A Compilation and Examination of How Current Data Science Practices Incapacitate at Risk Populations

Emily Christiansen
Department of Computer Science, University of Southern California, Los Angeles, California 90089, USA
(Dated: May 1, 2021)

As the use of data science techniques is in-arguably widespread across all industries, the use of flawed or incomplete data has lasting and often unseen effects to those who develop the systems. The origin and use of these methods can greatly affect the populations that are not creating them due to either intent to disenfranchise and discriminate or lack of awareness and knowledge of a particular blind spot in the data. In either case, the effects of using 'bad data' are plentiful and profound. In addition to this discussion, one of these flawed metrics of measuring representation if the Bechdel Test, using convolution neural network on the a dataset of posters of films returns a lackluster model with an accuracy of 42% but a very useful exercise and example of how giving proper context and motivations can help avoid common data science social pitfalls that often result in the damanging of others.

Keywords: Fairness, Bias, Neural Networks, Social Justice, Data Feminism

I. INTRODUCTION

While to most individuals the term data science is quite new and fresh, the ideas and effects of data science have been felt and known by communities across the world for years. Because of the inherent complexity of data analysis, even when it wasn't labeled as such, the work performed benefits those in power. The individuals "in power" in this case are also generally the same individuals who would gain from studies taking place. It is often the case where the question scientists and analysts set out to answer is in of itself flawed.

The goal of this paper is to provide the reader with the necessary vocabulary and real examples of data studies to effectively identify the attributes in data analysis, machine learning, and artificial intelligence that oppress groups either intentionally or unintentionally. Once armored with the data-specific vocabulary, a variety of cases and studies will be examined from the following fields: media and print, technology, and crime and law enforcement In each of these domains, some of the most glaring injustices performed by data science are compared to how a study or algorithm can be used with the principles of data feminism. Seeing how these cases differ can be a valuable tool when establishing one's own data driven system to ensure that no group or groups fall through the cracks or are inadvertently targeted.

Finally, to demonstrate quantitative analysis with data feminism, at the end of this paper we attempt to predict if a film passes the Bechdel test based on its poster using convolution neural networks. The accuracy of this model was only able to reach roughly 42% which is generally categorized as a failure. Nut this analysis proved a useful exercise in getting used to providing context and being critical to the social and global implications of the data that is being used.

II. BACKGROUND

As it will be discussed in later sections as well as in the vocabulary, providing context is an invaluable step in the analysis process. This informs the reader or user as to why certain steps are being taken, what is motivating choices, and what area of the matrix of domination may be prevalent (see below Vocabulary of Data Discrimination section). For these reasons it is becoming common practice among scholars of social data science and data feminism to provide a personal background and not shy away from the non-epistemic aspects of their analysis.

To continue with this method of sharing context, I will provide some key elements about my own background. I am a cisgender heterosexual white woman, I have lived in the state of California my entire life in upper middle class and upper class communities. I want to acknowledge that the lens I am seeing these papers, analyses, and algorithms through is that of a very high level of privilege. It is my desire to be as open an honest about my background as possible to better inform the reader as to how I draw conclusions or formulate theories. Providing this background is also essential in criticism of this paper, as it shows my blind spots and where my analysis can improve.

III. VOCABULARY OF DATA DISCRIMINATION

- Feminism: The term feminism in this context will be referring to the broad range of work, projects, and research performed in and around areas including, but not limited to, gender equality, racism, and able-ism. Also in reference to the positive forces and acts performed to create more just and equitable futures.[1, Chapter 1]
- Intersectionality: the additional facets of what constitutes an individual or group identity, the inter-

The Matrix of Domination

Table 1.1
The four domains of the matrix of domination

Structural domain Organizes oppression: laws and policies.	**Disciplinary domain** Administers and manages oppression. Implements and enforces laws and policies.
Hegemonic domain Circulates oppressive ideas: culture and media.	**Interpersonal domain** Individual experiences of oppression.

Chart based on concepts introduced by Patricia Hill Collins in *Black Feminist Thought: Knowledge, Consciousness, and the Politics of Empowerment.*

FIG. 1. FIG 1. The Matrix of Domination as displayed and described in *Data Feminism*.

section of all the elements such as race, class, sexuality, age, religion, and geography. How these experiences cross over one another to form a person's status, place, and perception of the world. [1, Chapter 1]

- Oppression: the act of repeatedly, intentionally, and systematically persecution and mistreatment of one group by another group, usually through power that the oppressor wields over the oppressed.

- Data Feminism: a way to think about data and data science in both what it can and cannot do. Data feminism is informed by intersectional thought as well as direct experience, the starting point of which is generally acknowledging that power is not evenly distributed across the world. Equally important is understanding that data feminism is not restricted to one gender, as for there to be a power imbalance there logically must be more than one gender.

- Open Data: An idea and movement that encourages and endorses the use of data to be made public in all contexts for any purpose. This movement is comprised of a vague network of governments, private individuals, and organizations.[1, Chapter 6]

- Zombie Data: The data sets are are simply released or made public with no documentation, references, or data dictionary of any kind to help those who access it in the future understand the set or its origin.[1, Chapter 4]

- Raw Data: This term as it is used colloquially is an oxymoron. Data is never truly raw, why certain questions are posed, how the questions are asked, under what conditions they are answered all influence the data collection process but are likely not taken into account themselves. Then when the data is "cleaned" more information about the data is removed. [2]

- Matrix of Domination: A conceptual and visual device used to explain how systems of power are configured and experienced. [3] Used to show how the intersection of gender, race, sexuality, ability, and geography results in unjust oppression or unearned privilege. Within the matrix of domination there are four main zones for which power can felt and wielded.
  - The Structural Domain: This domain is the organized form of oppression, concrete laws and policies that explicitly target or exclude groups of people - take for example the voting rights at the inception of the United States. [3]
  - The Disciplinary Domain: This domain describes the area of the indirect effects of the structural domain as well as the weight felt by those who are unable to work around the bureaucratic hoops established by the structural domain. [3]
  - The Hegemonic Domain: This domain is even more abstract than the disciplinary domain but neither the disciplinary domain nor the structural domain are possible without a thriving hegemonic domain, that being these structures and effects are only built upon the realm of thoughts and ideas in media and popular culture that are inherently flawed and oppressive in the first place. [3]
  - The Interpersonal Domain: The final domain is the most abstract but also the most motivating and impactful, the domain that churns about movement between these domains that materializes into change, which is the individual and personal experiences of oppression.[4]

- Privilege Hazard: The privilege hazard is a phenomenon that occurs when those in the most privileged areas and aspects of society such as those with good education credentials, accolades, and other desirable characteristics are “poorly equipped” to identify and recognize the plight of those who are oppressed. While this hazard begins on the interpersonal level, effecting the relationships and thoughts of the individual (more often the thoughts and feelings an individual doesn't have) this then builds into the other aspects of the matrix of domination. [5]
- Scarcity Bias: Often where these technologies are implemented initially is to make up for a scarcity bias, which revolves around the idea that there are not enough resources for everyone so technology can then be used in those places allowing the humans to keep either focus on the smaller picture.[6]
- The Three S's: Science, surveillance, and selling are the three dominant priorities focused on by corporations, governments, and universities that have been embracing, researching, and developing the fledgling data analyses to the mammoths they are today. These three words also showcase why in many cases there is missing data in important areas of life, but data is plentiful in others. [1, Chapter 1]
- Deficit Narratives: This is a narrative or story told about a particular subset of people that reduces culture, strengths, and creativity down to an oppressive simple statistic expressing the problems with this demographic.[1, Chapter 2]
- Co-Liberation: The idea of co-liberation revolves around the concept that all users and analysts alike suffer when one particular group is being discriminated against. Similarly to how data feminism isn't only about women, socially effective data science is not only about the groups that are traditionally oppressed.[1, Chapter 3]
- The God Trick: The God Trick is a perception specifically in data science visualizations that the reader is witnessing a fair and impartial data chart or representation because it appears that they have all the data. Often referring to a map that is showing entities “from above” when in reality these seemingly neutral visualizations enshroud the people displayed and the methodology used to collect the data. [7]
- Paradox of Exposure: This is often specifically when referring to the United States Census where resources are allocated based on population and other factors towards different districts. Those who would gain the most from these resources and funding are the least likely to respond, for instance, the undocumented immigrants who fear deportation [1, Chapter 4]

IV. EFFECT OF THE LACK OR PREVALENCE OF DATA FEMINISM IN DIFFERENT DOMAINS

Within each of the following sections, we will break down a paper or study and its contributions to the world of data science or data feminism. Each of the papers either uses data feminism and feminist techniques to justify their actions and decisions, or the papers (intentionally or unintentionally) use data science to oppress and remove power.

A. Technology

As the technology surrounding facial recognition has improved and become more streamline, the exposure the technology now has can be potentially concerning if the algorithms are not accurate across all demographics. The results found in *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification* where the most popular facial recognition algorithms consistently fail to correctly classify darker skinned females which an upper error rate of 34%.[8] The fact that this very popular facial recognition software, a system called Fitzpatrick which is used by dermatologists to identify skin tone and skin cancer, is still in use so widely with these glaring errors and discrepancies points to how the creators of these algorithms value the groups with darker skin tones.[9]

This failure in such a commonly applied technology as points to the privilege hazard held by the creators, because as we will discuss, the people who are classified correctly by the algorithm are almost always the same demographic as those who made the system. The privilege of being confident in these systems to not misidentify you as a wanted criminal or as a threat to national security. These are not problems that face the creators of these systems, but because the facial recognition softwares along with AI are being used in almost every aspect of large scale decision making, having data feminism principles evoked is essential.[10]

The civil liberties dangers of having flawed facial recognition software has been well documented, especially if used by law enforcement. People with darker skin tones have a higher risk as being identified and stopped by police.[11] This paper addressed the fact that these algorithms have to be made ”accountable to the public” and provide two contributions to this problem. First, they create a new data set of 1270 unique individuals such that the variety of skin tones is more balanced. This technique is not an uncommon route in data feminism unfortunately, as the data that would benefit or identified oppressed populations is not simply not in use but rather nonexistent. [12]

The second contribution is using the feminist concept of intersectionality to create a new scale of evaluating the faces based on a combination of skin tone and gender. Thus the categories became; lighter female, lighter male, darker female, and darker male. This allows the system to have a more fluid prediction process rather than try to classify gender and skin tone separately. This study also unfortunately has to fight against the paradox of exposure described in the vocabulary section. In order to make more accurate algorithms, we need more training data of a diverse group of faces. However, in a system that is still (clearly) error prone and flawed, individuals would be justified in a lack of trust in providing their faces for these algorithms.[1]

We have stated previously the importance of context and non-epistemic portions of analysis, how they can provide the motivation as to why certain questions had been posed. More over, embracing the non-epistemic portions of science is not anything particularly ground breaking. These values are what keep experiments and studies from being cruel or inhumane.[13] But how to structure this inclusion of non-epistemic values and context, specifically in this area of machine learning and AI? Proposed in *Model Cards for Model Reporting* published in the "Proceedings of the Conference on Fairness, Accountability, and Transparency". In this paper the authors propose a 'model card', a document describing the benchmark on which the model is evaluated. Most likely these benchmarks are along different cultural or phenotypic groups, or even between a combination of them forming intersectional groups. These model cards also inform the users of the model what the intended uses are, how to measure the performance, and provide ethical considerations and limitations.[14]

B. Crime and Law Enforcement

Where oppressive models have the most instant and long lasting impact is in law enforcement. While it is understandable as to the motivation for agencies and police to want to embrace the fledgling industry because the public's safety is never a waning concern. However, even well intended models have furthered the social equity divide which historically has always been troubled.[15] One type of algorithm that law enforcement has been implementing is a geography based crime prediction system. The software PredPol is based on seismic systems which uses a variety of different factors to predict a map with crime data.[10] Where a problem arises in with the prediction of what is called nuisance crimes, these are generally victim-less crimes that would go unreported such as public intoxication, disturbing the peace and vagrancy. A separate study on nuisance crimes in Miami-Dade County, Florida expressed how arrests for these crimes vary drastically depending on the cost of the properties nearby, primarily drug and alcohol violations in Black areas but the 'softer' alcohol and homelessness prosecutions in non-Hispanic white areas. [16]

Law enforcement agencies can choose whether or not to use these nuisance data, it does make for a more robust model but the inclusion of this data causes a very vicious feedback loop. The model works on the arrests made from its own predictions, so when police are told by the PredPol model to patrol certain areas where they had previous made the most arrests, these areas will predominantly be non-white, lower income areas. This then contributes to the over policing and the racial disproportionality in the US prison system. One of the most dangerous aspects of data science that is unaware of its errors or general lack of the data feminism components is the belief that because a scientific model is being implemented, that it is correct. [17]

This kind of 'algorithm arrogance' can blind law enforcement to the actual effects that these systems have. For this specific example, we will assume that the police *do not* want to use an algorithm that explicitly racially profiles potential offenders.[1]. Continuing, the PredPol algorithm is only accelerating the further policing of over policed lower income areas, effectively criminalizing poverty.[10] One potential solution is to include types of crimes that are seemingly ignored by this model which are 'white collar crimes'. These are clearly not victim-less crimes, criminality was one of the major root causes of the 2008 recession, along with negligence [18]. A study has been released that uses random forests to predict white collar crime and malfeasance based on this same PredPol model of using a geospacial format with a 90% accuracy. [19]

These law enforcement techniques don't need to be so punishing on the poorer communities and on communities of color, but until the methodology treats these nuisance crimes with the equal (or preferably less) severity as the white collar crimes, they will continue to feed the loop of poorly contextualized data. An easy 'knee jerk' argument that can be made is that these are still crimes that are being committed, a law has been broken and that these values aren't actually incorrect. A counter to this argument would be the obvious lack of any context with these values along with a lack of understanding of the importance of non-epistemic science. Numbers by themselves never 'speak for themselves', they have been curated by other individuals with opinions and motivations, there are heaps of context to understand about any individual data point[1]. Where data feminism is useful is opening the door to providing that context rather than it being the scientific status quo to present findings as facts.

C. Media and Print

When attempting to document and analyze the inequity of women in the UK film industry (or rather

[1] There is significant evidence from across the country and world that this is not the case, however for the case of introducing *new* data science techniques to the law enforcement field - this assumption can be made.

lack of women), paradoxically the analysis is primarily of men.[20] This conclusion was drawn in the paper *Data and Responsibility: Toward a Feminist Methodology for Producing Historical Stat on Women in the Contemporary UK Film Industry* by Natalie Wreyford and Shelley Cobb, the focus of which was to describe how a feminist methodology used in quantitative research benefited both the course of the process but also the results. These methods allow the authors to "provide a methodological framework for ensuring that quantitative research is rigorous, flexible, and aware of its own limitations."

The data analysis referenced in this study is the collection, parsing, and counting of women in the UK film industry's six primary roles: director, screenwriter, editor, cinematographers, producers, and executive producers. What makes this study of particular note is the steps that authors took to describe and contextualize the decisions made about the data and how to present the findings to the public. These feminist methods used in their analysis include working with incomplete or unstable data sets, especially in regards to labeling the working individuals in the film industry with a gender, race, and nationality. This is not an uncommon task when asking an algorithm to classify people, however the important step these authors take is addressing the power dynamic this forms. [20] While it is often said that classification of others is natural human instinct, it still activates the Hegemonic Domain of the Matrix of Domination.[21]

How the authors then address this activation of power is shifting their data collection to the much more tedious and time consuming work of scrubbing the individuals social media and other user-generated formats for their pronouns and other identifying characteristics such that the classification is from what they assign themselves. Because this is in the realm of film and media, this technique of looking to other social media sources for information is not unfounded because of the promotional aspect of this industry. This step is what sets these studies apart from classical data science, the combination of addressing a power dynamic and taking steps to resolve it. Compared to simply taking the easier route at the expense of the individuals being studied.

V. CONVOLUTION NEURAL NETWORKS AND THE BECHDEL TEST

The Bechdel Test is a very interesting metric that can be applied to almost any fiction work in media but specifically to cinema. A film passes the Bechdel Test if it passes the following criteria:

1. The movie has at least two women in it
2. They speak to each other
3. Their conversation is about something other than a man

This is not a metric whose goal is to give a binary answer as to if a movie is representing women well or not. Rather it is a way to measure the inequality between men and women in fiction and media on a larger scale and with a quick stroke. The goal of this algorithm is to provide a classification of the film, whether or not it would potentially pass, based on the most common poster of the film using convolution neural networks. Similar processes have been done attempting to classify a film's genre by its poster and this algorithm will function on a similar framework.

It should be noted that this is not a problem that we expect to get a robust and complete solution, as the concept itself is rather foolhardy. Instead this is a demonstration of the complicated current data science prediction techniques in the context of data feminism and to showcase simple tests like the Bechdel test that help individuals start evaluating media and popular culture through the lens of feminism.

A. Data Exploration

The data points for this algorithm was gathered from two different sources. The first being Kaggle, which has several thousand film posters from around the world, as well as a CSV file containing the corresponding titles, ratings, URL and IMDb ID values. The Internet Movie Database is a very common resource for analysis of film related data since all the films are classified by a unique ID number.[2] For the Bechdel data, this was drawn from a website where users submit a score of 0 to 3 for how many of the three criteria the film passes.[3] User of the site can also submit comments and corrections of other users submitted information, however this site is self policed and rather low fidelity. There is an unknown margin of error because these values are opinions provided by individuals. In a JSON file, each film is listed with its score, IMDb ID number, name, year, and ID specific to the website.

B. Data Preprocessing

The majority of the code work is based on a tutorial used to classify sport images and is not original to this paper.[4] This algorithm is working with film posters in their original JGP and RBG format so a significant amount of data preprocessing needed to occur before the films could be processed by the Convolution Neural Network (CNN). Before each poster is read into the system, however, the

---

[2] https://www.kaggle.com/neha1703/movie-genre-from-its-poster
[3] This website has a very simple and straight forward API to pull the data. https://bechdeltest.com/
[4] https://www.analyticsvidhya.com/blog/2020/10/create-image-classification-model-python-keras/

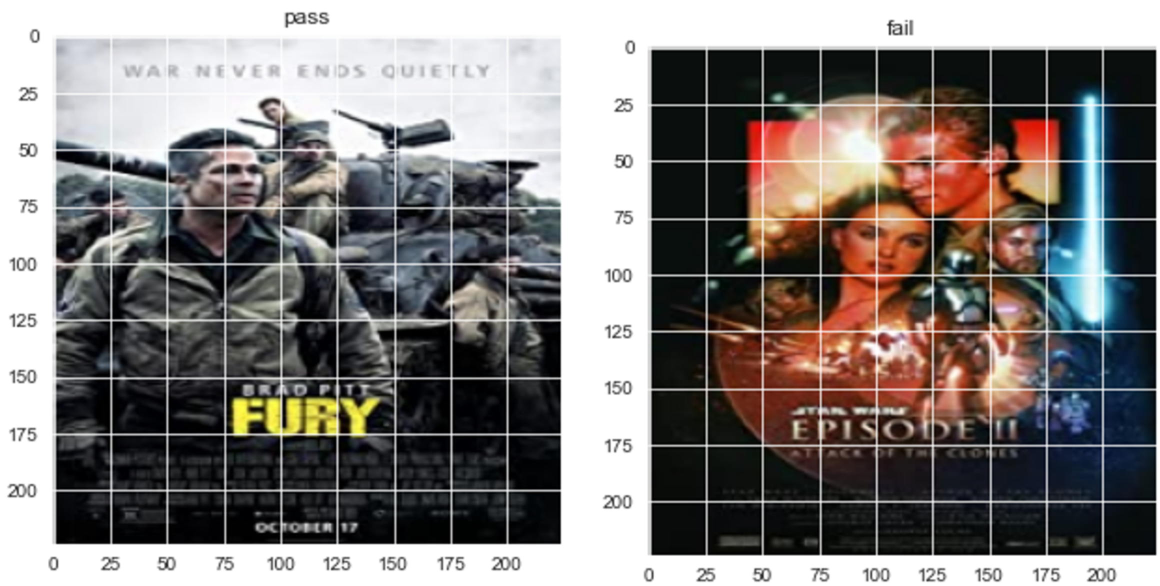

FIG. 2. FIG 2. The break down of two film posters after they have been warped to fit the 244-244 size, one that passed the Bechdel test and one that does not.

data that is available for film posters outweighs the data that is available on the Bechdel test, with less than 9000 reported films from the website, over 40000 URLs for posters from the Kaggle CSV file. Since there are certainly posters without a Bechdel test score and similarly a Bechdel test without a corresponding poster, we performed a check to see which IMDb ID numbers overlap in both data sets and that narrowed out datasets to roughly 7000 films. Each poster was pulled from the Internet with the listed URL, though there were (BLANK) HTTP errors, likely from out of date links so these films were removed from the data.

The Bechdel data had minor modifications, specifically removing the four classifications possible for a film, simplified to a pass/fail system because the values 1, 2, and 3 correspond to the sum of the criteria established rather than which criteria the film had, so this would cause a correlation between the scale of the values.

To effectively flatten the photos of the film posters, the standard technique for image processing was put into effect. That is, having a range of integers represent the intensity of each pixel and a third separate layer representing the RGB values since these are almost entirely color photos[22]. We then reshape all the posters so they are square (244, 244) and normalize it.

After several hours of attempting to minimize the computing power of training this model, the data set had to have been shrunk by roughly half, therefore the training data contains about 2500 posters and the testing about 600.

C. Model Selection and Feature Importance

The model selected for this analysis is a CNN primarily because it is the most popular and well suited for this color photo analysis because of how well it handles large dimensions of input and flattens them into a simple output[22]. Because of the aforementioned memory and computing power problems, the model itself was chosen to be relatively simple. There are three pairs of convolution and pooling layers using 'ReLu' activation, followed by dropout, flatten, and dense layers. For the optimization, an Adam optimizer was used along with Sparse Categorical Cross Entropy for the loss function.

D. Results and Conclusion of Analysis

After fitting the model to the broken down subset of the original data over the course of roughly an hour, the final accuracy was found to be roughly 48%. In conventional terms, this would be a failure because this is worse than assigning a classifier based on flipping a coin. However, again the goal of this analysis was not to make a train a model that correctly guesses if the film will pass the test but rather an exercise in exploring feminist ideas and testing in one of the most quantitative areas of data science. There are flaws and shortcomings to this process that will be described below.

E. Additional Notes and Considerations

Though the social implications of this model are relatively small, there are still data feminist related errors that should be addressed. Again, giving context and explanation whenever possible is tantamount to presenting findings and citing sources. The Bechdel test itself is presenting a very Western idea of femininity and Western priorities of women while the data set is pulling films on a global scale. Women in other countries may disagree with these criteria (again, it is a very flawed test) yet data from films the majority of Westerners have never seen is used to build this model. In order to remedy this, more research can be down to the specific demographics and areas where the data is drawn, because at this moment there is no geographical information on the films. It is unclear what percentage of the data is from the film industries of different countries and how this would effect the prediction.

VI. CONCLUSION

The spread of data science and quantitative analysis has been predictably reaching every field of personal and professional life. However, outdated ideas of epistemic science and blunt explanations without any context of reasoning are having increasingly detrimental effects on the population. Specifically, the population of people that are *not* writing and learning about these techniques. With the tools and vocabulary presented in this paper, the hope is to gather important examples and explain the necessity for non-epistemic reasoning and dialogue in all domains.

Many components go into these recurring themes of oppression and lack of empathy or understanding. Sometimes the algorithms and data tools are used purposefully against a population, but more often the effects are unintended, which is more dangerous. It was shown that with the unintentionally oppressive algorithms that the creators can then hide behind the intense math and computer science as well as proprietary software to not explain the models.

With the example analysis of predicting the Bechdel test outcome, a display was given of (a rather poor model as well as) a way to integrate the necessary context of a problem and motivation behind the choices made. This simple test of displaying a feminist idea also can introduce individuals to a simple way to analyze media and film themselves instantly since the Bechdel test is generally rather simple to take for an individual production.

All together, the tools, methods, and vocabulary of data feminism is a necessary inclusion in the social quantitative work going forward in the majority of the science community. Often these algorithms have unforeseen consequences and motivations, making these visible is one step closer to a more equitable society.

CODE AND DATA AVAILABILITY

Code is available at `https://github.com/97emilylc/final_project`.

Data is available at `https://bechdeltest.com/` and `https://www.kaggle.com/neha1703/movie-genre-from-its-poster`

ACKNOWLEDGMENTS

I would like to give a great deal of thanks to the teaching and grading staff in this DSCI-552 course during Spring of 2021. You feedback was fantastic and nearly made up for the online learning experience.

[1] C. D'Ignazio and L. F. Klein, *Data Feminism* (MIT Press, 2020).
[2] L. Gitelman, *"Raw Data Is an Oxymoron"* (MIT Press, 2013).
[3] *"Black Feminist Thought"* (Routledge, 1990).
[4] "your demand for statistical proof is racist", (2015).
[5] M. R. Mahoney, "whiteness and women, in practice and in theory", .
[6] V. Eubanks, *"Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor"* (St. Martin's Press, 2018).
[7] D. Haraway, "situated knowledges: The science question in feminism and the privilege of partial perspective", (1988).
[8] T. G. Joy Buolamwini, "gender shades: Intersectional accuracy disparities in commercial gender classification", (2018).
[9] T. B. Fitzpatrick, The validity and practicality of sun-reactive skin types i through vi, Archives of Dermatology **124**, 869–871 (1988).
[10] C. O'Neil, *Weapons of math destruction: how big data increases inequality and threatens democracy*, first edition ed. (Crown, New York, 2016).
[11] A. B. Clare Garvie and J. Frankle, he perpetual line-up: Unregulated police face recognition in america, (2016).
[12] C. Criado-Perez, *"Invisible women: exposing data bias in a world designed for men"* (Chatto Windus, 2019).
[13] K. C. Elliott and D. J. McKaughan, Nonepistemic values and the multiple goals of science, Philosophy of Science **81**, 1–21 (2014).
[14] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, E. Spitzer, I. D. Raji, and T. Gebru, Model cards for model reporting, in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19 (Association for Computing Machinery, New York, NY, USA, 2019) p. 220–229.
[15] S. Niedermeier and P. A. Cohen, *The color of the third degree: racism, police torture, and civil rights in the American South, 1930-1955* (University of North Carolina Press, 2019).
[16] N. Petersen, M. Omori, and R. Lautenschlager, (dis)order in the court: Examining neighborhood disorder prosecutions in miami-dade county, Justice Quarterly **35**, 1250–1279 (2018).

[17] A. J. Beck and A. Blumstein, Racial disproportionality in u.s. state prisons: Accounting for the effects of racial and ethnic differences in criminal involvement, arrests, sentencing, and time served, Journal of Quantitative Criminology **34**, 853–883 (2017).
[18] B. Berberoglu, *The global capitalist crisis and its aftermath: the causes and consequences of the Great Recession of 2008-2009* (Routledge, 2016).
[19] B. Clifton, S. Lavigne, and F. Tseng, Predicting financial crime: Augmenting the predictive policing arsenal, (2017).
[20] N. Wreyford and S. Cobb, "data and responsibility: Toward a feminist methodology for pro-ducing historical stat on women in the contemporaryuk film industry", (2017).
[21] G. Wood, "labels, welfare regimes and intermediation: Contesting formal power", (2007).
[22] A. Géron, *ands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (2019).