Bibliography

Andrus, M., Spitzer, E., Brown, J., & Xiang, A. (2021, March 1). *What We Can't Measure, We Can't Understand | Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. https://dl.acm.org/doi/abs/10.1145/3442188.3445888?casa_token=t1eX7ExNaPoAAAAA:6OkhCCFMx8y9GK6ll6UWAnwx6JWiauTpVI7n16Uc\

Buolamwini, J. (2019, February 7). *Artificial Intelligence Has a Racial and Gender Bias Problem | Time*. Time. https://time.com/5520558/artificial-intelligence-racial-gender-bias/

Communications, N. W. (2022, July 12). *Gender Bias in Search Algorithms Has Effect on Users, New Study Finds*. NYU. http://www.nyu.edu/content/nyu/en/about/news-publications/news/2022/july/gender-bias-in-search-algorithms-has-effect-on-users--new-study-

Cowgill, B., Dell'Acqua, F., Deng, S., Hsu, D., Verma, N., & Chaintreau, A. (2020, July 13). *Biased Programmers? Or Biased Data? A Field Experiment in Operationalizing AI Ethics | Proceedings of the 21st ACM Conference on Economics and Computation*. https://doi.org/10.1145/3391403.3399545

Heilweil, R. (2020, February 18). *Why algorithms can be racist and sexist*. Vox. https://www.vox.com/recode/2020/2/18/21121286/algorithms-bias-discrimination-facial-recognition-transparency

Niethammer, C. (2020, March 2). *AI Bias Could Put Women's Lives At Risk - A Challenge For Regulators*. Forbes. https://www.forbes.com/sites/carmenniethammer/2020/03/02/ai-bias-could-put-womens-lives-at-riska-challenge-for-regulators/?sh=34e830ea534f

Shrestha, S., & Das, S. (2022). Exploring gender biases in ML and AI academic research through systematic literature review. *Frontiers in Artificial Intelligence*, *5*. https://www.frontiersin.org/articles/10.3389/frai.2022.976838

Smith, G., & Rustagi, I. (2021, March 31). *When Good Algorithms Go Sexist: Why and How to Advance AI Gender Equity*. Stanford Social Innovation Review. https://ssir.org/articles/entry/when_good_algorithms_go_sexist_why_and_how_to_adv

* USE THIS NOTEBOOK FOR YOUR PAPER

Short Paper Assignment: The Impact of Gender Biases in Artificial Intelligence and Machine Learning Algorithms

When we think about the word “technology” today, the first thing that comes to mind due to the countless repetition of throwing this phrase around in news, work, school, etc. is Artificial Intelligence. Today, the number one trend is figuring out how we might implement Artificial Intelligence into our everyday lives, pushing us to ask ourselves questions: how might we leverage Artificial Intelligence in our company workflows? In the products we put out? In the services for the community? In school? Yet, with the rising implementation of Artificial Intelligence in recent years begs additional questions on ethics regarding gender imbalances. In this paper, I intend to discuss how gender biases might be ingrained into Artificial Intelligence algorithms—specifically in Machine Learning—and what the impacts of data science there are.

Artificial Intelligence has a young history with Machine Learning having an even younger history being a subcategory of Artificial Intelligence. Invented in the 1940s and gaining traction throughout the years up until today, Artificial Intelligence was invented by two male individuals. Their research would later support further developments in Machine Learning. While Artificial Intelligence encompasses the broader idea that machines or technology can exhibit intelligent behavior, Machine Learning—as Artificial Intelligence's subset—reflects the way machines can take on learning that allows algorithms to think for themselves based on the data implemented. Thus, Machine Learning has grown over the years from statistical use cases to implementation in our everyday lives; this is referencing anything from statistical analysis in advertising a product to facial recognition or virtual assistants. But, with these increased use cases comes a worrying factor: how bias might be mixed into the data. Heilweil discusses how Machine Learning is backed by historical data, and it is difficult to implement new data due to the capabilities of the system: Machine Learning learns data when we feed it to the algorithm overtime, hence, such historical biases being ingrained into the data despite evolving times *(Heilweil, 2020)*.

Data is not intentionally biased from the very beginning, but rather a product of many years worth of societally-breeded subconscious biases stemming from the media consume and what we are coached to believe in *(Shrestha & Das, 2022)*. Traditionally, with the presence of more men than women in the STEM fields, the presence of bias in data is unavoidable—unless society decides spontaneously and collectively to eradicate the patriarchy. The gender imbalance of professionals in Artificial Intelligence and data science related fields is discussed where only 22% of professionals in these fields are women *(Smith & Rustagi, 2021)*. Thus, the labels that go into the data inputted into these algorithms do not range far, often resorting toward binary factors of female or male, hence encouraging a lack of gender fluidity and both underrepresentation and overrepresentation of diverse parties *(Smith & Rustagi, 2021)*. This lack is further referencing that Machine Learning is often a product of historical data and its outcomes, thus allowing biases to seep through to today *(Cowgill et al., 2020)*. Those historical decisions include our hiring patterns, and the decisions we make to hire more men than women and assign male figures onto teams that develop these algorithms reap consequences; we are, in the end, a product of our own choices and actions—with evidence to back it up. While today calls into question the ethicality of incorporating Artificial Intelligence into company workflows to improve efficiency at the cost of many individuals losing their jobs, the past has reflected unethical behavior as well, calling into question the efficiency of Artificial Intelligence tools in hiring processes and the lack of unbiased evaluation in doing so *(Buolamwini, 2019)*. It is proven even from personal anecdotes of data scientists in the field regarding how big technology companies had previously utilized Artificial Intelligence tools that showed inherent biases in their algorithms toward passing males through the screening phases as opposed to females *(Buolamwini, 2019)*. This reflects a continuous loop of our choices: we are constantly stuck in this never-ending cycle of patriarchal behavior that even the digital worlds we create will continue to inhabit them and continue to limit change.

Because of our choices, we see this bias exhibited in a variety of tools utilized today with biased data science backing them. For instance, a case study of language models such as Google Translate has been studied, in which common translations will take the data of language inputted into the system and generate gender biased translations: automatically using the he/him/his pronouns when translating a sentence about an engineer *(Shrestha & Das, 2022)*. This is further seen in voice recognition tools, in which voice recognition tools commonly used in industries such as healthcare perform worse for women than they do men *(Smith & Rustagi, 2021)*. This not only poses to reflect past disadvantages women have had in opportunity, but it reflects future disadvantages women might have in advantage. This discriminatory behavior has long existed since the 1950s and has coined another case study regarding search algorithms that reveals how these algorithms commonly track societal gender inequality and bases the search results off these societal inequalities *(Communications, 2022)*.

Yet, while many sources reference these gender imbalances in Machine Learning and the effects it has on algorithms that reinforce even more gender inequality than before, the need to collect more inclusive data goes beyond just the impacts on workforce and personal lives. The impact includes life-threatening implications. An example referenced airbags and seatbelts. While supposedly life-saving tools, women reportedly have a higher chance of being unprotected during a crash and being harmed due to the creation of these products based on data collected about the physique of men. This causes increased danger for women—especially when pregnant—as they do not fit the standard measurements *(Niethammer, 2020)*. Other examples reference health apps, such as those that track heart attacks. While women are just as inclined to heart issues as men—if not more—many of these apps are trained with data sets that collect a history of male heart attacks and symptoms, while neglecting female heart attacks, thus making the results from these apps negligible for women. All these could severely lower female survival weights, which calls into question the ethical nature of Artificial Intelligence. How are we as a society supposed to advance and move forward with this life-changing technology if it is simultaneously life-threatening?

As a result, the matter of what figures in power do to combat these problems is called into question. As we ascend into the future, more and more regulations and diversity initiatives are being implemented in corporations to provide better opportunities for women and make gradual changes in the kind of data we feed Machine Learning algorithms. Already, initiatives are being taken to figure out ways to verify how data can be more inclusive: how can we go about our research to collect demographic data that reflects many and eliminates gender discrimination *(Andrus et al., 2021)*? This requires both change in implementing more diversity in data scientists in the field, but also change in partnering with gender experts to analyze the neutrality of Machine Learning datasets. While this by no means renders immediate change, it is somewhere to start as we delve deeper into the world of popular Artificial Intelligence tools today such as ChatGPT. Going forward, if we start implementing balanced practices now, the tools of the future will hopefully begin to shift gradually to reflect a more gender balanced society.

In [ ]:

```

```