

Thesis Project Portfolio

Scalable Ad-Targeting Technology using AWS
(Technical Report)

Examining the Fairness and Bias of ChatGPT
(STS Research Paper)

An Undergraduate Thesis

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Nitin Maddi
Spring, 2023
Department of Computer Science

Table of Contents

Sociotechnical Synthesis

Scalable Ad-Targeting Technology using AWS

Examining the Fairness and Bias of ChatGPT

Prospectus

Sociotechnical Synthesis

Targeted advertisements are irreplaceable tools for companies who wish to maximize their profits and receive the biggest return on their investment for advertising. However, in the process of creating these advertisements, companies sometimes end up using data that creates bias and aggravates already present issues. For example, a meta-analysis study in 2017 shows that people who come from low-income communities are more likely to smoke cigarettes (Casetta, 2017). If a cigarette company were to use this income data to send targeted advertisements, the issue of low-income smokers would be exacerbated and could cause more harm to this community. That is why it is essential for companies to use bias prevention tactics when creating targeted advertisements, such as keeping certain bias-forming data, such as data regarding gender and income, away from the advertisements. Like targeted advertisements, machine learning models heavily rely on previously collected data to analyze trends so that it can classify and predict unseen data by utilizing the previously found trends. However, the same bias that is present in the training data ends up getting propagated throughout the AI. A recently developed model, known as ChatGPT, has been rapidly gaining popularity after users have witnessed its proficiency in answering specific questions and its vast amount of knowledge and capabilities. Just as companies must be careful while selecting data for targeted advertising, though, OpenAI must be extremely cognizant of the data that is ingested by ChatGPT. If this issue isn't considered, it can lead to ChatGPT providing answers that are harmful for particular social groups. Overall, my two projects show that data is an extremely powerful resource, but when it is not properly managed, there can be serious negative effects on different social groups.

Currently, companies collect massive amounts of data on their users, but have no effective way to digest and utilize this information without investing millions of dollars into

infrastructure. Using AWS cloud technology, I developed an efficient ad-targeting solution that costs a fraction of the typical price. I leveraged AWS Simple Queue Service (SQS) to simplify and orchestrate the data analysis workflow by breaking down the problem into smaller parts. With the help of Java, these smaller processes are examined to match people to an advertisement based on information collected on them. Finally, these advertisements are pushed to an advertisement platform, such as Google or Facebook, and a person's account on these platforms will receive the advertisement. The finished product allowed companies to deliver targeted advertisements at a fraction of the cost, with a quick turnaround period, and no max cap on the advertisement quantity.

My STS research paper examines the fairness and bias of ChatGPT, a generative chat engine with widespread capabilities. My main method of analysis was discourse analysis to understand the techniques utilized by OpenAI and rate the chat engine on its fairness. Additionally, I used Actor Network Theory to model the actors who make up this system. This research demonstrated that ChatGPT, through its various data streams for training, ends up accumulating human bias as well as sampling bias due to difficulties in collecting varied and useful data. In turn, ChatGPT negatively affects specific communities due to inequality in accessibility due to ChatGPT's incapability to produce similar quality results for different languages.

Working on both these projects in conjunction greatly influenced the final project that I delivered in both cases. While working on the technical project, my research showed the importance of handling data carefully. I designed architecture to ensure that no engineer would have or need direct access to any of the data that was being used for targeted advertisements. Additionally, I tried implementing automatic data filtration procedures that would remove certain

columns of data which may be closely related to personal information. In my STS research project, I understood what types of architectural design patterns to look for. I knew when a design choice would have flaws and what these flaws would be. Looking for these design choices in ChatGPT helped me pinpoint where bias would arise in the system.

References

Casetta B, Videla AJ, Bardach A, Morello P, Soto N, Lee K, Camacho PA, Hermoza Moquillaza RV, Ciapponi A. (2017). Association Between Cigarette Smoking Prevalence and Income Level: A Systematic Review and Meta-Analysis. *Nicotine Tob Res.* 2017 Nov 7;19(12):1401-1407. doi: 10.1093/ntr/ntw266. PMID: 27679607.