Examining the Ethical, Practical, and Societal Implications of Data Monetization

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

Steven Pham
Spring, 2022

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

Signature _____ Date _____
Steven Pham

Approved _____ Date _____
Hannah Rogers, Department of Engineering and Society

Abstract

Data monetization is the act of generating measurable economic benefits from available data sources. The highest performing tech companies have adopted it to increase revenue, develop new products, and improve user satisfaction. The problem arises when tech companies monetize user-generated and private data. This paper applies the Social Construction of Technology theory to explore the implications of monetizing user data and reasons for its acceptance and rejection in society. The key takeaway of the paper is the unethical use of data monetization that exploits user addiction to Internet content and as well as the lack of user data protection.

Examining the Ethical, Practical, and Societal Implications of Data Monetization

Introduction

The monetization of data started in the dotcom era where the sale of data was more profitable than tangible goods themselves. User data generated from the Web was the golden ticket for tech giants (West, 2017). The most significant change during the dotcom era is how Internet content is created. The content the dotcom era, Web 2.0, is created by the users interacting with websites, as opposed to the websites themselves (Marr, 2015). A read-only site like the World Book Encyclopedia is an example of a Web 1.0 site whereas Facebook a Web 2.0 site. This change allowed Internet companies to track user web behavior and ultimately started the monetization of users' data that led to the invasion of privacy and the attack on self-determinism on the Internet. This paper seeks to explore the ethical, social, and political implications of the surveillance nature of big data through the relationship between technological advancements and regulations. There are four terms relevant to the subject: HTTP Cookie, Targeted Advertisement, Big Data, and Surveillance Capitalism.

What is a HTTP Cookie?

The HTTP cookie, created by Lou Montulli, is a technology that allows store data on a user's device to log various information. It was intended to check whether visitors to a website had previously visited the site so that websites can deliver preferential settings for the users (Kristol, 2001). However, advertising companies used Lou Montulli's invention to create the third-party cookie that enabled them to record users' browsing history and other private information. Advertising companies would offer ad revenue to websites, in exchange that they

get to place their ads and third-party cookies on the website. With a network of third-party cookies on the Internet, the advertisers have successfully established cross-site web trackers capable of recording users' browsing activities and behaviors. In the early days of Internet, web browsers had little to no measures against third-party cookies and its tracking nature. Still, using modern tracker blockers is comparable to "using an umbrella in a hurricane" (DuckDuckGo, 2021). The commoditization of user data has changed the priorities of Internet companies and created new privacy concerns for user data.

What is Targeted Advertising?

The HTTP cookie enabled targeted advertisement that delivers personalized ads based on the user's browsing history. When a user visits a cookie-embedded website and voluntarily gives up personal information through sign-ups and question, that information is recorded in an advertiser's database. The device of the user has a unique IP address that allows the advertiser to build a profile for the user. A business wanting to promote their products to a certain demographic will reach out to the advertising company for this list of potential customers. In a survey of 200 UVA students, over 80% have experienced social media targeted advertising. The invasion of privacy is most prevalent in sociodemographic targeting that delivers ads based on factors such as age, gender, and nationality. On the other hand, behavioral targeting profile users based on their actions and employ data mining and machine learning to identify characteristics of high-profit customers.

What is Big Data?

This mass collection of user data is referred to as Big Data, which is referred as data of individuals under the control of tech companies (Zuboff, 2015). Big Data is the application of statistical analysis to an unstructured dataset to draw insights through trends, patterns, and correlations (Tableau, 2022). Oracle (2022) identifies three characteristics of Big Data, the volume of data, the velocity or rate that the data is received, and the variety or type of data available such as text, audio, or video. The purpose of Big Data is to benchmark performance and derive insights through statistical analyses. It provides information that allow companies to innovate and enhance decision-making and automate business processes. I will breakdown the concept of Big Data using the Social Construction of Technology (SCOT) methodology in the Analysis section and explain the incentives that drive tech companies to monetize their users' data. But first, I will explain the driving factor of Big Data through the concept of *surveillance capitalism*.

LITERATURE REVIEW

Big Data has major influences on the economy and socio-political sphere. It possesses the power to change the status quo of institutional divisions with deep and invasive insights. Tech companies monetize its users and consequently change the profit incentive from customer satisfaction to volume data extraction. The process of collecting and analyzing real-time data is so endogenous to everyday business activities that the two domains are more or less conflated (Zuboff, 2015). The intention of targeting users' online behavior for monetization purposes is one aspect of *surveillance capitalism*. Coined by Harvard professor Shoshana Zuboff, *surveillance capitalism* is a logic of capitalist accumulation that…

> "…claims human experience as free raw material for translation into behavioral data [which] are declared as a proprietary behavior surplus, fed into advanced manufacturing process known as 'machine intelligence', and fabricated into prediction products that anticipate what you will do now, soon, and later" (Zuboff, The Age of Surveillance Capitalism, 2019).

Big Data has paved the way to surveillance capitalism that undermines online self-determination. Predicting a behavior like purchasing habits that highlights the ideal sequence of actions to get users to open their wallets is so profitable that companies are willing to excessively overspend to access this functionality. Big Data is as revolutionary as Henry Ford's assembly line and mass-production that gave rise to managerial capitalism. In her conclusion, Zuboff writes that as industrial capitalism exploited nature, surveillance capitalism exploited human nature (2019).

What are the Dangers that Social Media Users Face?

An example of *surveillance capitalism* is the change of cost based on anticipated behavior, such as airline companies' dynamic ticket pricing to maximize profits during the holidays, and enabling creditors to make lending decisions based on an individual's social network (West, 2017). Social media *surveillance capitalism* is the usage of user's action to train their next action to be monetizable (Saura, Palacios-Marques, & Iturricha-Fernandez, 2021). This strategy is known as positive intermittent reinforcement (PIR) in psychology that creates unconscious user habits by associating an action to an outcome like pulling down refreshing a newsfeed for new content (Hayes & Kelly, 2018). This action is akin to slot machines in Vegas and releases dopamine, a hormone that's released when a reward is expected. The access to

social media and availability of PIR actions can generate addiction and negatively impact user feelings. As described by Saura (2021):

> "The problem lies in the fact that users feel this addiction not towards the content published by their followers but are rather pushed by factors related to economies of action using social media to try to generate economic benefits from users' consumption of content".

The system of social media monetization leads users to betray their own autonomy. The unconscious behavior on social media is analyzed so that future actions can be predicted and once achieved, monetizable actions are substituted in, unbeknownst to the user. Figure 1 shows the groups that a social media user encounters by level of visibility or awareness of the group's existence.

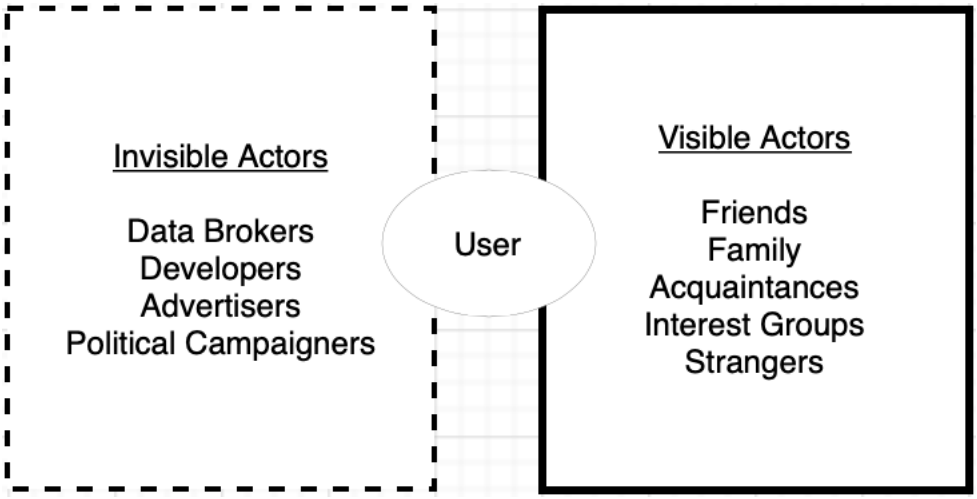

Figure 1. The visible and invisible social network (Bigo, Isin, & Ruppert, 2019)

Social media network like TikTok is an intermediary between the user, and the invisible actors. Through a series of economies of action shown in Table 1, TikTok is able to transform users into customers for the invisible actors.

Concept	Description
Tuning/Nudge	Tuning/nudge refers to any aspect of a choice architecture that predictably alters user behavior.
Herding	Herding relies on controlling key elements in a person-immediate context. Controlling context-aware data enables combining emotions, cognitive functions, vital signs, and so forth.
Conditioning	Conditioning is an approach to induce a certain behavior. Behavior modification should mimic the evolutionary process where naturally concurring behaviors are "selected" for success by environmental conditions; thus, conditioning is essential to the new science of massively engineered human behavior.

Figure 2: Key approaches to economies of action in social media networks (Zuboff, 2019)

None of which would be possible without the usage of artificial intelligence training algorithm that analyze user-generated content, behavior, and data to generate the blueprint to monetization (Saura, 2019).

ANALYSIS

In this section, I will apply the Social Construction of Technology (SCOT) method to the monetization of data. The SCOT method identifies the social groups, typically the producer and the users, with regards to a certain technology artefact which has different meanings to the different groups. A technology artefact has interpretative flexibility to different social groups, meaning that social groups attach different meanings to the artefact. The social groups can interact within the technological frame through which they arrive at shared meanings. In the process, the technology artefact is constructed and deconstructed through social interaction and

as time goes on it is acceptable by different social groups and stabilizes (James, 2021). Figure 3 breaks down the social groups of Big Data and their unique problems.

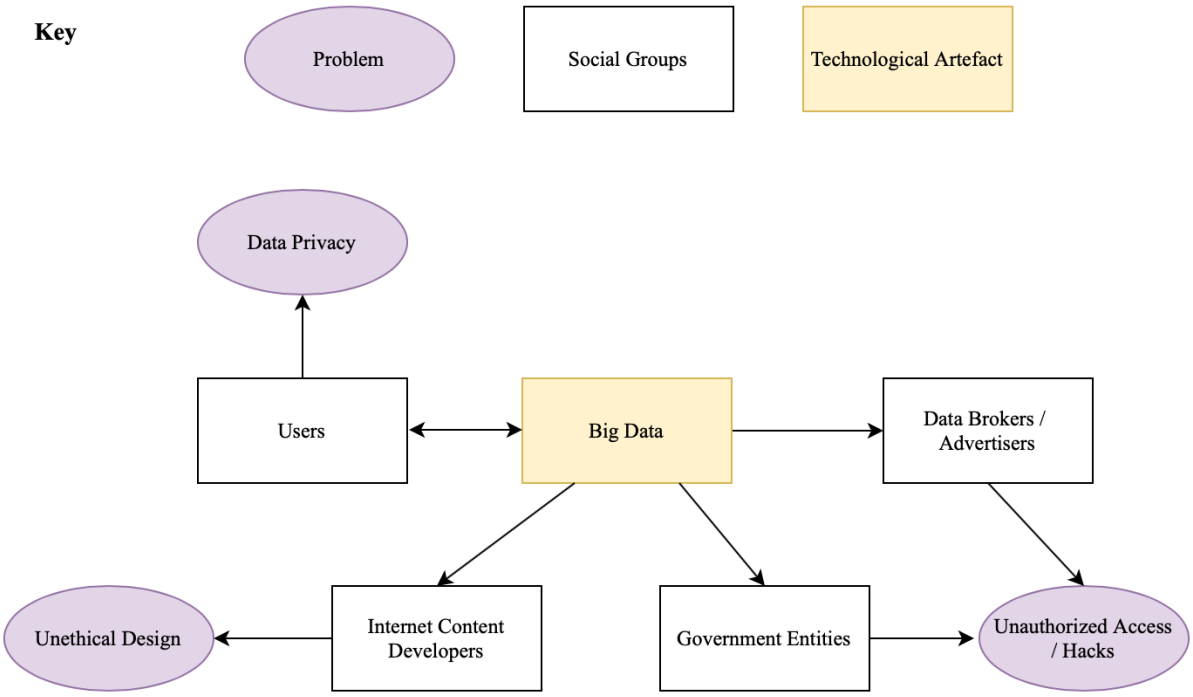

Figure 3: The Social Construction of Technology of Big Data with relevant social groups, the problems associated with each group(s), and the artefact (Pham, 2022).

Business optimization based on pattern recognition is entirely possible without Big Data. However, it is far less efficient and require more labor-intensive, manual statistical calculations and data collections. The introduction of Big Data brought opportunities as well as problems for different social groups. Private information is increasingly likely to be leaked to hackers and foreign governments. The point of stabilization – when social groups see the problem as being solved – is not reached. Developer do work to better secure the data, but as long as the data is still being collected, hackers will find a way to access the data.

What Regulations Have Been Proposed?

Surveillance capitalism would not exist if the HTTP cookie was responsibly used as intended by its creator. Recall that the cookie was created to make the Web more personal, comparable to going to a restaurant where the waiter remembers your order. The original creators, David Kristol and Lou Montulli saw the potential for misuse and invasive nature of their technology and outlined protocols and mechanisms to give users greater control over the cookie, allowing users to opt out and choose what information can be saved (West, 2017). Of course, this was not the case. The first implementation of cookies in a web browser included none of the features proposed. West (2017) states:

> "The incorporation of cookie technologies into web browsers and their connection to ad networks lay the groundwork for advertising too become the business model of the Internet, by creating a core infrastructure for commercial surveillance in Internet technologies on the eve of the Internet's most rapid period of consumer growth."

Unlike the United States, Europe privacy regulators were successful in restricting data collection by tech companies. The Data Protection Directive of 1995 outline clear requirements of unambiguous consent before tech companies can collect personal data. It was later succeeded by the E-Privacy Directive in 2002 that gives users the power to refuse the usage of HTTP cookies (West, 2017). However, the Snowden Revelation later reveals the framework of the European Union was rendered invalid and not robust. One of Snowden's leaked documents revealed the Squeaky Dolphin program, implemented by the British intelligence group GCHQ to monitor

specific social media features of certain population without the knowledge or consent of the company. The program went as far as targeting specific users with propaganda (Swider, 2014).

What is the Difference Between the EU and US Data Protection Policy?

According to an article published by the Arkansas State University in 2020, the core difference in user privacy between the European Union and United States is how legislations are prioritized (Arkansas State University, 2020) . The EU utilizes a top-down approach in the form of directives, a legal act enacted into a state's national law to achieve a particular result without explicitly stating the means of achieving that result. This allows nations to prioritize data protection within their own capacities and meeting the overarching goal as a whole. In 2018, the EU passed the General Data Protection Regulation (GDPR) that governs data protection for every member state. On the other hand, the US does not have a centralized system to regulate privacy. Instead, each state enacts its own privacy laws. On a federal level, the Federal Communications Commission (FCC) provide the legal frameworks but does not pass nor regulate privacy laws of its own (Arkansas State University, 2020). This bottom-up approach leads to multiple interpretations of what privacy practice is and isn't acceptable. The bottom-up approach lacks accountability and is susceptible to ambiguous definition in the exchanges of personal data between nation states.

What is the Implication of Bottom-Up Approach to Data Protection?

Due to a lack of accountability, tech giants in the US are able to monetize personal data at their discretion. This lack of accountability allows the US government to work with tech giants to surveil its citizens. According a report by ProtonMail (2020), an encryption email company,

"user data requests from the US government to Google have increased by 510% since 2010 and for Facebook data by 364% since 2013. The number of government requests made by the US is more than Canada, UK, New Zealand, Australia, France, Denmark, the Netherlands, Norway, German, Spain, Belgium, Italy, and Sweden combined. ProtonMail's CEO Andy Yen said in a statement:

> "It's no secret that companies like Google, Facebook, and Twitter have gathered vast amounts of personal data which users have provided unsuspectingly. As a result, big tech companies now retain more personal data than the most sophisticated government mass surveillance programs, and it is clear that governments do routinely ask for, and gain access to this data. Users may have consented that Google and Facebook can use their data for advertising, but many will be unaware that their personal data is also available to government." (ProtonMail, 2020)

How do Democracies Spy on Their Citizens?

An article from the New Yorker magazine by Ronan Farrow points out the usage of a Isareli-based commercial spyware, Pegasus, to spy on members of the European Parliament and other political figures. The article quotes Cristin Flynn Goodwin, a Microsoft executive who has led the company's efforts to fight spyware: "The secret is that governments are buying this stuff – not just authoritarian governments but all types of governments" (The New Yorker, 2022). The creator of the spyware, Hulio, admits that "almost all governments in Europe are using our tools" and that "[his company] has a monopoly in Europe" (The New Yorker, 2022).

CONCLUSION

What are the Learned Lessons of Past Data Protection Policy?

Legal roadblocks are only temporary measures to *surveillance capitalism*. A system that keeps tech giants and governments accountable will restore power to the Internet users. This new system must have new incentives and operate under a vision that prioritizes dignity and self-determinism. Similarly, tech companies have the responsibility to do right by its users. If the current state of privacy is to continue, the digital future will be continued to benefit tech giants and nation states. Zuboff (2014) observes that "big data is big contraband", meaning that the data that tech companies profit from is collected without the knowledge or consent of the user. She also compares the massive amount of raw data as a "wasteland", in that it is of value to the user, in that they possess no ability to transform it to actionable insights like tech giants, and thus they have no resistance in declaring it theirs (Zuboff, 2014). Will the government dethrone the data market with regulations, put tech companies out of business, and stop utilizing their technology to mass surveil citizens? Probably not. However, recognizing the current practices as unethical and calling for change is a step in the right direction.

What are the Ethical Perspectives of Surveillance Capitalism?

When analyzed from an ethical perspective, *surveillance capitalism* fails on many levels. Utilitarianism theory for example, is based on whether the result will benefit the majority. *Surveillance capitalism* benefits the tech companies and the invisible actors of the user. On the granular scale, one can argue that the majority are the stockholders of the tech company. But the population that tech companies such as Google and Facebook affect is beyond their stockholders. Similar to utilitarianism, the justice ethical principle states that the decision makers should focus

on actions that are fair to those involved. One can argue that the technology like the search engine and other functionality that big tech provides for free in exchange to gather user data is just. But the problem lies in the lack of transparency of what and how the data is collected. In this context, the usage of personal data without user knowledge to create a product that inhibits user autonomy is unjust. In the big tech companies' perspective, users' right to privacy is forfeited as part of the social contract, and their privacy agreement, when users agree to use their technology. However, this is all unbeknownst to the majority of the users, as these agreements are hidden behind layers of legal text walls. The dubious attempts to disclose these important points implies that big tech is unethical in their relationship with its users.

References

(2022). Retrieved from Tableau: https://www.tableau.com/learn/articles/big-data-analytics

(2022). Retrieved from Oracle: https://www.oracle.com/big-data/what-is-big-data/

Alexander Haslam, D. v. (2014). *Identity Metamorphosis and Groupthink Prevention.* Santa Cruz.

Bigo, D., Isin, E., & Ruppert, E. (2019). *Data Politics: Worlds, Subjects, Rights.* Routledge.

Demadis, C. (2021). *Stories are the new Storefront*. Retrieved from Social Media Today: https://www.socialmediatoday.com/spons/stories-are-the-new-storefront/597335/

Hayes, P., & Kelly, S. (2018). *Distributed morality, privacy, and social media in natural disaster response* . Technoloogy in Society.

Hutchinson, A. (2021). *SMT Expert Series: Nick Cicero Discusses the Latest Analytics Trends and Opportunities*. Retrieved from Social Media Today: https://www.socialmediatoday.com/news/smt-expert-series-nick-cicero-discusses-the-latest-analytics-trends-and-op/604624/

James, R. (2021). *The SOCIAL CONSTRUCTION of TECHNOLOGY* . Retrieved from YouTube: https://www.youtube.com/watch?v=cqjUqqANnK0

Krishnamurthy, P. (2019). *Understanding Data Bias*. Retrieved from Towards Data Science: https://towardsdatascience.com/survey-d4f168791e57

Kristol, D. M. (2001, November 1). Retrieved from ACM Digital Library: https://dl.acm.org/doi/10.1145/502152.502153

Marr, B. (2015, Feb 25). Retrieved from WeForum: https://www.weforum.org/agenda/2015/02/a-brief-history-of-big-data-everyone-should-read/

Michael Haenlein, E. A. (2020). Nativating the New Era of Influencer Marketing: How to be Successful on Instagram, TikTok, & Co. *Haas School of Business* .

Reuben Binns, M. V. (2017). Like trainer, like bot? Inheritance of bias in algorithmic content moderation. Oxford, United Kingdom.

Saura, J. R., Palacios-Marques, D., & Iturricha-Fernandez, A. (2021). *Ethical design in social media: Assessing the main performance measurements of user online behavior modification.* Journal of Business Research.

West, S. M. (2017, July 5). Retrieved from Sage Journals: https://journals.sagepub.com/doi/full/10.1177/0007650317718185#_i10

Zuboff, S. (2014). Retrieved from Opencuny: https://opencuny.org/pnmarchive/files/2019/01/Zuboff-Digital-Declaration.pdf

Zuboff, S. (2015). Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology*, 75-89. Retrieved from https://journals.sagepub.com/doi/pdf/10.1057/jit.2015.5

Zuboff, S. (2019). *The Age of Surveillance Capitalism.* Profile Books.