

Search Optimization: Data Refactoring to Enhance User Search Experience

(Technical Paper)

How have online marketing practices defined rights to data privacy in the digital age?

(STS Paper)

A Thesis Prospectus Submitted to the
Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia
In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Rehan Javaid

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

Signature _____ Date _____

Rehan Javaid

Approved _____ Date _____

Brianna Morrison, Department of Computer Science

Approved _____ Date _____

S. Travis Elliott, Department of Engineering and Society

Prospectus

Introduction

The World Wide Web has expanded its influence an unprecedented amount since its early days in the 1990s. Through its growth, the Web has adopted roles that have defined its use today—one such role being that of an online marketplace. Online retail has grown to become one of the most popular uses of the Internet today, with roughly 76% of U.S. adults shopping online and global e-commerce sales expected to reach \$5.5 trillion in 2022 (Barber, 2022).

As businesses made their transition to the digital world, their methods of marketing quickly followed suit. Digital advertising has since engulfed the browsing experience on the internet today. On any website we go, we are met with a diverse array of advertisements specifically catered to our interests and browsing habits at the time. However, this marketing shift that appears to have benefitted both consumers and businesses has not done so without controversial tactics at play.

Many of the major marketing strategies on the internet have developed around web tracking: an action which allows for anonymous tracking of a user as they browse between sites. Marketing firms use technologies such as that of the cookie— a file that stores data about a user—to provide businesses with potential customers to advertise their products to (Team, 2019). An average user is not told that their patterns are being tracked as they browse or that their data is profited from without consent (Patel, 2021). By tracking our paper trails as we browse websites on the internet, cookies—more specifically third-party cookies—collect increments of data about us in order to form a profile used for targeted advertising (CookieYes, 2022).

Increasing awareness of digital marketing malpractice resulted in a movement to advocate for digital privacy rights. Through government legislation such as the California Consumer Privacy Act (CCPA) in 2018 and most recently statements by tech giants *Apple* and *Google* to remove the ability to use third-party cookies from their platform browsers (Bergen, 2021), steps have been taken to actively uphold a user's right to the privacy of their online data.

While the recognition of a user's digital privacy online is an important step towards creating a more ethical internet, it begs the question as to if digital privacy rights would be defined to the extent that they are today had it not been for the controversial practices of the digital marketing industry. Furthermore, as the complete removal of third-party cookies from major browsers lurks on the horizon, it introduces a conversation of how we can better define online privacy rights to be proactive rather than reactive to further malpractices in data usage.

To dive into these topics, my STS project will involve researching the various techniques of data-usage and collection used in digital marketing for advertising today. Furthermore, I will conduct research into the various consumer-privacy laws that have been passed around the world to define digital privacy policies. This work will help to understand how online privacy came to be recognized today, and how legislation can assist in proactively protecting it from future data misuse. In my technical project, I will describe my experience optimizing a search engine for a consumer to browse products using ethical data practices. This work demonstrates how effective, fair data use can enhance a consumer's experience online without raising privacy concerns.

Technical Topic

After the best-performing quarter since launch, I decide to treat my employees to a getaway to plan for next quarter. I decide to take advantage of a third-party service to help plan

my event in order to relieve some of the stress. The process for making my selections works perfectly: I am able to search for different hotels in the destination of choice, view meeting-room spaces available in these hotels, and instantly make bookings through the service. However, just as I am checking out, I receive a notice that I am unable to book. All the time I invested browsing through my options is wasted, and I am forced to rethink my options. Any hindrance to a consumer's experience within an application can be frustrating and result in severe consequences on their likelihood of becoming a returning-user.

In the summer of 2022, event management solution company, Cvent, experienced a very similar issue with one of their products currently in development. Similar to the service described, Cvent's *Instant Book* product allows customers to search for hotels and instantly book meeting and guest rooms through their interface (Szymanski, 2022). As part of *Instant Book*, *Venue-Search* is the search engine that powers the ability to search for hotels in the process. When a search is performed, hotels and booking rates that match the selected search criteria are shown (Szymanski, 2022).

After testing this feature, however, Cvent noticed an unusual occurrence. Sometimes searches would disguise unbookable rates as bookable for users—an issue which could confuse and frustrate users upon trying to complete their booking. A deep dive into the issue revealed that invalid search results were an effect of missing data fields. In the third-party venue database used by Cvent, a set of venue-specific restrictions on the ability to book were available but not being utilized. Thus, searches did not take these restrictions into account and would display rates that did not pass them.

As an intern, I was tasked with spearheading the effort to ingest these new fields into the current data model for rates. I would then need to update the search engine logic to filter by these

criteria when returning venue rates. To do so, I had to perform back-end development in Java and work with *Elasticsearch* —the search engine framework that *Venue-Search* is built upon.

Performing this task involved gaining a rich understanding of the *Venue-Search* product, the structure of data stored in the venue database, and the *Elasticsearch* framework. The work that I did was written in the same codebase where the *Venue Search* product code was located, thus I often referred back to aspects of the *Venue Search* code when designing my solution.

I was able to successfully test my solution to the problem by running an instance of the *Venue Search* engine and ensuring that the correct results were obtained. I created dummy data: a fake rate that passed all of the restrictions for a given venue and a fake rate that did not pass all of the restrictions for a given venue. Running *Venue Search* on these data, I saw that the rate that passed the restrictions showed correctly and the other rate did not.

The main challenges I faced in designing my solution were understanding the new technologies that I had not previously worked with before: namely *Elasticsearch* and *Venue Search*. Furthermore, navigating the files that I needed to modify to create a working solution in a code repository as large as *Venue Search*'s was difficult in the beginning as well. Ultimately, however, after tackling these obstacles I was able to produce a correct working solution to Cvent's issue that will enhance the consumer experience of the *Venue Search* product.

This work relates to my STS topic due to the fact that both my technical topic and my STS topic surround the topic of data use. While my technical topic presents an ethical case of how data can be used to benefit the user experience, my STS topic dives into a case in which data use can harm a user's experience online by intruding on their privacy.

STS Topic

Advancements in data use and collection has curated personalized experiences with technology today. Smart technologies that learn from the interaction between its users have become commonplace through the many benefits they provide for the user experience. However, the detailed data required to fuel such technologies can often lead to significant conflicts with a user's privacy.

Our experience on the internet today is a prime example of such personalized experiences with technology. Today, cross-site tracking via the use of third-party cookies is one of the main ways advertisers are able to personalize ad recommendations online (Gozman, 2022). In a research study from Statista conducted on data from 2021, roughly 51% of marketing participants stated that “third-party cookies were very important for their current marketing strategy as they made up a majority of the data their company used” (Statista, 2022). In addition, methods of data collection like *deep packet inspection*, *history sniffing*, and *scraping* have also been commonplace practices of data collection assisting the formation of the profile made about a user for targeting (Christiansen, 2011). This data does not need permission to be collected.

In the 2021 fiscal year alone, Meta made \$115 billion in ad revenue (Lin, 2022) and Google made \$209.49 billion (Statista, 2022). Digital marketing has etched its mark as a major source of profit not only for businesses but for the big tech companies hosting their ads as well. With the profit opportunities in digital marketing fueling the digital economy today, threats to change the way that advertisers are able to target their customers raise concerns for all parties involved in digital marketing.

These threats loom on the horizon. As awareness of such malpractice in data collection spread, a movement for digital privacy rights on the internet began to take center-stage. Legislation such as the General Data Protection Regulation (2018) published in the European

Union defined a set of rights for the user online (Kant, 2021), while U.S. legislation such as California Consumer Privacy Act (2018) required businesses to inform users when their personal data is being collected (Gavejian et al, 2022). Key research analyzing the state of user privacy and marketing practices online also had substantial impact. Sarah West’s 2017 account of *data capitalism* brought to light to imbalance between the user and the “actors who have access and the capability to make sense of information” (West, 2017). In another instance, on the topic of third-party cookies, Reuben Binns and Elettra Bietti discuss the third party tracking industry and the “unique challenges for privacy and fundamental rights” that it has brought forth (Binns & Bietti, 2020). As calls for online privacy grew louder, *Apple* and *Google* made public claims that they plan to completely eradicate the use of third-party cookies on their browsers in upcoming years (Bergen, 2021).

In the past two decades, substantial progress has been made towards defining privacy rights for a user’s data online. However, given that there are currently no federal laws covering privacy of all types of data (Klosowski, 2021), there is still much more work to be done in the space of digital privacy. As the age of a “cookieless” internet approaches in an Internet subject to effects of *data capitalism*, digital marketers will find ways to navigate the new environment that may continue to use personal data without user knowledge. Understanding how privacy rights on the internet got to this point — the marketing technologies at play—will help determine whether defining privacy rights as a preemptive defense to the data misuse occurring was possible and if we can use preemptive digital privacy rights to defend against unethical data practices in the future.

In order to conduct this analysis, I will primarily rely on the social construction of technology (SCOT) theory developed by Trevor Pinch and Wiebe Bijker (1987). The SCOT

theory discusses how human action shapes technology, and further that the success of a technology is dependent on the relevant social groups that interact with it. I will apply a SCOT analysis in order to understand why certain marketing technologies came to serve a prominent role in data-collection based on the study of internet user behavior. This will help me determine both how and why marketers chose to use specific technologies for information gathering. Furthermore, I will use SCOT to see how social changes towards these technologies through the understanding of data misuse influenced marketing technology thereafter. Part of my analysis involves seeing how the recognition of digital privacy seen through legislation influenced marketing technology after the fact, thus SCOT will prove helpful in understanding the impact of changes in the social awareness of the internet user.

Research Question and Methods

The research question that I will be examining is: “How have online marketing practices defined rights to data privacy in the digital age?” The ramifications of analyzing this question are significant in determining why privacy rights developed in response to data malpractices by the digital marketing industry rather than as preemptive defense against them. As we navigate a new “cookieless” era of the internet, further data privacy violations can unfold as marketers harness the power of new technologies to collect personal data. Determining if there is a robust way to define the rights over our online data prior to such violations occurring will lead to an overall safer internet experience moving forward.

The methods utilized to perform my research will involve a literature review and a policy analysis. As part of my literature review, I will conduct research from business intelligence sources on the techniques used for user-data tracking catered towards business owners. I will analyze these sources in order to gain an understanding of the types of technologies that are used

for user data collection. Specifically, I will analyze web postings from different business intelligence groups that explain how companies can collect data on their target customers today. I will also conduct a historical analysis of digital marketing practices and calls for user privacy online. I will examine academic journals detailing the historical digital marketing practices on the internet and the responses to these practices in order to see how marketing techniques for data collection and calls for data protection changed before and after the legislation I will analyze was enacted. Furthermore, I will conduct a policy analysis. I will analyze four key legislations that have been passed in regards to defining user privacy rights online. These are the California Consumer Privacy Act (CCPA) of 2018, the General Data Protection Regulation (GDPR) of 2018, the 2020 EU Digital Services Act, and lastly the 2021 Draft Online Safety Bill. An analysis of these legislation will allow me to determine what online privacy rights have been defined, why they have been defined, and how these policies have affected digital marketing tactics today.

Conclusion

To conclude, I believe the research question is important to answer in order to understand if online privacy rights could serve their role better as a preemptive defense to data malpractice rather than as a reactionary measure. Determining how the online privacy rights we have today came to fruition is crucial in figuring out the best way to define our rights moving forward. As technology becomes more intelligent and data collection tactics become less transparent, having a clear set of privacy rights that define the extent to which our personal data can be used becomes increasingly important. I believe that the analysis conducted of the research question will reveal that digital marketing practices fostered the birth of digital privacy rights through the use of personal user data without permission, and furthermore that current legislation to handle

this misuse has proven effective in forcing marketers today to respect user privacy online, advocating for the establishment of clear, widespread digital privacy rights today.

References

- Barber, R. (2022, September 09). Online shopping statistics, facts & trends in 2022. Cloudwards. Retrieved October 27, 2022, from <https://www.cloudwards.net/online-shopping-statistics/#:~:text=Key%20Takeaways%3A,hit%20%245.5%20trillion%20in%202022.&text=Around%2076%25%20of%20U.S.%20adults%20shop%20online.&text=Over%20half%20of%20consumers%20prefer%20shopping%20in%20a%20physical%20store.&text=Smartphones%20are%20slowly%20becoming%20the%20preferred%20way%20to%20order%20online.>
- Bergen, M. (2021, April 26). When do third-party cookies end? Apple (AAPL) google (GOOG) ending practice. Bloomberg. Retrieved October 27, 2022, from <https://www.bloomberg.com/news/articles/2021-04-26/how-apple-google-are-killing-the-advertising-cookie-quicktake>
- Binns, R., & Bietti, E. (2020). Dissolving privacy, one merger at a time: Competition, data and third party tracking. *Computer Law & Security Review*, 36. doi:10.1016/j.clsr.2019.105369
- Can cookies identify you personally? (2022, July 05). CookieYes. Retrieved October 27, 2022, from <https://www.cookieyes.com/knowledge-base/cookies-101/can-cookies-identify-you-personally/>
- California Consumer Privacy Act, Cal. 1798.120. (2018)

Christiansen, L. (2011). Personal privacy and internet marketing: An impossible conflict or a marriage made in heaven? *Business Horizons*, 54(6), 509-514.

doi:10.1016/j.bushor.2011.06.002

Draft Online Safety Bill 2021.

EU Digital Services Act 2020.

Gavejian, J., Lazzarotti, J., Silver, D., & Paisan, S. (2022, April 29). California Consumer Privacy Act, California Privacy Rights Act faqs for covered businesses. Retrieved October 27, 2022, from <https://www.jacksonlewis.com/publication/california-consumer-privacy-act-california-privacy-rights-act-faqs-covered-businesses>

General Data Protection Regulation 2018.

Google: Advertising revenue 2021. (2022, July 27). Statista. Retrieved October 27, 2022, from <https://www.statista.com/statistics/266249/advertising-revenue-of-google/#:~:text=In%202021%2C%20Google%27s%20ad%20revenue,and%20apps>

Gozman, V. (2022, September 13). Council post: The slow death of third-party cookies. Forbes. Retrieved October 27, 2022, from <https://www.forbes.com/sites/theyec/2022/09/12/the-slow-death-of-third-party-cookies/?sh=5888cfdd4026>

How much privacy do you have online? (2022, April 12). University of Dayton. Retrieved October 27, 2022, from <https://onlinelaw.udayton.edu/resources/how-much-privacy-do-you-have-online/>

Internet safety: Understanding browser tracking. (n.d.). GCFGlobal. Retrieved October 27, 2022, from <https://edu.gcfglobal.org/en/internetsafety/understanding-browser-tracking/1/>

- Kant, T. (2021, October 24). A history of the data-tracked user. MIT Press. Retrieved October 27, 2022, from https://thereader.mitpress.mit.edu/a-history-of-the-data-tracked-user/#_edn15
- Klosowski, T. (2021, September 06). The State of Consumer Data Privacy Laws in the US (and why it matters). New York Times. Retrieved October 27, 2022, from <https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/>
- Lin, Y. (2022, August). Facebook ad revenue (2015–2022). Oberlo. Retrieved October 27, 2022, from <https://www.oberlo.com/statistics/facebook-ad-revenue#:~:text=In%20comparison%2C%20ad%20revenues%20from,97.2%25%20of%20its%20total%20revenue>
- Patel, V. (2021, November 2). Third party cookies are being blocked - what you need to know. Initialze. Retrieved October 27, 2022, from <https://www.initialze.com/blog/2021/10/third-party-cookies-are-being-blocked/>
- S. (2022, February 08). Reliance on cookies in advertising in the U.S. 2021. Statista. Retrieved October 27, 2022, from <https://www.statista.com/statistics/1222230/reliance-cookie-advertising-usa/>
- Schomakers, E., Lidynia, C., & Ziefle, M. (2021). The role of privacy in the acceptance of Smart Technologies: Applying the Privacy Calculus to Technology Acceptance. *International Journal of Human–Computer Interaction*, 38(13), 1276-1289.
doi:10.1080/10447318.2021.1994211
- Szymanski, B. (2022). CVENT instant book: Book simple meetings online. Cvent. Retrieved from <https://www.cvent.com/en/supplier-venue/instant-book>

Team, F. (2019, September 10). 4 important facts about the history of online privacy protection.

FigLeaf. Retrieved October 27, 2022, from <https://figleaf.com/blog/perspectives/4-important-facts-about-history-of-online-privacy-protection/>

West, S. M. (2017). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society*, 58(1), 20-41. doi:10.1177/0007650317718185