How Sociotechnical Factors Create Gender Biases within Google's Search Engine Algorithms

A Research Paper submitted to the Department of Engineering and Society

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

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Kiana Wilkins

Spring 2023

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

Advisor

Kent Wayland, Department of Engineering and Society

Introduction

When looking at search engine results, what factors are considered when determining one link is “better” than the other? Today's search engines are a vital source of information that has a significant influence in shaping our understanding and beliefs around the world. Every internet search uses an algorithm to decide the best results based on the user's query. Search engine algorithms are believed to be neutral, but many sociotechnical factors affect these algorithms along with their outcomes. Algorithms are not flawless tools since they are created, maintained, and influenced by people who already have their own sets of values and innate biases. The common sentiment that search engine algorithms are impartial, can potentially negatively alter one’s beliefs since search engines are heavily relied on as a main source of information. Throughout history, marginalized groups have been negatively affected by discrimination in search engines due to its perpetuation of false stereotypes and lack of transparency.

My research will primarily focus on analyzing Google’s utilization of search algorithms since Google owns 91.42% market share of the global search engine market (Chris 2022). This paper aims to discuss *how sociotechnical factors create gender biases within Google’s search engine algorithms.* To address this research question, I gathered material on how Google’s search engine algorithms work to provide a holistic view of what contributes to its functionality. I also referenced scholarly literature that analyzes gender biases within search engines along with what social and economic factors influence these algorithmic biases. To propose a potential solution to Google’s search engine bias I propose solutions based on the specific contributing factors presented in the results.

Background

To comprehend the significant influence of gender bias in daily life through search engines, it is essential to know the history of search engines and society's deep history with gender inequality, which persists today due to a multitude of factors. In America and globally women are discriminated against because of cultural and societal norms that stem throughout history. The GSNI, or Gender Social Norms Index, is a tool used to evaluate the societal beliefs that hinder gender equality in various domains such as politics, education, and work. It encompasses information from 75 nations, which includes more than 80 percent of the global population. According to the GSNI, almost 90% of both men and women hold some sort of bias against females (United Nations 2020). The survey is conducted using a set of 22 questions that are designed to measure five key dimensions of gender norms: political, economic, educational, intimate partner violence, and discriminatory norms. The survey is administered to a nationally representative sample of individuals in each country. The sample is selected using a multistage sampling technique, and the survey is conducted using face-to-face interviews or computer-assisted telephone interviews. This shows that the world still deals with issues of gender inequality because it is so deeply ingrained in humans today. These inequalities are still reflected in the workplace and technology.

There is a significant gender gap within the workforce especially in the tech industry. Women make up 47% of all employed adults in the US and yet they hold only 28% of computing and mathematical roles. According to a survey, nearly four out of ten women working in the technology industry believe that gender discrimination poses a hindrance to their career advancement (McCain 2022). Google along with other big tech companies is not immune to this issue of gender socialization. Google has issues of diversity, which directly affect the products

they create. Companies like Microsoft and Google consistently state that around 30% of their employees are women, but their representation in leadership or technical roles is even lower. Similarly, startups also have disappointing figures when it comes to gender diversity in their workforce. This is due to a multitude of factors a couple being early socialization of women's roles in society and discrimination during hiring processes due to implicit gender biases (Khan 2023). Lack of representation can lead to the exclusion of marginalized groups in technology, such as algorithms used by search engines.

Search engines were invented to help people find information on the internet quickly and efficiently. Before the advent of search engines, locating specific information on the internet was a time-consuming and tedious process that required sifting through vast amounts of data. Search engines revolutionized this process by using complex algorithms to crawl and index web pages, allowing users to search for and retrieve relevant information with just a few clicks. Information is now so easily accessible through search engines that it's important to be mindful of how this technology operates.

Search algorithms take into account various elements when evaluating results that may not always be known to the regular user. Google claims that in order to provide the most beneficial information, their search algorithms take into consideration a variety of factors and indicators, such as the terms used in your search query, the relevance and user-friendliness of web pages, the credibility of sources, and your location and preferences. The importance assigned to each factor changes depending on the type of query being conducted ("Ranking Results," n.d.). Despite this statement of showing some factors that affected their algorithms, there have been numerous instances where Google's search engine has been critiqued for its lack of transparency and its biased search results.

In recent years this issue of gender bias in search algorithms has become increasingly prevalent. A well-known example of algorithmic bias in Google's search engine was brought to the public's attention by communications scholar, Safiya U. Noble. To find websites about black role models for her young stepdaughter and nephews, Safiya U. Noble typed "black girls" into Google, expecting to be taken to sites with historical context, pop culture for kids, or instructional content. Instead, HotBlackPussy.com was the first result to pop up in their search. On the first results page, there were multiple other identical sites. Black girls were not characterized by their history, interests, or aspirations, according to Google, they met the definition of porn (Noble 2012). Many people may conclude that the public's search habits are the cause of this failure but experiments with Google search and scholarly literature related to search engines oppose this view claiming that developers can practice ways to prevent biased outcomes.

Although the large datasets and search habits of users are a major contributor to these gender-biased results programmers still have the power to decide what to do with these datasets. On a more detailed level, it is people who create, gather, and classify the information that is included in datasets. Additionally, these people are responsible for choosing which datasets, variables, and regulations the algorithms use to acquire knowledge and make decisions (Smith 2022). It's the programmer's role to put an effort to not continue to perpetuate these unfair biases.

These biased Google search results also influence important processes such as hiring decisions. It's found that Google's image results tend to have biases, such as predominantly featuring men when searching for terms like "engineer" and predominantly featuring women when searching for "cleaner." A recent study done by Vlasceanu, & Amodio, which will be

discussed later on in this paper, emphasizes how damaging these biased search results can be. Even when searching for gender-neutral terms, the results can still skew towards men, which can perpetuate gender biases that then, in turn, affect job hiring for women.

After being equipped with a brief overview of how search engines and algorithms work, the state of gender inequality today, and Google Search's history with gender bias, it should provide a solid foundation for my research of how sociotechnical factors create gender biases within Google's search engine algorithms.

Methods

The overall design of my research on gender bias in Google's algorithms is primarily diagnostic, where I analyze the underlying causes of Google's biased search results and how to address them. I gathered evidence with the goal of providing significance to the reader as well as a solid understanding of how Google's algorithms work. Most of my gathered research and literature are found through research databases that provide access to a wide range of academic journals, magazines, newspapers, and other sources. When gathering evidence, I included search terms such as "Google", "search engine", "algorithm", "gender bias", etc. The majority of my sources are secondary sources. I analyze case studies on discrepancies regarding gender in Google's search results and analyze what economic and social factors contribute to these algorithmic outcomes. With knowledge of how Google's search engine operates, the social and economic factors that influence its results, along with a thorough analysis of its algorithmic gender bias, gives the audience a holistic view of the sociotechnical problem. Lastly, I discuss previous research on how algorithmic bias is mitigated and will use it as a framework for a possible solution to addressing gender bias in Google's search results. A limitation of this

research is the fact that direct access to the algorithms that Google uses is not accessible to the public. A lot of the evidence analyzes the results more than the actual code itself.

The Influence of Societal Gender Bias on Search Results

There have been prior studies that looked into gender disparities within Google's search engine algorithms. Google's search engine algorithms are programmed to identify patterns in massive datasets, but the issue is that these datasets often reflect pre-existing biases in society. To evaluate the existing gender inequality in society, researchers analyzed data from the Global Gender Gap Index (GGGI), which measures the extent of gender disparities in 153 countries in areas such as economic participation, educational achievement, and other indicators. The researchers then conducted a search for the gender-neutral term "person" in Google Images in 37 countries, using the primary language of each country. Three months later, they repeated the experiment in 52 countries, including 31 from the previous search. The results showed that countries with higher levels of gender inequality, as reported by the GGGI, produced more male-dominated Google Image search results, indicating a connection between societal gender disparities and algorithmic outcomes, the authors noted (Vlasceanu, & Amodio 2022). These results highlight that Google's search engine algorithms mirror and propagate the existing levels of inequality in society.

After having shown the correlation between algorithmic search output and gender inequality in society, the study also investigated if these search outputs can influence people's thinking and choices in ways that align with the already existing societal inequality. In this study, participants were randomly put into two groups. The first group received search results that were biased towards traditional gender roles (e.g., women in caregiving roles, and men in leadership

roles), while the second group received gender-neutral search results. Each group was given image search results of four unfamiliar professions: chandler, draper, peruker, and lapidary. The participants' initial attitudes toward gender roles and gender equality were measured using a survey. Then the participants' behavioral responses to the biased search results were measured using a decision-making task. Participants were presented with hypothetical scenarios that involved gender-related decisions (e.g., hiring a job candidate, or choosing a student for a scholarship). The participants were then asked to make decisions based on the information provided. The results of the study showed that exposure to biased search results increased participants' endorsement of traditional gender roles and reduced their support for gender equality. In the decision-making task, participants in the biased search group were more likely to make choices that were consistent with traditional gender roles (Vlasceanu, & Amodio 2022). These findings suggest that Google's algorithmic search results can contribute to the spread of gender inequality in society by shaping people's beliefs and decision-making processes.

Influence of PageRank with Google Search

Since we are unable to directly access the search algorithm Google uses, it's difficult to pinpoint direct evidence of gender bias coded into the algorithm, but we do have knowledge of how their algorithms work in general and one of the algorithms that are utilized. Google's search engine primarily uses an algorithm called PageRank to determine the ranking of web pages. In general, the PageRank algorithm works by assigning a numerical score to each web page based on the number and quality of other web pages that link to it. The score is calculated recursively, with each page's score being influenced by the scores of the pages that link to it. The algorithm assumes that a web page is more important if it is linked to other important pages, and less

important if it is only linked to less important pages (Jain 2021). A study published in Nature in February 2022 looked at how the PageRank algorithm can introduce or amplify biases. Researchers created five models of societies with different levels of similarity between people, and each model had 20 websites assigned a rank as part of the majority or minority. The study found that the algorithm can either reduce, replicate or increase biases depending on the model. If people in a group are similar, the algorithm tends to favor the dominant views, while if people are diverse, minorities are represented more. (Espín-Noboa 2022). This means that pages that are already popular and have many incoming links are more likely to be ranked higher, while pages with fewer incoming links are less likely to be ranked higher. This creates a positive feedback loop where already popular pages continue to be ranked higher, while less popular pages remain lower in the ranking, even if they have relevant and valuable content. Biases in PageRank can also be amplified when used in recommendation systems, where the algorithm suggests content to users based on their past behavior or preferences. If the algorithm has already ranked certain pages or resources higher due to their popularity, it may continue to recommend them to users, even if there are other relevant resources that are less popular. This showcases that the use of network-based ranking and recommendation algorithms like PageRank used by Google can perpetuate and amplify biases in the network structure.

The Impact of Search Engine Optimization on Search Results

Companies and professionals who are aware of how PageRank works try to use this knowledge to their advantage through a process called Search Engine Optimization (SEO). The purpose of SEO is to make advertisements or websites appear at the top of the results page for a specific search query (Abdallah 2017). There is a study that reveals that a major portion of pages

appearing in Google search results is likely to have been optimized. This was discovered through the examination of three sets of data containing a combined 1,914 queries and 256,853 results. (Lewandowski 2021). This finding suggests that utilizing SEO techniques is essential for achieving visibility in search engines. Therefore, SEO professionals have the power to influence companies ranking on Google's search results pages. What is concerning is that usually, the main goal for an SEO professional is to figure out how to make their content appeal to the search algorithm, rather than questioning whether the algorithms are fair or whether their work might be reinforcing harmful biases.

There are multiple ways these SEO practices influence Google's search engine gender bias. SEO practices often involve using specific keywords to improve a website's ranking on Google. However, these keywords often reflect gender biases in language. For example, SEO professionals may use gendered language to target specific audiences such as "sports cars for men" or "home decor for women" which reinforces the idea that certain products and interests are gendered. When these keywords are used more frequently on websites, it can reinforce gender stereotypes and biases in search results. This gendered language is a tactic used by SEOs to successfully gain higher rankings on Google's search engine results.

The SEO industry is also very male-dominated and lacks diversity, as does many other fields in the tech industry. According to a study done in 2022 based on 636 SEOs, 69% of published SEO authors are male (Infante 2022), which is a huge gender gap. This lack of diversity can contribute to gender biases within the SEO industry as well as in SEO practices, such as choosing keywords or phrases that reflect male-dominated industries or interests. As a result, search results may continue to favor male-centric content. This implies that a lack of diversity in SEO can reinforce gender biases in search results.

Gender Bias in Advertisements and User Data

Over the years Google has manipulated search results by making ads look identical to regular search results. This makes it harder for users to distinguish whether a result is manipulated or not. Many users have complained about ads taking up the majority of their results page depending on what they search. There is a lack of transparency in Google's search engine in terms of what economic factors influence users' search results. Evidence implies that most resources for search engine development are allocated in accordance with market reasons or technological/scientific concerns. In the minds of search engine developers, fairness and representativeness are not the primary determinants of search engine quality (Van Couvering 2007). When advertisers use Google Ads to target specific demographics, such as gender, they may inadvertently perpetuate gender stereotypes and biases by relying on assumptions or generalizations about certain groups. Without transparency, it is difficult for users to know how advertisers are targeting them or what biases may be present in the ads they see.

Many are aware that Google collects information on us based on what we do on the Web to target users with certain ads and web search results. What is not clear is exactly how Google utilizes its users' information. To look into how user information such as gender may affect Google Ad results, a group of scientists at Carnegie Mellon University constructed a computerized testing system named AdFisher, that created multiple male and female job applicants. The tool created 17,370 fabricated profiles that solely accessed job seeker websites and viewed approximately 600,000 advertisements, which the team monitored and analyzed. From this study, it was found that advertisements promoting coaching services for high-paying jobs were displayed to males more frequently than to females (Datta 2015). AdFisher also revealed that Google's "ads settings," a tool that enables users to review and modify the

"interests" that Google has identified for them, do not consistently reflect sensitive information that may be used to target them. For instance, when browsing websites designed for individuals with substance abuse problems, the system displayed numerous ads for rehabilitation programs, yet Google's transparency page remained unaffected. This finding shows that the ad settings tool Google provides gives a false sense of security in terms of allowing the user to believe they have control over what Google does with their activity on the web.

Overall, the lack of transparency with Google Ads can contribute to gender bias by perpetuating stereotypes and limiting the representation of diverse perspectives in advertising. It is important to promote transparency and accountability in advertising practices to address gender bias and promote more inclusive and equitable representation in advertising.

Discussion

The aim of this research was to investigate how sociotechnical factors contribute to the creation of gender biases within Google's search engine algorithms. By analyzing the underlying mechanisms of the search algorithms, we explored how gender biases are perpetuated, and the potential implications of these biases. The results revealed that sociotechnical factors play a significant role in the creation of gender biases within Google's search engine algorithms. Specifically, the algorithms learn from the patterns of data generated by human search queries and user behavior. However, since these patterns are influenced by existing societal stereotypes and biases, the algorithms perpetuate and reinforce gender biases.

Our findings suggest that gender biases in Google's search engine algorithms are not inherent, but rather, are the result of various factors. Many claim that SEO professionals and users are the reason for Google’s search results being gender biased but from the findings above

it’s evident that Google’s algorithms have just as much, if not more power than the individuals who are utilizing the tools they provide to impact a website’s ranking. If Google wanted to change or lessen the amount of gender-biased results there are they could. Due to the underrepresentation of individuals with an in-depth understanding of the history of gender bias and inequalities in the big tech industry, there is a tendency to develop products without sufficient scrutiny of their potential effect on underrepresented individuals. If software engineers at Google are not held accountable for the design of their algorithms, then who is responsible? Part of the problem does rely on the algorithm in terms of its lack of initiative to lessen these biases and just settle with the biased data it relies on.

Since this issue of gender bias in Google’s search results is so multifaceted there are several steps, Google could take in order to mitigate these biases. Firstly, it could improve its algorithms by identifying and focusing on eliminating these gender biases in the datasets it uses. As seen earlier in the paper Google’s search results reflect the existing levels of gender disparities in society, so if Google’s search algorithm relies on society's biased data its results will also be biased. There is previous research on ways algorithmic bias could be mitigated. Orphanou et al. (2022) argue that bias can enter the algorithmic process at various stages, such as data collection, algorithm design, or testing, and that addressing bias requires a multi-faceted approach. Based on the results above the PageRank algorithm used by Google can introduce or amplify biases. This highlights the potential for the algorithm to spread societal biases as well as lessen them. From these findings, it would be beneficial if Google altered its algorithm and data to account for unfair gender biases when producing results. Beside just diversifying the data sets it's important to constantly monitor the algorithm itself as well as make sure all the stakeholders involved are also diverse since they are a part of many stages when designing the algorithm.

These findings suggest that Google's search engine algorithms can perpetuate and amplify existing gender and societal biases. This demonstrates the need for continued research into the ethical and social implications of its search engine algorithms, as well as efforts to mitigate these biases.

There is also a need for diversity in the SEO industry that influences Google's search engine results. It is important for SEO professionals to try and avoid gendered language and imagery so that this constant perpetuation of gender biases can lessen. Queries that should be gender-neutral should not have such skewed results of gender.

Lastly, another step to mitigating the spread and reinforcement of gender bias from Google's search results is to increase transparency in how Google's search algorithm uses user information for both search results and ads. This can in turn help reduce biases in advertising and promote more equitable representation. If people are more aware of how their data is used, they can potentially have more control over how they choose to utilize this tool. Google could share data on search queries, user behavior, and other relevant factors to identify patterns that could indicate gender bias. Once these patterns are identified, Google can work to adjust its algorithm to remove any unintended gender bias. Transparency can also help build trust and confidence in Google's search engine among users. By being open and clear about its processes, Google can demonstrate its commitment to fairness and impartiality, which can help address any concerns about gender bias.

Conclusion

There is clear evidence of gender bias in Google's search engine algorithms from various socio-technical factors. Google's search engine algorithms reflect pre-existing biases in society,

which stem from massive datasets they use that are influenced by societal beliefs around gender. Additionally, biased algorithmic search results could increase people's endorsement of traditional gender roles and reduce their support for gender equality. The PageRank algorithm used by Google can introduce or amplify biases depending on the model. Companies can also manipulate the system to climb the rankings of top search results using SEO practices. All these findings support the claim that Google's search algorithm can propagate gender inequality in society by shaping people's beliefs and decision-making processes, including when it comes to hiring decisions. Based on the research above, Google does have the capability of improving its search algorithms to reduce rather than spread or enhance gender stereotypes in the world. To address this issue, Google should increase transparency and accountability by providing more information about its algorithms' functioning and bias mitigation efforts.

Works Cited

Abdallah, F., & Anis, I. (2017). A Survey on Search Engine Optimization (SEO). *International Journal of Computing, Communication and Instrumentation Engineering*, *4*(2). https://doi.org/10.15242/IJCCIE.AE0417136

Berlatsky, N. (2018, February 21). *Opinion | Noah Berlatsky: How search algorithms reinforce racism and sexism*. NBC News. https://www.nbcnews.com/think/opinion/google-search-algorithms-are-not-impartial-they-are-biased-just-ncna849886

Bradford, K. (2022, August 7). Google Shows Men Ads for Better Jobs. The Good Search. https://tgsus.com/diversity/google-shows-men-ads-for-better-jobs/

Chris, A. (2023, March 28). *Top 10 Search Engines In The World (2023 Update)*. Reliablesoft.Net. https://www.reliablesoft.net/top-10-search-engines-in-the-world/

Datta, A. (2015). *AdFisher: Detecting Discrimination and Opacity in Online Behavioral Advertising*. https://fairlyaccountable.org/adfisher/

Espín-Noboa, L., Wagner, C., Strohmaier, M., & Karimi, F. (2022). Inequality and inequity in network-based ranking and recommendation algorithms. Scientific Reports, 12(1), Article 1. https://doi.org/10.1038/s41598-022-05434-1

Infante, L. (2022, January 15). Gender Gap in SEO Publishing: 2022 Study. Retrieved from https://www.lidia-infante.com/post/gender-gap-in-seo-publishing-2022-study

Jain, U., Mishra, A., Jaganathan, B., & Shukla, P. (2021). Study and analysis of category based PageRank method. Wireless Networks, 27, 1–16. https://doi.org/10.1007/s11276-021-02617-y

Khan, C. (2023, February 24). Why Are There So Few Women in Tech? https://careerfoundry.com/en/blog/career-change/women-in-tech/

Lewandowski, D., Sünkler, S., & Yagci, N. (2021). The influence of search engine optimization on Google's results: A multi-dimensional approach for detecting SEO. *13th ACM Web Science Conference 2021*, 12–20. https://doi.org/10.1145/3447535.3462479

McCain, A. (2022, October 31). 40 Telling Women In Technology Statistics [2023]: Computer Science Gender Ratio. *Zippia*. https://www.zippia.com/advice/women-in-technology-statistics/

Nations, U. (2020). 2020 Gender Social Norms Index (GSNI). In *Human Development Reports*. United Nations. https://hdr.undp.org/content/2020-gender-social-norms-index-gsni

Noble, S. U. (2012). *Missed Connections: What Search Engines Say About Women*. Bitch Magazine, 6.

Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York University Press.

Orphanou, K., Otterbacher, J., Kleanthous, S., Batsuren, K., Giunchiglia, F., Bogina, V., Tal, A. S., Hartman, A., & Kuflik, T. (2022). Mitigating Bias in Algorithmic Systems—A Fish-eye View. ACM Computing Surveys, 55(5), 87:1-87:37. https://doi.org/10.1145/3527152

Ranking Results – How Google Search Works. (n.d.). Google Search - Discover How Google Search Works. Retrieved May 12, 2023, from https://www.google.com/search/howsearchworks/how-search-works/ranking-results/

Smith, G., & Rustagi, I. (2022, July 9). *Mitigating bias in Artificial Intelligence*. Berkeley Haas. Retrieved May 1, 2023, from https://haas.berkeley.edu/equity/industry/playbooks/mitigating-bias-in-ai/

Van Couvering, E. (2007). Is Relevance Relevant? Market, Science, and War: Discourses of Search Engine Quality. Journal of Computer-Mediated Communication, 12(3), 866–887.

Vlasceanu, M., & Amodio, D. M. (2022). Propagation of societal gender inequality by internet search algorithms. Proceedings of the National Academy of Sciences, 119(29), e2204529119. https://doi.org/10.1073/pnas.2204529119