices in developing and maintaining algorithm hygiene.

What can we do to combat algorithmic bias?

There are things that all three of these groups (developers, companies, and institutions) are already doing, but we need to do more.

The problem with the current approach is that they are mostly reactionary. It needs to be more proactive.

When problems occur, we tend to put “band-aids” on problems until algorithms can be changed, if they can be changed at all. Truly though, these problems should not have occurred at all, and are preventable. Many times, band-aids are never fixed and become permanent but incomplete solutions to a problem.

We know how to circumvent this approach, because we're already doing it in cybersecurity. In cybersecurity, experts are hired to purposefully break into systems and find gaps in security. The same can be applied to the problem of algorithmic bias.

What can individuals do to combat algorithmic bias?

The most important thing that an individual can do is to change the way that they think about algorithms and technology. We, as humans, have a tendency to hold technology up as morally superior and think that it stands above human bias. But the truth is, the technology that we have cannot exist without data that is produced through exclusion, bias, and inequity.

The tendency to lift technology, innovation, and technology CEOs and to think that "they can do no wrong," or that "they work faster, and therefore, better than humans," is inherently flawed. And always wanting the newest, shiniest thing, the newest technology, and wanting it fast, perpetuates the neverending fast-paced cycle, and the idea that technology is better than humans, and not fallible, contributes to more algorithmic bias and more issues in technology, overall.

Individuals should report problems when they see them.

This is an incomplete solution, because these systems are often audited by real people. Content moderators at social media platforms like Tik Tok and Facebook have talked about the explicit, violating, and sometimes violent behavior they've had to screen on their platforms, even going so far as to sue these companies for the mental strain and post traumatic stress care caused.

Moreover, these decisions are left to personal assessment by content moderators. Do we want these individuals to be our moral compass? And how do we fix this problem on a global scale? Especially when the content is not in English.

We need assurances that once something is reported and removed, it cannot be put up again.

Luckily, there are some companies working on solutions to this problem. Twitch, a global streaming platform, bans IP addresses rather than user ids making it more difficult for problematic users to rejoin the platform.

Still, there are benefits to reporting problems on technology platforms when you see them. The primary of which is that it holds tech companies accountable to their users.

What can technologists do to combat algorithmic bias?

Developers must learn to incorporate bias mitigation and detection techniques at every stage of the software development lifecycle, from analysis and planning to development to testing and maintenance.

Development teams can institute bias testing as part of their regular quality assurance and maintenance plans, as well as include regular data quality checks throughout the development process. It is important to remember that more data does not necessarily mean good data.

What can companies do to combat algorithmic bias?

Institutions and third parties can audit algorithms.

Auditing algorithms is crucial in mitigating algorithmic bias, and in reducing pressure on content moderators. Many, if not most, machine learning algorithms, which rely on big data, are black-boxes. However, there are many institutions, corporate, academic, and non-profit that are looking at frameworks that can audit algorithms and provide insight into an algorithm's behavior, even without technical expertise.

The AI Now Institute at New York University has introduced a framework that governmental institutions can use to create Algorithmic Impact Statements (AIAs) which are able to evaluate the potential detrimental impact of an algorithm. This is not a new concept and has been created in the same vein as environmental, privacy, data, and human rights impact statements. Algorithms, by nature, are premeditated. And formal and regular auditing is not only the best practice for mitigating algorithmic bias, but in the best interests of both users and corporate institutions.

Corporations can take steps without formal audits too. There are easy processes that can be incorporated into the regular testing step of the software development life cycle. For example, teams can compare outcomes for different groups in order to check for anomalous results. Or create simulations of predictions and compare them to the actual results produced by an algorithm.

What can governments do to combat algorithmic bias?
Before we talk about what governments can do, it's important to understand that there are several regulatory challenges that governments face when thinking about algorithmic bias:

- AI regulation is reactionary and not ubiquitous.
- AI regulations have practical enforcement challenges.
- Gathering evidence for accountability poses risks for remediation.
- There are conflicting business forces and objectives at play.
- There are model development and process challenges.
- Many believe models, algorithms, and technology companies are "too big to audit."
- There are community and organizational challenges.
- Still, there are methods through which we can regulate algorithms. And, the practical challenges make it even more critical to do so.
- What can governments do to combat algorithmic bias?
- There are several steps governments can take in order to combat algorithmic bias from a regulatory perspective. The first is a change to Section 230. Section 230 is a 1996 law that allows Internet Services companies (including social media companies) to host user-generated content without being legally responsible for libelous speech or illegal content posted by their users.

Experts like Frances Haugen say that reforming Section 230 would make companies like Facebook responsible for their ranking algorithms and discourage engagement-based ranking. A.I. systems are not able to and likely will never be able to identify all instances of illegal content, and making companies responsible for their user's content would force companies to rethink algorithmic feeds.

Governments can also update non-discrimination and civil rights law to apply to digital practices and regularize algorithmic hygiene standards.

The solution to algorithmic bias in our technology cannot be strictly data driven.

Many groups (including developers, companies, and institutions) have worked to reduce discrimination present in their algorithms and datasets, but this will never solve the problems of human nature, that each of these datasets will always be gathered by and tested with humans with bias. Avoiding bias is not possible without understanding how and why bias appears and that any human monitoring these systems will also have their own inherent biases.

All of the problems related to algorithmic bias are inextricably linked.
When we stop thinking about technology as infallible and as better than humans, we stop being overtaken by what is shiny and new, and we are better able to hold technology companies accountable for the issues present in their technologies.

When technology companies move slower and take more time to consider the consequences of their work, when they think more about bias, and confront their own biases internally, and apply these mindful practices to both technologies that already exist, and to new technologies, then we get better algorithms that do not compromise on fairness and that serve all users in the best way possible.

When regulatory bodies force companies to innovate ethically and work to understand the true impact that technology has on demographic groups, they are not stifling innovation. Rather, they are creating innovation, the constraints posed allow for the development of improved, superior technologies which do not jeopardize users or morality.

Resources

- Adler, P., Falk, C., Friedler, S., Rybeck, G., Scheidegger, C., Smith, B., & Venkatasubramanian, S. (2016). Auditing Black-box Models for Indirect Influence.
- Benjamin, R. (2019). Race After Technology.
- Chen, J., Storchan, V. & Kurshan, E. (2021). Beyond Fairness Metricks: Roadblocks and Challenges for Ethical AI in Practice.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women.
- Dickey, M.R. (2017). Algorithmic accountability.
- Dickey, M.R. (2020). Twitter and Zoom's algorithmic bias issues.
- Eisenstat, Y. (2019). The Real Reason Tech Struggles With Algorithmic Bias
- Feldman, M., Friedler, S., Moeller, J., Scheidegger, C. & Venkatasubramanian, S. (2014). Certifying and removing disparate impact.

- Friedler, S., Scheidegger, C., & Venkatasubramanian, S. (2016). On the (im)possibility of fairness.
- Georgetown Law Center on Privacy and Technology. (2016). The Perpetual Line-Up: Unregulated Police Face Recognition in America.
- Heilweil, R. (2020). Why algorithms can be racist and sexist.
- Hill, K. (2020). Wrongfully Accused by an Algorithm.
- Lee, N.T., Resnick, P. & Barton, G. (2019). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms.
- Lindsay, R. (2021). I Designed Algorithms at Facebook. Here’s How to Regulate Them.
- Noble, S. (2018). Algorithms of Oppression: How Search Engines Reinforce Racism.
- O'Neil, C. (2016). Weapons of Math Destruction.
- Pepitone, J. (2015). Can Resume-Reviewing Software Be As Biased As Human Hiring Managers?.
- Roberts, L.M. & McCluney C.L. (2020). Working from Home While Black.
- Simonite, T. (2018). When It Comes to Gorillas, Google Photos Remains Blind.
- Smith-Strickland, K. (2015). Computer Programs Can Be as Biased as Humans.
- Twitter. (2021). Sharing learnings about our image cropping algorithm.
- Vartan, S. (2019). Racial Bias Found in a Major Health Care Risk Algorithm.

Who's Responsible for Algorithmic Bias was a two-year research project with the Diversity in Data Graduate Specialist Program at Rutgers University. This project was originally presented as a series of talks and interactive workshops at the university during the 2021-2022 academic school year. Recordings of these workshops can be found here. The project was overseen by Ryan Womack, Data Librarian at Rutgers, https://ryanwomack.com.

This project was developed by Katherine Lee. Katherine Lee has a Masters in Computer Science (in Artificial Intelligence and Machine Learning) from Rutgers University and a Bachelors in Computer Science from Bryn Mawr College. Her research interests are primarily focused around ethics issues in technology and human-computer interaction, particularly algorithmic bias and fairness, data privacy, accountability in technology, accessible technology, and the role that developers play in building out a safer, more ethical future. She can be reached here.