Scholars Journal of Arts, Humanities and Social Sciences OPEN ACCESS
Abbreviated Key Title: Sch J Arts Humanit Soc Sci
ISSN 2347-9493 (Print) | ISSN 2347-5374 (Online)
Journal homepage: https://saspublishers.com

Testing Data Feminism in India

Shubhankar Kashyap[1*], Avantika Singh[2]

[1]Ramjas College, Delhi University, Delhi, 110007, India
[2]Lady Shri Ram College, Delhi University, Delhi, 110007, India

DOI: 10.36347/sjahss.2021.v09i10.006 | **Received:** 03.09.2021 | **Accepted:** 19.10.2021 | **Published:** 25.10.2021

***Corresponding author:** Shubhankar Kashyap

Abstract **Original Research Article**

Catherine D'Ignazio and Lauren F.Klein in their book Data Feminism argue that data models reflect existing power structures and social hierarchies. We aim to test this hypothesis in India on Instagram. Instagram is one of the most accessible and politically engaging social media platforms in India which makes its data models appropriate subjects for our study. The research question of our study is "Do Instagram data models disproportionately prioritise accounts that publish majoritarian feminist content over intersectional feminist content in India?" The paper employs two methodological approaches; an experimental set-up under controlled setting for primary data collection through a positivist sociological approach kept under time bound observational study and secondary data qualitative analysis. This paper first analyses the biases and preconceived notions which cloud digital data models. It further elaborates upon the concept of Data Justice which acknowledges historical inequalities and power differentials amongst communities that drive data collection. The paper attempts to test this hypothesis through an experiment. The experiment includes creation of two Instagram accounts dedicated to two different forms of feminism. Account "A" would publish content ascribing to popular feminist ideals i.e. non-intersectional and Account "B" would publish intersectional feminist content. The creation of new accounts is critical for establishing a causal relationship between data models and disparity in accounts growth, as a pre-established follower count would affect the accounts engagement. For a period of three months, both the accounts will employ the same strategies to increase user engagement/reach. The impact of these strategies on metrics such as follower count, post likes, post reshares, post comments, profile visits, frequency and duration of story views, and duration of post visibility would be documented through a weekly monitoring system. The data collected for each metric through this system would be graphed to determine a trend line to illustrate the conclusion for the hypothesis. The findings of the experiment provide important theoretical and practical implications for the development of more equitable data models.

Keywords: Data, data bias, machine learning, feminism, algorithms, social media.

INTRODUCTION

Automation is a contemporary consequence of advancement in technology and is becoming increasingly pervasive in the lives of people. Machine learning algorithms drive this force of automation. This paper explores the complex relationship between machine learning algorithms that are fundamental to user driven applications like social media and microblogging websites and social conditions such as marginalization on the basis of identity markers, stereotypes, prejudice and structural discrimination. This paper uses the discourse around Data Feminism and Data Justice with emphasis on innate data biases and algorithmic anomalies to understand how these situations occur and how adverse their actual consequences are. To ground the validity of the arguments made by existing data ethic academics like Catherine D'Ignazio, Lauren F. Klein, Cathy O'Neil and Sarah Boechter, we test the hypothesis that existing power structures and structural inequities are reflected, adopted and developed upon by machine learning models using a simple experimental model that involves studying account reach and user interactions of two Instagram accounts which espouse two different ideas of feminist theory and cater to a divergent group.

Alexandra Olteanu, a post-doctoral researcher at Microsoft Research, US and Canada (How We Can Solve Big Data's Bias Problem, n.d.) lists down ethical considerations and how there exist practical problems which prevent academia from coming up with solutions for latent bias in machine learning algorithms. Their work determines that there are two ideas of fairness, where the idea is to cluster at least two individuals on the basis of some similarity. The challenge then becomes as to what attributes should be included and what type of mathematical function must be used to eventually predict why some attributes would have more propensity for error then others. This is further extrapolated to include “group fairness” to measure aggregate error rates. The second approach they talk about is “outcome fairness” which needs to be extended to “process fairness” i.e. whether the process of data treatment involves representative involvement to remove individual bias. The process of improvement also demands a critical appraisal of historical data with inherent bias driven models, discriminatory against certain groups.

This paper aims to tackle some of these concerns. It critically examines the work of academics who give numerous instances of historical data across platforms and systems (Examples: welfare systems in the UK, prison systems in USA, social media corporations) being inherently biased and oftentimes discriminatory. It also contextualizes the reasons behind these model biases with respect to human intervention in development and implementation by examining the composition of developers’ rooms, justice systems and state authorities, thereby establishing a clear line of causation. Furthermore, this paper involves experimental analysis using machine learning tools that help to determine significance of differentiability between two lines of thought and cluster groups on the basis of their interactions with the social media interfaces propagating them. Through this, insights on group error identification via clustering may be obtained. This paper also aims to maximise “process fairness” by balancing out perspectives on content created and posted on these accounts.

METHODOLOGY

The paper employs two methodological approaches: positivist quantitative approach and secondary data qualitative analysis. For primary data collection, a controlled experiment was conducted under time bound observational study. Positivist quantitative approach was selected for the paper in order to examine the cause and effect relationship between social media algorithms and determine a numerical value for it. The experiment’s results create reliable insights from the data gathered (Merton, R. (1968). Social theory and social structure. New York, NY: Free Press (Original work published 1949). Secondary data sources used for literature review includes analysis of secondary data sources including existing journals, academic papers, published books, and newspaper articles.

LITERATURE REVIEW

Data and Bias

Data Bias is defined as an overarching phenomenon, arising due to certain innate characteristics of ML/AI models, wherein a given dataset is not representative of the actual population or phenomena of study (Krishnamurthy, 2021, para 3). The dataset either omits valuable variables or incorrectly prioritises the effect of some over the others, thereby leading to inaccurate study of predictive variables. These biases end up interfering in the working of models of all scales; whether they’re simple regression models or frameworks for complex big data and IoT applications. Bias in statistical models are oftentimes measured but ignored (Singh, 2018) since they don’t affect the degree of precision, unlike the sampling error. However their qualitative implications are prolonged and far reaching, especially due to specific biases that we shall highlight in this paper. We encounter three types of biases in these models (Krishnamurthy, 2021, para 4) that directly affect the conclusions of this paper. Response or Activity Bias (generated due to user feedback to results of a typical model, examples include Amazon reviews, Wikipedia entries etc), Selection bias due to feedback loops (bias affecting generation of data used to train models due to consecutive selection via feedback loops, examples include recommender systems, content and ad personalization) and Societal bias (Inherent bias in content produced by humans, examples include usage of racial and/or gender stereotypes in social media content or curated news articles).

Biases in traditional ML models would not have become problematic had it not been for a particular approach to data treatment; this approach termed as “data fundamentalism” (Crawford, The Hidden Biases in Big Data, 2018, para 1) is the notion that correlation always indicates causation and that massive data sets and predictive analytics always reflect objective truth. Both Krishnamurthy and Crawford quote the Hurricane Sandy Twitter study (2012) and the Google Flu Trend mishap. The first (Crawford, The Hidden Biases in Big Data, 2018, para 3) aimed to study tweets regarding the hurricane to obtain findings on communal responses in New York City. Their findings were a mix of expected (grocery shopping hauls at night) and surprising (nightlife picking up a mere day after) insights, largely because the majority of tweets came from Manhattan and not from the more affected areas of Breezy Point and Coney Island. Hence, Twitter data failed to convey the experiences of people outside the privileged group its model was prioritising, leading to massive misrepresentation and biased conclusions. Similarly, Google Flu Trends massively overestimated flu rates in the USA in the 2013 flu season as 11%, almost double of the Centre of Disease Control’s measure of 6%. These examples sufficiently explain how data models are inherently subjective.

These biases become pervasive across retributive justice and welfare systems as they become increasingly automated. In Automating Inequality: How

High Tech Tools Profile, Police and Punish the Poor (2018), Virginia Eubanks studies automated systems which dictate welfare provisions, housing and child protection services. From designing predictive algorithms to determine risk of child abuse and neglect by scoring new-borns to predicting “employment sustainability” by factoring in gender and race (Book Review: Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor by Virginia Eubanks, 2021), the state deliberately uses inherently biased variables to create a prejudiced automated decision maker, free from accountability, from whom seeking recourse is a long and tedious process with a massive human cost. The people denied these benefits never seek justice because they can barely comprehend the systems they have to challenge and the perception of objectivity in tech often allows state departments to not deal with these complaints seriously. This idea is corroborated by other academics as well. In Weapons of Math Destruction (2016), Cathy O’ Neil refers to mathematical and algorithmic models, which aim to quantify important traits such as prison conviction rates, as “Weapons of Math Destruction” or WMDs. They are effectively proprietary with complex math calculations, inaccessible to lay persons and affect large numbers of people (Lamb, 2016). They have biases such as sexism and racism encoded in their algorithms and can cause massive crisis; O’ Neil gives the classic example of the 2008 financial crisis with subprime mortgages as the WMD in question and mortgage reliant rural African American

This paper grounds this idea specifically for social media and lifestyle apps which not just dictate our social interactions but also the amount of privilege and opportunity marginalized people with aspirations can leverage to secure employment and status. In Technically Wrong: Sexist Apps, Biased Algorithms and other Threats of Toxic Tech, Sara Wachter-Boettcher provides numerous accounts of how impersonal models have affected the mental and emotional health of its users. Whether its Facebook’s reminder algorithms that can’t differentiate between happy and tragic experiences and hence issue potentially triggering reminders to its users or Apple’s Siri which is unsure of how to respond to alarming requests concerning self-harm or assault, the problematic tones continue to persist and are not removed unless faced by massive outcry.

Furthermore, the author also explains where these biases stem from by explaining the composition of a typical developers’ room. The developers’ room for “Glow”, a well known app for menstrual health is a classic example for this. Effectively developed solely by cis het white men, the app is riddled with prejudice of how people who menstruate across the gender spectrum view their sexual experiences. Moreover, the potential to incorporate progressive opinions at the very outset is also minimal because these rooms blatantly disregard any alternative voice and invalidate lived experiences as evidenced by many accounts of minority software developers in Silicon Valley. This history of minority exclusion has a long drawn structure in tech companies, such as recruitment from institutions such as Stanford, Harvard and MIT instead of historically Black and Hispanic-serving institutes, the creation of a “corporate culture” that involves traditionally masculine activities like sports and BBQs and many other subtle policies that prevent certain groups from making decisions.

All these ideas boil down to a single line of reasoning, there is an imminent need to make data algorithms unbiased as their intrusive presence in our lives entail greater and more dangerous ramifications.

Data Justice

Data is ubiquitous. It is present in all corners of the internet by all the sites that any internet stroller visits. Discourse around data is around its technical efficacy and potential. However, the current discourse lacks a perspective that focuses on social justice in the realm of data, data authorities, data regulations and data collection (Dalton *et al*., 2016). More than 4.66 billion people actively use the internet (Statista, 2021a). Surveillance using data mechanisms disportadly affect the poor with administration systems releasing law enforcements over areas occupied by lower income households (O’Neil, 2016). Data surveillance is also used specifically to curb illegal immigrants by tracking their movement (Taylor, 2015).

Political and social awareness around data issues are hard to emerge because of its seemingly invisible impact. One of the key reasons behind lack of political action regarding data issues is that even if privacy violations are a concern, digital platforms are a necessity for most people which means digital users cannot afford to risk their access to digital platforms (Turow *et al.*, 2015). Redressal for data abuse is hard to formulate due to the connected nature of data on the internet and lack of authority that can be held accountable (Taylor, 2017).

Essentially, there are three approaches to conceptualise data justice. The first approach explains the exacerbation of power inequalities via data (Johnson, 2014). The second approach elaborates upon the utilization of data to ensure representation of the poor within it (Heeks and Renken, 2016). Lastly, the third approach explores the effect of data surveillance on organisations working for social justice (Dencik *et al*., 2016).

The first approach argues that data used for administration is gathered through a normative lens that leaves out marginalised groups. Thus, data systems largely occupy privileged groups which leads to the formation of inequality in administrative data (Johnson, 2014).

Heeks and Renken (2016) proposed data justice within the context of international and human development. They argue that while Sustainable Development Goals recognise data and justice as two separate concepts, scholarly research must focus on its intersection. They aim to develop a structural approach towards data that incorporates social and political justice. They refer to the United Nations Declaration of Human Rights (UN General Assembly, 1948) in order to vouch for the connection between the right to data ownership and its fair use.

The third approach emphasises the danger of dataveillance to curb protests and activism against states. It describes the specific data systems that prefer those who are in positions of power. Anti-surveillance and social justice activism have scope for collaboration under this formulation of data justice (Dencik *et al*., 2016).

The design of data justice cannot be uniform for all since the way data is used, deployed and gathered varies on the basis of states. The concept must take shape according to the local context it is used in. However, there has been a debate regarding data regulations hindering the efficacy of development agencies in the developing world. The underlying principle of the debate argues that development agencies should be able to equip data if it leads to communal development (Taylor, 2017). On the contrary, some scholars argue that a rational approach to data justice that is framed according to the needs of the people rather than their rights is achievable and necessary (Taylor, 2017).

Individual rights based approach towards data justice fails to recognise that data injustice does not exist on an individual level but is rather perpetuated collectively (Taylor *et al*., 2017). Data companies gather data to dissect profiles according to groups rather than individual traits.

Taylor (2017) sets preconditions for a data justice mechanism. It must address the nuance and vastly intricate tasks performed by big data systems. Secondly, it must be designed to attend to the benefits and negative potential of data due to its unparalleled capacity under which it influences the digital world currently. This implies an approach must have three pillars: visibility; digital engagement and disengagement; and bereft of data discrimination. Visibility requires protection of privacy and representation of those who are overlooked during data gathering including but not limited to marginalised groups. Data engagement protects data gathered for developmental and administrative purposes from commercial exploitation. The agency over how and what data is being used must rest within the person whose data is being procured. Data discrimination becomes harder to identify as data systems become more complex thus, methods must be created in order to reduce data bias along with government driven accountability measures for its enforcement.

Data Inconsistency

Data inconsistency is usually not spoken of in the context of biases. Rather, it is oftentimes used to refer to multiple data tables in a database which provides the same kind of information (Data Redundancy and Data Inconsistency Hurts Your Business, n.d.). However, data inconsistency is the first impact of data bias on as it directly associates with database design. This phenomenon emerges from a space of data anomalies, evidenced by repeated or identical information. Increase in anomalies and their overarching impact lead to data redundancy. Redundancy stems directly from poor initial relational database design wherein information is scattered and not accurate. Furthermore, the faults of the initial designs exacerbate if the framework is extended to include more variables.

Inconsistency is compounded by redundancy but is differentiated from regular anomalies as it focuses on content as well and not purely design. This is where human propensity for putting forth information in out there compounds existing database problems (Data Redundancy and Data Inconsistency Hurts Your Business, n.d, para 6).

Problems associated with data inconsistency are not purely technical and can’t be fixed purely through data integration. Semantic and subjective questions continue to linger. In the absence of intentional interventions, a trained machine learning model does amplify undesirable biases in the training data (Hooker S, 2021, para 1). A rich body of work has examined these inconsistencies to produce bias relating to race, gender and geo diversity in the machine learning models (Barocas S. Hardt M. Narayanan A, 2019).

How data inconsistency and its ramifications are viewed will influence the mitigatory practices that we will undertake. These include de-biasing of the data pipeline through re-sampling and re-weighing, understanding the sensitive features responsible for problematic bias and providing comprehensive labels for protective attributes and all proxy variables (Hooker S,2021, para 4). However, these approaches pose some major problems which include complexity in labelling all sensitive features and protected attributes in large domains of data (Hooker S, 2021, para 5). Moreover, multi-dimensional data i.e. inclusive of images, songs, and textual content have layers of data points, making it hard to comprehensively label every variable. Even if labelling does happen at scale, algorithms can still leverage proxy labels to reconstruct the forbidden label. Furthermore, standardisation in the variables does not really happen either, thereby leading to innate inconsistencies.

If we can't guarantee complete eradication of bias and inconsistency, we have to look at the interaction between data and the model design choices that have been made. Hence, the only way to combat inconsistencies is through understanding the model design (Moving beyond algorithmic Bias Is a Data Problem, 2021) which includes architecture, loss function, optimizer, hyper parameters. Model design choices made to maximize test-set accuracy do not hold static other properties we care about such as robustness and fairness (Moving beyond algorithmic Bias Is a Data Problem, 2021, para 8). They operate with fixed parametric constraints and introduction of new criteria set off a chain of new trade-offs. Notions of fairness in ML models often coincide with how underrepresented minorities are in the given models, leading to amplification of biases (Moving beyond algorithmic Bias Is a Data Problem, 2021, para 9). Facial analysis data sets for instance reflect a preference or importance for lighter skinned people with far higher model error rates for dark skinned women (Buolamwini, J., & Gebru, T, 2018). Furthermore, models trained on data sets with limited geo diversity show sharp degradation on data drawn from locales (S. Shankar, Y. Halpern, E.Breck, J.Atwood, J. Wilson, D.Sculley, 2017). In both these examples, the algorithmic bias a model learns can be attributed to over and under representation of a protected attribute within a data set category. Most real world data, especially the one obtained via user input have a skewed distribution similar to the aforementioned. The skew in feature frequency leads to disparate error rates on the underrepresented attribute, which is usually protected.

However, there is scope to further better design choices. For instance, the widespread use of compression and differential privacy techniques in sensitive domains like health care diagnostics (Moving beyond algorithmic Bias Is a Data Problem, 2021, para 12). Here understanding the distribution of error is paramount in understanding and mitigating potentially adverse harm to human welfare. These results have to be undertaken with caution before their usage in sensitive domains, this mechanism definitely provides a valuable roadmap to reduce harm. For instance, auditing for problematic biases in large data sets and usage of compression identified exemplars (CIEs), a human-in-the-loop tooling which examines data points disproportionately impacted by compression. Hence, data inconsistencies help in identifying potential data biases by conclusively trying to deal with algorithmic biases.

Differentiating between Majoritarian Feminism and Intersectional Feminism

The term "intersectionality" was coined by professor Kimberlé Crenshaw in 1989 to shed light on the plight of black women's suffering in the post Civil Rights Act era (Crenshaw, 1989). Intersectionality refers to the recognition of different individual and identity characteristics intersecting with each other to create unique lived experiences for individuals with overlapping identities. The concept is critical to understanding specific problems marginalised groups face rather than framing their obstacles uniformly on the basis of one characteristic. Crenshaw (1989) posits that the legal framework in the United States lacks the acknowledgement of intersectionality which creates further obstacles for underprivileged groups amongst minorities. Crenshaw refers to the DeGraffenreid v. General Motors case of 1976 to highlight the shortcomings of legal policies that are founded on single identity discrimintaion ultimately fail to protect the most vulnerable. Intersectional feminism espouses gender equality that is based on acknowledging different identities of women and thus responding equitably to it (Coaston, 2019).

Another key principle of intersectional feminism is to reflect on social, economic and legal structures that are built on the exploitation and discrimination of marginalised groups. Crenshaw argues that these structures are not merely influenced by a few bad individuals but rather a system that eliminates any meaningful growth for the underprivileged.

On the contrary, mainstream feminism claims that while gender inequality is a problem of the status quo, it must be resolved within the structures that exist today instead of targeting the deep rooted discrimination. Mainstream feminism is primarily occupied by women who exercise privilege due to their identity traits in specific contexts. White feminism in the US, for example, centers the experience of white women in the feminism discourse while disregarding the problems faced by women of color (Moon & Holling, 2020). This form of mainstream feminism focuses on individual gains over collective issues which impacts the interpretation of feminism that occupies the popular discourse. The critique of white feminism emphasizes its oversimplification of gender equality. Intersectional feminists argue that white feminists conceptualise equality in its same formulation for them and women of color which is erroneous due to the different struggles faced by women of color compared to white women (MacIntyre, 2013). White feminists are also accused of supporting the capitalist structure that often exploits women of color and poor women (Beck, 2021). The issues at the center of discourse vary for white feminists and intersectional feminists. Intersectional feminists focus on racism, problems facing queer people and poor women while white feminists often focus on succeeding individually within the capitalist structure.

White feminism is also critiqued to be exclusionary towards women of color and women from lower economic sections that do not share the same goals as women with white privilege. Shyerl Sandberg (2013) writes that if women toil enough in their workplace while exercising their dominance, they will achieve success. Lean in feminism was decried for putting the onus of success on individuals while balancing their home lives. The approach was critiqued by various feminists (Gibson, 2018) to turn the blame of failure towards women, which adversely affects women, rather than focusing on the structures responsible for discriminatory problems in the workplace.

Mainstream feminism does not restrict itself to a single identity group, it is based upon privilege holding sections within the movement in different landscapes. In India, savarna feminism occupies the mainstream discourse as it focuses on the experiences and challenges solely faced by upper-caste women that alienates the problems of lower castes (Dhanaraj, 2018). As Devika (2020) puts it,

"*Savarna refers to the privileged caste-communities that, from pre-colonial times, controlled land and other material resources and ritual practices, and continued to do so to a large extent even later. Avarna refers to those oppressed groups that laboured for the savarna and were subjected to degradation through such practices as untouchability and unseeability, and whose exclusion from social power continues in different ways despite these groups having achieved economic presence and education*."

Savarna feminists due to their caste privilege exercise social power over lower caste feminists which leads to erasure of lower caste women's problems and lower caste women facing caste oppression perpetuated by the savarna feminists themselves (Dhanaraj, 2018).

Experiment Design

Research Design

We make two Instagram Accounts showcasing content of two separate lines of feminist discourse – Exclusionary Feminist Discourse and Intersectional Feminist Discourse. 2 posts every day are posted in two time frames i.e. 19:00-21:00 hours and 01:00-02:00 hours. Insights from each account are recorded on a Google Sheet. Insight variables are Account Reach, Impressions, Profile Visits, Followers, Total Likes, Total Shares, Total Comments and Total Saves. Each individual post, its type and post specific insights are also recorded on a sheet.

Analysis involves consideration of all cumulative values of all variables and focuses extensively on mapping weekly trends and not individual insights. The experiment was performed over a period of 29 days.

Assumptions

1. Viewer is indifferent to the qualitative nature of the content
2. Instagram insight computational models that are devised through AI (Marr, 2018) perform accurate weekly mapping of page insights, with no omissions of crucial data points.

Our dataset consists of insights obtained via Instagram analytical tools for two accounts with differing kinds of feminist content. This data set comprises 812 observations with 16 variables. The primary aim is to examine these variables and visualize key data with regards to accounts reached, impressions achieved, no of likes, no of shares, no of saves and no of comments. We assume that account reach and cumulative impressions are the primary metrics which allow us to measure data algorithms operating in isolation.

Steps of Analysis

Our approach towards analysis is three-pronged, before that, we extract variables from the data set and encode them to perform easy analysis. Encoding is as follows:

Variable	Urban Feminist	Feminist Union
Accounts Reached	X1	Y1
Impressions	X2	Y2
Profile Visits	X3	Y3
Followers	X4	Y4
Likes	X5	Y5
Comments	X6	Y6
Saves	X7	Y7
Shares	X8	Y8

Tests performed

1. Performing t.test to determine whether there exists any significant difference in the values of X and Y variables
2. K - means clustering to understand distribution of account reach and impressions variables

Hypothesis

Ho: There exists no significant difference between true values of Xi's and Yi's i.e. (p value > 0.05) (i = 1,2,3,4,5,6,7,8) H1 : There exists a significant difference between true values of Xi's and Yi's i.e. (p value > 0.05) (i = 1,2,3,4,5,6,7,8)

Data Set Importation

```
library(readxl)
Paperfinal_df <- read_excel("Paper2.xlsx", col_types = c("numeric","numeric", "numeric",
"numeric", "numeric","numeric", "numeric", "numeric", "skip","numeric", "numeric", "numeric",
"numeric","numeric", "numeric", "numeric", "numeric"))
library(shiny)
library(leaflet)
library(RColorBrewer)
library(leaflet.extras)
library(lubridate)
##
## Attaching package: 'lubridate'
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
library(gridExtra)
library(tidyr)
library(dplyr)
##
## Attaching package: 'dplyr'
## The following object is masked from 'package:gridExtra':
##
##     combine
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
library(gridExtra)
library("cluster")
X1_convert = Paperfinal_df[["X1"]]
X2_convert = Paperfinal_df[["X2"]]
X3_convert = Paperfinal_df[["X3"]]
X4_convert = Paperfinal_df[["X4"]]
X5_convert = Paperfinal_df[["X5"]]
X6_convert = Paperfinal_df[["X6"]]
X7_convert = Paperfinal_df[["X7"]]
X8_convert = Paperfinal_df[["X8"]]
Y1_convert = Paperfinal_df[["Y1"]]
Y2_convert = Paperfinal_df[["Y2"]]
Y3_convert = Paperfinal_df[["Y3"]]
Y4_convert = Paperfinal_df[["Y4"]]
Y5_convert = Paperfinal_df[["Y5"]]
Y6_convert = Paperfinal_df[["Y6"]]
Y7_convert = Paperfinal_df[["Y7"]]
Y8_convert = Paperfinal_df[["Y8"]]
```

Performing T Test

```
t.test(X1_convert,Y1_convert)
##
## Welch Two Sample t-test
##
## data: X1_convert and Y1_convert
## t = 1.3879, df = 43.451, p-value = 0.1722
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -22.87985 123.98330
## sample estimates:
## mean of x mean of y
## 266.2759 215.7241
t.test(X1_convert,Y1_convert, alternative = "greater")
##
## Welch Two Sample t-test
##
## data: X1_convert and Y1_convert
## t = 1.3879, df = 43.451, p-value = 0.08612
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
## -10.66363 Inf
## sample estimates:
## mean of x mean of y
```

266.2759 215.7241
t.test(X8_convert,Y8_convert)
##
Welch Two Sample t-test
##
data: X8_convert and Y8_convert
t = 1.4549, df = 35.812, p-value = 0.1544
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.3941858 2.3941858
sample estimates:
mean of x mean of y
1.2758621 0.2758621
t.test(X8_convert,Y8_convert,alternative = "greater")
##
Welch Two Sample t-test
##
data: X8_convert and Y8_convert
t = 1.4549, df = 35.812, p-value = 0.0772
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
-0.1605468 Inf
sample estimates:
mean of x mean of y
1.2758621 0.2758621

Interpretation (T Test for i = 1,8)

We perform 2 tailed and greater than test to conclude that though the mean value of X coordinate is greater, the p value >0.05, hence we accept null hypothesis and conclude that there is no significant difference between Xi's and Yi's

t.test(X2_convert, Y2_convert)
##
Welch Two Sample t-test
##
data: X2_convert and Y2_convert
t = -0.23529, df = 53.299, p-value = 0.8149
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-93.59168 73.93650
sample estimates:
mean of x mean of y
287.5862 297.4138
t.test(X3_convert,Y3_convert)
##
Welch Two Sample t-test
##
data: X3_convert and Y3_convert
t = -1.3863, df = 50.17, p-value = 0.1718
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.8576957 0.3404544
sample estimates:
mean of x mean of y
1.827586 2.586207
t.test(X4_convert,Y4_convert)
##
Welch Two Sample t-test
##
data: X4_convert and Y4_convert
t = 1.3093, df = 30.793, p-value = 0.2001
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.1154749 0.5292680
sample estimates:
mean of x mean of y
0.24137931 0.03448276
t.test(X5_convert,Y5_convert)
##
Welch Two Sample t-test
##
data: X5_convert and Y5_convert
t = -0.28424, df = 43.341, p-value = 0.7776
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.953594 1.470836
sample estimates:
mean of x mean of y
2.344828 2.586207
t.test(X6_convert,Y6_convert)
##
Welch Two Sample t-test
##
data: X6_convert and Y6_convert
t = 0.95901, df = 43.484, p-value = 0.3429
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.2280422 0.6418353
sample estimates:
mean of x mean of y
0.4137931 0.2068966
t.test(X7_convert,Y7_convert)
##
Welch Two Sample t-test
##
data: X7_convert and Y7_convert
t = -0.34192, df = 48.867, p-value = 0.7339
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.4743245 0.3363935
sample estimates:
mean of x mean of y
0.3103448 0.3793103

Interpretation (T Test for i = 1,8)

We perform 2 tailed and greater than test to conclude that though the mean value of X coordinate is greater, the p value >0.05, hence we accept null hypothesis and conclude that there is no significant difference between Xi's and Yi's

```
t.test(X2_convert, Y2_convert)
##
##  Welch Two Sample t-test
##
## data:  X2_convert and Y2_convert
## t = -0.23529, df = 53.299, p-value = 0.8149
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -93.59168  73.93650
## sample estimates:
## mean of x mean of y
##  287.5862  297.4138
t.test(X3_convert,Y3_convert)
##
##  Welch Two Sample t-test
##
## data:  X3_convert and Y3_convert
## t = -1.3863, df = 50.17, p-value = 0.1718
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.8576957  0.3404544
## sample estimates:
## mean of x mean of y
##  1.827586  2.586207
t.test(X4_convert,Y4_convert)
##
##  Welch Two Sample t-test
##
## data:  X4_convert and Y4_convert
## t = 1.3093, df = 30.793, p-value = 0.2001
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1154749  0.5292680
## sample estimates:
##  mean of x  mean of y
## 0.24137931 0.03448276
t.test(X5_convert,Y5_convert)
##
##  Welch Two Sample t-test
##
## data:  X5_convert and Y5_convert
## t = -0.28424, df = 43.341, p-value = 0.7776
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.953594  1.470836
## sample estimates:
## mean of x mean of y
##  2.344828  2.586207
t.test(X6_convert,Y6_convert)
##
##  Welch Two Sample t-test
```

```
##
## data:  X6_convert and Y6_convert
## t = 0.95901, df = 43.484, p-value = 0.3429
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2280422  0.6418353
## sample estimates:
## mean of x mean of y
## 0.4137931 0.2068966
t.test(X7_convert,Y7_convert)
##
##  Welch Two Sample t-test
##
## data:  X7_convert and Y7_convert
## t = -0.34192, df = 48.867, p-value = 0.7339
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.4743245  0.3363935
## sample estimates:
## mean of x mean of y
## 0.3103448 0.3793103
```

Interpretation (i = (2,3….,7)

For all the variables from X2 to X7, the p value > 0.05, hence we accept the null hypothesis and conclude that there is no significant difference between Xi's and Yi's

K Means Clustering

K means clustering model enables us to partition our data set into clusters of high, low and medium weekly reach and impressions to determine the impact presence of each account in the given cluster:

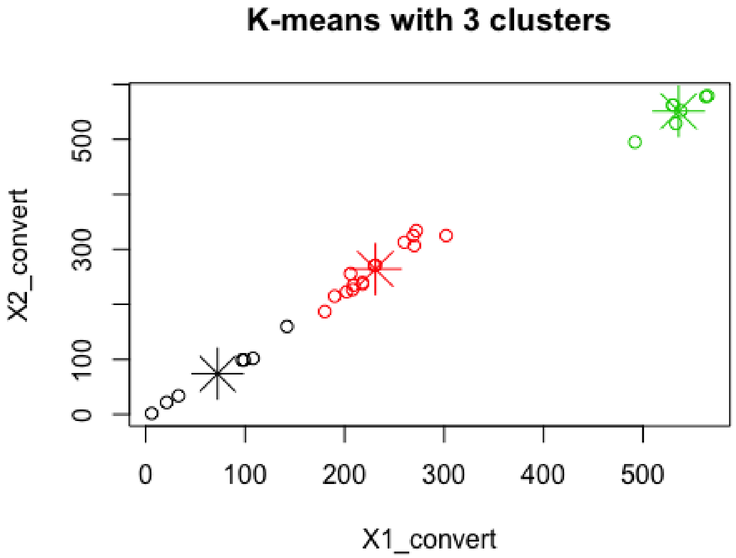
```
cbind(X1_convert,X2_convert) -> uf_df
kmeans1.re <- kmeans(uf_df, centers=3, nstart =20)
kmeans1.re
## K-means clustering with 3 clusters of sizes 7, 15, 7
##
## Cluster means:
##   X1_convert X2_convert
## 1   72.28571    74.0000
## 2  231.00000   264.3333
## 3  535.85714   551.0000
##
## Clustering vector:
##  [1] 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 2 2 2 2 1 1
##
## Within cluster sum of squares by cluster:
## [1] 34945.429 47893.333  8966.857
##  (between_SS / total_SS =  94.6 %)
##
## Available components:
##
## [1] "cluster"     "centers"     "totss"      "withinss"     "tot.withinss"
## [6] "betweenss"    "size"        "iter"        "ifault"
cbind(Y1_convert, Y2_convert) -> fem_df
kmeans2.re <- kmeans(fem_df, centers=3, nstart =20)
kmeans2.re
## K-means clustering with 3 clusters of sizes 4, 11, 14
##
## Cluster means:
##   Y1_convert Y2_convert
## 1    33.2500    54.0000
## 2   306.5455   440.2727
## 3   196.5000   254.7143
##
## Clustering vector:
##  [1] 1 1 1 1 3 2 2 2 2 3 3 2 2 2 2 2 2 2 3 2 3 3 3 3 3 3 3 3 3
##
## Within cluster sum of squares by cluster:
## [1] 20084.75 15128.91 46676.36
##  (between_SS / total_SS =  89.7 %)
##
## Available components:
##
## [1] "cluster"     "centers"     "totss"      "withinss"     "tot.withinss"
## [6] "betweenss"    "size"        "iter"        "ifault"
```

Cluster Plots

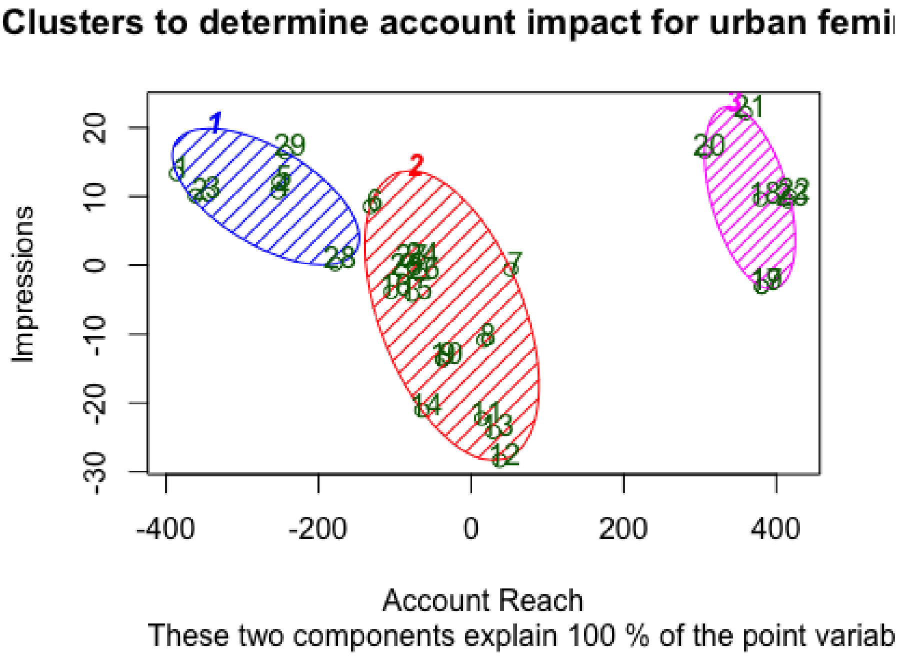
We create plots to understand this cluster distribution

Urban Feminist

plot(uf_df,col = kmeans1.re**$**cluster,main = "K-means with 3 clusters") **points**(kmeans1.re**$**centers, col = 1**:**3, pch = 8, cex = 3)

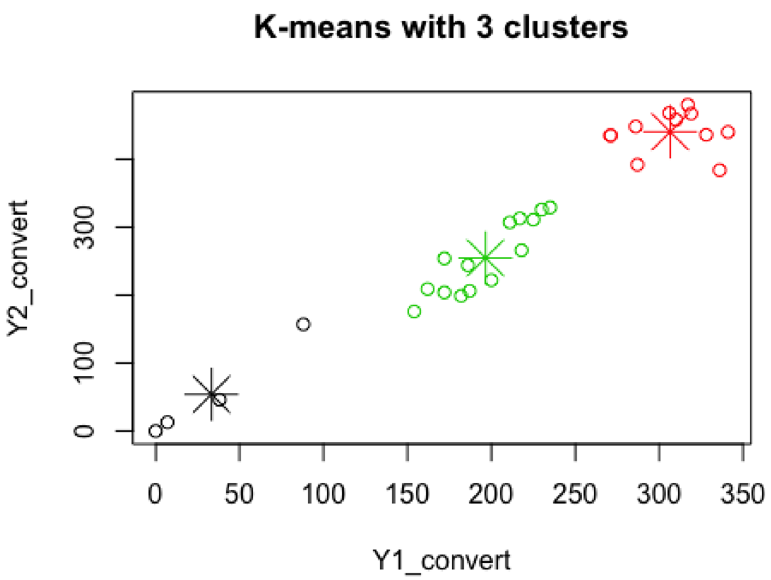


y_kmeans <- kmeans1.re**$**cluster
clusplot(uf_df, y_kmeans, lines = 0, shade = TRUE, color = TRUE, labels = 2, plotchar = FALSE, span = TRUE, main = **paste**("Clusters to determine account impact for urban feminist "), xlab = "Account Reach", ylab = "Impressions")

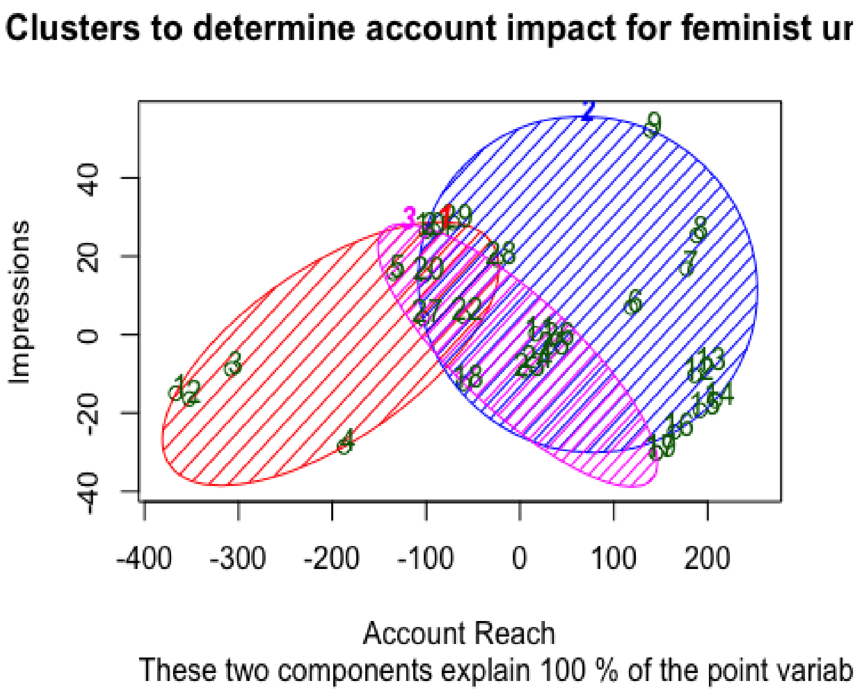


Feminist Union

plot(fem_df,col = kmeans2.re**$**cluster,main = "K-means with 3 clusters") **points**(kmeans2.re**$**centers, col = 1**:**3, pch = 8, cex = 3)




```
y1_kmeans <- kmeans1.re$cluster
clusplot(fem_df, y1_kmeans, lines = 0, shade = TRUE, color = TRUE, labels = 2, plotchar = FALSE, span = TRUE, main = paste("Clusters to determine account impact for feminist union"), xlab = "Account Reach", ylab = "Impressions")
```



Interpretation: K-Means Clustering

K-Means Clustering is an iterative algorithm that tries to partition the dataset into K-pre-defined distinct non-overlapping subgroups (clusters) (K-Means Clustering - an Overview ScienceDirect Topics, n.d.) where each data point belongs to only one group. It tries to make the inter-cluster data points as similar as possible while also keeping the clusters as different (far) as possible.

1. For the Urban Feminist account, we find clear demarcations between the clusters 7, 15 and 7 observations in cluster 1, 2 and 3 respectively. The difference between cluster means is also very high. The second cluster has the highest variability. The compactness percentage of the cluster is however

94.6% showing that we had exclusive high, medium and low reach and impression days (Fonseca, 2019). This clearly showcases that there is the propensity for a directed push by general algorithms to push the account for more viewership, engagement and attention

2. In the Feminist Union account, we find overlapping clusters with 4,11 and 14 observations respectively. Traditionally clusters created by K-means clustering are disjoint with zero commonality, but it is being increasingly accepted that (Dhillon et al, 2019) real world data sets are far from disjoint and overlaps are often present. This shows there was low propensity for directed algorithmic approach to increase reach and impressions and low, medium and high activity clusters represent a spectrum rather than extremities in activity.

Limitations of the Experiment

Pre-Analysis

1. Data collection period was short (29 days) and insights were limited to those computed by "Insight" and "Analytics" computational models operated by Instagram TM
2. Content creation and uploading followed designated timelines with mutually agreed time frames, however timeline discrepancy error may still exist
3. Content specifics were mutually decided and creation was alternated between the experimenters, however, individual creator bias may still exist
4. InstagramTM insight computation models provide weekly and not daily insights i.e., they map weekly trends in account reach, impressions, follower count, likes, shares, saves and comments and not daily accurate numbers, thereby creating weekly estimations of directed push, rather than daily estimations.
5. Model and conclusions do not account for user preference, subjective approach towards posted content.

Post-Analysis

1. T-Test assumes common scale, assumes population to be normally distributed. Welch T-Test removes (S. 2020) limitation of common variance requirement unlike Student's T-Test, nevertheless, it frames conclusions using normality assumption.
2. K- means clustering involves specification of cluster value at the outset, thereby creating inherent bias. Clustering is sensitive to initial conditions, different initial conditions such as shifting centroid, changing cluster constitution may affect outcome

CONCLUSION

Instagram has 1 billion active users every month (Statista, 2021b) which makes it a popular platform to advance social justice agendas. The platform's algorithms responses become vital to study in order to gauge the efficacy of the movements and awareness techniques.

Data bias has significant implications in the discipline of data science. For a more equitable representation of people, data needs to reflect those on the margins and not just those occupying the privilege. It becomes increasingly important for scholarly research to focus on equity since data used to formulate and implement policies hurts poor neighbourhoods disproportionately, especially policies related to the criminal justice system.

This paper explains how the data models we staunchly believe in reflect and in fact exacerbate inherent societal biases, instead of mitigating them, unlike what is commonly believed and expected out of purely engineered systems. Furthermore, this paper also attempts to conclusively determine the cause of such bias by analysing the sources they may emerge from.

This paper also aims to ground the arguments of reflection in an experimental model. Although this model is not able to effectively provide us with staunch conclusions as to whether these biases exist, it provides two relevant insights. First, there is an attempt to statistically determine significant differences between reception of Urban Feminist and Feminist Union Account via Welch's T Test which provides some propensity for significant difference with a potentially larger data set and second, there is scope to cluster data groups to understand how data patterns indicate targeted reach and impressions. This paper concludes that there is a plausible likelihood of algorithmic push for content that is more broadly accepted as mainstream.

REFERENCES

- Beck, K. (2021). *Self-Optimization in the Face of Patriarchy: How Mainstream Women's Media Facilitates White Feminism*. Shorenstein Center Discussion Papers, United States of America.
- *Book Review: Automating Inequality: How High-Tech Tools Profile, Police and Punish the Poor by Virginia Eubanks*. (2021, June 11). LSE Review of Books. https://blogs.lse.ac.uk/lsereviewofbooks/2018/07/02/book-review-automating-inequality-how-high-tech-tools-profile-police-and-punish-the-poor-by-virginia-eubanks/
- Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.
- Coaston, J. (2019, May 28). *Intersectionality, explained: meet Kimberlé Crenshaw, who coined the term*. Vox. https://www.vox.com/the-

highlight/2019/5/20/18542843/intersectionality-conservatism-law-race-gender-discrimination

- Crenshaw, K. (1991). Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color. *Stanford Law Review*, *43*(6), 1241. https://doi.org/10.2307/1229039
- Dalton, C. M., Taylor, L., & Thatcher (Alphabetical), J. (2016). Critical Data Studies: A dialog on data and space. *Big Data & Society*, *3*(1), 205395171664834. https://doi.org/10.1177/2053951716648346
- *Data Redundancy and Data Inconsistency Hurts Your Business*. (2021). BizData. https://www.bizdata.com.au/blogpost.php?p=costs-of-data-redundancy-and-data-inconsistency
- Dencik, L., Hintz, A., & Cable, J. (2016). Towards data justice? The ambiguity of anti-surveillance resistance in political activism. *Big Data & Society*, *3*(2), 205395171667967. https://doi.org/10.1177/2053951716679678
- Devika, J. (2020). The Defence of Aachaaram, Femininity, and Neo-Savarna Power in Kerala. *Indian Journal of Gender Studies*, *27*(3), 445–470. https://doi.org/10.1177/0971521520939283
- Dhanaraj, C. T. (2018, November 18). MeToo and savarna feminism: Revolutions cannot start. the privileged, feminist future must be equal for all. *Firstpost*. https://www.firstpost.com/india/metoo-and-savarna-feminism-revolutions-cannot-start-with-the-privileged-feminist-future-must-be-equal-for-all-5534711.html
- Fonseca, L. (2019, August 16). *Clustering Analysis in R using K-means - Towards Data Science*. Medium. https://towardsdatascience.com/clustering-analysis-in-r-using-k-means-73eca4fb7967
- Gibson, C. (2018, December 20). The end of leaning in: How Sheryl Sandberg's message of empowerment fully unraveled. *Washington Post*. https://www.washingtonpost.com/lifestyle/style/the-end-of-lean-in-how-sheryl-sandbergs-message-of-empowerment-fully-unraveled/2018/12/19/9561eb06-fe2e-11e8-862a-b6a6f3ce8199_story.html
- Heeks, R., & Renken, J. (2016). Data justice for development. *Information Development*, *34*(1), 90–102. https://doi.org/10.1177/0266666916678282
- *The Hidden Biases in Big Data*. (2018, January 16). Harvard Business Review. https://hbr.org/2013/04/the-hidden-biases-in-big-data
- *How We Can Solve Big Data's Bias Problem*. (2019). Knowledge@Wharton. https://knowledge.wharton.upenn.edu/article/big-data-ai-bias/
- Johnson, J. A. (2014). From open data to information justice. *Ethics and Information Technology*, *16*(4), 263–274. https://doi.org/10.1007/s10676-014-9351-8
- *K-Means Clustering - an overview | ScienceDirect Topics*. (n.d.). Science Direct. Retrieved June 30, 2021, from https://www.sciencedirect.com/topics/computer-science/k-means-clustering
- Krishnamurthy, P. (2021, May 6). *Understanding Data Bias - Towards Data Science*. Medium. https://towardsdatascience.com/survey-d4f168791e57
- Lamb, E. (2016, August 31). *Review: Weapons of Math Destruction*. Scientific American Blog Network. https://blogs.scientificamerican.com/roots-of-unity/review-weapons-of-math-destruction/
- MacIntyre, T. M. (2013). *White Feminist Fragility: from Part of the Problem to Radical Allyship* (Thesis). https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0390302
- Marr, B. (2018, December 12). The Amazing Ways Instagram Uses Big Data And Artificial Intelligence. *Forbes*. https://www.forbes.com/sites/bernardmarr/2018/03/16/the-amazing-ways-instagram-uses-big-data-and-artificial-intelligence/
- Moon, D. G., & Holling, M. A. (2020). "White supremacy in heels": (white) feminism, white supremacy, and discursive violence. *Communication and Critical/Cultural Studies*, *17*(2), 253–260. https://doi.org/10.1080/14791420.2020.1770819
- *Moving beyond âalgorithmic bias is a data problemâ*. (2021, April 9). ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2666389921000611
- O'Neil, C. (2017). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* (Reprint ed.). Crown.
- S. (2020, December 17). *Welch's Test for Unequal Variances*. Statistics How To. https://www.statisticshowto.com/welchs-test-for-unequal-variances/
- S. Barocas, M. Hardt, A. Narayanan (2019**)** *Fairness and Machine Learning.* https://www.fairmlbook.org
- Sandberg, S. (2013). *Lean In: Women, Work, and the Will to Lead* (1st ed.). Knopf.
- Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J., & Sculley, D. (2017). No classification without representation: Assessing geodiversity issues in open data sets for the developing world. *arXiv preprint arXiv:1711.08536*.
- Singh, S. (2018, October 9). *Understanding the Bias-Variance Tradeoff - Towards Data Science*. Medium. https://towardsdatascience.com/understanding-the-bias-variance-tradeoff-165e6942b229
- Statista. (2021a, April 7). *Worldwide digital population as of January 2021*.

https://www.statista.com/statistics/617136/digital-population-worldwide/

- Statista. (2021b, June 29). *Instagram: distribution of global audiences 2021, by age group*. https://www.statista.com/statistics/325587/instagram-global-age-group/#:%7E:text=With%20roughly%20one%20billion%20monthly,million%20million%20Instagram%20users%20each
- Taylor, L. (2017). What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society*, *4*(2), 205395171773633. https://doi.org/10.1177/2053951717736335
- Taylor, L., & Broeders, D. (2015). In the name of Development: Power, profit and the datafication of the global South. *Geoforum*, *64*, 229–237. https://doi.org/10.1016/j.geoforum.2015.07.002
- Turow, J., & Hennessy, M. (2015). The Tradeoff Fallacy: How Marketers are Misrepresenting American Consumers and Opening Them Up to Exploitation. *SSRN Electronic Journal*. Published. https://doi.org/10.2139/ssrn.2820060
- UN General Assembly. (1948). *Universal Declaration of Human Rights*. United Nations.
- Wachter-Boettcher, S. (2018). *Technically Wrong: Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech* (1st ed.). W. W. Norton & Company.
- Whang, J. J., Hou, Y., Gleich, D. F., & Dhillon, I. S. (2019). Non-Exhaustive, Overlapping Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(11), 2644–2659. https://doi.org/10.1109/tpami.2018.2863278.