On Bias:

How Algorithmic Inequality Propagates and Perpetuates Cultural Discrimination

Tyler LaBonte and Stephanie Lampotang

University of Southern California

World's Stage, SB Hacks IV

MLH Ethical Tech

June 15, 2018

PART ONE: ALGORITHMIC BIAS AND ITS RESPONSES

Following the American crime wave of the 1980s, University of Colorado statistics professor Tim Brennan and correctional officer Dave Wells collaborated on an algorithmic criminal risk assessment system, called the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS). Now used in justice departments across the nation, COMPAS is consulted regularly to determine the probability of a criminal reoffending, and thereby to inform decisions regarding sentence length and pretrial release. However, there is one monumental defect with the software—it is biased against black people (Angwin and Larson).

The Pulitzer Prize-winning investigative journalism nonprofit ProPublica completed a comprehensive analysis of the algorithm in 2016. When they compared the COMPAS scores of more than ten thousand offenders with their actual reoffense rates, they found that "blacks are almost twice as likely as whites to be labeled a higher risk but not actually reoffend," while whites "are much more likely than blacks to be labeled lower risk but go on to commit other crimes" (Spielkamp). The developers of COMPAS disputed these accusations as expected, but the fact is that this algorithm contributed to the wrongful imprisonment of blacks due to its internal bias, perpetuating the pervasive and well-documented stereotype that black people are more suspicious and dangerous than whites with the same criminal backgrounds.

With the world's increasing dependency on automation and machine learning, algorithmic bias is one of the great challenges of our time. As Google's AI chief John Giannandrea stated, "The real safety question, if you want to call it that, is that if we give these systems biased data, they will be biased" (Knight: "Forget"). When it comes to cultural issues, every person is biased whether they are aware of it or not; differences in location, background,

and cultural exposure lead individuals to possess inherently short-sighted perspectives ("Implicit"). It follows that algorithms designed by humans will also possess cultural bias, which results in unexpected—and unethical—consequences.

Algorithmic bias is a far-reaching challenge, which does not only affect the field of law enforcement. In our technologically-saturated society, where more decisions are made via algorithm than by humans, algorithmic bias is present *everywhere*. Will Knight, senior AI editor at the *MIT Technology Review*, states "Algorithms that may conceal hidden biases are already routinely used to make vital financial and legal decisions. Proprietary algorithms are used to decide, for instance, who gets a job interview, who gets granted parole, and who gets a loan." A few controversial examples in the news recently include racially-biased facial recognition algorithms classifying Asians as perpetually blinking, and gender-biased natural language processing (NLP) algorithms showing women less advertisements for high-paying jobs than males with similar experience (Crawford: "Artificial", Knight: "Biased").

The tech sector is well aware of the cultural and socioeconomic ramifications of algorithmic bias; however, very few preventative measures have been taken except in the most obvious places. Technological pundits like Giannandrea and academic publications such as the *MIT Technology Review* have assisted in spreading awareness about the issue, and the 2017 keynote speech at a top machine learning summit, the Conference on Neural Information Processing Systems (NIPS), was dedicated to algorithmic bias (Crawford: "The Trouble"). Many members of the tech sector are likely to immediately grasp the ethical implications of algorithmic bias as they become aware of it, so several initiatives have already been established in government and universities. In 2016, the Obama Administration published the National

Artificial Intelligence Research and Development Strategic Plan, urging researchers to "design [algorithms] so that their actions and decision-making are transparent and easily interpretable by humans, and thus can be examined for any bias they may contain, rather than just learning and repeating these biases" (National). On the academic side, the Center for Artificial Intelligence in Society at the University of Southern California and the AI Now Institute at New York University are two prominent centers for discussion and research about the sociocultural impact of AI. In spite of their funding and expertise, these initiatives generally have little to show in terms of solutions to algorithmic bias—on the whole, they are more focused on reducing algorithmic bias in their organization than attacking the root causes of the problem.

The general public, however, is woefully uneducated about the dangers of algorithmic bias, but the veil has been partially lifted in recent years thanks to outreach efforts by top researchers. In contrast to the extensive coverage by academic publications, mainstream media outlets have yet to incorporate this powerful cause of social inequality into their news cycle. Most evidence for the impact of algorithmic bias is contained in research journals, out of reach and budget for the average person, instead of in easily-accessible nonfiction books—until 2016, when Harvard mathematics Ph.D. Cathy O'Neil authored *Weapons of Math Destruction*. Longlisted for the 2016 National Book Award for Nonfiction, *Weapons* is a powerful introduction to the ethical consequences of biased algorithms in "an unusually lucid and readable look at the daunting algorithms that govern so many aspects of our lives" ("Kirkus"). Despite the success of *Weapons*, it takes more than one book to develop public awareness; a nontrivial increase in the technical complexity of mainstream media is the unlikely but optimal solution.

There is evidently a monumental discrepancy in the awareness and understanding of the tech sector and general public regarding algorithmic bias, perhaps due to the underlying reasons being difficult to control. First and foremost, algorithmic bias is a subject of inherently technical nature. In order to understand the implications of the issue, the public first has to be educated about the basics of general computer science, which does not come easily. University of Illinois professor of computer science, Karrie Karahalios, "showed that users generally don't understand how Facebook filters the posts shown in their news feed. While this might seem innocuous, it is a neat illustration of how difficult it is to interrogate an algorithm" (Knight: "Forget"). This sentiment is incredibly evident in Congress as well; during Facebook CEO, Mark Zuckerberg's testimony in April 2018, lawmakers asked ignorant questions regarding how Facebook sustained their business model, to which Zuckerberg famously replied, "Senator, we run ads" (Abramson and Ducharme and Gajanan). The cluelessness of several otherwise-educated Congressmen highlights the difficulty of the general public to understand technical issues, even when such algorithms crucially affect their daily lives.

Furthermore, algorithmic bias is an incredibly recent issue, which means there has only been a couple years at most for seminal research to be disseminated to the public. While systems like COMPAS have received attention for their inequality, there are many more examples that have yet to be discovered, quietly contributing to cultural discrimination right under our noses. Kate Crawford, principal researcher at Microsoft, summarized this issue in a 2017 interview with the *MIT Technology Review*: "It's still early days for understanding algorithmic bias. Just this year we've seen more systems that have issues, and these are just the ones that have been investigated" (Knight, "Biased"). The recency of algorithmic bias also means that researchers

have no good solutions to share with the masses. Giannandrea suggests that such bias is inherently a reflection of bias in Big Data, for which a solution would require a paradigm shift in data collection regulations and take a significant amount of time (Knight, "Forget"). O'Neil's solution, through her company ORCAA, is to audit existing algorithms for accuracy, bias, and fairness; while a good idea in theory, the process of auditing inherently includes the auditor's internal biases, and may even make the problem worse if a biased algorithm is improperly "certified" (O'Neil). The recency of algorithmic bias combined with its technical nature explains the discrepancy between tech sector and general public reactions; while the underlying causes are nontrivial to solve, it is imperative that effective action is taken—lives and livelihoods depend on it.

Algorithmic bias may be a big problem, but it often appears in small places. Our hack, *World's Stage*, is a web-based video discovery interface: seemingly innocuous, but controlled beneath the surface by black-box algorithms that are just as susceptible to cultural bias as COMPAS. By considering the ethical implications of our project going viral, we hope to provide illustrative examples of the complexity and omnipresence of algorithmic bias, as well as offer a few solutions to mitigate the consequences.

PART TWO: EFFECTS OF ALGORITHMIC BIAS ON *WORLD'S STAGE*

World's Stage (tmlabonte.github.io/SBHacks2018) is a web application we built in 36 hours at the University of California at Santa Barbara hackathon, designed to celebrate cultural diversity by allowing users to search international hit songs, then displaying a map with the most popular dance video to that song over a diverse set of the world's cities. Through it, we have discovered that culture is less of a barrier to musical innovation than we thought; we have found

bhangra dances set to latin music and ballet set to hip-hop—unexpected yet beautiful new dances to familiar music. In order to provide this service, we combined the YouTube and Google Maps APIs in a modular JavaScript program. Our crawler takes the input song, then scours YouTube for top dance videos from selected locations. From there, our program places links with video thumbnails over the major cities on Google Maps. The result: an intuitive interface that encourages discovery and appreciation of world cultures through music and dance.

Social media applications like Facebook and Twitter optimize their interfaces for maximum engagement and retention, aligning newsfeed content with what has already been liked or retweeted—the perfect formula for a monoculture. These algorithms make it dangerously easy for users to remain sheltered in an isolated bubble of their own culture and values. Our antithesis to the modern social media environment is *World's Stage*, where users are not only encouraged but required to take a worldwide perspective. This creates new spaces for undiscovered videos repressed by the Facebook-like algorithms which promote only already successful content. However, underneath the pretty splash screen, *World's Stage* is really just elegantly combining several existing algorithms, which are all inherently biased. While a valiant effort to break down cultural boundaries with dance, it is clear that if *World's Stage* ever went viral, its algorithmic bias could be counterproductive, even destructive, to its purpose.

First and foremost, *World's Stage* has a powerful possibility to be used as a tool for cultural scorn, rather than discovery and appreciation. While a vast majority of user experiences with the application may be positive and informative, the algorithmic bias of the modern newsfeed means that if a negative experience goes viral, that experience will become strongly associated with the application. For a real-life example, consider the Twitter controversy that

made national headlines when Keziah Daum, a senior at a Utah high school, posted prom photos of herself wearing a traditional Chinese dress called a cheongsam. While Daum, who is not Chinese, said that the "beautiful" dress "really gave [her] a sense of appreciation and admiration for other cultures and their beauty"—exactly what we are trying to accomplish with *World's Stage*—outrage was sparked when Jeremy Lam commented "My culture is NOT… your prom dress" (Schmidt). Lam's comment has since been retweeted over forty thousand times, and despite the many positive comments defending Daum's choice, it was Lam's comment and the outrage it sparked that made the daily news. If *World's Stage* reached that level of popularity, it would be inevitable that someone will take offense to how a video portrayed their culture, potentially going viral and offending many others. Then, in the worst case, people may begin to use our application as a tool to seek out insulting cultural videos and highlighting them on social media, thereby spreading hate instead of acceptance.

Second, we the developers introduced bias into the *World's Stage* algorithm when we defined what it meant to be a "top city" and a "top video." *World's Stage* faces the challenge of ranking not just users' video content, but even their countries and hometowns. We took time during the hackathon to discuss possible choices for distributing our beacons across Google Maps: do we spread the locations out geographically or choose the ones with the highest populations, and how forgiving should our cutoffs be? Yet, for all our deliberation, we eventually just settled on going down a list of largest cities and selecting locations such that they looked aesthetically distributed on our map. All of our options would leave someplace behind (for example, Montreal is too close to New York, and Santiago is not as populous as Rio de Janeiro). This bias immediately cuts a huge majority of world cultures from our application for

various arbitrary reasons, especially small cultures and those in dense areas—thus going against our goal of embracing cultural diversity.

We had a similar discussion regarding what constitutes a "top video," eventually settling for most-viewed because it is the easiest to scrape from YouTube. However, this bias for popular content is the same as the social media newsfeeds that we were trying to break away from. The cycle of the most influential or popular videos staying in power was shifted, but not broken. Just because a video has millions of views does not necessarily mean it is the best representation of its culture, and members of that culture may feel slighted if we featured it in *World's Stage*. The fact that we only display one video is a problem in itself; perhaps the good that we do bringing cultures together is negated by the singularity of each culture's voice. The presence of a "top video" invites the idea that a city or entire culture or can be represented by a single four-minute long dance cover—which in turn can lead to the strengthening of cultural bias.

Finally, *World's Stage* can easily be cheated to gain extra ad revenue from YouTube. By mis-tagging videos with less popular locations to gain unfair exposure, perhaps by uploading their videos via proxy server, content creators can redirect attention that should be given to the lesser-known culture to their own channel. Thus, not only will these creators be taking advantage of others' cultures to make money—the very definition of cultural appropriation—but also circumventing the intended purpose of our application. While this technique does not necessarily deal with algorithmic bias as per Part One, it is still a stunningly simple example of how such algorithms can be exploited for money at the expense of users and developers.

PART THREE: MITIGATION OF ALGORITHMIC BIAS IN *WORLD'S STAGE*

The effects of algorithmic bias analyzed in Part Two all lead to thorny, multifaceted problems; while none of the effects can be completely resolved, we can mitigate their consequences by implementing more ethical alternatives.

The possibility for *World's Stage* to be used as a tool for cultural scorn fundamentally arises from human nature—people's tendency to ridicule unfamiliar or scary cultural traits and practices. Instead of trying to eliminate this inherent bias (a truly impossible task), we aim to remove the factors that allow people to think this way. One of the largest contributors to this kind of scorn is dehumanization: perceiving *World's Stage* dancers as actors on a screen rather than mothers and fathers, sisters and brothers. When Jeremy Lam posted his comment that begat national outrage, he surely was not thinking of Keziah Daum as a high school senior stressed out over choosing a beautiful prom dress, but rather as a caricature appropriating his culture out of spite. To combat this, we will invite featured videographers to submit a short paragraph detailing what life is like in their culture and how their dance embodies its values, sharing the personal and cultural significance of their video with the world at large. Thus, *World's Stage* users will be cognizant that they are interacting with *real* members of *real* cultures, decreasing opportunity for dehumanization. While this solution will likely require outreach and translation efforts on our part, it will be worth it to reinforce the mission of *World's Stage* as a platform for positive cultural awareness.

Humanity's inherent cultural bias may be impossible to eliminate, but we can heavily reduce developer-introduced bias by changing our definitions of "top cities" and "top videos" to be objective and culture-neutral, thereby becoming more inclusive of worldwide users. One

elegant solution is to introduce a rotation of featured cities on our map, rather than settling for our current arbitrarily-chosen defaults. It is fitting that, as cities are always growing and changing, our featured cities could also change—one day highlighting Los Angeles and Beijing, the next Amsterdam and Bogota. And, if a certain city is not available in the rotations, a new feature that allows the user to customize their map by searching for chosen locations would add to the mitigation of developer-introduced algorithmic bias.

Currently, our algorithm for selecting the "top video" from each location is solely based on view count, but in order to even the playing field and give different content a chance to be discovered, videos should receive a ranking that takes into account both views *and* how recently it was uploaded. Anything that had had its fair share of the limelight would become old news and give way to upcoming material. Together with our city rotation algorithm, this change would serve to develop an atmosphere of cultural discovery in *World's Stage*, where no city or video is given undue attention or unfairly ignored.

With regards to the risk of mis-tagged locations, we would take preventative measures by partnering with YouTube to verify relative location through location services at the time of upload. This method of verification can also be manipulated, but we hope that it would act as a deterrent, dissuading users from mis-tagging videos. If that solution is not sufficient alone, we could develop a machine learning algorithm to detect whether a video's location is accurate. We would need to utilize a convolutional neural network which analyzes each video and outputs a probability vector of possible locations. If the tagged location does not have a high probability—for example, if the video was filmed in a desert but tagged as Moscow—the video would be flagged and sent to a human operator for thorough review. However, this solution also

has its downsides; it only works for videos filmed outside, and with the strength of today's video editors, it would be simple to doctor a video background so it is certified by the algorithm.

World's Stage brings to life its Shakespearean namesake as a powerful application for celebrating cultural diversity; while its algorithmic bias combined with popularity reveals potentially devastating ethical consequences, judicious use of clever technological solutions can greatly mitigate their impact. On a broader scale, *World's Stage* is representative of the majority of today's algorithms: seemingly innocuous and beneficial, yet gravely susceptible to bias. As humanity advances in our technological prowess—and destructive potential—the need for ethics becomes critical. Soon, everything from nuclear weapons to cancer medicine will be directed by algorithms, perhaps in the form of general artificial intelligence. It is imperative that the computer scientists of tomorrow be held to strict ethical standards because, quite literally, the fate of humankind is at stake.

Works Cited

Abramson, Alan, Jamie Ducharme, and Mahita Gajanan. “7 of the Most Awkward Moments from Mark Zuckerberg’s Testimony to Congress.” *Fortune Magazine.* Time, Inc. 11 April 2018. Web.

Angwin, Julia, and Jeff Larson. “Machine Bias.” *ProPublica*. 3 May 2016. Web.

Crawford, Kate. “Artificial Intelligence’s White Guy Problem.” *The New York Times*. 25 June 2016. Web.

Crawford, Kate. “The Trouble with Bias.” *Conference on Neural Information Processing Systems (NIPS).* 5 December 2017. Web.

“Implicit Bias Review.” *Kirwan Institute for the Study of Race and Ethnicity.* The Ohio State University. November 2017. Web.

“Kirkus Review: Weapons of Math Destruction.” *Kirkus Reviews*. 19 July 2016. Web.

Knight, Will. “Biased Algorithms Are Everywhere, and No One Seems to Care.” *MIT Technology Review*. 12 July 2017. Web.

Knight, Will. “Forget Killer Robots—Bias Is the Real AI Danger.” *MIT Technology Review.* 3 October 2017. Web.

National Science and Technology Council. “National Artificial Intelligence Research and Development Strategic Plan.” *US Government*. October 2016. Web.

O’Neil, Cathy. “O'Neil Risk Consulting & Algorithmic Auditing.” Web.

Schmidt, Samantha. “It’s Just a Dress.” *Washington Post*. 1 May 2018. Web.

Spielkamp, Matthias. “Inspecting Algorithms for Bias.” *MIT Technology Review*. 12 June 2017. Web.