Freelance Web Design: Building a Custom Website for a Budding Local Business

The Illusion of Choice in the Internet Age

A Thesis Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Brian Tufts

Spring 2022

ADVISORS

Kent Wayland, Department of Engineering and Society

Daniel Graham, Department of Computer Science

The Mutual Shaping of the Targeted Advertising Industry

General Research Problem:

How are AI, the targeted advertising industry and society being mutually shaped?

Targeted advertising has become an increasingly large portion of the global economy over the last decade and plays an integral part in the fast growing $200 billion e-commerce economy (U.S. Department of Commerce, 2021). This has all been with little to no regulation or analysis of the effectiveness of the industry. The industry deals with an old concept known as human data, which is any data relating to a particular individual, in a relatively new way. The general understanding of targeted advertising is that it uses artificial intelligence and human data to target advertisements at particular people who are likely to click on the advertisement or otherwise do what the advertisement wants. The growth of targeted advertising has led to many discussions about the ethics of human data collection and whether the industry needs to be regulated. Not from the reference of the data collection aspect, but rather the way in which targeted advertising is being used to manipulate the habits of the public. However, for this to be regulated, we must first research whether it is in fact manipulating people and removing their choice without them being conscious of it.

Usage of AI in Social Sorting for the Purpose of Targeting Advertisements

How can AI be designed to limit discrimination based on uncontrollable factors?

With the increase in the popularity of targeted ads there is naturally an increase in the development of the technology that supports it: artificial intelligence. There is a lot of research done on the ethics of artificial intelligence in many fields including the field of targeted

advertising. This has led to companies such as Facebook no longer allowing advertisers to target advertisements based on what they claim to be sensitive topics including race, political affiliation, and gender (Bond, 2021). However, this may not be enough to completely remove the unethical discrimination that exists within the artificial intelligence algorithms. The algorithms are simply fed data and are able to make connections based on that data and decide whether or not a particular person is likely to interact with an advertisement. To determine if this process is discriminatory, we must take a look at the actual data for places where discrimination can take place. One study performed an analysis with Facebook and found that there were carefully curated attributes that closely correlated with "sensitive" topics presenting an avenue of discrimination for advertisements (Speicher et al., 2018).

My first instinct would be to break down every attribute in the data and determine if there is a possibility for discrimination using this attribute. However, this process would be quite tedious, get very complicated over time and ultimately reduce the effectiveness of the targeting. Instead, a different approach might be introducing randomness into the artificial intelligence algorithms to diversify the groups they select to be targeted by certain advertisements. This is a relatively simple idea that could have profound implications if applied in real world advertisement targeting. The goal is to force an algorithm to occasionally go against what it would otherwise decide based on some random variable. For example, say an algorithm decides that a person is not likely to interact with an advertisement, the random variable will then have a chance to reverse that decision with some probability. Specifically, I will use machine learning software such as TensorFlow or Scikit Learn to simulate a targeted advertising algorithm. There will be some metrics created to quantify the success of said algorithm as well as the bias it portrays based on uncontrollable or sensitive traits. Success in this case will be defined by how

accurate the groupings created by the algorithm are. In other words, how much more likely someone in this grouping is to click on a specific ad for that group than a random person shown the same ad. Next, I will introduce the random variable to it and again quantify its success and bias. The random variable can then be manipulated to find the optimal point that maximizes the effectiveness while minimizing the bias. If the effectiveness and bias do not have a linear negative correlation with each other then there should exist some random variable that will reduce bias without impacting the effectiveness of the algorithms. However, there are some constraints to this experiment, namely the limited amount of data that I would have access to compared to large targeted advertising companies such as Google or Facebook, yet acquiring data from public voter registries should provide a good baseline for the validity of the technique.

The Illusion of Choice in the Internet Age

Does Targeted Advertising Create the Illusion of Personal Choice?

There has been substantial research surrounding the field of targeted advertising, most of which is within the context of how privacy invasion is performed to make the field possible. There is a clear reasoning why the privacy of an individual could be perceived as invaded when their personal data is collected for the purposes of targeting advertisements. This data is collected in almost every aspect of the average American's life in the internet age. Okonkwo introduces the concept of behavioral surplus, which is any data that has been collected on a user that is not directly utilized by the application. This behavioral surplus was the stepping stone to what is now known as surveillance capitalism, the umbrella term for products such as targeted advertising, alluding to the fact that the collection of human data on the internet began well before the inception of targeted advertising (Okonkwo, 2021). Okonkwo goes on to state that Google was

the first to stumble upon this idea of "behavioral surplus," and in their search for a way to profit from this discovery, targeted advertising was born. Since then, just about every company has found a way to turn their product(s) into data harvesting tools to make a profit. This entire process is well researched; however, the broader implications of targeted advertising have been accepted to be something that "beyond merely strip-mining our intimate inner lives, it seeks to shape, direct and control them," which is not clearly understood by most Americans (Pal and Crowcroft, 2019). Since there is a lack of research that studies the control that targeted advertising has over society and its implications, there is a need for this gap to be filled when arguing for increased regulation of human data collection. There is arguably a precedent set by other corporate laws in America that regulation isn't done simply for regulations sake, but rather to prevent action encouraged by capitalism that may negatively impact society. Therefore, it is important that we answer the question of whether targeted advertising does in fact negatively impact society through controlling habits and creating a so-called illusion of choice.

The first step to answering this research question is simply showing that personal data collection of an individual by companies is near inescapable in the modern era. This can be done by exploring everyday technologies that people cannot reasonably be expected to function without and how their personal data is collected by said technologies. One of these technologies is cookies, and just about every website has a popup requesting your consent to use them. Cookies are used by almost 40 percent of websites today, but have been on a slow decline since 2021 (W3techs, 2022). Many websites do not require the use of cookies, but offer a much better user experience by allowing websites to save user preferences, logins, and details such as a shopping cart on web stores. Another technology that is ingrained in our society is mobile phones. Mobile phones "enable the gathering of a wealth of data on location, motion,

communications content, in-application activities, and other traces of mobile use that can be pieced to gather, understand and influence users' behavior" that was not available previously (Martin and Shilton, 2016). Mobile phones allow us to stay connected with people anywhere from simply texting and calling to social media sites. One would find it very difficult to function in modern society without a mobile device with at least the ability to make phone calls. Martin and Shilton go on to describe how mobile devices also offer a platform for harvesting data to a number of new actors. The actual personal data that is collected from an individual's mobile devices is very dependent on the individual, however, the rich opportunity for data collection is obvious and someone cannot reasonably expect to operate in modern society without a mobile device. There are a number of other technologies that could be explored in the context of an individual's inability to avoid their personal data being collected and utilized for targeted advertising that will be important to explore when attempting to answer the research question. The sources for this will primarily be usage statistics of different technologies, but could also consist of primary sources explaining why they cannot reasonably expect to function without a particular technology.

As a result of using these technologies, companies now have access to aspects of your personal data. When accepting cookies, a Web site can use them to track which pages you visit on the site (Kristol, 2001). This may not seem like a user is sacrificing much personal data, however, sites such as Google can record everything that you look up and sites like Amazon can keep track of what items you view. Therefore, this allows them to create a profile of your interests which has a direct use for targeted advertising. Mobile phones offer access to applications that can have a number of data harvesting capabilities. An important subsection of mobile applications for the purposes of personal data collection are social media platforms such

as Facebook, Instagram, and LinkedIn. In a study of teen social media usage, four questions were asked and found that a majority of teen social media users posted their real name, interests such as movies, music or books, their birth date, and relationship status to their profile they use most often (Madden et al. 2013). This personal data that is willingly shared on social media platforms again has a direct usage in targeted advertising. While the claim to an invasion of privacy may be much weaker due to the nature of willingly sharing this information, there is still a lack of understanding on how that personal data is actually used. These are some examples of the data that can be collected and used for the purposes of targeted advertising and my research will explore more personal information that is collected and used. Some important sources for this information will be research surrounding targeted advertising and the methods for targeting offered by advertising companies such as Google and Facebook.

The last step in answering this research question will involve exploring the effectiveness of targeted advertising as a method for controlling habits. If targeted advertising is shown to be effective then it can be regulated. This is a precedent that we can see in antitrust laws which "prohibit conduct by a single firm that unreasonably restrains competition by creating or maintaining monopoly power" (Federal Trade Commission). The key here is that the law only prohibits conduct when the firm is unreasonably restraining competition, not just the ability to unreasonably restrain competition or the fear that they might. This means that to regulate the targeted advertising industry it must actually be effective in its attempts to control people rather than just attempting to. Targeted advertising works by using data analytics, machine learning and predictive algorithms to convert data into purchases by interposing between choice and said outcome (Okonkwo, 2021). Thus, the overall goal of targeted advertising is to manipulate a consumer into purchasing or otherwise performing an act that the advertisement wants. This

means that if targeted advertising were to be effective, it would mean removing the choice from the consumer by controlling them to do something they may not have done otherwise. Now that we know that the overall goal of targeted advertising is manipulation, if we can show its effectiveness in practice then we will be able to answer the overall question of whether targeted advertising creates an illusion of personal choice. One type of source that could be used to evaluate the effectiveness of targeted advertising is case studies. A notable example of this is an experiment that attempted to shift Republican views on climate change and found that "the campaign increased Republicans' understanding of the existence, causes and harms of climate change by several percentage points" (Goldberg et al. 2021). Another potential source of information will come from small businesses that can quantify the impact of using targeted advertising. Lastly, we can look at the flow of money into the industry of targeted advertising as an implication of success. In order to use this idea, we will have to dive into the information available to consumers as well as economic theory.

As shown, there is a well-defined pathway to answering this research question. The first step is showing that there are certain technologies that people in the internet age cannot reasonably be expected to operate without. If they must use these technologies, then they have no ability to escape the collection and usage of their data in targeted advertising. With that connection made, targeted advertising will have been shown to be challenging to escape. If targeted advertising is shown to be effective at manipulating behavior in practice, then the inability to escape it will inherently mean it will have some control over your choices. If you are unaware of its control of your choices, as most people are, then targeted advertising will have successfully created the illusion of choice. Therefore, there exists some potential for negatively impacting society which is grounds for regulation.

Conclusion

Targeted advertising is becoming an increasingly popular tool used by companies to sell their products. For this to work, it requires mountains of human data so the right ads can be seen by the right people. This targeted process introduces the possibility of discrimination by utilizing details of a person such as gender or race allowing certain groups of people to see ads that others can't. More research needs to be done on how to remove this problem without hindering the ability of the targeted advertising algorithms to sort people into targetable groups. Despite that issue, there are many people who think that targeted advertising should be regulated anyway. The ability to regulate the practice of targeted advertising falls on the answer to whether targeted advertising is truly effective in its goal. Thus, this research will examine how targeted advertising is being used to manipulate human behavior and if people are consciously aware of this process so that we can make an educated decision on regulation.

Bibliography

Bond, S. (2021, November 9). *Facebook scraps ad targeting based on politics, race and other 'sensitive' topics*. NPR. Retrieved March 31, 2022, from https://www.npr.org/2021/11/09/1054021911/facebook-scraps-ad-targeting-politics-race-sensitive-topics

Federal Trade Commission. Monopolization Defined. Federal Trade Commission. https://www.ftc.gov/advice-guidance/competition-guidance/guide-antitrust-laws/single-firm-conduct/monopolization-defined

Goldberg, M. H., Gustafson, A., Rosenthal, S. A., & Leiserowitz, A. (2021, June 14). *Shifting Republican views on climate change through targeted advertising*. Nature News. Retrieved May 12, 2022, from https://www.nature.com/articles/s41558-021-01070-1

Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., & Beaton, M. (2020, August 17). *Teens, social media, and privacy*. Pew Research Center: Internet, Science & Tech. Retrieved May 12, 2022, from https://www.pewresearch.org/internet/2013/05/21/teens-social-media-and-privacy/

Martin, K., & Shilton, K. (2016, April 13). *Putting mobile application privacy in context: An empirical study of user privacy expectations for mobile devices*. Taylor & Francis. Retrieved May 12, 2022, from https://www.tandfonline.com/doi/full/10.1080/01972243.2016.1153012

Okonkwo, V. (2022, January 24). *Between privacy, autonomy and surveillance capitalism: Exploring a trade-off of priorities*. SSRN. Retrieved May 12, 2022, from

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3970537

Pal, R., Crowcroft, J. (2019, July 1). *Privacy trading in the surveillance capitalism age viewpoints on 'privacy-preserving' societal value creation*. ACM SIGCOMM Computer Communication Review. Retrieved May 12, 2022, from

https://dl.acm.org/doi/pdf/10.1145/3371927.3371931

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F. N., Arvanitakis, G., Benevenuto, F., Gummadi, K. P., Loiseau, P., & Mislove, A. (2018). Potential for Discrimination in Online Targeted Advertising. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, 81(1).

U.S. Department of Commerce. 2021. "Quarterly retail e-commerce sales 4th Quarter 2021." Press Release, February 18.

https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

W3Techs. (2022, May 12). *Usage statistics of persistent cookies for websites*. W3Techs. Retrieved May 12, 2022, from

https://w3techs.com/technologies/details/ce-persistentcookies