Bias in Machine Learning and Diversity

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Srujan Joshi

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Sean M. Ferguson, Department of Engineering and Society

Bias in Machine Learning and Diversity

In our digital age, more data is being produced than ever before. This has caused a massive surge in the popularity of Machine Learning, a subset of Artificial Intelligence (although, as in this paper, the two terms are used interchangeably) where algorithms are trained to learn from data, identify patterns and make decisions with minimal human intervention. Machine Learning has numerous applications such as recommendation engines, targeted ads, self-driving cars, voice assistants, image recognition and language translation to name to a few and has brought about immense technological and social progress in the world.

That isn't to say Machine Learning is without its issues. One of the most prominent topics when it comes to the discussion of Machine Learning is that of bias in Machine Learning. This bias can be defined in many ways ranging from narrow formal statistical definitions to much wider legal and normative definitions. No matter how you choose to define it, the central idea is that of Machine Learning systems producing undesirable results leading to unfair decision making. Bias in Machine Learning can have serious negative impacts for those affected.

Take the example of image recognition technology. While this technology has many positive use cases like improving global agricultural output through satellite imagery data or detecting malignant tumors in CT scans, there are some applications more prone to bias where certain demographics might question the use of these algorithms. For example, black communities who already feel like they are the primary targets of a surveillance state may be worried about the rollout of facial recognition in security cameras.

The issue of solving bias in Machine Learning is a complex one with both technical and non-technical solutions. In this paper I will be providing a survey of the state of bias in Machine Learning and measures to mitigate this bias through a socio-technical decolonial critique

framework, with an emphasis on the lack of Diversity in the AI workplace as a factor in propagating bias.

Example Instances of Bias in Machine Learning

Given that it is hard to understand exactly how an ML model reaches a certain decision, most often bias reveals itself only in the form of inequality of outcomes, after the fact. In this section I will look at some concrete examples of bias in Machine Learning solutions to drive home the point that bias in Machine Learning has very tangible real-world impacts. Applying a decolonial framework allows us to contextualize these examples of bias with socio-political, historical and cultural background (Mohamed, Png, & Isaac, 2020, p. 667).

One of the most harrowing examples of bias in ML is that of Northpoint's COMPAS software, which is used to make prison sentencing decisions across the United States. The COMPAS software calculates a defendant's "risk score" for committing crime again in the future based off answers to 137 questions that are either answered by defendants or pulled from their criminal records. Furthermore, the specific methodology behind calculating risk scores is considered a company secret. So, to the public, the algorithm is a black box. Although the scores generated are not legally binding, they have considerable influence in swaying court decisions, and have resulted in a disproportionate number of people of color getting locked up in the Prison system (Angwin, Larson, Mattu, & Kirchner, 2016).

In 2016, Amazon rolled out same-day delivery for its Prime customers in select ZIP codes across the country. The decision was powered by the extensive Machine Learning algorithms at Amazon's disposal. According to a Bloomberg analysis that compared Amazon-same day delivery areas with US Census Bureau data, it was found that the service excluded predominantly black ZIP codes to a large degree (Ingold & Soper, 2016).

In 2017, Uber introduced an AI powered security solution called the "Real-Time ID Check" which had drivers periodically upload selfies of themselves, which would be checked by a Machine Learning algorithm against photos on file to verify that the right person is behind the wheel of the car. It was reported that this software was falsely flagging transgender drivers. The consequence for these drivers was several days of missed work (Melendez, 2018). This is just one of the many examples of Facial Recognition software not working as intended with minority groups.

The online advertising space has been one of the cornerstones of the Internet since its inception. In recent times, online advertising platforms have tapped into Machine Learning techniques to increase revenue by delivering personalized, targeted ads that users are more likely to engage with. These platforms offer many ways in which advertisers can target or exclude groups of people seeing their ads. In some cases, though, even if it isn't the advertiser's intention to selectively target a certain group, ad delivery software can be more likely to serve certain ads to people belonging to certain demographics. For example, Facebook's ad delivery program was found to have been biased in its delivery when it comes to certain races and gender categories (Ali, et al., 2019). Google's competing ad delivery software, AdSense, has also been found to be similarly biased against racial minorities (Sweeny, 2013).

How does Bias arise in Machine Learning?

The short answer to that question would be the data that is used to train the Machine Learning system. Let me elaborate on that. Training data determines what an ML model learns. For various reasons, training data is usually imperfect. Datasets might be incomplete, missing entries for certain data points or attributes. Datasets might contain incorrect or outdated data. Datasets might not be representative of certain populations. Datasets can be tainted by cultural

biases during the data collection process. On top of this, the process of data collection can be opaque. All to say, datasets are only as good as the process in which they are created.

For example, for most image and speech recognition tasks, training data is generated by having humans manually label massive repositories of image or speech files. In such cases it is not hard to imagine human bias creeping into the dataset. Another common data collection method is to harvest data from users' activity on user facing internet applications such as YouTube and Instagram for example. This can be done by the owners of the applications themselves, Google and Facebook in this case, in which case the data is collected "behind the scenes" as users go about interacting with the application. Data can also be scraped from applications by third parties aiming to aggregate information from publicly available sources. The problem here is that such data might only be representative of the subsection of the population that actively uses these applications. Not to mention the fact that this completely excludes populations which do not have unconstrained access to the internet and smart phones (Campolo, Sanfilippo, Whittaker, & Crawford, 2017; West, Whittaker, & Crawford, 2019).

A comical, yet disturbing, example of bad data leading to a bad model is when Microsoft released its experimental chatbot called "Tay" which was designed to talk like a teenager and was trained on tweets from Twitter and messages from GroupMe. Tay was designed to get better at conversation over time as more users interacted with it on Twitter. After Tay was released on public Twitter however, as Tay tried to "learn" how to converse better, it got attacked with hateful content by certain Twitter users. Within 24 hours, Tay went from posting harmless images of kittens to spewing fringe conspiracy theories and disturbing antisemitic rhetoric (Lee, 2016).

It could also be the case that the data a model is trained on is corrupted by historical or societal biases as a result of which it goes on to uphold and propagate historical unfairness and stereotypes. A classic example of this is the use of automated decision systems and predictive policing systems by government agencies that enforce the law. The data that these systems are trained on is usually compiled from historical criminal records, and thus the data inherits all the racial biases that have been appeared in the criminal justice system through the ages. Models trained on such data are used to inform future decisions on policing and criminal sentencing, and thus continue the cycle of injustice. (Richardson, Schultz, & Crawford, 2019)

Another problem is the potential disparity between the assumptions that were made when an ML model is designed and the context in which it is deployed in the real world. For instance, ML-based mapping applications usually provide indirect routes to users to accomplish traffic load balancing. Such a system cannot tell whether a user is commuting to work, going on a casual joyride or making an urgent visit to the hospital. Decontextualized assumptions like this can disadvantage non-consenting and unaware populations with no way of providing actionable feedback to the model to allow it to correct its predictions (Bird, Baracas, Crawford, Diaz, & Wallach, 2016).

There are also cases when the idea behind the model or algorithms itself is biased. For example, there has been a slew of research into facial recognition algorithms which, based only on still images and no other context, automatically infer certain characteristics about people such as race, gender, IQ, social status and criminality (Han & Jain, 2014) (Wu & Zhang, 2016). Take for example, Faception, a company whose core product is a facial recognition tool which purportedly can determine whether someone is an Academic Researcher, Professional Poker Player or a Terrorist just based off a still image of their face (Faception, 2021).

The above-mentioned types of bias in ML models do not occur separately, but an ML solution often contains a mixture of these. The main idea is that if the data or methodology of creating and deploying an ML model is biased, this bias will be incorporated into the predictions that are made by the model, and ultimately into the decisions taken based off those predictions.

Bias and the Diversity Problem in Tech

In the next section of the paper, I detail my analysis of the lack of diversity in the Pipeline of AI creation and how this is inherently linked to bias in ML. For this discussion I will focus on the organizations where ML models themselves are made. Most of the work on ML is done by companies in the tech industry, with the rest of the work being done in academic research labs run by a handful of universities around the world. It is no secret that the tech industry is a hostile environment for women and minorities, and Universities are only marginally better when it comes to this regard. In terms of decolonial theory, these elite, exclusive institutes are the centers of power or metropoles, whereas the minority groups that are excluded make up the periphery.

To illustrate this, I will point out a few examples where discrimination in the tech industry has come to light. An investigation by the Department of Labor unveiled the fact that Google was underpaying women when compared to men working the same job (Kolhatkar, 2017). Apple's board rejected proposals to increase diversity at Apple, calling them "unduly burdensome" on the company (O'Brien, 2016). A black former employee of Facebook spoke up about how the company had created a hostile environment for people of color, and how he was discouraged from participating in Black activism by the company (Luckie, 2018). Women employees of Tesla recall being catcalled and whistled at whilst at work in what they called the "predator zone". According to an anonymous study titled "The Elephant in the Valley", sixty six

percent of women said that they had been sexually approached in an inappropriate manner by male co-workers and forty percent of women said that they wouldn't report such incidents for fear of losing their jobs (Kolhatkar, 2017).

Currently the best, albeit imperfect, measure of a company's diversity comes from the EEO-1 component report. The EEO-1 is an annual government-mandated report that all private employers with a hundred or more employees must submit to the government. The reason why EEO-1 forms are a more accurate measure of company diversity compared to company diversity reports, which usually only include misleading percentages, is because they include actual numbers of employees segregated by demographics such as line of work, gender and racial identity. Unfortunately, by law, companies do not have to reveal the contents of their yearly EEO-1 reports to the public. In the recent past though larger companies like Facebook, Google and Nvidia have released their EEO-1 reports to prevent public backlash (Rangarajan & Evans, 2017).

Most companies, however, do not release their EEO-1 data to the public. In fact, most companies don't disclose numbers related to diversity full stop. Of the handful of tech companies that do disclose internal statistics on employee diversity, they rarely present their raw numbers, and instead to resort to using flowery PR speak with vague charts and figures to gloss over their actual lack of diversity. Some companies like Oracle have even publicly pushed back against laws which require companies to report diversity-related statistics to the government, with the argument that internal company employee statistics are a "trade secret" (Holman, 2019).

Even EEO-1 reports are not perfect. They have come under scrutiny for deficiencies which mainly revolve around the categories in the form being too broad and all-encompassing. For example, the "professional" category does not differentiate between tech and non tech

workers, which makes it easy for companies to report higher than actual diversity numbers by lumping various lines of work together. The form also only has two categories of gender and a limited number of options for the race category. EEO-1 forms were a step in the right direction for their time but have since become stale as measures of diversity.

Even amongst the tech giants such as Google, Facebook and Microsoft, who do release detailed yearly reports (Google, 2021) (Facebook, 2021) (Microsoft, 2021) with concrete numbers on the Diversity in the Workforce, the numbers don't paint a flattering picture. Women only comprise 24.6%, 24.8% and 30.9% of the workforce at Google, Facebook and Microsoft respectively while only 2.9%, 2.1% and 5.6% of the workforce consists of people of color. These forms also fail to present data on other oppressed groups such as gender minorities or trans workers.

With tech companies hardly disclosing any diversity statistics of their own accord, studies on gender statistics in the tech industry have had to resort to other means to compile diversity data. These includes harvesting data from employment websites such as LinkedIn, Glass Door, conference feedback forms, and internet surveys. Using such methods, a 2018 survey by WIRED and Element AI revealed that only around 13% of researchers in AI identified as female. Google was found to employ 641 "machine intelligence" specialists of which only 60 identified as female. An analysis of author data from the most popular machine learning conferences revealed that only 12% of contributors to papers presented at these conferences identified as female (Simonite, 2018).

It is not as though the tech industry is unaware of its lack of diversity. Some in the industry have dismissed efforts to improve diversity as unnecessary and have instead tried to defend the status quo by coming up with the PR speak to do so. One such example is Facebook's

push for “Cognitive Diversity” which pushes for diversity of thought above other conventional diversity metrics (West, Whittaker, & Crawford, 2019). According to this argument, the aim of the company should be to maximize the diversity of thought in the room. This argument focuses on individual identity while ignoring power dynamics, race and social hierarchies. So, by this argument, a room full of white men with different thought patterns would be more diverse than a room which had representatives from different genders and ethnicities who all thought the same way. Although the company claims that this is to deliver more equitable technological solutions, in practice, this has served to further justify the exclusion of minority groups in tech.

Another line of reasoning that the tech industry uses to justify the lack of diversity in tech is that there is not enough hirable talent in unrepresented groups. The blame is thus shifted to the education system for not providing minority groups with the requisite skills to secure jobs in tech. Facebook for example has publicly criticized the Public American Education System in this manner (Wells, 2016). This argument is shallow and misleading. It turns out that there is a large disparity between the number of STEM graduates from minority groups and the number of tech workers from minority groups. For example, Black and Hispanic people account for 6% and 8% of STEM graduates but only make up 1% and 3% of the tech industry respectively (Daniels, 2019).

Even with increased inclusivity efforts from tech companies these days, the tech industry can still be hostile for even for those minorities who do make it in. A 2020 study conducted by Accenture and Girls Who Code found that half of young women would give up their job in the tech industry by the age of 35. The same study also revealed that only 21% of women thought they could thrive in the tech industry, with black women reporting an even lower number at 8% (Accenture, 2020).

This systemic inequity and bias in the field of technology is tightly related to the creation and adoption of biased ML models. In an environment where people belonging to racial and gender minorities do not feel like they belong and their voices don't matter, it's not hard to see how bias can creep into the pipeline of ML solution creation.

Beyond Localized Diversity

The next section I examine ways in which AI bias can be mitigated, focusing on solutions that seek to eliminate bias not only in specific datasets but in the entire landscape of AI creation, from problem definition to data collection to model creation to the workplace where solutions are crafted to the impact on the various stakeholders on the ground. As suggested by the theory of decolonial AI, ML solutions should be employed in a way which acknowledges and reconciles imbalances in power (Mohamed, Png, & Isaac, 2020).

In order to apply ML in ways which deliver equitable outcomes, we must first acknowledge bias in ML is not a purely technical problem with a technical solution. To the contrary, bias in ML is an artifact of the bias present in the institutions that are stakeholders in the creation of ML solutions, and in order to mitigate this bias, we need to take a holistic approach applying both social and technological solutions.

With regards to increasing diversity in the AI creation pipeline, it is not sufficient just to increase diversity in entry-level positions at tech companies. It is also not sufficient to treat diversity as an issue localized to one specific industry or area. We need more diversity across the board with people from minority and underrepresented communities in the room making decisions and being involved in all the stages of AI solution creation, from problem definition to algorithm creation to real-world application. In this regard we still have a long way to go. Take for example, a 2017 report from Forbes which revealed that of the 16 Fortune 500 companies

that reveal detailed diversity statistics, 72% of the leadership is made up of white men. Another study in 2020 showed that 90% of all Fortune 500 CEOs are white men (Zweigenhaft, 2020).

Diversity in terms of numbers alone is not enough. Even with a diverse workforce and leadership, we need to educate and sensitize those who design and deploy ML solutions about the biases and inequities that arise from bad data and bad algorithms. Having ML solution designers take a more contextualized view when designing ML models may allow for bias to be seen more extensively. Looking ahead, the prerequisites to be considered an AI specialist might be expanded to include not only technical knowledge regarding AI solutions but also an understanding of and a sensitivity towards all the different kinds of societal biases that may creep in during the AI workflow whether that is during data collection, model evaluation or product deployment.

Organization-wide procedural changes are also required when it comes to handling data collection and subsequent data use. To this end, there has been research into practices and tools that can be adopted by organizations to promote transparency around data collection pipelines. For example, researchers have put forward the use of "Decision Provenance" methods. This involves keeping track of the origin of data, maintaining the privacy of the individuals the data pertains to, and recording any changes made to the data along its journey from the point of collection to the integration into the dataset. Such practices seem crucial to adopt since the modern-day data collection process, as mentioned earlier in the paper, involves collecting data from user facing applications on client devices, subsequent transmission to company servers, and in some cases also ends in the processed data being passed onto third parties. The adoption of Data Provenance methods would assist all the stakeholders involved in the ML pipeline through

increased transparency and would enable better government regulation of data collection (Singh, Cobbe, & Norval, 2019).

"Datasheets for Datasets" is another proposal in a similar spirit which aims to introduce a standardized format for datasheets to accompany datasets. The idea is that there should be a universal standard when it comes to the metadata associated with a dataset. This metadata would have information regarding the actual data in the dataset, such as the source of the data, the parties involved in data collection, the motives behind data collection, the methods used to collect the data, and the envisioned use cases of the data (Gebru, et al., 2021).

Another solution is to improve AI Ethics education within the tech community. Such efforts have primarily revolved around changes to the higher education curriculum in STEM subjects, especially Computer Science. Researchers have pointed to a current lack of mutual support between CS and the humanistic social sciences. There have been calls for educating future ML practitioners to take a more holistic approach when it comes to designing ML solutions. This includes identifying the right problem, consulting all the stakeholders involved in a project, calling upon non-technical experts especially when dealing with high social impact problems and relying on advice from voices in communities that might be underrepresented (Raji, Scheurman, & Amironesei, 2021).

Conclusion

In this paper, I provided a survey of the state of bias in Machine Learning from a socio-technical standpoint. I went over some examples of bias to illustrate the gravity of the situation. I then examined the different kinds of bias that can be exhibited by Machine Learning problems. This was followed by a deep dive into the lack of diversity in the tech industry. I ended the paper by looking at ways in which bias can be remedied at all stages of the AI creation pipeline.

Although one might think that bias in Machine Learning is a purely technical problem, upon a deeper analysis, one comes to realize that bias in ML is just a symptom of not only a deeply ingrained systemic bias across the tech Industry but also the power structure and dynamics at play in broader society. This is in line with the framework of decolonial AI theory put forward by Shakir Mohamed, Marie-Therese Png and William Isaac. Any effort towards promoting equitable outcomes in ML needs us to be cognizant of the "the hierarchies, philosophy and technology inherited from the past" (Mohamed, Png, & Isaac, 2020, p. 677). Thus, as this paper illustrated, mitigating bias in ML entails changing these power dynamics to prioritize the interests of society as a while, especially those who are currently oppressed such as racial and gender minorities, and not just the interests of those in power.

References

Accenture. (2020). *RESETTING TECH CULTURE.*

Ali, M., Sapiezynski, P., Bogen, M., Mislove, A., Korolova, A., & Rieke, A. (2019, September 12). *Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes.* Retrieved from arXiv: https://arxiv.org/pdf/1904.02095.pdf

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine Bias*. Retrieved from Propublica: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Bird, S., Baracas, S., Crawford, K., Diaz, F., & Wallach, H. (2016, October 2). *Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI.* Retrieved from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2846909

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 Report.* New York: AI Now Institute.

Daniels, J. (2019, April 3). *"Color-blindness" is a bad approach to solving bias in algorithms.* Retrieved from Quartz: https://qz.com/1585645/color-blindness-is-a-bad-approach-to-solving-bias-in-algorithms/

Elizabeth, D., Tiku , N., & Timberg, C. (2021, November 21). *Facebook's race-blind practices around hate speech came at the expense of Black users, new documents show.* Retrieved from The Washington Post: https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/

Facebook. (2021). *Facebook Diversity Update.*

Faception. (2021). Retrieved from https://www.faception.com

Freire, A., Porcaro, L., & Gomez, E. (2021). Measuring Diversity of Artificial Intelligence Conferences. *2nd Workshop on Diversity in Artificial Intelligence (AIDBEI).*

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daume III, H., & Crawford, K. (2021, December 1). *Datasheets for Datasets.* Retrieved from arXiv: https://arxiv.org/pdf/1803.09010.pdf

Google. (2021). *2021 Diversity Annual Report.*

Google. (n.d.). *Responsible AI practices.* Retrieved from Google AI: https://ai.google/responsibilities/responsible-ai-practices

Gu, J., & Oelke, D. (2019). Understanding Bias in Machine Learning.

Han, H., & Jain, A. K. (2014). Age, Gender and Race Estimation from Unconstrained Face Images. *MSU Technical Report*. Retrieved from http://biometrics.cse.msu.edu/Publications/Face/HanJain_UnconstrainedAgeGenderRaceEstimation_MSUTechReport2014.pdf

Holman, J. (2019, February 13). *Silicon Valley is Using Trade Secrets to Hide Its Race Problem*. Retrieved from Bloomberg Quint: https://www.bloombergquint.com/business/silicon-valley-is-using-trade-secrets-to-hide-its-race-problem

Ingold, D., & Soper, S. (2016, April 21). *Amazon Doesn't Consider the Race of Its Customers. Should It?* Retrieved from Bloomberg: https://www.bloomberg.com/graphics/2016-amazon-same-day/

Kolhatkar, S. (2017, November 20). *The Tech Industry's Gender Discrimination Problem.* Retrieved from The New Yorker: https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem

Lee, P. (2016, March 25). Retrieved from Microsoft Blog: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/

Luckie, M. S. (2018, November 8). *Facebook is failing its black employees and its black users.* Retrieved from Facebook: https://www.facebook.com/notes/3121422367961706/

Melendez, S. (2018, August 9). *Uber driver troubles raise concerns about transgender face recognition*. Retrieved from Fast Company: https://www.fastcompany.com/90216258/uber-face-recognition-tool-has-locked-out-some-transgender-drivers

Meta. (n.d.). *Facebook's five pillars of Responsible AI* . Retrieved from Meta AI: https://ai.facebook.com/blog/facebooks-five-pillars-of-responsible-ai/

Microsoft. (2021). *Global Diversity and Inclusion Report* .

Microsoft. (n.d.). *Face API*. Retrieved from Microsoft Azure: https://azure.microsoft.com/en-us/services/cognitive-services/face/

Mohamed, S., Png, M.-T., & Isaac, W. (2020). Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence. *Philosophy and Technology*.

O'Brien, S. A. (2016, January 15). Retrieved from https://money.cnn.com/2016/01/15/technology/apple-diversity/index.html

Raji, I. D., Scheurman, M. K., & Amironesei, R. (2021, March 3). *"You Can't Sit With Us": Exclusionary Pedagogy in AI Ethics Education.* Retrieved from Association for Computing Machinery: https://dl.acm.org/doi/pdf/10.1145/3442188.3445914

Rangarajan, S., & Evans, W. (2017, October 19). *Hidden figures: How Silicon Valley keeps diversity data secret*. Retrieved from Reveal: https://revealnews.org/article/hidden-figures-how-silicon-valley-keeps-diversity-data-secret/

Richardson, R., Schultz, J. M., & Crawford, K. (2019, May). *DIRTY DATA, BAD PREDICTIONS: HOW CIVIL RIGHTS VIOLATIONS IMPACT POLICE DATA, PREDICTIVE POLICING SYSTEMS, AND JUSTICE.* Retrieved from NYU Law Review: https://www.nyulawreview.org/wp-content/uploads/2019/04/NYULawReview-94-Richardson-Schultz-Crawford.pdf

Simonite, T. (2018, August 17). *AI Is the Future—But Where Are the Women?* Retrieved from WIRED: https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/

Singh, J., Cobbe, J., & Norval, C. (2019). Decision Provenance: Harnessing data flow for accountable systems. *IEEE Access*.

Stathoulopoulos, K., & Mateos-Garcia, J. (2019, July). *Gender Diversity in AI Research.* Retrieved from Nesta: https://media.nesta.org.uk/documents/Gender_Diversity_in_AI_Research.pdf

Sweeny, L. (2013, January 28). *Discrimination in Online Delivery.* Retrieved from arXiv: https://arxiv.org/pdf/1301.6822.pdf

U.S. Equal Employment Opportunity Commission. (n.d.). *EEO-1 Data Collection.* Retrieved from U.S. Equal Employment Oppoprtunity Commission: https://www.eeoc.gov/employers/eeo-1-data-collection

Vincent, J. (2016, March 24). *Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day*. Retrieved from The Verge : https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Wells, G. (2016, July 14). *Facebook Blames Lack of Available Talent for Diversity Problem.* Retrieved from The Wall Street Journal: https://www.wsj.com/articles/facebook-blames-lack-of-available-talent-for-diversity-problem-1468526303

West, S. M., Whittaker, M., & Crawford, K. (2019). *DISCRIMINATING SYSTEMS Gender, Race, and Power in AI.* New York: AI Now Institute. Retrieved from https://ainowinstitute.org/discriminatingsystems.pdf

World Economic Forum. (2021). *Global Gender Gap Report.*

Wu, X., & Zhang, X. (2016, November 21). *Automated Inference on Criminality using Face Images.* Retrieved from arXiv: https://arxiv.org/pdf/1611.04135v2.pdf

Young, E., Wajcman, J., & Sprejer, L. (2021). *Where are the Women?* The Alan Turing Institute.

Zweigenhaft, R. (2020, October 28). *Fortune 500 CEOs, 2000-2020: Still Male, Still White.* Retrieved from The Society Pages: https://thesocietypages.org/specials/fortune-500-ceos-2000-2020-still-male-still-white/