Consumer Data Collection

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

Rohit Chhatre

Spring 2021

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

Prof. S. Travis Elliott, Department of Engineering and Society

Introduction

Nowadays, so many people have smartphones and other technology devices where they can provide their personal information to these different platforms. Although some data is encrypted and kept secure, like social security numbers and exact credit card numbers, other information is leaked through to the companies and they have access to all of our personal data. My technical project focuses on creating an application to promote cashless transactions between donors and panhandlers. Through figuring out goals for my app, one was to gather information on how much people donate to those causes. This made me think about consumer data collection in general which led me to choose it as my topic for my STS research. A lot of people and the media harp on all the negatives of collecting people's data through app usage, but I believe there are positives that people do not consider. Using the Social Construction of Technology (SCOT) framework, I will analyze the topic of data collection, data analysis, and the selling of data. In choosing the SCOT framework it is important to consider its differences with its counterpart, technological determinism. This framework suggests that technology is the driving factor in shaping society. On contrast, SCOT argues that technological innovation occurs as a response to different societal forces and is shaped by human actions. Technological determinism would say that the innovation of this technology influences how society develops; however, SCOT is the best methodology to analyze this topic because it shows that there are many different societal factors that influence how consumer data collection is used and how the technology keeps developing. The main factors of the SCOT framework are relevant social groups, flexibility of interpretation, and the technological frame (YOUSEFIKHAH 2017). By analyzing the different stakeholders involved and by showing that there are positives within all the scrutiny, it is clear

that perhaps spreading greater knowledge about this technology could foster greater acceptance and could be beneficial for all parties.

STS Background & Analysis

Consumer data collection has been a hot topic recently, as more and more of our world is transitioning online. Everyone has the chance to post content to the web and provide information about themselves. This leads companies who manage these platforms, like social media, email, and other online applications, to have control over the personal information of so many customers. It has become a common trend for companies to track and analyze the data that comes through their sites. There has been a lot of negative backlash on these practices, like we have seen with Facebook, but I believe that there are also a lot of positives to tracking online data. For my research, I could look at this topic through many viewpoints. SCOT tells us that we need to analyze all the different social groups that play a role in developing this technology.

Social Groups

With the first tenant of SCOT, it is important to analyze all the relevant social groups. From an economic standpoint, these companies that collect data make so much money from selling it to other companies to use. We have even seen products like the Amazon Alexa go down in price because it is clear that these companies aren't selling the actual product, rather they are providing a platform as a means for collecting user information. In terms of social media, it is critical for these platforms to be free, so if companies want to increase profits, they have to decide between spamming users with more advertisements or selling user data. Some of the uses of data collection could be malicious but there are a lot of positive uses for the data as

well. Many complex technical aspects go into actually tracking user data in an application and some of this data is actually useful, like when predicting interests. Sometimes these sites know what you want even before you, as a consumer, knows what you want. So many of my friends have bought different items off of the "recommended" list or catered advertisements that show up on their feed. Due to the many intricacies of retrieving the data, this data is filtered pretty well and is also great to feed into different Machine Learning algorithms that we use for other technologies. The three main reasons why online learning is crucial is because of its volume, velocity, and variety (Hunt 2017). There is just so much data out there online. This volume can never be replicated in terms of a survey or physically collecting the data. Also, filtered data from the application is usually better than data collected from surveys because it is most "raw". People tend to not include inherent biases in a physical survey but they can't hide these biases online. The velocity of data is also important. Online data is constantly updated due to new data always coming in. This means that our algorithms can be constantly changing to fit the new data which will be better in the long run, in terms of adapting to a changing society. The variety of online data may be the most important, though. When it comes to physical collection of data, there may be human biases that are unaccounted for, like lack of representation of race or gender for example. This bias would be counteracted by the usage of online learning. These online learning models are the foundation of technologies like self-driving cars, online shopping, and voice recognition technology. Due to all these benefits, consumers could be pleased that their information is used in other technologies that could in turn benefit themselves. With the rise of companies investing more and more money towards online learning research, there may be a higher economic incentive for companies to use consumer data for just causes. These are the

relevant stakeholders that could turn companies away from selling user information in a malicious manner.

So why is it such a big deal for companies to take this data? Privacy has always been a big talking point when it comes to consumer data collection. For instance, from a political standpoint, people are concerned with companies being able to "spy" due to all the tracking that happens behind the scenes. As an example, there has been a scandal with TikTok having access to location data and other inner phone workings when maybe they shouldn't be authorized to that information. They have gotten into a $650 million lawsuit after companies found out they had access to the Notes app in iOS (Allyn 2020). With access to this very personal information, it could be very harmful if it gets into the wrong hands. TikTok has been labeled as a "Chinese Spyware" company (Doffman 2020) because all of this data that they have collected has allegedly been sent to China or the Chinese Communist Party. Because of this backlash, United States has forced government employees and their families to delete the app. There has also been a push for American companies to buy the app, like Microsoft, so "data on American citizens would remain in U.S. borders" and data backed-up or outside the country would be deleted after being transferred (Allyn 2020). Some people say that the data collection fraud shown in TikTok doesn't warrant the deletion of the app because other domestic technology companies also follow these malicious practices. Even within the United States, there have been many private corporations that have made people question whether they are collecting data in a fair manner. For example, Facebook has faced millions of dollars in lawsuits because of "unethical practices" (Solon & Farivar 2019) in terms of recording and selling data. They had a huge data breach in which 50 million accounts got leaked, which equated to almost a third of the users in North America (Graham-Harrison & Cadwalladr 2018). For such a big company with information on

so many people to leak all this data is definitely alarming. In these cases, it is valid for people to have concern for their privacy and companies should also be kept in check on how the data they collect is stored and managed.

Interpretive Flexibility

Another factor of SCOT that can be analyzed with this topic is interpretive flexibility. This states people can appropriate technologies, or artifacts, differently in order to find the most uses for it (YOUSEFIKHAH 2017). In this case, there are so many different possible uses for data collection and data analysis. Today, there are many things we take for granted which were built on preexisting data sets. There are many examples such as language translation tools, mobile traffic services, digital mapping technologies, instant spell checkers, and spam and fraud detection tools (Thierer 2013). The list just keeps growing. This data can provide a deeper understanding of the overall market, because it can pick up on people's wants and tendencies. Companies are able to capitalize on this and are able to know consumers better than anyone else. This can be due to an increase in the amount of data that can actually be collected from people. Newer phone and computer models have allowed companies to have access to more and more of consumer information such as location data, voice data, and cross-application data. The consumer database has improved as a result of increasing consumer information. With data analysis, companies can also find better marketing strategies in order to get people to buy their items because they now have more insight on their market. It is very prominent especially with airline and hotel companies. These companies make the most profit off of repeated consumers who travel all the time. It is vital that travel firms have correct user data to make sure companies know which people they need to retain. For example, a hotel chain like Marriott will offer more

upgrades and amenities to a regular customer over a one-time customer because the regular is more likely to come back to book another Marriott stay. Airline companies, like Alteryx, have saved millions of dollars by analyzing flight routes to maximize fuel efficiency ("Reviving the travel industry through data analytics" 2020).

Data analysis can also help in identifying and preventing racist, threatening, or other malicious content that is on the web. These strategies have been used by companies and government agencies to track down potential terrorist attacks or other potential harmful scenarios. A solution called predictive analysis has been used recently in order to identify content on the web that could be likely harmful. This technique has been used to find probable warning signs which can help us predict the possibility of a school shooting, as an example (Parkman n.d.). Right after the 9/11 attacks and the rise of global terrorism, predictive analysis was used to target potential suspects who were posting suggestive content or pledging their loyalty online to various terrorist groups. On the consumer end, it provides greater personalization of content (Truyo n.d.). Firstly, there are catered advertisements which encourage people to spend money on things they most likely enjoy. Furthermore, on social media sites, different user recognition technologies help to display only relevant content to that specific person based on their usual interests in general social media content. Lastly, as mentioned before, the data that is collected from consumers is ideal for feeding into different online learning algorithms because of its optimal volume, velocity, and variety.

Technological Frame

The third factor of SCOT that is important is technological frame. There are three domains within technological frame: the nature of technology, technology strategy, and

technology in use (YOUSEFIKHAH 2017). One part that is currently missing is a bit of the first domain. Users know some of the functionality of the technology but may not fully understand the full capabilities and functions of the technology. If all the different stakeholders in this issue consider the capabilities of the technology, it can be optimized so that it can satisfy all parties. The second domain is super important in helping combat the negative light that consumer data collection is in right now. A lot of the stakeholders like the technology companies may not necessarily have the right motivation for implementing these technologies. Due to the capitalist nature of the United States, these companies may have set up data collection so that they make the most profit (like we have seen in the travel industry) and users may also not know the full extent to which their data is being collected. This may not bode well for all the users of these different platforms. For companies to make data collection a beneficial technology for all, their strategy would need to be modified (like we have seen with online learning). Another factor which may not be as prevalent is that developers of other technologies in other companies should also conscious on how data collection works through the usage of existing technological platforms. Facebook invented their signature "Graph API" (Solon & Farivar 2019) which is an Application Programming Interface (API) that developers use in order to utilize in order for different programs to interact with one another. Facebook has been able to collect data through these services as well which means that they can not only collect data from users on their own platform but are enabled to collect data on any platform that uses their API. This scrutiny came about from other companies, like Spotify and Yahoo, who were complaining that their data was being stolen by Facebook through the usage of these APIs ("Facebook's data-sharing deals exposed" 2018). This means that there is also a conflict between different companies between the technology collected, not just between a user and a company. Bigger companies with more

data are at an advantage in this scenario because they are the ones controlling the vast majority of user information. Lastly, people know how to use these technologies and generally are also contributing to the problem. People willingly put all of their personal information on these sites, so they should expect their data to go somewhere behind the scenes. Facebook many times during their lawsuits has claimed that people have accepted the privacy agreement that shows up when you first download the app and agree to sharing your data that is collected on the app. It is important for users to recognize this and post only the content in which they are okay being public.

Overall, using SCOT we can see that there are many social factors that influence the technology of consumer information collection. The different players need to know how the technology works to build greater acceptance. By laying out all of the societal factors that go into the development of this technology, there may be a way for this technology to be more universally accepted and used in a positive way so that some of the downsides that people harp on can be overcome.

Conclusion

Consumer data collection has many positives and negatives. With inspiration from my technical project of collecting information on how much people donate to panhandlers, I was interested in all the other positive uses of collecting user data. In my case, the government has little to no data on how much money homeless people actually receives and how much money people are spending to these causes so the data that my application will retrieve will actually be helpful. Nowadays, society is consumed on all the negatives, focusing on breach of privacy and malicious intent on the usage of this information, even though there are a lot of positive uses of

this technology. Using the SCOT analysis, it is clear that there are many different societal factors that play into this technology and different people have used this technology in different ways. These different sectors should accept that this technology exists and that it can be improved within the different sectors in order in order to fit the needs of all the different stakeholders involved. The benefits of this technology to all the different social groups are clear, but the greediness of certain groups and the power of those groups to enact their own agenda has resulted in the lack of trust in consumer data collection and analysis. By allowing more people to understand the implications of this technology, there could be a greater acceptance for the usage and continued development of consumer data collection.

References

Allyn, B. (2020, August 04). Class-Action Lawsuit Claims TikTok Steals Kids' Data And Sends It

To China. Retrieved December 9, 2020, from

https://www.npr.org/2020/08/04/898836158/class-action-lawsuit-claims-tiktok-steals-kids-

data-and-sends-it-to-china

Doffman, Z. (2020, August 11). Is TikTok Seriously Dangerous-Do You Need To Delete It?

Retrieved December 9, 2020, from

https://www.forbes.com/sites/zakdoffman/2020/07/11/tiktok-seriously-dangerous-warning-

delete-app-trump-ban/?sh=15a54f252b0e

Facebook's data-sharing deals exposed. (2018, December 19). Retrieved December 9, 2020, from

https://www.bbc.com/news/technology-46618582

Graham-Harrison, E., & Cadwalladr, C. (2018, March 17). Revealed: 50 million Facebook profiles

harvested for Cambridge Analytica in major data breach. Retrieved December 9, 2020,

from https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-

influence-us-election

Hunt, X. (2017, October 17). Online learning: Machine learning's secret for big data. Retrieved

December 9, 2020, from

https://blogs.sas.com/content/subconsciousmusings/2017/10/17/online-learning-machine-

learnings-secret-big-data/

Parkman, B. (n.d.). Using Predictive Analysis to Address School Safety. Retrieved from

    https://bluelightllc.com/wp-content/uploads/blue-light-the-safe-campus-white-paper.pdf

Reviving the travel industry through data analytics. (2020, August 18). Retrieved March 12, 2021,

    from https://revenue-hub.com/reviving-the-travel-industry-through-data-analytics/

Solon, O., & Farivar, C. (2019, April 18). Mark Zuckerberg leveraged Facebook user data to fight

    rivals and help friends, leaked documents show. Retrieved December 10, 2020, from

    https://www.nbcnews.com/tech/social-media/mark-zuckerberg-leveraged-facebook-user-

    data-fight-rivals-help-friends-n994706

Thierer, A. (2013, September 16). Relax and Learn to Love Big Data. Retrieved December 9,

    2020, from https://www.usnews.com/opinion/blogs/economic-intelligence/2013/09/16/big-

    data-collection-has-many-benefits-for-internet-users

Truyo. (n.d.). The Importance of Data: The Top Benefits of Collecting Customer Data. Retrieved

    October 20, 2020, from https://insights.truyo.com/consumer-data

YOUSEFIKHAH, S. (2017). Sociology of Innovation: Social Construction of Technology

    Perspective. *Ad-Minister*, *30*, 31–43. https://doi.org/10.17230/ad-minister.30.2