May 2023

“The Same Information Is Given to Everyone”: Algorithmic Awareness of Online Platforms

Meghan Lindsey Dowell
University of Wisconsin-Milwaukee

“THE SAME INFORMATION IS GIVEN TO EVERYONE”: ALGORITHMIC AWARENESS OF ONLINE PLATFORMS

by

Meghan Lindsey Dowell

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Information Studies

at

The University of Wisconsin-Milwaukee

May 2023

ABSTRACT

"THE SAME INFORMATION IS GIVEN TO EVERYONE": ALGORITHMIC AWARENESS OF ONLINE PLATFORMS

by

Meghan Lindsey Dowell

The University of Wisconsin-Milwaukee, 2023
Under the Supervision of Professor Nadine Kozak

After years of discourse surrounding the concept of "filter bubbles," information seekers still find themselves in echo chambers of their own thoughts and ideas. This study is an exploratory, mixed methods analysis of platform privacy/data policies and user awareness of the personal and usage data collected and user awareness of how platforms use this data to moderate and serve online content. Utilizing Bucher's (2018) framework to research algorithms through the black box heuristic, this project learns how users inform themselves about data collection and use policies, and their awareness of algorithmic curation. The algorithmic systems that return search results or populate newsfeeds are opaque, black boxed systems. In an attempt to open the black box, this dissertation analyzes the privacy and data policies of the top three platforms by traffic in the United States – Google, YouTube, and Facebook – to first learn how they describe their data collection practices and how they explain data usage. Then a cross-sectional survey provides user perception data about what personal data is collected about them and how that data is used, based on the privacy policy analysis.

The findings of this dissertation identify a need for algorithmic literacy and develop a new frame for the ACRL's Information Literacy Framework to address algorithmic systems in

information retrieval. Additionally, the findings draw attention to two subgroups of internet users – those who believe they do not use search engines and those who use only privacy-focused search engines. Both groups require additional research and demonstrate how online information retrieval is complicated through multiple points of access and unclear methods of information curation.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES ix

Chapter 1: Introduction 1

1.1 Background, Context, and Theoretical Framework 5

1.2 Problem Statement 10

1.3 Research Questions 11

1.4 Rationale, Relevance, and Significance of the Study 12

1.5 Structure of Dissertation 14

Chapter 2: Literature Review 15

2.1 Information Retrieval 16

2.2 Black Box Heuristic of Algorithmic Systems 30

2.3 Sociotechnical Systems of Information Retrieval 39

2.4 User Beliefs 47

2.5 Policy 51

2.6 Conclusion 64

Chapter 3: Methodology 68

3.1 Methodology 69

3.2 Data Collection 72

3.3 Data Analysis 76

3.4 Limitations 91

Chapter 4: Findings 92

4.1 Snapshot of Platform Privacy and Data Policies 92

4.2 Survey Instrument and Population Characteristics 98

4.3 Understanding Users' Awareness of Algorithmically Delivered Content 100

4.4 Usage Characteristics 106

4.5 Demographic Variables and Platform Knowledge 130

4.6 Conclusions 138

Chapter 5: Discussion 142

5.1 User Information Practices 143

5.2 User Awareness 145

5.3 Platform Privacy and Data Policies 153

5.4 Implications 157

Chapter 6: Conclusions 172

6.1 Overall Findings 173

6.2 Key Findings 175

6.3 Contributions 178

6.4 Limitations 180

6.5 Future Research 180

References 183

Appendix A: Qualtrics Documentation 207

Appendix B: Survey Instrument 208

Appendix C: Privacy Policy Analysis..... 264

LIST OF FIGURES

Figure 1: Yangyuen et al. Collaboration Perspective between User and System for Information Retrieval 21

LIST OF TABLES

Table 1: Research Question Matrix 68

Table 2: Policy Analysis Categories 77

Table 3: Policy Analysis Influences Coding Schema 78

Table 4: Policy Analysis Collected Coding Schema 82

Table 5: Variable Summary Table for Personal Information 89

Table 6: Information Type by Platform 93

Table 7: Google – Frquency of Influnces 95

Table 8: Google – Frequency of Personal Data Collection 95

Table 9: Facebook – Frequency of Influences 97

Table 10: Facebook – Frequency of Personal Data Collection 97

Table 11: Age 99

Table 12: Race 99

Table 13: Gender Identity 100

Table 14: Political Affiliation 100

Table 15: Education 100

Table 16: Overall Customization 101

Table 17: Actions Influencing Search Results 101

Table 18: Actions Influencing YouTube 102

Table 19: Actions Influencing Facebook 102

Table 20: Personal Data Collected by Google 103

Table 21: Personal Data Collected by YouTube 104

Table 22: Personal Data Collected by Facebook 105

Table 23: What is the purpose of your search? 105
Table 24: What is the purpose of your use of YouTube? 106
Table 25: What is the purpose of your use of Facebook? 106
Table 26: Crosstabulation of Search Engine Use Frequency by Awareness of Personal Information Influencing Search Results 108
Table 27: Search Engine Use 110
Table 28: Search Engine Access 111
Table 29: Google Policy Reading Habits 112
Table 30: Crosstabulation of Self-Reported Reading of Privacy Policy and Actions that Influence Search Results 112
Table 31: Crosstabulation of Self-Reported Reading of Privacy Policy and Personal Data Collected About the User – Google 114
Table 32: YouTube Policy Reading Habits 116
Table 33: Crosstabulation of Self-Reported Reading of Privacy Policy and Actions that Influence Video Results 117
Table 34: Crosstabulation of Self-Reported Reading of Privacy Policy and Personal Data Collected About the User – YouTube 119
Table 35: Facebook Policy Reading Habits 121
Table 36: Crosstabulations of Self-Reported Reading of Privacy Policy and Actions that Influence Facebook Newsfeed Results 121
Table 37: Crosstabulation of Self-Reported Reading of Privacy Policy and Personal Data Collected About the User – Facebook 123
Table 38: Search Engine Bias 125

Table 39: Do two people get the same search results if they entered the same search terms at the same time on Google? 126
Table 40: Why? 126
Table 41: Search Result Belief Frequency 127
Table 42: Source Reliability 128
Table 43: Chi-square test results examining the relationship between demographic variable and source of information reliability 136
Table 44: Search Result Belief Frequency 149
Table 45: Platform Comparison of “Actions taken by other users” 157
Table 46: Proposed Legislation 2019-2022 164
Table 47: Google & YouTube Policy Analysis 264
Table 48: Facebook Policy Analysis 268

ACKNOWLEDGMENTS

Many thanks and much gratitude are given to my chair and advisor, Dr. Nadine Kozak. Her expertise and guidance not only made me a better writer, but she also showed me how to be a great colleague. Additional thanks to my committee: Dr. Dietmar Wolfram, Dr. Iris Xie, Dr. Michael Zimmer, and Dr. Nicholas Proferes. Their feedback encouraged me to go further in this research and for that, it is better.

I could not have completed this journey without the support of writing partners, friends, and family. Maria, you have been a steady and encouraging virtual presence. My lovelies (Katie, Nicola, and Sandra) and biggest cheerleaders, thank you for the encouraging notes and giant bags of gummy worms. Mom, thank you for believing in me and knowing just how frequently to ask if I'm done yet. Finally, Alan – thank you – for loving me, for supporting me, for knowing I could finish even when I doubted myself.

Chapter 1: Introduction

When I first started this project search engines were the focus of my research. As an academic librarian I worked firsthand with students and individuals who were seemingly unaware of how and why they received the results they did while searching Google. But information retrieval is larger than Google as it is embedded into our lives though all platforms including Facebook and YouTube. This pervasiveness was evident during the 2016 and 2020 elections and the COVID-19 pandemic. Information seekers found themselves entrenched in highly polarized discourse on important topics because of how platforms curate information sources for delivery to them. As an information literacy professional, I wanted to learn more about how platforms inform users about the functionality of the platform to allow me to better instruct patrons about searching and interacting with these platforms.

Over the last decade the number of internet searches has doubled, and Google is averaging 3.5 billion queries a day, globally (Statista, 2021). YouTube is the second largest search engine with 3 billion searches per month and on average users viewing 30 minutes of videos per session (GMI Blogger, 2022; Statista, 2022). Facebook has over 1.9 billion daily active users (*Meta - Q1 2022 Earnings,* 2022). Individuals are spending time serendipitously finding and intentionally gathering information from these platforms that affect their daily lives. My work as a librarian, confirms anecdotally Pasquale's (2016) statement that, "users lack both the ability and incentive to detect manipulations as long as they are getting 'good enough' results," which leads to the mountain of evidence compiled by scholars and observers attributing beliefs and action to information found during internet searches (p. 83). I see this when teaching first-year undergraduate and graduate student information literacy sessions. Class begins with the question "where does information come from?" and without fail, one of the students' first

answers is Google. After a few more responses like parents, books, and the library, I come back to Google and ask, “what is Google”? After a few strange looks I hear “search engine” or “the internet”. We then discuss Google’s business model and the construction of search results, which leaves most of the students surprised. These experiences in the classroom, coupled with the proliferation of mis/disinformation over the last several years has led to my interest in learning about the public’s awareness of algorithmically delivered content.

One of the major topics of concern to information studies scholars is the relationship between information retrieval and information policy, specifically concerning algorithmic awareness and literacy. In recent years, a growing body of scholarship in this area has begun to attend to the societal impacts of algorithmic filtering in search results and newsfeeds. Other studies have developed algorithmic folk theories to give language to how platforms operate in everyday life. This previous work around algorithmic power, the oppressive nature of search algorithms, and user studies has laid the foundation for my research on the user awareness of how the collection of users’ personal information impacts their platform experience through algorithmic filtering. I chose one point of contact between the platform and user – the agreement to the Privacy policy as the information medium. To either confirm or rebut the idea that users and platforms are highly intertangled, there is a focus on the users’ awareness of how their personal information, stated in the privacy policies, influences their online experience. The findings illustrate users are unaware of how their personal information is used to curate their online experience which reinforces perceptions and previous research of algorithmic curation and provides a framework to support information literacy instructors.

What users might not know about information retrieval platforms is that they, according to Halavais, “distill the social behaviors of their users, the socially defined structure of the World

Wide Web, and our collective creation of knowledge" (2018, p. 73). The concept of filter bubbles is not new to information scholars; however, searchers are falling victim regularly to the echo chamber of their online lives (Powers, 2017). This study is an exploratory, mixed methods analysis of platform privacy/data policies and user awareness of the personal and usage data collected and user awareness of how the platforms use this data to moderate and serve online content. Centering what users know about how their personal data is collected and then used to influence their platform experience is not necessarily a privacy question, but one about the symbiotic relationships between the user and the platforms. Whether users know it or not, they are constructing their online experience and the algorithmically delivered content is not neutral or unbiased.

This project utilizes Bucher's (2018) framework to research algorithms through the black box heuristic. I use this framework to learn what users know about the black boxed algorithmic systems of Google, YouTube, and Facebook. As such, there are three guiding methodological tactics employed: disregard the epistemological limit of the "black box" and identify what is known (privacy policies), apply a phenomenological approach to the perception of algorithms through tacit or practical knowledge (other studies), and "interrogate the configurations of strategic unknowns" (survey data) (p. 63). In Bucher's (2018) work of analyzing black box research, she found that scholars were creating limits of what was researchable by adhering to the framework of the black box – that the contents of the black box are unknowable. However, through careful consideration of what is known about the black box, one can begin to parse together a shadow structure to identify what is known and unknown for research.

With this framework in mind, this study employs the following four guiding questions: What is the general public's awareness of what influences algorithmically delivered content?

How do users understand the personal data collected on them by the platforms? How does time spent on platforms impact awareness and knowledge? How do platform policies inform users about how algorithmically determined content is displayed and what personal data is used to influence the information? This research seeks to take a recent snapshot of the public's awareness of Google, YouTube, and Facebook and their processes of content delivery. The potential exists for the dataset to inform future research using interview questions or focus groups, in addition to the goals of this dissertation.

The contributions of this study are threefold. It builds on the critical algorithm studies work of Noble (2018), Bucher (2018, 2021), and Pasquale (2016) wherein they elaborate the harmful effects to society due to algorithmic oppression. I contribute to this work as I identify the common elements of what is unknown to the average user and elaborate on known issues around algorithmic curation of search results and newsfeeds, for example learning that users are aware about the personal data collected while using the platforms but unaware of how that data is used to shape their online experience. These findings further discuss how the act of black boxing of search algorithms compounds issues of information poverty and inhibits searchers' efforts to understand the information retrieval system. The searchers' perspectives on algorithmic awareness and filtering inform the work of information literacy professionals by updating the curriculum and content to break down the mysticism of algorithms and information retrieval. The final contribution identifies how current legislation does not address these issues of algorithmic literacy and would create further ambiguity on how platforms disseminate content.

As previously stated, this research builds on the work of critical algorithmic studies scholars who have identified major societal impacts of algorithms in search, but also the financial market, criminal justice system, and education among others (Benjamin, 2019; Bucher,

2018; Gillespie, 2018; Noble, 2018; Pasquale, 2016). These researchers have built a foundation for my research through digital ethnographies, case studies, policy evaluations, and interviews. I contribute to their work and knowledge base by adding generalized user awareness of algorithmically delivered content. Additionally, while investigating usage of the studied platforms, two subgroups of participants arose – one that believes they do not use search engines and the other who only uses privacy-focused search engines. Further connections are made between awareness and demographic characteristics to give personal context to these societal impacts and begin to understand how to address these serious issues through education and information policy.

This chapter provides an overview of the literature related to the research problem and how the black box theory provides the theoretical framework to the study. Then a brief overview of the research design and methodology that details how the results address the research problem. The chapter concludes with a roadmap of the dissertation chapters.

1.1 Background, Context, and Theoretical Framework

The importance of this research lies in understanding the public's awareness of curation in their search results and newsfeeds. The research narrows focus on content delivery algorithms, specifically the algorithmic filtering of results based on a hundred or so signals of personal information, including past searches, links clicked, contact lists, emails received, and geolocation information retrieval platforms collect on their users (Bucher, 2021; *Discover How Google Search Works*, 2020; Hillis et al., 2012; Pariser, 2011; Zimmer, 2008). The research is focused on the collection and use of personal data in algorithmic curation because of the perceived neutrality of these platforms when delivering information (van Dijck & Poell, 2013). This study focused on platforms with high traffic in the United States and employing algorithmic filtering to

their search results, autoplay, and newsfeeds; and platforms that have a precedent of algorithmic bias and disinformation. According to April 2021 data, and still true in 2023, these include Google (1), YouTube (2), and Facebook (3) (*Top 100: The Most Visited Websites in the US*, 2021, 2023). In addition to having high traffic in the United States, these platforms are the focus of congressional hearings, antitrust lawsuits, disinformation campaigns, and selective censorship of content (*Facebook, social media privacy, and the abuse of data*, 2018; *Complaint, United States v. Google LLC*, 2020; Jungherr & Schroeder, 2021). Furthermore, in the United States, as of December 2020, Google held 87.8% of the search engine market share; compared to 6.45% on Bing, 3.05% on Yahoo!, and 2.3% on DuckDuckGo (*Search Engine Market Share United States Of America*, 2020). However, it's broadly applicable as most search engines collect personal data on their users unless they have built their business model around privacy, such as DuckDuckGo and StartPage. To acknowledge the opaqueness of the algorithm, inherent consent, magnitude of power, and overall ubiquity of search and social media, the following literature contextualizes the theoretical approach to the dissertation.

The theoretical framework utilizes mutual constitution and black box theory. Mutual constitution provides the language for a symbiotic relationship between the user and the technology (Leonardi, 2009; Sawyer & Jarrahi, 2014). Sawyer & Jarrahi (2014) define mutual constitution for information systems as both humans and technologies having some sort of agency, but the actions are not deterministic. In information retrieval platforms, the human's agency is the choice to engage with the platform and how often it is used. For the platform this agency is within the algorithmic system and how it is programed to use the given information to curate or return information. The idea of mutual constitution describes the coevolution between the technology and how it is socially situated. Google's evolution from PageRank results to

including images to the Knowledge Pane is an example of how users and the platform are mutually constitutive. This allows for further investigation into how technology and society are intertwined and how user behavior shapes the technology. However, the user is not always aware of their impact, such as the "programmed sociality" of search engines and social media, meaning how algorithms affect one's social circle without explicit knowledge of the user (Bucher, 2018). Programmed sociality acknowledges the relationship between the human and the technology, but highlights the issues of users being unaware of how algorithmic systems curate information.

Ashby (1999) defines black box theory as an application to identify "systems whose internal mechanisms are not fully open to inspection" (p. 86), Bucher (2018) goes on to modernize saying, "the black box is not the exception but the norm" (p.59). Platform privacy/data policies explicitly state what personal information is used to curate the online experience for the user, both demonstrating social constructivism and making some effort at letting users peek inside the black box. Black boxing allows investigation into the unseen relationship between the human and nonhuman actors and by acknowledging that search firms protect their proprietary algorithms as trade secrets. Bijker (2001) describes this as the "technological frame" that "technology is constructed by a combination of enabling and constraining interactions" (p. 27). There are two sides of the debate surrounding black boxed algorithms, either the interpreted opaqueness of the algorithm creates a situation where the thing is unknowable, and only the societal impacts thereof are seen or the black box as a metaphor is dangerous while researching the algorithm because it prevents large and difficult questions from being asked (Bucher, 2018; Pasquale, 2016). This study acknowledges the usefulness of black box theory during the analysis of Google, YouTube, and Facebook' privacy/data policies to

investigate the clarity of their statements and also pushes it forward to pull apart the box by using Bucher's (2018) work to make the unknown known through the analysis of user knowledge.

Bucher (2018), employing black box theory, outlines three methodological approaches to "(un)knowing algorithms" in order to research them effectively. The first is to reverse engineer the known parts of how the algorithm operates using the example from ProPublica's work on criminal justice algorithms and the bias toward Black men (Angwin et al., 2016). The second is a phenomenological approach to learning the unknown through the perceptions of the algorithm. The third is examining the configuration or situated practice of algorithms not through the black box, but the process of black boxing. This dissertation employs the phenomenological approach of the perception of algorithms by the public and the practice of black boxing in the data and privacy policies. Algorithms are cultural artifacts that influence culture, or as Seaver suggests algorithms are culturally situated both *in* and *as* culture (2017, 2019). This shift in the perception of algorithms from benign calculators of data to cultural influencers creates the opportunity to discuss the effect Google, YouTube, and Facebook have on public opinion.

In the mid-20th century, the information age created an information-rich public. There are constraints on access to information and capabilities of the user. The working definition of information poverty in information studies is "that situation in which individuals and communities, within a given context, do not have the requisite skills, abilities, or material means to obtain efficient access to information, interpret it and apply it appropriately" (Britz, 2004; Marcella & Chowdhury, 2018). As Britz (2004) argues, information poverty is a social justice issue and if the users' search results are undergoing censorship through user-data driven filtering, then the user is not accessing all available information on the topic and cannot interpret fully the information presented. A just society requires access to information. The objective of this

research is to make progress toward defining facets of information poverty when connected to information retrieval platforms and algorithmic filtering. The First Amendment (FA) not only protects the freedom of speech, but also the freedom of inquiry to access needed information without interference from the government. However, in a review of First Amendment lawsuits brought against information retrieval platforms, courts found in favor of the search engine giving search engine speech FA protections (Balkin, 2018; Ballanco, 2013). Search engine speech includes the work of the algorithmic systems and the search engine result page. The idea of freedom of inquiry is at odds with algorithmically filtered search results as the searcher assumes all possible information is presented while the search engine has the right to curate the search engine results. There is a clash of rights. Users' rights to information access conflict with platforms own FA protections.

The FA lawsuits brought against YouTube and Facebook are Terms of Service disputes. These cases often are regarding the removal of posts due to user speech and/or perceived censorship, and not how the algorithm delivers content (*Loveland v. Facebook*, 2021). The First Amendment protects users from the United States government silencing their speech. Digital platforms have the right and ability to determine what speech is allowed and disallowed on their sites. Historically, if public protections to limit access to specific types of information needed to be put into place, Congress drafted new legislation, however, this was often difficult. Congress proposed and passed two bills, the Communications Decency Act (CDA) and the Child Online Protection Act (COPA) prior to the Children's Internet Protection Act (CIPA). While the intent of COPA is to prevent children from viewing obscene or harmful content on the internet, COPA never took effect because the Supreme Court found it unconstitutional due to the strength of the

First Amendment. CIPA passed when the legislation narrowed to include only public schools and libraries receiving federal funds for telecommunications access.

State and federal governments have begun to draft legislation that would ban or limit the use of personal information for algorithmically delivered content. In the fall of 2019, the U.S. Senate introduced a bipartisan bill titled the Filter Bubble Transparency Act. The bill's goal was "to require internet platforms give users the option to engage with a platform without being manipulated by algorithms driven by user-specific data" (S.2763, 2019). Democrats introduced a second bill in the summer of 2020 to the House of Representatives to amend the Federal Election Campaign Act of 1971 to change the regulation of targeted political advertisements using online behavior data and require platforms to maintain public records on the purchasing of political advertisements (H.R.7012, 2020). In January of 2022, Congress put forth the Banning Surveillance Advertising Act of 2022 that seeks to prevent digital advertisers, such as Facebook and Google, from using personal data to target advertisements (S.3520, 2022; H.R.6416, 2022). During the same month, Congress also proposed a bill to require platform governance to be more transparent through shortening the policies and making them easier to understand (S.3501, 2022; H.R.6407, 2022). The bills are examples of information policy that consider the information literacy of internet users and acknowledge the potential harm caused by algorithmic filtering. Even though none of the aforementioned bills made it out of committee; if Congress is beginning to prioritize algorithmic information policy, what do we know about their constituents' perceptions of how search platforms function?

1.2 Problem Statement

As search platforms became publicly available in the 1990s, scholars expected information would become more freely available and easily accessible (Lawrence, 2000).

However, as information retrieval platforms developed, and search algorithms evolved, how users received information became less clear to the user. Online marketing firms indicates American users of the internet do not understand the functionality of the search engine and subsequently the results they encounter (BrandVerity, 2020). In 2012, 66% of internet users found search engines a fair and unbiased source of information (Purcell et al., 2012). This idea that Google is neutral source or a "general consensus" of public opinion appeared in Tripodi's (2018) work on identity-based searching; interviewees reported using the search engine as a fact-checker and say, "the first information we see is what I'll remember and I'll keep with, and I'll assume it's true" (p. 28).

The majority of the literature in information studies surrounding algorithmic filtering has centered the technology and its effects as the method of inquiry through digital ethnographies and through the development of folk theories surrounding algorithmic curation. This project studies the knowledge of users of search and social media platforms, not the platforms' functionality. While scholars dedicate entire issues of journals to understanding the technicality of the algorithm and the power that drives the social consciousness of the algorithm; by digging into the opaque nature of search and social media platforms through policy analysis, this dissertation aims to gain an understanding of the users' awareness and knowledge of the platforms.

1.3 Research Questions

Building on what is known about the perceptions of algorithmic curation, including the influence Google, YouTube, and Facebook have over the distribution of information and ideas, this dissertation provides insight into what the public know about personal data collection and their awareness of how search results and newsfeeds are curated to reflect their worldview.

The exploratory mixed methods study seeks to answer the following research questions:

Research Question 1 (RQ1): How do platforms (Google (1), YouTube (2), Facebook (3)) govern and inform their users with regard to using collected data to filter the users' online content?

Research Question 2 (RQ2): What do United States-based adult users of these platforms know about the personal and usage data being collected and their awareness about how the platforms use this data to moderate and serve online content?

Research Question 3 (RQ3): Are there any associations between usage characteristics of United States-based adult users of these platforms and users' awareness of algorithmic curation using personal information?

Research Question 4 (RQ4): How does user awareness about algorithmically delivered content vary based on demographic characteristics, including age, race, education level, and political affiliation?

1.4 Rationale, Relevance, and Significance of the Study

The methodology of this dissertation, fully detailed in chapter three, is a mixed methods study employing content analysis of privacy/data policies and descriptive statistical analysis of cross-sectional survey data. The chosen platforms' policies were analyzed for how they address personal data collection and the impact of those data on search results or newsfeeds. The platforms' policies informed some of the survey questions to provide direct analysis of user awareness. The target population of this study is anyone living in the United States, over the age of 18, who used the internet in the last 30 days. The timeframe of 30 days includes those who use

the internet sporadically, however, because the distribution of the online survey requires the internet, those who do not use the internet are excluded.

The study uses a sample that is representative of demographic averages in the United States based on 2020 Census data. To achieve a representative sample, a panel was purchased from Qualtrics because they guarantee demographic averages and completeness of data to ensure the most generalizable results. The questions are a combination of Likert scale, yes or no, and short answer to best capture the awareness and knowledge of the searcher. Data analysis employs open coding to find patterns and outcomes of the user data and content of privacy/data policies and descriptive statistics to synthesize the survey results.

The novelty of this dissertation lies in centering the users' awareness of how platforms use personal information to deliver an online experience, their perceptions of algorithmic filtering on information retrieval platforms based on the privacy policies of the platforms, and connections between demographic characteristics. The findings contribute to the field by providing guidance to information and algorithmic literacy standards and through confirmation of previous studies that include algorithmic curation folk theories with quantitative data based on user awareness. Ultimately, the findings offer value to information policy makers, information literacy practitioners, and information and communication technology scholars. An informed public is required for a democratic society. As academic and public libraries provide information literacy instruction, users have access to increase their capabilities of finding and interpreting information as well as expand their critical thinking and analysis. However, by only centering on information literacy as the solution to algorithmic filtering and bias, we are relying on an individual solution to a systemic problem.

1.5 Structure of Dissertation

Chapter 2 presents a review of the current research from information studies and allied disciplines that center the dissertation by using SCOT and black box theory to understand the users' relationship with search engines. This chapter also provides a historical overview of information policy surrounding the public facing policy of the internet and search engines. Chapter 3 describes the methodology, research design, and procedures for data collection and analysis for the collected user data and public policy for this study. In chapter 4 I report the findings from both the content analysis of the privacy and data policies, as well as the survey questions. The survey findings are reported using descriptive and inferential statistics. Chapter 5 provides a discussion of the findings, and the theoretical, practical, and methodological implications of the study. Finally, chapter 6 concludes the dissertation with a summary of the key findings, contributions to the field, limitations, and opportunities for future research.

Chapter 2: Literature Review

A group of friends are standing on a street corner discussing their options for dinner. One pulls out their phone, searches “tacos” and five options appear within a ten-block radius. A family watches the news about local elections of city council members. The child opens their laptop to search the names of the candidates and finds a video of the city council debate. Three people are having a debate about a Black Lives Matter protest. One uses their phone to search for an article to prove their point, another does the same, the third cites an article a friend posted on Facebook. They have opposing results from their search engines that only support their point of view. People have multiple needs for information such as information for relocation, entertainment, health information, local civics, and others. The method of finding information online has become nearly universal; people’s understanding of algorithmically delivered content is less known. At the heart of this dissertation is the desire to gain a greater understanding of how users perceive the functions of information retrieval platforms and the subsequent results when seeking information. In recent years, legislators introduced multiple bills in Congress aimed at regulating search engines and social media platforms due to the algorithmic manipulation of search results, increasing transparency of policies, and banning surveillance advertising based on the collection of personal data (S.3520, 2022; S. 4066, 2020; S.2763, 2019). Ultimately, I seek to learn about platform users’ awareness and understanding of the algorithmic curation of results on information retrieval platforms: Google, YouTube, and Facebook. The study takes place during a time where the evening news, newspapers, and documentaries focus on search engine bias, the influence of social media, and the rampant spread of mis/disinformation (Kantayya, 2020; Orlowski, 2020; Nawaz, 2022). The dissertation captures public awareness regarding the role information retrieval platforms take in a democratic society.

The literature review provides contextual and historical frameworks to the research problem. Utilizing Ashby's (1999) definition of black box theory which examines how researchers understand a problem that cannot be seen or known, in this case algorithmic curation, combined with Bucher's (2018) approach provides the framework for this dissertation. Bucher argues that researchers should not let the black box stop us from approaching it to learn knowledge but should rather identify the parts that can be known and those that cannot. Then we should use the current beliefs and perceptions of the black box to build from and add to while, finally, interrogating the configurations of what is known of the black box (algorithmic systems) and how the users perceive the black box.

The structure of the literature review first acknowledges and analyzes the models of information seeking behaviors and practices which help provide a base for information poverty. Next is an exploration into the evolution of socio-technical systems of information retrieval and the multiple facets of internet searching to understand the construction of search engines and social media newsfeeds. This leads to a discussion of algorithmic systems through the lens of the black box heuristic. Folk theories of algorithmic curation provide the language for how the average internet user perceives the functionality of algorithmic systems and content delivery. And finally, how information policy has impacted how platforms operate and how users interact with the platforms. Throughout the chapter and dissertation, the discussion centers searchers, and searchers are people seeking information, either actively or passively, regardless of platform.

2.1 Information Retrieval

The complex processes of information retrieval come together in three facets: the information need, the information searcher, and the information environment (Knight & Spink, 2008). The interdependence of the three facets creates an unlimited number of variables when

considering information retrieval. The user's cognitive processes inform the information need and the environment of their search. For the purposes of this research, Google, YouTube, and Facebook are the information environments even though due to algorithmic filtering, each user's environment is slightly different. As we approach the black boxed information retrieval platform, we need to identify what is known about the process. Information behavior models provide a starting point for how searchers approach a problem and the iterative process that follows.

2.1.1 Information Needs, Practices, and Retrieval

Research or information seeking begins with an information need. This section outlines several information seeking models and aligns them with user-oriented information retrieval platform research. It is easy to oversimplify information seeking behavior when the information system is accessible at your desk, in your pocket, or on your wrist. There are multiple factors to consider when a person has an information need and while some are visible, others are cognitive and not as easily identified. To have a full understanding of users' perspectives of the algorithmic filtering of search engine results or newsfeed pages, it is necessary to conduct a more granular investigation of information behavior models. I used Xie (2008, 2010) and Knight and Spink's (2008) work on web search to identify the following information behavior models. The purpose for the investigation into information seeking models is to identify the missing contributors to web search and algorithmic curation. The following models do not specifically acknowledge the multiple factors and data points that construct search results or newsfeeds, only that there is an information system. The behavior of the system is just as important as the behavior of the information seeker.

Wilson: Model of Information Behavior

Wilson's first model of information behavior, developed in 1981, focuses on the user and their information need which then determines the information seeking behavior. The user has two choices of information seeking behavior, making demands on an information system or demands on an information source; this model lacks, however, an explanation of how the user interacts with the system or the source. Today, what might be defined as a system (search engine, library discovery layer, application, social media) was not well defined or delineated from a human information source (Knight & Spink, 2008; Wilson, 1981; Xie, 2008). The aspect of the searcher's needs is well-defined in Wilson's model as he takes into consideration the physiological, affective, and cognitive needs within their social role; as well as addressing the personal and environmental barriers to information seeking behavior (Wilson, 1981).

Kuhlthau: Information Search Process

This model describes the user's perspective of the information search process (ISP) and focuses on the "activity of finding meaning from information" (Kuhlthau, 1991, p. 361). Any one of these experiences informs the other and moves the search along the six stages of information seeking. ISP investigates the users' cognitive behaviors before, during, and after the search process to better understand how searchers identify their information need, locate sources, and use the information found. Through the analysis of several other studies, Kuhlthau's (1991) ISP model builds a framework of a logical sequence to information seeking behavior. The framework outlines the feelings, thoughts, actions, and tasks the searcher experiences. Selection identifies the topic to be searched and the presumed location of that information – an internet search, a library visit, or a community inquiry. Exploration requires the searcher to make sense of, and evaluate, the information found that can leave the searcher confused or frustrated. Formulation leads to clarity for the searcher when their feelings of confusion make way to feelings of

confidence and understanding. Collection of new information leads to the expansion of the topic and the further building of confidence in the search process and their information consumption. Presentation is the culmination of their search process and is generally met with relief, satisfaction, or the summarization of their findings (Kuhlthau, 1991). While the six stages are sequential, they are not linear. Users often find themselves in a cycle within the stages before finalizing their search process in an iterative and interactive search process (Bates, 1989; Knight & Spink, 2008; Xie, 2010). Kuhlthau's model lacks the recursive behavior of a searcher when stages need to be repeated as their feelings, thoughts, actions, and tasks interact with their information need and how that informs their future information seeking decisions. Kuhlthau's research centers the searcher and their cognitive abilities. Cognitive abilities are not a direct correlation with intellectual ability. The information need influences the cognitive ability, and the cognitive ability influences the information seeking behavior. Searchers react differently when seeking information that is sensitive or worrisome due to privacy issues, like health information, rather than less sensitive information, like information relevant to a trip (Libert, 2015). It is the searcher's abilities that create variance in the information retrieval process and the reaction to information received.

Bates: Berrypicking Model

The berrypicking model improves on Kuhlthau's model by recognizing the user has a cognitive response to each query and source, including the revisions needed to queries to satisfy their search (Bates, 1989). This model captures interactive IR approaches in the most realistic way. Interactive information retrieval recognizes the current information needs of the user might change as they process information and utilize additional information retrieval processes to achieve their goal. There are four elements of actual user-centered information retrieval captured

by the berrypicking approach: the evolution of search queries, information is sought in pieces rather as a whole, searchers employ multiple search techniques, and access a variety of sources (Bates, 1989; Xie, 2008).

Belkin, Oddy, and Brooks: Anomalous States of Knowledge

The Belkin, Oddy, and Brooks (1982) study on anomalous states of knowledge (ASK) found that information retrieval is not often a linear progression of ideas to source materials. The anomalous state is created by having a problem where a person's known knowledge cannot be applied, the cognitive dissonance creates a lack of trust in their information seeking processes and does not allow them to adequately explain their information need (Xie, 2008). The process of ASK information retrieval is that the searcher describes the information they are looking for, rather than identifying a specific information need. The information found then contributes to ASK until the searcher identifies the piece of information they need for a satisfactory search.

Knight and Spink: Theoretically based, Contextual, Macro Model

Knight and Spink (2008) move the discussion forward as they developed a model for web-based information retrieval behavior. This model combines theoretical models (some previously mentioned) and claims that the user's information needs, and cognitive style, impacts the search strategies. They do not equate cognitive style with intellectual ability, rather with the preferred methods of the user at the time of the need (Knight & Spink, 2008). Additionally, what this model addresses that the others do not is the system feedback when choosing information seeking where the searcher is browsing or navigating interactions. However, this model does not identify system feedback during the information searching behavior as the search formulates a

query. One reason for this might be that Google had not fully launched autocomplete to both Google.com and their apps until 2008. Prior to this, autocomplete was an opt-in feature and not widely used (Garber, 2013).

The previous IR models indicate the multiple pathways searchers take when seeking new information. The cognitive state, information need, and choice of strategies the searcher undertakes inform how searchers use and understand search engines. When search strategies are limited to few or single information systems, the cognitive state of the searcher remains more confident, and trust is built with that system. However, none identify how the system influences what the user sees in their search results or newsfeeds.

Yangyuen et al.: Collaboration Perspective between User and System for Information Retrieval

The collaboration perspective model incorporates how the user and system interact for information is the most comprehensive IR model (Yangyuen et al., 2020). It clearly indicates some of the information systems use to curate and display the results.

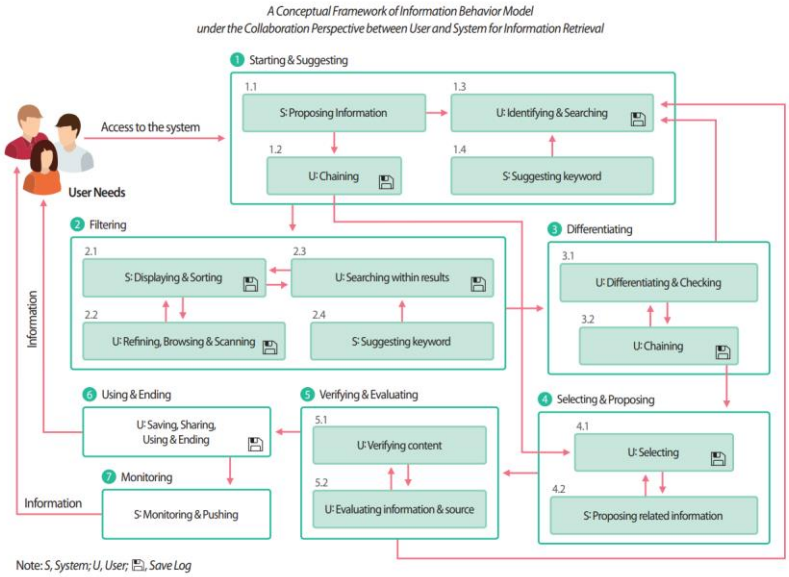


Figure 1: Yangyuen et al. Collaboration Perspective between User and System for Information Retrieval

The model moves beyond the unidirectional information retrieval process between the user and system and incorporates real-time actions of the system to curate and personalize search results. The inclusion of system-suggested keywords and the chaining of information is a more nuanced model of IR. The authors of this model made large improvements to information behavior models, and acknowledge more work is needed to incorporate the users' personal information into how the system displays and sorts. The next section outlines how information poverty impacts a searcher's ability to find and accept new information and how trust in information retrieval systems, specifically search engines, contributes to the cognitive abilities of the searcher.

These IR models, including the Collaboration Perspective, perpetuate viewing the systems as black boxes. The focus is placed on how the user determines their search activity and words or phrases, not how the system processes those queries which ultimately leads to the user not knowing why they need to modify their search or how the information is presented to them.

2.1.2 Information Poverty

The ability to access, understand, and use information is powerful. Childers' (1975) foundational work, *The Information-Poor in America*, examines the many ways citizens are information poor. He identifies groups of marginalized and disadvantaged people and links how different information needs require a different set of skills to navigate the information available. The study identified the information poor as people of color, economically poor, geographically rural, the elderly, incarcerated, and disabled. Those experiencing information poverty can identify with one or multiple groups. The areas of information need that tend to compound the experience of information poverty, according to the study, include medical, household, consumer, housing, employment, welfare, legal, political, transportation, education, and

recreation questions. The 1970s information landscape looked quite different from today where those in need of information either had to call, travel, or send a letter to a local library or government agency to inquire about their information need and formalized information literacy education was not commonplace in K-12 schooling. Information seeking generally involved speaking with another person about a potentially private or sensitive issue. If the person did not travel or lived in a rural area, they turned to their community to locate a solution or answer.

Chatman (1996) examines the inner workings of community information sharing through the development of six propositions of information poverty. First, people who are information poor perceive themselves to be devoid of any sources that might help them. Second, information poverty is partially associated with class distinction. That is, the condition of information poverty is influenced by outsiders who withhold privileged access to information and the perception of the information poor that information is not freely available. Third, information poverty is determined by self-protective behaviors that are in response to social norms. Fourth, both secrecy and deception are self-protecting mechanisms due to a sense of mistrust regarding the interest or ability of others to provide useful information. Fifth, a decision to risk exposure about our true problems is often not taken due to a perception that negative consequences outweigh benefits. And sixth, the information poor selectively introduce new information into their community. A condition that influences this process is the relevance of that information in response to everyday problems and concerns (Chatman, 1996, pp. 197–198). The prior propositions are in the context of interpersonal relationships and choosing who to trust, creating what she called their information world. As access to the internet became more freely available both in the library and at home, community information sharing widened to include information from search platforms

and social media, and as such the definition of information poverty evolved. It also created a need for individuals to become more information literate.

The Information Literacy Competency Standards for Higher Education, finalized in 2000, defined information literacy as the ability to "recognize when information is needed and have the ability to locate, evaluate, and use effectively the needed information" (ACRL, 2000, p. 2). Discussing information poverty as a social justice issue, Britz defines the phenomenon as a "situation in which individuals and communities, within a given context, do not have the requisite skills, abilities or material means to obtain efficient access to information, interpret it and apply it appropriately" (2004, p. 194). Two similarly defined phenomenon take care to draw attention to the skills required to be information literate. Without access and the capability to understand and use information, the information poor remain unable to navigate societal issues. Additionally, even as the information landscape changed to include the internet, access to reliable, fast internet is still lacking in 25% of American homes and 17% of Americans use their smartphone as their primary internet access at home ("Demographics of Internet and Home Broadband Usage in the United States," 2019).

Information-based rights are needed for "the successful implementation and protection of our civil, political, cultural, economic and social rights" (Britz et al., 2012, p. 106). These rights include access to the internet, free software, open educational resources, and open access scholarship. Access must be combined with educational opportunities to learn about the information that people are accessing, and the differences between types of information available and how they are produced. At times, this also means the reeducation of searchers who are taught and trained by instructors, family, or friends into thinking one way about an information source without critical analysis and understanding of that source. For example, after a discussion I had

with first-year undergraduate students about Wikipedia and explaining the protocols in place for editing and contributing, I asked, “what did you learn” and they anonymously respond by writing on slips of paper. One student responded with, “thanks for telling me Wikipedia isn’t the devil’s search engine.” A ten-minute conversation provided the opportunity to build on the student’s capability to evaluate information sources by illuminating facets of an information source they were already familiar with. This definition is explored further through Sen’s capability approach after accessing information. In this argument is the idea that access to information is not enough to overcome information poverty, the user must have the capability to use and understand the information (Britz et al., 2012).

The definition of information poverty and information literacy centers the person with the information need as the solution to information poverty. They all but overlook the infrastructure that upholds the systematic issues that cause information poverty, because becoming information rich is not only understanding the information it is also understanding how the information is delivered during the retrieval process. The definition of information literacy became more nuanced in 2015 as “a set of integrated abilities encompassing the reflective discovery of information, the understanding of how information is produced and valued, and the use of information in creating new knowledge and participating ethically in learning communities” when the ACRL approved the new Framework for Information Literacy for Higher Education (Framework), built of six frames with accompanying knowledge practices and dispositions (ACRL, 2015, p. 8). The six frames are “authority is constructed and contextual,” “information creation as a process,” “information has value,” “research as inquiry,” “scholarship as conversation,” and “searching as strategic exploration” (ACRL, 2015). The Framework received both praise and criticism due to the less prescriptive nature of the language. The authors of the

Framework provided a more critical lens of the information landscape and allowed for more customized approaches to teaching information literacy. Two of the six frames, "information creations as a process" and "information has value" acknowledge the systems in place by focusing on how the platform indicates the type of information available, however, it still places the onus of understanding a complex system of information production and retrieval on the searcher. If the goal is to eradicate information poverty, then focus needs to also be placed on the structures that keep it in place, not only providing instruction to improve the searchers' capabilities. Public libraries also provide structured instruction; however, they generally focus more on digital literacy (how to access and use technology) rather than information literacy (how to access, evaluate, and use information) (Public Library Association, n.d.).

An implication of algorithmic filtering is information poverty. The act of filtering the search results and newsfeeds removes access to information without the knowledge of the searcher (Dutton et al., 2019). In library and information studies a solution to information poverty is through information literacy instruction. Determining how users critically evaluate search engine results pages (SERP) gives background on how information literacy can attempt to address the skills needed to have a critical eye. In a recent study to learn how algorithms impact the selections of users searching for health-related queries, Bakke (2020) found that the order of the search results influenced which source searchers chose more than the source itself.

Medical videos on YouTube receive attention due to the ease of access to information about sensitive topics. However, these videos require a considerable amount of evaluation to not endanger the user or their acquaintances. In the United States, people often seek medical information on YouTube due to a lack of access to or affordability of healthcare. During the COVID-19 pandemic, the need for accurate medical information increased. In a comprehensive

study of YouTube videos about COVID-19 prevention and treatments, researchers found 55% of the sampled videos have very poor or poor quality and 40% of the videos to be misleading (Sawant et al., 2021). As these videos are algorithmically delivered based on users' activity on and off the platform, the video results provided might not be from public health officials, but rather those unqualified to administer such advice.

The work to recognize one's own information poverty is difficult because you cannot know what you do not know. However, there are feelings and issues connected with information poverty that users might be able to recognize, even if they were not aware as to why they were feeling them. Early definitions of information poverty developed from Childers (1975) work around the information needs of historically excluded populations. Chatman further developed the definition of information poverty to include how groups of people searched for and exchanged information and sources. Britz worked to highlight the social injustice of information poverty through lack of resources and skill to use and apply information. Even as definitions of information poverty and information literacy change, access to information stays at the center of the problem, both the infrastructure to access and the understanding of what the user is accessing and how they access it.

2.1.3 Algorithmic Literacy

Additional criticism of the Framework identifies the lack of specific language to attend to algorithmic literacy (Brenneur-Garel, 2021). Koenig's (2020) algorithmic literacy framework incorporates three levels of practice: "basic understanding of how algorithms function, how they are used to reinforce traditional power structures, and how humans can recognize and act upon their own agency when interacting with algorithmic" systems (p. 3). Algorithmic literacy is less

about the technicality of the mathematical or computer code, but recognizing how their functionality affects individuals, groups, and culture.

Oeldorf-Hirsch and Neubaum (2022) situate the human-algorithm interaction into three categories: knowing, feeling, and doing. To know about an algorithm is to be aware of its presence and the impact it has on information retrieval. Users can feel algorithms through first-hand experience and either having positive or negative reactions to the outcome of the interaction with the platform. The researchers describe *doing algorithms* as the assessment work of analyzing algorithmic literacy outcomes. For users to be more knowledgable about algorithmic systems they first must be curious about what algorithms do and why, it is not enough to just be aware of their existence. Further, users need to be motivated to engage with algorithms, to understand them better and to investigate why they receive the results or posts they do (Oeldorf-Hirsch & Neubaum, 2022). Users used to feel or notice something about their search results or newsfeed before they were aware of the system behind constructing their experience. Now, "the algorithm" has become commonplace vernacular for explaining why friends' photos are not as frequent in a newsfeed or why the same advertisements are seemingly on every platform.

Platforms share few details about how their algorithmic systems work, leaving users to learn through direct experience, discussing with friends and colleagues, or formal education (DeVito et al., 2018). We know that algorithmic literacy has predictable gaps in knowledge when looking at users through a socioeconomic lens which leads to a greater need for public education (Cotter & Reisdorf, 2020). Working to close those gaps, in January 2023 New Jersey became the first state to require information literacy instruction in kindergarten through grade 12 (Sitrin, 2023). The New Jersey curriculum reflects much of the ACRL Framework and, as currently written, lacks specific algorithmic literacy (NJ S.B. 588 & NJ A.B. 4169).

2.1.4 Trusting the System

Broadly defined, trust is the relationship between one party to rely on another party when risk or loss is involved, and in online platforms when trust is established parties share more information (Buchanan & Benson, 2019). Studies found users' trust in search engines waiver between high to undecided based on the ranking of recognizable links and the content of their query (Haas & Unkel, 2017; Nakamura et al., 2007). In 2012, 73% of search engine users in the United States said the information found on search engines is accurate and trustworthy (Purcell et al., 2012). Building on a 2007 study, researchers replicated eye-tracking methodology to identify the change in trust users found in Google's SERP (Pan et al., 2007; Schultheiß et al., 2018). The 2018 study found the trust of users declined over 10 years. However, the findings of Schultheiß et al. differed from Pan et al. in that while the viewing, or eye-tracking, of the source is influenced by the position of the result, the click factor is influenced by the relevance, not the position on the SERP (2007; 2018). These studies equate trust with viewing time and click rates which are difficult to differentiate from close reading while evaluating a claim or site information. As users become more "code-dependent" on information retrieval platforms and social media, they find it difficult to image a world without the invisible assistance (Raine & Anderson, 2017). The dependency on algorithmically curated information creates a sense of trust between the user and the platform. Users begin to feel that the platform *knows* more about their request or interests than they do themselves.

Trust in a traditional social setting relies on equal give and take from at least one other person, a feeling of vulnerability, and the context of the relationship. Whereas online trust is a feeling of confidence that the platform does not exploit the user's vulnerabilities (Bauman & Bachmann, 2017). In a multidisciplinary review of literature, researchers found there are social

and technological factors of online trust. The most common social factors that influence trust are word of mouth, social presence, and culture. The technological factors included perception of privacy and website design (Bauman & Bachmann, 2017). Trust is not only about the feeling of security when interacting with the platform, but also how the user feels about the SERP or order of newsfeed or autoplay and how the cultural relevance of the platform. The cultural dominance of Google is well documented in *The Googlization of Everything* (Vaidhyanathan, 2011) and *Google and the Culture of Search* (Hillis et al., 2012); these researchers explain the history of how Google rose to cultural dominance and the social, political, and economic impacts on society. Google relies on its social and cultural dominance to maintain trust in the search engine.

Facebook is unique in that it combines the trust of social setting with that of an online environment. Facebook's users are more likely to trust and share information on the platform if a trustworthy friend shared, liked, or commented on the post (Buchanan & Benson, 2019). However, Facebook users are just as susceptible to the modification of trust-based relationships due to network influences. The interactions between users, posts, and groups add an additional layer of communication with people outside of the user's known friend group due to post interactions (reactions, comments, shares) because a Facebook user does not only experience influence from their chosen contacts but from the larger network of connections and groups, furthering the concern of trust during passive information consumption (Lanel & Jayawardena, 2020).

2.2 Black Box Heuristic of Algorithmic Systems

When Ashby (1956) explored Black Box theory in *An Introduction to Cybernetics* he recognized that black boxes exist all around us and provided guiding questions when researching the processes within them. These questions ask how the researcher should proceed when met

with a black box, what contents of a box are discoverable and what are fundamentally undiscoverable, and what methods should be used when researching black boxes (Ashby, 1956). Ultimately black box theory applies to a process where the inputs and outputs are known, but the construction of the output is unknown. With the frequency black boxes appear in everyday life, Ashby emphasizes they are the norm not the exception. The black box heuristic is not a deterministic obstacle, but the lens with which to investigate the unknown.

For many years, to investigate the black box meant attempting to understand the physical components of technology (Von Hilgers, 2011). Disassembling the black box, by taking apart the components, or becoming an expert in the field are methods to understand the seemingly unknowable technology (Winner, 1993). Today's black boxed technologies are not as tangible as pieces and parts to a transistor, and they are protected through intellectual property (IP) rights, but they have more power over our lives than ever before because they are making banking and housing decisions on behalf of humans. Decision-making algorithms exist in all parts of daily life from a simple web search to seeing credit card offers from a bank (Pasquale, 2016). The investigation into black boxed algorithms is challenging because of the IP rights and the inability to fully understand the interconnectedness of the platform's algorithmic systems.

Seaver (2019) expands the understanding of algorithms in critical algorithmic studies by recognizing there is not one algorithm at a time on a platform. Platforms contain largely networked algorithmic systems that reject the technical/cultural divide because "algorithms are not simply means of interpreting culture, they are productive of culture" (Bucher, 2018, p. 150; Seaver, 2019). Seaver argues that making algorithms transparent does not solve the problem of making them known. Knowledge of the process does not equate with the cultural consequences

of the algorithm. The oversimplification of the Google algorithm applies a more deterministic understanding of the black box approach (Seaver, 2019).

Regarding algorithms, Bucher (2018) answers Ashby's question about effective methodologies when researching the contents of a black box by recommending reverse engineering of what is known of the unknown. This type of research is demonstrated in tracking outcomes of algorithmic processes; one such example tracks the number of male or female applicants recommended to the next round in hiring processes (Kleinberg et al., 2018). Another methodological tactic is a phenomenological approach to unknown knowns. This method highlights tacit and experiential knowledge of how algorithms function. Bucher's final method asks the researcher to interrogate the configuration of the unknown through the work of identifying contradictions between use of the platform and what is known through document analysis. These contractions are what the user assumes to be true of an algorithmic system after use, not necessarily the precise technological function.

Seaver (2019) reiterates facets of information poverty when discussing the construction of algorithms by labelling the developers as the insiders and those interacting with them as outsiders. Applying the idea of an algorithmic system changes the dichotomy of insider and outsider as both are applicable in the cultural understanding of the algorithmic system. Furthering this idea is Hunt and McKelvey's (2019) work by aligning algorithms with a form of cultural policy, who argue that algorithms regulate to "increase social isolation and diminish public culture by restricting the salience of cultural expressions" (p. 313). Social isolation takes on the form of an echo chamber which are the effects of algorithmic systems curating online content for the users.

2.2.1 Social Power of Algorithms

The social power of algorithms alludes to the infiltration of algorithms and algorithmic systems into everyday life through automated decision-making processes in organizations, institutions, and governments and the politics of algorithmic sorting, ordering, and predicting (Beer, 2017). Sorting and ordering require the power to name the structure and system; naming information "is not simply representation of information but is also the construction of that information" (Bowker & Star, 2000; Olson, 2002, p. 6). This power comes with inherent biases of both the searcher and developer. Examples of naming bias in library and information science are in subject headings and classification systems; the language marginalizes searchers and makes browsing difficult (Olson, 2001). Search engines work to match search terms given by the user to find relevant results, but if they cannot find matches, they infer meanings through natural language process to those words constructed by the developers; the developers hold the power to build ontological dictionaries of implied meanings of search terms (Brock & Shepherd, 2016).

Algorithmic systems work as procedural enthymemes to persuade users to trust the logic of the platform; an example is how most users of Google believe the first page of search results holds the most relevant or important information introducing a sense of greater agency to the platform by the user (Brock & Shepherd, 2016). Algorithmic power manifests through what Pasquale (2016) describes as the black box society where "authority is increasingly expressed algorithmically" (p.8). Those holding power rely on the difficulty people have understanding the algorithmic system to maintain the status quo. As algorithmic systems become responsible for decision making, the power and persuasion of the firms continues to influence the culture of the users (Beer, 2017).

Mis/disinformation spreads quickly online through media manipulation of trusted platforms. While the algorithmic systems are black boxed, groups have identified methods of gamifying content distribution through the use of bots (Walsh, 2018). For example, over a six-month period Facebook Pages known for posting misinformation received six times more clicks than Pages known for posting trustworthy information because of the practices used by Edelson et al. (2021).

2.2.2 Mutual Constitution of Algorithmic Filtering

Mutual constitution is the premise that users and platforms both have agency and that they are reliant on each other for optimal functionality. The reliance is not deterministic, it is part of the co-evolution of the platform (Sawyer & Jarrahi, 2014). The needs of the platform and user are intertwined, breaking the dichotomy between early information retrieval as determinant on the system or user (Leonardi, 2009). This reliance creates systems that are not opaque to the user, even though they are providing the platform with information needed to perform well.

There is a rich body of scholarship that explores the consequences when social processes are reliant on algorithm-driven information systems (Benjamin, 2019; Bucher, 2018; Gillespie, 2014, 2017; Noble, 2018; O'Neil, 2016; Pasquale, 2016; Robertson et al., 2018; Seaver, 2019). Search algorithms and the subsequent search results are influenced in multiple ways, two of these are through user data (which links users click, how long users stay on the page) and website metadata (linking to or from a page, number of times a keyword appears on the website). An example of these at work happened in 2003 when Dan Savage responded to Senator Rick Santorum's controversial comments about homosexuality. After Savage authored a *New York Times* op-ed, his readers suggested a contest to name a sex act after Santorum. Once decided, Savage purchased the domains santorum.com and spreadingsantorum.com. Due to the naming of

the websites, the metadata and content on the website, and the linking from multiple news outlets and bloggers spreadingsantorum.com was soon the top Google result when searching "Santorum" (Gillespie, 2017). This was a problem, not only for Santorum who was running for re-election but also Google to down rank a website highly ranked in search results due to algorithmic filtering. A search in January 2021 showed spreadingsantorum.com as the fifth result, below the fold, with the query "santorum sex act."

The Savage/Santorum example is that of purposeful influence on search results using algorithmic outcomes of search results. However, there are other examples of passive influence over search results through inherent biases of the developers and engineers. Another example of Google modifying their algorithm is provided in Safiya Umoja Noble's (2018) book *Algorithms of Oppression.* She identifies multiple ways search engines uphold racism and misogyny in their results. What initiates the research is a simple search for "Black girls" that returns results primarily containing pornography instead of activities or content for Black girls. In contrast, an image search of "beautiful" displays not nature or art, but scantily clad white women (p. 22). Upon finding these results, Noble (2012) wrote a piece for *bitch* magazine that outlined how women, especially women of color, are represented in search results. Due to this article and continued pressure through social media, Google ultimately changed the algorithm to down rank pornographic sites when searching "Black girls," while espousing the neutrality of their search algorithms (Crane, 2014; Noble, 2018).

The type of algorithmic filtering that bolsters the searcher's confirmation bias is referred to as a filter bubble (Pariser, 2011). Filter bubbles develop when search platforms use personalization signals to enhance the users' experience. On the surface, personalization signals are convenient and provide topical sources for the searcher's location or previous searches.

Selection power in information retrieval assumes that the user has the final selection from the search results, but when filter bubbles remove opposing viewpoints or fail1 to challenge the searcher with additional sources the final selection is given to the algorithmic systems (Warner, 2007). What the user views as convenience or excellent results is personalization. Halavais (2018) posits search has infiltrated society through the ubiquity of the service and how search platforms have situated themselves as indispensable through the personalization algorithms that focus on "the individual and their own networked connections" rather than the "place, community, or topic;" including social media (p. 144).

Sociotechnical systems like search and social media require data inputs to deliver their results, especially those that are personalized. As such, datafication of the user needs to be explored. Datafication is "the requirement, not just the possibility, that every variation in the texture of human experience be translated into data for counting and processing" (Couldry, 2019). The black box heuristic not only applies to the process of curating results and newsfeeds, but also the data that feeds the process. However, as we see in chapter four, platforms remove some of the opacity when explaining what information they collect on their users to "deliver our Products, including to personalize features and content (including your ads, Facebook News Feed, Instagram Feed, and Instagram Stories)" (*Facebook Data Policy*, 2021) and "to deliver our services, like processing the terms you search for in order to return results" (Google, 2021). Datafication of people in a digital world often replicates how the physical world excludes and marginalizes populations instead of democratizing them (Dencik et al., 2018). This is relevant and highly problematic as users see search engines as an unbiased form of information (Purcell et al., 2012).

While the datafication of users occurs, the opposite is also true of the platforms they are using. Data voids are the phenomenon where a query does not have enough associated results. Data voids can occur during breaking news, new or old terminology, fragmented concepts, and problematic queries (Golebiewski & boyd, 2019). When media manipulators identify data voids they work to associate specific websites with the queries or keywords. Search engine Bing has long had a problem with data voids resulting in misinformation in their search results (Bush & Zaheer, 2019). During the COVID-19 vaccine roll-out, vaccine conspiracy theorists were looking for information that confirmed their opinion that the vaccine was dangerous. When searching Google, they claimed censorship because websites containing that information were not in the top results. However, they were successful in finding such material using DuckDuckGo. DuckDuckGo largely sources search results from Bing (Results Sources, 2022; Urman et al., 2022). The differences in these algorithmic systems demonstrated inconsistencies in information retrieval and how the power of source reliability is dependent on the user.

The initial premise of *If…Then* is that "algorithms have become a key site of power in the contemporary mediascape" (Bucher, 2018, p. 32). The power of the algorithm and the firms that develop them are ensconced in our daily lives as "autonomous decision making" (Diakopoulos, 2015, p. 400). As users of platforms become more reliant and entrenched in the services, awareness of the algorithms' power wanes and users become normalized to the power. Bucher continues the analysis of algorithms to state the multiplicity they hold, as "concealed behind a veil of code" and "impenetrable" (2018, p. 42). The unpacking of this multiplicity is where Bucher (2018) argues that the heuristic of the black box does a disservice to how algorithms are investigated; that by creating a scenario where something is unknowable it prevents researchers from asking difficult questions. Bucher explains three approaches to research black boxed

algorithms. First, ignore the fact that the whole of something is unknowable by breaking it apart into what can and cannot be known and investigate through that lens. Second, use the perceptions or beliefs of algorithms as a starting point. And finally, black boxes change and evolve over time providing opportunities to find the seams of the box and learn its history (Bucher, 2018).

It is impossible to hold search firms accountable with either citizen action or public policy if the difficult questions are not asked. The call for transparency of algorithms is loud and frequent, from both scholars and legislators (for example: Beer, 2017; Bozdag, 2013; Pasquale, 2016; S.2763, 2019). As Pasquale states, "you can't form a trusting relationship with a black box" (Pasquale, 2016, p. 83). The black box metaphor creates an atmosphere of the algorithm being unknown, similarly when a plane crashes and officials are searching for the black box so that they might know what happened prior to the crash. Without opening or accessing the box, the cause is unknown. The issue is not that, as Winner (1993) suggests, we might open the black box to find it empty, it is that labeling it as such creates an aura that continues the myth of the unknowing. An example of this phenomenon happening is data voids in search results. Data voids occur when search queries return few to no results because the request is not searched often, or the language has not been associated with the data (Golebiewski & boyd, 2019).

The power and the politics of algorithms, specifically search algorithms, have been investigated since the foundational work of Introna and Nissenbaum (2000) who discuss why the politics of search engines matter. They stress that search engine designers are making decisions for searchers everywhere without their knowledge or consent.

Introna and Nissenbaum's argument that the firm is the creator of the social process; "These choices are embedded in human-interpreted description criteria, in crawl heuristics, and in

ranking algorithms" (Introna & Nissenbaum, 2000, p. 175). They voice concerns over the search engine giving preference to paid advertisements over other, non-paid content and voice their support for a search engine that raises the voices of the "typically unheard" and promotes "broadly inclusive interactivity" (2000, p. 180).

2.3 Sociotechnical Systems of Information Retrieval

Exploring sociotechnical systems begins the process of opening the black box, to understand how it is constructed, and how to study its impacts. Sociotechnical systems are made up of many moving parts that interact with each other at both predictable and random intervals. For the purpose of this study, I focus on the people, procedures and processes, and policy.

2.3.1 Web History

Public web search first became available in 1990 with Archie; prior to this the availability of web search was limited to academia, government, and corporations. Archie created a searchable database of files stored on FTP (file transfer protocol) sites (Seymour et al., 2011). Other searchable databases appeared quickly after such as Gopher, Veronica, Jughead, W3Catalog, and Wandex. These databases searched the title level of webpages only. Early search engines retrieved information based on indexes, thesauri, and notion ranking with Boolean operators. The first full text search engines launched in 1994: WebCrawler and Lycos. While these became widely known by the public, it was AltaVista that served as a prototype for all future search engines (Seymour et al., 2011). In early designs, search engines acted as information retrieval systems in the purest form. They searched the titles and text of webpages and results included webpages based on how many times the keywords appeared on the website. In 1998 Page and Brin not only developed an iterative algorithm that curated SERP that provided ranking of websites outside of the relevance of keywords, but they also launched a

popular search engine. PageRank ranked the quality of a website based on calculations similar to academic citation ranking and considered the number of links a website has pointing to and from the webpage (Brin & Page, 1998). When Brin and Page published their paper on PageRank in 1998, they revolutionized search algorithms because PageRank was more than keyword or index searching. As the rate of information online increased exponentially, search engines struggled to maintain searchability of the web's content because the indexing web crawlers could not keep up. Google quickly dominated search through a clean design and accurate results for novice searchers.

After PageRank, Google's search algorithms evolved and started to include more personal data about the end user in their results. Page and Brin spoke publicly about their future goals for search; Page in particular has an interest in the perfect search engine, which allows for feedback loops from the end user that includes personal information such as results clicked, text of email messages, and location (Hillis et al., 2012; Zimmer, 2008). In 2004, Page and Brin discussed search being a part of the brain. Acknowledging this claim borders on search cyborgs, and Google appears to continue search development with this goal in mind (Hillis et al, 2013, p. 55). One such product, called Google Instant was nicknamed Miss Cleo referencing the popular TV psychic (Hillis et al., 2012, p. 56). Google Instant is the development that introduced the predicted search terms in the drop-down box. At the time of deployment, the user had to opt-in for the feature; now, it is fully embedded into the Google search algorithmic system. Further developments made to the algorithm include consideration to the popularity of a topic or event-based data from the user to produce search results (Hillis et al., 2012, p. 66). As search algorithms developed, research to identify and understand the societal impacts is necessary to illuminate how a proprietary, seemingly free search engine is harmful to its users.

YouTube only existed a year prior to its acquisition by Google in 2006. The video platform grew quickly from simply a video sharing platform to providing substantial income to content creators. The platform is regarded as a search engine, a social media site, a news medium, and entertainment space (Soukup, 2014). As the video platform changed, governance had to change to keep up with the ways users and creators were using the platform. In 2007, Viacom filed a lawsuit against YouTube claiming Digital Millennium Copyright Act violations as content creators used music in videos without the permission of the rights holder (Seidenberg, 2009). YouTube also had to change to keep up with their changing demographic as users became younger with access to smartphones and tablets. The platform launched a child-friendly version, YouTube Kids which restricted content that violated FCC rules and regulations concerning children's programming (Alghowinem, 2019). Pertinent to this research, YouTube's autoplay feature became the site debate surround the radicalization of users by playing video after video with specific points of view, however several studies have proven this not to be true (Hosseinmardi et al., 2021; Ledwich & Zaitsev, 2020).

The founding and development of Facebook is widely known and captured in film in the 2010 biopic, *The Social Network*. Since its launch as Facemash in 2003, a Harvard student version of Hot or Not, the platform has grown to 1.96 billion daily active users (DataReportal, 2022). The last 20 years have seen many highs for Facebook as people flocked to the social media site once it opened to the public in 2006. The company changed the way people communicate with one another and disrupted the digital advertising market (Bucher, 2021). The "move fast and break things" culture at Facebook resulted in several issues, like the Cambridge Analytica scandal, the creation and deletion of fake accounts, and the foray into fact checking after the 2016 presidential election.

2.3.2 Designing the Perfect Search Engine

It took Google only four years to reach and maintain a majority of the search engine market share. Vaidhyanathan (2011) captures this rise to cultural dominance in *The Googlization of Everything*, which warns readers of the dangers of allowing a private company to control several facets of one's online life as documented through Google's purchase of YouTube and Blogger, and development of Google Drive and Chrome. Google's move into semantic search, the effort to read or understand the intended query, keeps the company with the largest market share in search. Google's mission statement, "to organize the world's information and make it universally accessible and useful," demonstrates the desire to be the gatekeeper of information. One of the ways Google does this is through building the/a perfect search engine.

Zimmer (2008) investigates the privacy challenges of perfecting a search engine through the lens of contextual integrity. Google, an advertising company, relies on the perception of a free service to collect user information to sell more targeted advertisements (Vaidhyanathan, 2011). The collection of personal data disrupts user understanding of information flows from the user to the product. Google's Privacy Policy states the collection of personal information builds better services by delivering personalized content and advertisements, processing search terms, and developing new services (Google, 2021). However, Google does not specifically indicate when personal information impacts search results to provide more personalized and filtered information. What is thought to be private information, such as email content, contact lists, or calendar events, is then used to filter search engine results of public information. The use of social network information provides an underlying social aspect to information seeking not necessarily known to the searcher.

2.3.3 User-Centered Design

Early online information retrieval systems focused heavily on the collection of information being searched (Wolfram & Xie, 2002). Users navigated to different websites and databases based on information need. For example, Yahoo! required users to select the subset of information, such as stocks, and search within that collection. Digital libraries and academic databases still use this model. When Google launched with a clean interface and a single search box, IR system design shifted to center the user instead of the collection. When Spink (2002) conducted user-centered research on a meta-search tool, Inquirus, they found that participants appreciated the vast information content included but struggled to find relevant information based on their keyword search partially due to unranked results. Which partially explains the fast success of Google.

Google Search is an information retrieval system originally built on PageRank. When designed, the PageRank algorithm not only used keyword matching, but it also determined relevance using a calculation based on academic citation ranking and used the number of links a website has pointing to and from the webpage (Brin & Page, 1998). As an example, one of the reasons Wikipedia is often in the first five search results is because of the frequency of links to other Wikipedia pages and the links to other websites in the bibliography section. As with any algorithm, PageRank relies on data supplied by humans who build or edit the websites as well as the users' clicks on results after their searches. Search engines are no longer simple indexes of websites, they "combine [their] own knowledge of available content with user queries to provide recommendations to the users" (Grimmelmann, 2013, p. 4).

The problem that algorithmically filtered SERPs create for the end user is the appearance that the search process is linear, and that "correct" results are easy to come by. The evidence of

this disconnect is noticeable when recent high school graduates begin their first year of college. Academic databases are not as intuitive to search because the students are familiar with the personalized results from Google. The search engine appears to have read their mind when they receive the answer to a short question (Miller, 2013). If they try this same technique in a database, they are often disappointed with the limited results that are not as relevant. To a novice researcher, it appears that Google is providing more accurate results to their query regardless of content because it feels relevant and easy to the user.

In their simplest form, search engines are a conduit between the searcher and the indexed web and provide the ability to access information, which has radically changed the information landscape. Eighty-one percent of adults living in the United States own a smart phone and 90% use the internet, which means information retrieval is more accessible than ever before ("Demographics of Internet and Home Broadband Usage in the United States," 2019).

Access to information has implications in both information retrieval and policy. Searchers can only access what they have available to them. And to most searchers, that will be what is freely available on the internet, or before the website imposes a paywall. The person with an information need chooses the path of least resistance. The prevalent convenience of search engines in pockets and smart speakers creates a scenario where searchers do not need to consider different information systems, they choose the option closest to them. Search engines are the gateway to information on the internet and commonly feature an answer module or knowledge panel that seemingly "answer" questions. A recent eye-tracking study focusing on answer modules found that 14% of searchers, when presented with useful information in the knowledge panel, did not click on any results compared to a search without a knowledge panel, where less than 2% of searches received zero clicks (Wu et al., 2020).

2.3.4 Social Search

Social constructivism provides a lens to understand that search engines have created and mirror a reality in which the searchers are searching (Mager, 2010, 2012). The utilization of personal information influences the curation of the search results and newsfeeds that create a push and pull between the platform and the user. An example of this is the development of Google Images. In the year 2000, Jennifer Lopez wore a green Versace dress to the Grammy Awards and the query quickly became the most popular search on Google. The problem was that the results were text based, and not an image of the celebrity in the dress. The artifact is Google, the social groups are people who want to see a picture of the dress, the problem is that the picture was not readily available, and the solution is that Google evolved to return search results as images. Schmidt admits that Jennifer Lopez in the dress was the impetus for the creation of Google Images (2015).

Developers and engineers play an important role in the creation of technology, from the design of the product to the functional code, code that learns from users. The knowledge and biases of the developers are baked into the final product. While the technology industry rejects government regulation, they either ignore or are comfortable with the code regulating their platforms (Lessig, 2000, 2003). This can look like a technical solution that is in fact enforcing cultural policy decisions (Hunt & McKelvey, 2019). Code is inherently political and therefore can never be neutral (Winner, 1980). However, developers do not necessarily claim the algorithms as their own, rather they see themselves working on a product as a whole and argue they are not responsible for the final outcome (Seaver, 2017). Personalization techniques in search engines and social media shape the reality of the searcher both individually and as a group, just as the collection of personal information by these platforms is representative of

society has shaped search and social media. The technological frame of search and social media implies the embedding of how the developers categorize or tag content and the perceptions of potential users which "intertwine the social and the technical" (Sismondo, 2010, p. 102).

Building on Bucher's (2018) theory of "programmed sociality" and expanding the notion to search algorithms, social search is an activity that indicates interacting with others while searching (Halavais, 2018). Programmed sociality is the notion that our social interactions online are dictated by algorithms, in addition to explicit decisions by the user. The searcher might not be aware of this interaction unless they are well versed in the curation of search results. Google Search meets this definition of social search as the algorithm that produces search results is informed by other people through PageRank, geographic region, and user data collected from clicks (Pariser, 2011). Based on user preference, Google Search is a "sociable system" because they "perform best when they fulfill the needs of the community" (Halavais, 2018, p. 74). However, as we come to understand in studies of information poverty, the community needs to trust these systems or else they ignore the information.

The opponents of social constructivism argue several points including the "total disregard for the social consequences of technical choice," the decisions surrounding the "relevant" social groups, the idea that technology can be created without a social problem, and the lack of principles to help people understand the possibilities of technology (Winner, 1993, p. 368-371). I appreciate Winner's critical analysis of social constructivism and even agree that there is merit in discussing social consequences of technical choice, however, technology does not exist without human input and social problems would eventually become known to influence the technological change. Search algorithmic systems have an opportunity to be the technology that is held accountable. Searchers are making a technical choice to continue to use

Google and there are social consequences to that choice. As scholars and journalists find problematic search results to basic queries, search engines edit their code to correct the specific problem, even while claiming the autonomy and neutrality of the algorithm (Crane, 2014; Gillespie, 2017; Noble, 2018). However, as Noble (2018) increased awareness about the racist results when searching "black girls" and the search results were corrected, the search for "beautiful woman" still returns mostly white, blonde, skinny women.

2.4 User Beliefs

People are complex actors within socio-technical systems. They become dependent on the convenience of a platform quickly and they also abandon platforms for social and political reasons (Raine & Anderson, 2017; Sweney, 2022). The social and political reasons for leaving a platform are often informed by their level of awareness and understanding of algorithmic systems.

2.4.1 Algorithmic Understanding

The definition of algorithm has evolved over time to add the human elements to what was once defined as a mathematical process that takes in some data as input and produces other data as output. The use of the term algorithm has expanded outside of computer science to incorporate a multidisciplinary understanding in critical algorithmic studies (Seaver, 2017, 2019). The critical study of algorithms recognizes this process as both *in* and *as* culture, meaning that algorithms were *in* culture as an abstract being, created by mathematicians or computer scientists to do a job. However, the conversation changed to view algorithms *as* culture not only because they deal with cultural objects "but because they are composed of collective human practices" (Seaver, 2017, p. 5). Gillespie suggests algorithms are "complex sociotechnical

assemblages" that combine the process of data, calculation, and relevance with their proprietary nature as they are the property of the designer or corporation (2016, pp. 22, 25).

Understanding algorithmic systems is often difficult to describe or conceptualize when so much of the process is hidden. Folk theories provide community-based perceptions on a phenomenon. There is a large corpus of algorithmic folk theories that are platform or functionality specific. For this literature review, I focus on platform specific folk theories pertaining to the consumption of information as opposed to how content creators perceive the algorithms impacting the distribution of their work. This literature provides a baseline into sense making processes, social cues, and feedback loops pertaining to algorithmic understanding (DeVito et al., 2018).

In a study investigating user perceptions of Facebook's newsfeed, researchers developed six folk theories pertaining to how and why content appears in the newsfeed: passive consumption, producer privacy, consumer preferences, missed posts, violating expectations, and speculating about the algorithm. Overall, their respondents identified an "entity" that determined how, when, and why content appeared in their newsfeed and that the "entity" required intervention to force important posts into their feeds (Rader & Gray, 2015). The finding of knowing "something" is impacting their experience but not fully understanding what and how it functions is similar to the Powers' (2017) study about news consumption. Users are not fully unaware of an algorithmic presence, but do not fully understand the impacts of its presence.

A smaller interview-based research project asked participants to explain how content ended up in their newsfeed. First, researchers asked participants if they were aware of algorithmic filtering and less than half were aware of algorithmic filtering of their newsfeeds. This work resulted in four folk theories, developed by those who were aware of algorithmic

curation: personal engagement, global popularity, format, and narcissist theories (Eslami et al., 2016). These theories simplify algorithmic curation to what the user clicks or comments on, how strongly the topic is trending, giving preference content type (like giving preference to images over text), and finally showing posts that mirror oneself.

The users in the previous studies appear ambivalent but cautious at the use of algorithmic curation in their newsfeeds. However, that is not always the case as some users fight back over the changes social media platforms make to their newsfeeds (DeVito et al., 2017). Not only were users paying attention to press releases about an algorithmic change to the Twitter timeline, but they also created a trending hashtag to raise awareness of the issue. DeVito et al. (2017) did not interview participants but rather developed ten folk theories from tweets containing the #RIPTwitter hashtag. This study's sample is very aware of algorithmic filtering and its impacts, unlike previous studies. The folk theories ranged from outrage to resignation to questioning the platform's economic motivations. What the findings indicated beyond Twitter is that users who spend a significant amount of time on a platform do begin to notice changes to the algorithmic systems as they happen, and often have reactions to them.

A more recent study in Norway uses folk theories to identify the effects of datafication on the user (Ytre-Arne & Moe, 2021). Sixty percent of this study's population were aware of algorithms and their impacts on social media and/or the internet. Perhaps because of the high level of awareness, five thorough folk theories were developed from their data: algorithms are confining, practical, reductive, intangible, and exploitative. The first three address the functionality of the algorithms' use on the platform while the last two begin to attend to the less recognizable factors of algorithmic curation.

A study testing five folk theories about economic motivations and algorithmic operations during internet use, Dogruel (2021) found most users to be motivated to learn about how algorithms function even if their technical skills are low. Regardless of their main focus of algorithmic operation, users overwhelmingly believed that popularity of a website or page is why it ranked higher in search, regardless of platform.

There appears to be a gap in the literature analyzing how users believe their personal information is used by the platform. Except for a study using TikTok as the research site which recognized an identity-based folk theory. Participants in this study found that some users believe that specific social identities were removed from their For You Page (Karizat et al., 2021). The dialogue surrounding TikTok's algorithm has created an impression that TikTok knows the user better than the user knows themselves (Cotter et al., 2022). Algorithmic systems are utilizing personal data to make these recommendations and personalization to platform content.

2.4.2 User Beliefs on Personal Data

Much of daily life is tied up in using various apps, platforms, and online services that the control of personal data collected is overwhelming and inevitable (Hargittai & Marwick, 2016). A general sentiment around online privacy is the idea that so long as the online behavior is good, the user does not care if their information is private, as they have nothing to hide (Solove, 2007). The collection of personal data is pervasive on all platforms. At times the user is aware of what information they are providing while completing a form during sign-up or if they read the cookie disclosure pop-up when visiting various websites. However, much more is collected about the user while they are using the platform or service. The idea of "if you're not paying for a product, you are the product" can be traced back to the 1970s to a video broadcast created by artists bringing criticizing television (Serra, 1973). This sentiment has been replicated many times over.

In Fiesler and Hallinan's (2018) work to find patterns in how users felt about online privacy they collected comments from people on articles about unroll.me's data breach. They identified "you are the product" as the rationale for users needing to be more aware of their online activity. The overwhelming attitude identified by the users is that users should be aware of data collection practices of the platforms and services they use. And if a user does not like the practices of the service or platform, it should be avoided. However, even if the data collection practices are known by users there is a general ambivalence to the potential data sharing or breaches, except for location data (Zimmer et al., 2018).

2.5 Policy

Information and communication technology policies both complicate and structure sociotechnical systems. This section outlines a brief history of internet policy, platform governance, the application of the First Amendment to algorithms, and the uniformity of privacy policies.

2.5.1 Information Policy

Information policy is an investigation of the "laws, regulations, and doctrinal positions – and other decision making and practices with society-wide constitutive effects – involving information creation, processing, flows, access, and use" (Braman, 2011, p. 3). The following focuses specifically on analyzing how information policy addresses the algorithmic systems of online interactive platforms. The web of policy issues surrounding the internet, search engines, and social media sites remains tangled, with each facet impacting the other, which has political and ethical ramifications. Access to information is a political issue, especially when private corporations create infrastructure to search, find, and read the needed information (Braman, 2006).

2.5.2 The Power of the Political System

Access to and freedom of information are closely tied together when searching for information. Search engines act as the gatekeeper of the internet. The politics and policy of the search engine are necessary to evaluate when researching the search structure. The political economy of Google has been thoroughly investigated; however, the research often focuses on the organization as a whole and not its specific services (Fuchs, 2011). This is likely due to the difficult nature of understanding a concept that is so opaque and can change from user to user. However, the implications of searchers living in filter bubbles are also damaging to society. Personalization works against the free form inquiry needed when investigating a topic.

For algorithmic filtering of SERPs to operate, the search engine must be collecting data about the searcher. However, the collection, storage, and access of data is not transparent to the user. Search histories are important and private artifacts of an individual. Richards (2015) equates searching to thinking – the act of searching is the external cognitive process of processing thoughts. This idea is eerily similar to Chrome's slogan of "browse the web as fast as you think." The process of searching and learning is included in the freedom of thought and is a part of intellectual privacy (Richards, 2015). The settings of Google Search allow for some modification of what is tracked, but what if the searcher had to opt into surveillance instead of opting out. The research identifies all the ways in which our data is being used for our benefit and alludes to the dangers, but it does not address the level of knowing of a typical searcher.

In *Change of State: Information, Policy, and Power* Braman (2006) outlines four forms of power: instrumental, structural, symbolic, and informational. The different forms of power manipulate and shape human behavior: instrumental power impacts the material world through physical force; structural power impacts the social world through rules and institutions; symbolic

power impacts the material, social and symbolic worlds through ideas, words, and images; and information power impacts the informational bases of instrumental, structural, and symbolic power (Braman, 2006). Concrete evidence of users' perspectives of algorithmic filtering would explain the informational power of the search firms and identify areas where users can take back some of that power. Additionally, the evidence informs practitioners of information studies on how best to communicate privacy and personalization settings to users. It is not the intention of this research to weaponize the ignorance of the user to further manipulate their informational power.

Search engines uphold hegemonic order as they are "embedded in social relations and social organizations throughout society" and because they allow for the passive consent of use as a normal common-sense activity rather than coercion (Mosco, 2009, p. 206). The structure of the search platform is such that each search on Google is the user consenting to their Terms and Services to use the site. Google clearly states they earn the majority of their profits from advertising sales and that while they do not sell personally identifiable information to advertisers they provide generalized reports for personalized ads (Google, n.d.).

A critique of Google's political economy is a useful device for understanding their place in upholding hegemonic power. While Google appears to be a free service to its user, the company accesses, stores, analyzes the personal data and internet activity of its users. The organization sees this exchange as fair and reasonable instead of exploiting power and labor (Fuchs, 2011). The appearance of a free service hides the commodification of the user that supports the structure of the platform. The distribution of information flow is another hidden facet of search that capitalizes on its user. When signed in, Google tracks information from the user through different devices and platforms; a search conducted by Google on your home

computer influences search results on your work computer. Similar tracking happens across apps such as YouTube, Drive, Maps, and Gmail while using Chrome or Google on a different browser (Nield, 2019).

To maintain the default search engine on various browsers, Google outbids and holds contracts with those browsers. Firefox ended their contract with Yahoo! early and contracted with Google in 2017 as the default search engine (Lardinois, 2017). The United Kingdom's Competition and Market Authority reported Google paid Apple nearly $1.5 billion to maintain its status as the default search engine in Safari in 2020; similar figures also were reported in the United States (Statt, 2020). This buying power of search prominence, specifically mobile browsers where Google holds 90% of the market share, is one of the reasons cited in the 2020 antitrust lawsuit brought forth by the U.S. Department of Justice (*Complaint, United States v. Google LLC*, 2020).

Facebook has had its share of government investigating. In 2018, Mark Zuckerberg testified in a congressional hearing about the platform's use and abuse of user data (*S.Hrg. 115-683*, 2018). This hearing came after the Cambridge Analytica scandal during the 2016 presidential election. More recently a whistleblower came forward with information and testified to the Senate Committee on Commerce, Science, and Transportation that Facebook was targeting younger users and not putting safeguards in place to which protects them from harmful content (Haugen, 2021).

2.5.3 Internet Regulation and Governance

Internet regulation is the laws and structure in place through the government or government agencies like the Federal Communications Commission (FCC). Internet governance are those systems in place internally that govern the norms and flows of information on the

platforms. This section outlines both regulation and governance of the internet and how the lack of regulation has put the power of control into the hands of the organizations that provide access and content on the web. Regulations of the internet exist at the access point for the user – how ISPs provide access to the internet, the interference of content, and how they can charge for the service. There are additional regulations set in place through jurisprudence on content restrictions for minors, copyright, and First Amendment rights.

Some scholars and observers argue for less regulation for search engines, and specifically their algorithmic systems as they believe competition will regulate. While calls for transparency of the algorithm are plentiful, an argument made by search films is that competition drives innovation and transparency increases the likelihood of bad faith actors gaming the system as seen in spamming and Google bombing (Granka, 2010).

In 1995 Netscape released the browser cookie protocol to help identify users and terms of agreements while computers and servers interact. After broad acceptance, Netscape effectively created a user tracking "law" for anyone interacting with a browser and as tracking increased, user privacy decreased (Lessig, 2003). The provision of internet privacy laws in the United States has largely been left up to state governments, California for example, which passed the California Consumer Privacy Act (CCPA) that gives consumers some control over their online personal information. In contrast to the U.S. case, the European Union passed the General Data Protection Regulation (GDPR) in 2018 which gives individuals power over their own personal data and requires internet platforms to plainly disclose the data they collect.

Internet governance investigates the "mechanisms of control" within the platforms and can include algorithmic ordering, privacy policies, and the regulatory contexts binding them (DeNardis, 2020, p. 4). Governance comes from within an organization through terms and

conditions, privacy policies, boards of directors, and other proper use policies in place. The effectiveness of governance varies by firm. The internal policy-making creates black box implications as well through the algorithms by "translating its norms and rules into inscrutable, often proprietary systems that cannot be publicly scrutinized" (Hunt & McKelvey, 2019, p. 319). With the addition of machine learning, the algorithms are not just difficult to understand by the user, but also the developers (Hunt & McKelvey, 2019).

The Telecommunications Act of 1996 (The Act) overhauled the Communications Act of 1934 and classified telecommunication services as the telephone system and information services as cable and broadband services in addition to private data networks used by large corporation to connect computers, faxes, and other technology that allowed for dial-up internet services (Ehrlich, 2014). This distinction dates back to 1966 when the FCC "launched *Computer Inquiry* to explore the broad range regulator and policy problem generated by" computer technology, data processing, and communication services to consider hybrid communications and hybrid data processing services ("The FCC Computer Inquiry," 1972). To allow for continued regulation, the FCC made a distinction between hybrid communications and hybrid data processing. This also meant that those organizations could continue to improve upon their basic communication systems, while exploring additional service models. This decision is not without fault as the regulation of these hybrid models prevented competition and relied on "the good faith of the carriers" as it is nearly impossible for the regulation of data processing ("The FCC Computer Inquiry," 1972, p. 202).

Most notably, The Act excluded information services from common carrier laws that require the organization transporting the goods to remain impartial to what is being transported. The lack of fast, reliable internet coverage in the United States is why net neutrality debates are

persistent. The debate over the definitions of telecommunication service and information service continued into the early 2000s when then FCC Chairman Michael Powell ruled cable companies (telecommunication service) offering internet access be reclassified as an information service, stating "that simply because an 'information service' contained a 'telecommunications' component that did not make it a 'telecommunications service' subject to public utility common carrier regulation," specifically unbundling of services (Ehrlich, 2014, p. 11).

Net neutrality at the federal level has been a common phrase since 2015. It began in earnest in 2009 with the American Recovery and Reinvestment Act that included stimulus funding to extend broadband and the inclusion of common carrier laws to ensure equitable distribution of data through the internet. In 2015, the Obama-era FCC passed regulations to prevent broadband providers from blocking, throttling, or discriminating against "lawful" internet content and classified the service as a telecommunications service utility (Ruiz & Lohr, 2015). In 2017, Trump appointed Ajit Pai, the former associate general counsel of Verizon Wireless and an opponent of net neutrality. FCC Chairman Pai oversaw the FCC vote that repealed the Obama-era regulation and reclassified the internet as a Title I information service from a Title II common carrier (Shepardson, 2018). Seven states – California, Colorado, Maine, New Jersey, Oregon, Vermont, and Washington – and the territory of Puerto Rico have passed net neutrality legislation in an effort to protect their citizens from potential throttling by internet service providers (ISP) (Morton, 2021). The broad classifications of the internet impact innovation, price, and speed but there are specific sections of The Act that regulate in much more specific ways.

Section 230(c)(1) states "no provider or user of an interactive computer services shall be treated as the publisher or speaker of any information provided by another information content

provider" (*Section 230 of the Communications Decency Act*, n.d.). Kosseff (2019) claims these twenty-six words created the internet we know today by laying the ground for social media companies and other similar platforms. Section 230 protects ISPs as well as other "interactive computer services" including social media, or any platform that publishes third party content from liability for what is said or posted (*Section 230 of the Communications Decency Act*, n.d.). Section 230 essentially encourages platforms to maintain and enforce their own user policies. The success of Google and other search engines is not the content they create, but the access they provide to third party sites. Section 230 protects search engines from litigation from users and web content providers who do not agree with their results or rankings. In addition to Section 230 protection, several lawsuits filed against Google set precedent for users' First Amendment rights and copyright infringement. As Section 230 works to keep an open internet by protecting ISPs and platforms from the speech of their users, the Online Copyright Infringement Liability Limitation Act (OCILLA) protects online service providers (OSP) from liability for direct copyright infringement by their users as part of the Digital Millennium Copyright Act (DMCA) (H.R.2281, 1998). OCILLA only provides a "safe harbor" if the OSP has governance in place to remove the infringing material, notify offenders, and remove repeat offenders from their services. As Section 230 protects corporations from the liability of user speech and OCILLA protects OSPs from the liability of copyright infringement, how are the rights of the users protected or are they forfeited by using the platform? In the aforementioned laws search engines play a role in the distribution and findability of third-party content and copyrighted materials while also enjoying the protections of the same laws. Later in 2011, the Stop Online Piracy Act (SOPA) proposed a requirement that search engines remove access to sites containing pirated materials, opponents of this measure claimed censorship of the internet (Kozak, 2018).

Prior to 2000, cases citing search engines fell into two categories. First, disputes from website providers who sought fair treatment from search engine features that would increase their popularity. Second, companies sought retribution from search engines for brand-related infringements (Gasser, 2005). In the years 2000-2005, the courts read more cases investigating malfeasance in sponsored links and banner ads by those using copyrighted or trademarked content within their ads. In *Field v. Google Inc*., the plaintiff claimed copyright infringement by those clicking on a cached link and downloading copyrighted materials (412 F. Supp. 2d 1106 (D. Nev. 2006)). The court claimed Google remained passive in the search process and that it was the issue of the search engine *user* not the search engine *provider* (Gasser, 2005). *Search King, Inc. v Google Technology, Inc* is one of the first cases where search engine operators claim their algorithms, in this case PageRank, are protected opinions under the First Amendment (CIV-02-1457-M (W.D. Okla. May. 27, 2003)).

State and federal governments have begun to draft legislation that would ban or limit the use of personal information for algorithmically delivered content. In the fall of 2019, the U.S. Senate introduced a bipartisan bill titled the Filter Bubble Transparency Act. The bill's goal is "to require internet platforms give users the option to engage with a platform without being manipulated by algorithms driven by user-specific data" (S.2763, 2019). Democrats introduced a second bill in the summer of 2020 to the House of Representatives to amend the Federal Election Campaign Act of 1971 to change the regulation of targeted political advertisements using online behavior data and require platforms to maintain public records on the purchasing of political advertisements (H.R.7012, 2020). In January of 2022, Congress put forth the Banning Surveillance Advertising Act of 2022 that seeks to prevent digital advertisers, such as Facebook and Google, from using personal data to target advertisements (S.3520, 2022; H.R.6416, 2022).

During the same month, Congress also proposed a bill to require platform governance to be more transparent through shortening the policies and making them easier to understand (S.3501, 2022; H.R.6407, 2022).

2.5.4 Algorithmic Systems and the First Amendment

Intellectual freedom is a two-sided phenomenon when discussing algorithmic filtering. First there is the users' freedom of inquiry to consider and on the other side the protections given by the First Amendment to the organizations that create the algorithms that provide search results. First, this section discusses the searcher's intellectual freedom and the aspirational rights set forth by the American Library Association (ALA) and the United Nations. Second, it discusses the historical use of the First Amendment in cases brought against search firms regarding their search algorithms.

The American Library Association (ALA) succinctly defines intellectual freedom as "the right of every individual to both seek and receive information from all points of view without restriction" (American Library Association, 2007). The Library Bill of Rights upholds intellectual freedom through collection development and programming. ALA provides direction and assistance for those libraries accepting federal telecom assistance, providing minors access to the internet with filters in accordance with Children's Internet Protection Act (CIPA). However, the ALA does not believe CIPA is constitutional as impedes access to library patrons (Jenner & Block, 2001). The United Nations provides a more nuanced approach to intellectual freedom through Article 19 in the Universal Declaration of Human Rights: "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" (United Nations, 1948). Aspects of algorithmic filtering do not allow for

"all points of view without restriction" as filtering does restrict information due to the personalization filters created for each user (American Library Association, 2007). Restriction takes on different meanings depending on the medium of the information required such as attempting to access a print book in a location that does not have a public library. Accessing information on the internet can appear without restriction, so long as you have an internet connection. However, algorithmic filtering of search results impacts how the searcher views the information and what is served. As discussed earlier, cognitive abilities inform the information seeking behavior of the searcher and the knowledge of methods used to obtain information. Search engines' lack of forthrightness creates a level of opaqueness to the search process.

The phenomenon of secret algorithms is known as the "black box" (Pasquale, 2016; Winner, 1993). In *The Black Box Society*, Pasquale (2016) describes how Google gives broad outlines of their search algorithms to include relevance and importance, however, the organization does not give specifics to prevent websites from taking advantage of the algorithm to change their page's ranking. Additional factors included in Google's search algorithm are natural language processing (NLP) that trains artificial intelligence (AI) programs the meaning of terms to better serve relevance results (*Google's Search Algorithm and Ranking System - Google Search*, n.d.). The First Amendment upholds the organization's desire to keep their algorithms opaque and secret. The black box is created through intellectual property rights and the First Amendment protections.

An aspect of intellectual freedom is First Amendment protection – what speech is and is not protected by the Constitution. Historically, search engine speech is a form of protected speech that includes the platforms' algorithm and the search results (Balkin, 2018; Bracha, 2014; Gasser, 2005). Precedent has been set by the courts to consider search platforms as editors of

their search results, like newspaper editors deciding the layout and order of the newspaper or that the search engine is considering contextual information from the user to provide a relevant opinion in the form of search results (Bracha, 2014). Balkin (2018) continues the breakdown of First Amendment (FA) rights as they have shifted from individual protections to those of the corporation and capital for quickly growing and high-valuation tech and media companies. What makes this distinction most interesting is that companies are using the FA to protect their business ventures and profit lines; as an example, the regulation of how firms collect, use, or sell user data is an infringement of their FA rights as data is speech.

However, Bracha (2014) argues that there is another perspective of search engine speech that does not qualify for First Amendment protection and that is search engine speech is a function. Functional speech situated in the social practice of the search engine as the algorithmic system provides a functional service that requires multiple facets of input to return results. The function of the search engine speech is to help users find specific information.

Balkin identifies three problems for the current algorithmic society. First, private governance creates a scenario where platforms have "become governors of their spaces, and not merely facilitators of communication" (p. 997). Second, the precedent of new-school speech regulation that aims to protect corporations and their data flows. Third, the private surveillance of users allows corporations to collect endless amounts of data to use, distribute, or sell for their benefit. This is especially problematic when data breaches like Facebook/Cambridge Analytica happen that allowed for users to be targeted based on political affiliation (Balkin, 2018). Instead, what platforms have instituted is widespread moderation (Gillespie, 2018). When platform's responsibility of content moderation is at the discretion of human moderators, problems arise as the application of community guidelines and Terms of Service policies differ from user to user,

from infringement to infringement. On social media platforms, human workers moderate flagged content as the ultimate decision makers. The atrocities these workers are forced to see as part of their daily work are detailed in Roberts' *Behind the Screen* (2019). Search engines use machine learning as their form of moderation, which begins to remove even the developer in the understanding of how data is processed to achieve the search results pages (Seaver, 2019).

2.5.5 Privacy Policies

There is no federal law that requires platforms to have a privacy or data policy, however, there are several other laws that make it necessary for companies to provide a disclosure for the type of data they collect, how they share it, and how to delete data. For example, the California Consumer Privacy Act requires privacy policies to include information on the Right to Know, the Right to Delete, the Right to Opt-Out of Sale, and the Right to Non-Discrimination (*California Consumer Privacy Act (CCPA)*, 2018). Similarly, the European Union's (EU) General Data Protection Regulation (GDPR) has implications for American companies who do business in the EU. American users see the impacts of this when visiting websites and having to accept or decline the use of cookies or other tracking mechanisms. Additionally, there are federal laws that require companies to manage how data is collected and from whom. For example, the Children's Online Privacy Protection Act (COPPA) requires users to be at least 13 years old when signing up for a service or social media platform. As such, privacy and data policies have become common practice and follow similar formats.

Researchers in a 2016 analysis of the privacy policies of Facebook, Google, Twitter, and LinkedIn found that Facebook collected the most personal and user information (Zeadally & Winkler, 2016). In this study, they identified that Google did not collect information about user-

generated content, payment information, and off platform activities. Since then, Google launched Google Pay (replacing Google Wallet) in 2018 and now collects those three areas of user data.

Obar and Oeldorf-Hirsh (2020) call the reading of privacy policies and terms of service the "biggest lie on the internet." In their research signing students up for a fake social networking site, they found 74% of participants skipped reading the policies all together and those who did click on them spent less than five minutes reading the policies. Pew Research finds that one in five people report they always or often read the privacy policies of the platforms they use but the majority of those only skimming the contents (Auxier et al., 2019). Privacy policies tend to be lengthy and written in a language that is not accessible to the average user which is the primary reason the TLDR Bill was proposed, to simplify the language of platform policies.

Ibdah et al (2021) identified that users who read privacy policies are motivated by the service credibility and type, type and amount of data collected by the service, service popularity, readability of policy, location and interface design of the websites, their own habits, and recommendations from others. These factors vary based on user experience and need. Researchers found that 77% of the participants claim to have some experience reading or interacting with privacy policies, and another 12% believe privacy policies are unnecessary (Ibdah et al., 2021).

2.6 Conclusion

Information needs are specific and personal to the searcher. The cognitive and procedural processes a searcher takes requires some knowledge of either the system or the community and because of this, facets of information poverty appear when the searcher does not have the capability or access to find, understand, and use the information found. And while information literacy is often the proposed solution to information poverty, the guiding frameworks of

professional organizations ignore the systematic issues that uphold information poverty, instead focusing on the searcher to bear the weight as the solution. Google, YouTube, and Facebook are nearly ubiquitous for information retrieval, as the top three visited sites in the United States. The library and library databases have become the *other* place to look for information. As search becomes a daily interaction, users trust the system explicitly as it provides fast, useful information.

Information retrieval (IR) and the models used to study the flow of information have remained unchanged for many years. The models still presume human like behavior of the system without the inputs of user data, natural language processing, and linking of content by the algorithmic systems. One of the first IR models to include system suggestions of keywords, does not identify how the system might suggest those keywords and from what data the system is building those suggestions (Yangyuen et al., 2020). The IR models perpetuate facets of information poverty and continue the idea of black box systems because they do not explicitly acknowledge the unseen data to provide the user with information.

The public influences information retrieval platforms' development through the frequent searching of a term, and the use or lack of use of a service. At the same time, platforms normalize collecting personal information for the use of their systems. The implicit agreement of personal data for platform use is part of the black boxed process. Black boxes describe an unseen or not fully understood process as inputs become outputs. Private companies are complicit in the black box metaphor for their algorithms, as they hold trade secrets and proprietary algorithms. The metaphor, while useful for description, also highlights the need for deeper inquiry into the sociotechnical system. Through the theoretical and methodological application of Bucher's (2018) framework of interrogating black boxed algorithms, I identify what is known of the

algorithmic systems, gather the perceptions of the processes, and interrogate the cracks of the box. By learning where the cracks exist, I can work toward the further opening of the black box through practical applications like information and algorithmic literacy.

The platforms studied in this dissertation (Google, YouTube, and Facebook) are user-centered sociotechnical systems of information retrieval. There are social elements within all of the platforms due not only to the interactive elements of the platform but also the mutual constitution of its development. However, these platforms have a history in claiming neutrality when issues arise with their algorithmic systems (Crane, 2014). The situating of these platforms as sociotechnical systems allows for the investigation of how the search results and newsfeeds are constructed by the user data provided to them. The algorithmic filtering or personalization is a large part of the success of these platforms. Users should be more aware of how algorithmic filtering can distort their online experience.

Previous research on user beliefs of algorithmic systems demonstrates the complexity of human interaction with sociotechnical systems. Depending on how users are asked about their awareness or knowledge of algorithms, their beliefs about them change. While some studies found users to be more pragmatic about how algorithmic systems impacted their online lives, other identified how algorithms are confining, reductive, or intangible while navigating online information sources (Yre-Are & Moe, 2021); other studies identified external factors as users considered how algorithms functioned, such as global popularity and personal engagement (Eslami et al., 2016).

While there is no federal law that stipulates the requirements or contents of privacy policies, several states and foreign entities have passed legislation that have created similarities in the format and contents. The study of information policy is the first step in learning about how

users are informed about personal data collection and use by the platform. Privacy policies and terms of service are lengthy documents filled with legalese and complicated structure (Obar & Oeldorf-Hirsh, 2020). This is likely why users do not read them or lie about reading or skimming the policies (Auxier et al., 2019). As algorithmic literacy is typically understood as three functions, knowing, feeling, and doing, the first point of knowing an algorithmic system is typically the platforms' policy which is why the research of this study begins there. After which, the dissertation study seeks to learn about users' awareness levels of the algorithmic systems through their frequency of use and demographic characteristics.

Chapter 3: Methodology

Employing Bucher's (2018) black boxed algorithmic systems research methodology, this dissertation study first analyzes privacy and data policies from Google, YouTube, and Facebook to identify what is known about the black boxed system. Then using that analysis, I built a survey to learn users' awareness of personal data collection and use by the information retrieval platforms based on their privacy and data policies. This work aims to identify the gaps in information and algorithmic literacy of the platforms' users. The survey was distributed to a sample representative of the United States population and analyzed through descriptive statistics and open coding methods. This dissertation aims to answer the following research questions:

Research Question 1 (RQ1): How do platforms (Google (1), YouTube (2), Facebook (3)) govern and inform their users with regard to using collected data to filter the users' online content?

Research Question 2 (RQ2): What do United States-based adult users of these platforms know about the personal and usage data being collected and what is their awareness about how the platforms use this data to moderate and serve online content?

Research Question 3 (RQ3): Are there any associations between usage characteristics of United States-based adult users of these platforms and users' awareness of algorithmic curation using personal information?

Research Question 4 (RQ4): How does user knowledge about algorithmically delivered content vary based on demographic characteristics, including age, race, education level, and political affiliation?

Table 1: Research Question Matrix

Research Question	Data Collected	Data Analysis
RQ1	Data/privacy policies of platforms	Content analysis for language explaining construction of results/news feeds, open coding
RQ2	Cross-sectional survey; knowledge tests, open-ended questions	Descriptive statistics, chi-square tests, open coding
RQ3	Cross-sectional survey; knowledge tests, open-ended questions	Descriptive statistics, chi-square tests, open coding
RQ4	Cross-sectional survey; knowledge tests, open-ended questions	Descriptive statistics, chi-square tests, open coding

3.1 Methodology

3.1.1 Research Design

This dissertation uses a mixed method design to provide a comprehensive analysis of the research problem. I address Research Question 1 through a content analysis of the platforms' data/privacy policies that address algorithmic filtering and/or the construction of search results and news feeds (Herring, 2010). Research Question 2 adds the user knowledge of these platforms and their functionality to provide context to future policy proposals that aim to hold platforms accountable by increasing transparency. The data also provides a greater understanding of how users view their interactions with platforms through a cross-sectional or snapshot survey of

platform users' perspectives of algorithmic filtering and analysis using descriptive statistics. The analysis of the results will not only discuss the connections between users' knowledge and the privacy policies of platforms, but also how the frequency of use of these platforms impacts algorithmic awareness.

The strength of a mixed methods study is that it provides a comprehensive analysis of the research problem where the contradictions or incongruent findings are explained or further probed (Creswell & Creswell, 2018). A weakness of this design is that this is an evolving topic that will continue to grow and expand. Due to that nature, this research is a snapshot in time and viewed as historical data. The strength of this research allows for continued studies in this area to maintain an understanding of users' perspectives of algorithmic filtering and to provide additional data to policymakers working on legislation at the local, state, and federal level.

RQ1

To narrow platforms for this study, the chosen three include those with high traffic in the United States and employ algorithmic filtering to their search results, autoplay, and news feeds; and where users search for information and the platforms have a precedent of algorithmic bias and disinformation. According to April 2021 data, these include Google (1), YouTube (2), and Facebook (3) (*Top 100: The Most Visited Websites in the US*, 2021; 2023). The sample includes the current public facing privacy/data policies that inform users of how their data is used to algorithmically filter the information they receive. In addition to having high traffic in the United States, these platforms are the focus of antitrust lawsuits, disinformation campaigns, and claims of selective censorship of content (*Complaint, United States v. Google LLC*, 2020; Jungherr & Schroeder, 2021).

RQ2, RQ3 & RQ4

The target sample for this study is search engine and social media users. Since 90% of Americans access and use the internet daily, finding participants who are familiar with the technology is not difficult (Anderson et al., 2019). However, achieving a randomized cross sample of search engine users, representative of current population percentages creates challenges with convenience or snowball sampling. To participate, individuals must be over 18 years old, have an internet connection, and have used a search or social media platform in the last 30-days. Qualifying questions are at the beginning of the survey to disqualify participants as needed.

This study samples eligible participants by purchasing a panel of respondents from Qualtrics, University of Wisconsin-Milwaukee's (UWM) preferred survey response program. Qualtrics adheres to the population guidelines set by the researcher and for this study represents national averages of age, race, and gender (U.S. Census Bureau, 2021). The service guarantees clean and complete results of the required sample size. Qualtrics maintains and recruits from a traditional, double-opt-in market research database use for corporate and academic market research. For example, Qualtrics has relationships with airlines which allows participants to complete surveys in exchange for frequent flyer miles. Documentation of their European Society for Opinion and Marketing Research (ESOMAR) and Institutional Research Board (IRB) clarifications are in Appendix A. The population size of internet users in the United States is 295.38 million people, using Pew data that 90% of Americans are internet users (Anderson et al., 2019). Using the equation to calculate sample size with a 95% confidence level includes a .5 standard deviation and a 5% margin of error, the survey sample should reach at least 384 responses (Connaway & Radford, 2017, p. 148). With 396 survey responses collected, the total

cost of the panel was $1,980 which was partially funded by UWM's School of Information Studies Doctoral Research Funds.

The rationale behind the choice of using a participant panel is the potential problems with collecting data through cold call surveys which can have a low response rate. The research is meant to answer questions about understanding and awareness of algorithmic filtering to make generalizable claims. A campus sample has limited demographics available. Similar issues persist with a snowball sample, since selection bias reflects the researcher's own demographics when using their social networks for recruitment and presents a major concern with sampling from social media in that it excludes people who are not online and not on social media (Baltar & Brunet, 2012). With any sampling method using data collection via an online survey, the sample excludes people who do not use the internet and favors those who want to provide their opinion on a given topic.

3.2 Data Collection

RQ1

The researcher saved snapshots of the privacy and data policies in the WayBack Machine in May 2021 of Google (1), YouTube (2), Facebook (3) as the top three platforms based on monthly traffic. In addition to web archiving, PDFs of each of the policies are saved in the researcher's files. The focus on only privacy and data policies is because the platforms' Terms of Service point back to the privacy policy when addressing any data collection.

All three platforms have several avenues of communication with their users to distribute information about how and why they collect personal information, such as separate explanatory websites, videos, and press releases. The choice to only include the official privacy/data policies

is that the information replicated itself and upon account creation and subsequent use or changes, they are the documents the user is asked to agree with. The analysis of the privacy and data policies provided the framework for survey questions that answer research questions two and three.

RQ2, RQ3 & RQ4

UWM's Institutional Review Board (IRB) (#22.140) approved the questionnaire before it was distributed through Qualtrics. The survey questions are varied and include yes/no/unsure, multiple choice, Likert, and free text questions. Survey research has several benefits that include generalizability, cost-effectiveness (when compared to interviews and other types of qualitative methods), and reliability. The standardization of the questions and the format creates a consistency of the participant experience and helps with data analysis. Common issues with survey data collection are the response rate and potential bias in the responses due to those people who are more opinionated about the topic being more willing to participate. Inspiration for survey questions comes from several previous studies. The structure of some of the knowledge questions draws inspiration from Proferes' (2015) dissertation on users' understanding of Twitter's information norms. The questions presented create a foundation of the users' knowledge of how platforms operate both technically and financially. Most questions focus on asking participants to self-rate their awareness or knowledge of algorithmic filtering and the functionality of search or social media platforms as Likert scale questions tend to provide stronger results than asking an open-ended question about what they think they know (Hargittai, 2009).

The questionnaire consists of 60 questions divided into four sections. Where possible I use radio buttons, Likert scales, and true false questions rather than dropdown menus to prevent

endorsing a specific answer and creating more hand-eye movement for the participant (Toepoel, 2017). The questionnaire uses pages for organization instead of scrolling and uses occasional logic to ask follow-up questions as marked in Appendix B. The consent agreement is the only required question in the survey.

The first part includes demographic information about the participants of the study. These questions provide data on groups of users to compare how their behavior differs from other groups (Xie, 2008). Age, race, and ethnicity bracketing is directly correlated with the U.S. Census data to ensure quotas (Dimock, 2019). I took guidance for gender identity from the Human Rights Council foundation to ensure inclusivity while also capturing important data for historically marginalized genders (*Self-Identification of LGBTQ Employees*, n.d.). The remaining questions in this section ask for highest level of education completed, occupation, and political affiliation.

The second section first asks about internet connection types, then determines internet literacy by asking the familiarity with common words associated with internet use asks about platform. The literacy test is borrowed from recent studies on users' experience with fitness trackers (Ibdah et al., 2021; Zimmer et al., 2020) and modified to include additional social media platform specific terms. Platform specific sub-sections investigate the participant's usage and purpose of search engines and platforms, asking questions about frequency of use and time spent on the platform and the purposes of use such as information on a particular topic, looking up news on an event, or finding entertaining content.

Section three examines the awareness of participants about platform functionality and contents of the data/privacy policies – specifically how platforms algorithmically filter search results and news feeds. Questions in this section are developed from the content analysis of the

aforementioned policies and by drawing on a similar study about Google News (Powers, 2017). Again, this section has platform specific sub-sections of Google, YouTube, and Facebook. This section provides yes or no questions with a few follow-ups for participants to elaborate if they have interacted with, for example, Google's Terms and Conditions or privacy policy. The purpose of this section is to gain a better understanding of specifics of the functionality of search while also capturing attitudes toward search. The business model questions draw from Proferes' (2015) on Twitter's messaging to users. It is within this section users are asked whether actions influence search results and newsfeeds and what personal data is collected on the user while using the platforms. I tested these questions in another study about algorithmically curated advertisements on Facebook.

Finally, the fourth section focuses on internet literacy by asking participants to evaluate the accuracy and trust of different types of information sources (Dutton et al., 2017, 2019). Question 58 allows for direct comparison to a 2012 Pew Research Center study on internet and search engine use by asking if search engines are fair and unbiased sources of information (Purcell et al., 2012). The final question asks if two people get the same search results if they entered the same search terms at the same time on Google, with the ability to explain their answer as a follow-up.

The survey organization has mechanisms in place to verify the consistency of the survey results such as automatically replacing respondents who finish less than half the survey completion length and providing the researcher with an opportunity to view the data to request response replacements due to quality issues (Appendix A). The survey underwent useability testing for clarity to prevent misinterpretations of the questions and approval through UWM's Institutional Review Board.

3.3 Data Analysis

RQ1

A textual analysis of companies' public-facing privacy and data policies to identify the language that informs users on how algorithmic filtering is employed in the construction of their search results and news feeds through the use of open coding. Each phrase is given a category to further identify the language used to inform the user of the platform's privacy policy. To reduce bias and ensure consistency in coding, this study employs inter-coder reliability where a colleague codes a portion of the responses to find cohesiveness in the process. Intercoder reliability is calculated using Cohen's Kappa formula (Kim, 2017).

3.3.1 Coding Schema

Platform privacy and data policies were analyzed and categorized by the information collected while using the platform and what influences the platforms' algorithmically delivered content through open coding. The decision to identify the influences comes from the platforms' policies. Google's Privacy Policy states that they "use the information we collect to customize our services for you, including providing recommendations, personalized content, and customized search results" (*Privacy & Terms – Google*, 2021) and Facebook's Data Policy states, "We use the information we have to deliver our Products, including to personalize features and content (including your ads, Facebook News Feed, Instagram Feed, and Instagram Stories) and make suggestions for you (such as groups or events you may be interested in or topics you may want to follow) on and off our Products" (*Facebook Data Policy*, 2021). Since the policies drive what is known about the platform and the platforms do not provide a more specific breakdown of use, the categories were developed under the assumption that personal data collected could and would be used for algorithmically delivered content, including

advertisements are included in both search results and newsfeeds. The reason for this assumption is the phrasing at the start of the privacy and data policies. Facebook's Data Policy states, "We use the information we have to deliver our Products, including to personalize features and content (including your ads, Facebook News Feed, Instagram Feed, and Instagram Stories) and make suggestions for you (such as groups or events you may be interested in or topics you may want to follow) on and off our Products" (*Facebook Data Policy*, 2021). Google's Privacy Policy states that they "use the information we collect to customize our services for you, including providing recommendations, personalized content, and customized search results" (*Privacy & Terms – Google*, 2021).

The actions were simplified into the following categories for each area – influence and collection – as seen in Table 1. These categories then were used in questions 38, 39, 47, 48, 56, and 57 in the user survey.

Table 2: Policy Analysis Categories

Influences Results/Autoplay/Newsfeed (Author Generated)	**Type of Data Collected (Author Generated)**
Your actions on the platform	Interactions with Platform (clicks, likes, comments, subscribes)
Your actions on other websites	Browser, application, or device in use
Details about your online session	Location
Details about your smart phone usage	Address book contacts
Actions taken by other users	Usage of Platform products and features

Actions taken by platform engineers/editors/curators	Purchases or financial transactions through Platform products
Data details between applications	Details about your online session
Details about accounts across devices	Details about your smart phone usage
	Information from advertisers, app developers, and games from when you are not using Platform
	Information from non-Platform apps and advertisers
	Voice and audio information when you use audio features

Tables 2 and 3 contain the coding schema for the policy analysis. The first column contains the categories developed in the open coding process, the middle column contains the definition of the category, and the third column provides a single policy clause as an example. The platform labels are [G-YT] for Google-YouTube and [FB] for Facebook. The full policy analysis table is available in Appendix C.

Table 3: Policy Analysis *Influences* Coding Schema

Influences Results/Autoplay/Newsfeed (Author Generated)	**Definition**	**Policy Clause Example**
Your actions on the platform	Any action taken on the websites or apps – clicks, likes, comments, tags,	[G-YT] We also collect the content you create, upload,

	uploads, direct messages (sent/received), purchases, search terms	or receive from others when using our services.
Your actions on other websites	Any action taken on other websites or apps, typically collected using cookies and/or pixel tags – clicks, likes, uploads, time spent on websites, purchases	[G-YT] The activity information we collect may include: terms you search for, videos you watch, views and interactions with content and ads, voice and audio information when you use audio features, purchase activity, people with whom you communicate or share content, activity on third-party sites and apps that use our services, Chrome browsing history you've synced with your Google Account
Details about your online session	Usage of internet connected applications on a computer or smart phone, programs open,	[FB] Information about operations and behaviors performed on the device, such as whether a window is

	operating system, internet service	foregrounded or backgrounded, or mouse movements (which can help distinguish humans from bots).
Details about your smart phone usage	Details about application usage, location, data storage, cell tower, data usage, messages, emails, purchases, search terms, etc.	[G-YT] Your location can be determined with varying degrees of accuracy by: GPS, IP address, sensor data from your device, information about things near your device, such as Wi-Fi access points, cell towers, and Bluetooth enabled devices
Actions taken by other users	Any actions taken by other users on platforms, websites, applications – including their likes, comments, clicks. Or combination of actions – order of videos watched, search term and click combination, content of post liked	[FB] We collect information about the people, Pages, accounts, hashtags and groups you are connected to and how you interact with them across our Products, such as people you communicate with the most or groups you are part of.

Actions taken by platform engineers/editors/curators	Modifications of metadata, performance of application or website, cookie type, pixel tag location, server log refresh, information retention	[G-YT] We use various technologies to collect and store information, including cookies, pixel tags, local storage, such as browser web storage or application data caches, databases, and server logs.
Data details between applications	Details shared between applications on smart phones and computers through caches, pixel tags, cookies, etc.	[G-YT] We use various technologies to collect and store information, including cookies, pixel tags, local storage, such as browser web storage or application data caches, databases, and server logs.
Details about accounts across devices	If signed into accounts on multiple devices, any actions taken on one device apply to the other device. This is most visible with search terms and autocomplete.	[FB] We also collect contact information if you choose to upload, sync or import it from a device (such as an address book or call log or SMS log history), which we use for things like helping

		you and others find people you may know and for the other purposes listed below.

Table 4: Policy Analysis *Collected* Coding Schema

Type of Data Collected (Author Generated)	**Definition**	**Policy clause**
Interactions with Platform (clicks, likes, comments, subscribes)	Any action taken on the websites or apps – clicks, likes, comments, tags, uploads, direct messages (sent/received), purchases, search terms	[FB] We collect the content, communications and other information you provide when you use our Products, including when you sign up for an account, create or share content, and message or communicate with others. This can include information in or about the content you provide (like metadata), such as the location of a photo or the date a file was created.
Browser, application, or device in use	If signed into accounts on multiple devices, any actions taken on one device apply to	[G-YT] The information we collect includes unique identifiers, browser type and

	the other device. This is most visible with search terms and autocomplete.	settings, device type and settings, operating system, mobile network information including carrier name and phone number, and application version number.
Location	The location of the user through WiFi, GPS, IP address, cell towers	[G-YT] Your location can be determined with varying degrees of accuracy by: GPS, IP address, sensor data from your device, information about things near your device, such as Wi-Fi access points, cell towers, and Bluetooth enabled devices
Address book contacts	People you've emailed, direct messaged, chatted with	[FB] We collect information about the people, Pages, accounts, hashtags and groups you are connected to and how you interact with them across our Products, such as people you

		communicate with the most or groups you are part of.
Usage of Platform products and features	Any action taken on the websites or apps – clicks, likes, comments, tags, uploads, direct messages (sent/received), purchases, search terms, camera filters, video stop/start time, sharing of videos, posts, links	[FB] It can also include what you see through features we provide, such as our camera, so we can do things like suggest masks and filters that you might like, or give you tips on using camera formats. Our systems automatically process content and communications you and others provide to analyze context and what's in them for the purposes described below.
Purchases or financial transactions through Platform products	Purchases of in-app products, games, rentals, or any other financial transaction on the Platform	[FB] If you use our Products for purchases or other financial transactions (such as when you make a purchase in a game or make a donation), we collect information about the

		purchase or transaction. This includes payment information, such as your credit or debit card number and other card information; other account and authentication information; and billing, shipping and contact details.
Details about your online session	Usage of internet connected applications on a computer or smart phone, programs open, operating system, internet service	[G-YT] We use various technologies to collect and store information, including cookies, pixel tags, local storage, such as browser web storage or application data caches, databases, and server logs.
Details about your smart phone usage	Details about application usage, location, data storage, cell tower, data usage, messages, emails, purchases, search terms, etc.	[FB] Information such as the name of your mobile operator or ISP, language, time zone, mobile phone number, IP address, connection speed and, in

		some cases, information about other devices that are nearby or on your network, so we can do things like help you stream a video from your phone to your TV.
Information from advertisers, app developers, and games from when you are not using Platform	Modifications of metadata, performance of application or website, cookie type, pixel tag location, server log refresh, information retention	[FB] Advertisers, app developers, and publishers can send us information through Facebook Business Tools they use, including our social plug-ins (such as the Like button), Facebook Login, our APIs and SDKs, or the Facebook pixel. These partners provide information about your activities off Facebook—including information about your device, websites you visit, purchases you make, the ads you see, and how you use their services—whether or

		not you have a Facebook account or are logged into Facebook.
Information from non-Platform apps and advertisers	Any action taken on other websites or apps, typically collected using cookies and/or pixel tags – clicks, likes, uploads, time spent on websites, purchases	[G-YT] In some circumstances, Google also collects information about you from publicly accessible sources.
Voice and audio information when you use audio features	Voice commands, questions, search phrases, responses, frequency of use, type of use.	[G-YT] The activity information we collect may include: terms you search for, videos you watch, views and interactions with content and ads, voice and audio information when you use audio features, purchase activity, people with whom you communicate or share content, activity on third-party sites and apps that use our services, chrome browsing history you've

		synced with your Google Account

3.3.2 Research Hypotheses

RQ3

Descriptive statistical analysis and crosstabs are used to explore the knowledge and awareness among respondents as a whole, representative group in the survey data (Connaway & Radford, 2017). Chi-square tests are used to determine statistical significance within the relationships of variables. Open coding is employed during the analysis of the open-ended survey questions. The following three research hypotheses are explored to fully answer research question two:

H_1: There is an association between time spent on the platform and the users' awareness of algorithmic curation using personal information.

H_2: There is an association between reading or skimming the privacy policy and the users' awareness of algorithmic curation using personal information.

H_3: There is an association between those privacy-focused search users and their awareness of algorithmic curation using personal information.

RQ4

Descriptive statistical analysis and crosstabs are used to explore differences among respondents based on demographics and platform use in the survey data, as identified in the five research hypotheses below (Connaway & Radford, 2017). Chi-square tests are used to determine

statistical significance within the relationships of variables. The crosstabs are segmented by usage of platform and platform knowledge and drill down to compare knowledge trends where statistically impactful.

H_{4A}: There is an association between age and users' awareness of algorithmic curation using personal information.

H_{4B}: There is an association between gender and users' awareness of algorithmic curation using personal information.

H_{4C}: There is an association between race and users' awareness of algorithmic curation using personal information.

H_{4D}: There is an association between education and users' awareness of algorithmic curation using personal information.

H_{4E}: There is an association between political affiliation and users' awareness of algorithmic curation using personal information.

H_5: There is an association between specific demographic variables and specific sources of information reliability.

As this is exploratory research, these are high-level hypotheses meant to identify associations between a number of variables. The following table summarizes and identifies "personal information" in the aforementioned hypotheses.

Table 5

Variable summary table for personal information

	Google	YouTube	Facebook
Overall personalization using personal information	personal information to customize search results	personal information to customize video results	personal information to customize newsfeed

Personal information collected	Interactions with search results	Interactions with videos (likes, comments, subscribes)	Interactions on Pages, in Groups, and hashtags
	Browser, application, or device in use		
	Location		
	Address book contacts		
	Usage of Google products (YouTube, Maps, Drive) and features		Usage of Facebook products (Instagram, Messenger, WhatsApp) and features
	Purchases or financial transactions through Google products	Purchases or financial transactions through Google products	Purchases or financial transactions from Facebook products (Instagram, Messenger)
	Details about your online session		
	Details about your smart phone usage		
	Information from advertisers, app developers, and games from when you are not using Google	Information from advertisers, app developers, and games from when you are not using YouTube	Information from advertisers, app developers, and games from when you are not using Facebook
	Information from non-Google apps and advertisers	Information from non-YouTube apps and advertisers	Information from non-Facebook apps and advertisers
	Voice and audio information when you use audio features		Frequency of communication to people and groups
Personal information used to influence	Your actions on Google	Your actions on YouTube	Your actions on Facebook
	Your actions on websites other than Google	Your actions on websites other than YouTube	Your actions on websites other than Facebook
	Details about your online session		
	Details about your smart phone usage		
	Actions taken by other users		
	Actions taken by Google engineers/editors/curators	Actions taken by YouTube engineers/editors/curators	Actions taken by Facebook engineers/editors/curators
	Data details between applications		
	Details about accounts across devices		

3.4 Limitations

Even with best efforts in place, textual analysis is subjective to the interpretation of the researcher. The policies analyzed represent the accurate policy at the time of collection and survey distribution. These policies are subject to change at any time to represent the current practices of the platforms. For example, both companies updated their policies in early 2022 with increased transparency on the data collection and potential use.

While the sample of the survey is representative of the United States population, there are concerns self-reported survey data contains the risk that participants misrepresent their knowledge. Surveys administered by Qualtrics contain some mechanisms for catching and removing participants who are not reading the questions, which is helpful for the reliability of the data.

Chapter 4: Findings

The previous chapter outlined the methodology used to obtain the following results. This chapter answers the three research questions in succession. The first section examines the privacy policies of Google, YouTube, and Facebook; the second section describes the survey instrument and the participant population; the third section provides the overall analysis of user knowledge; and the fourth section analyzes user awareness and knowledge through demographic characteristics.

4.1 Snapshot of Platform Privacy and Data Policies

This section answers the first research question (RQ1) which asks, how do platforms (Google (1), YouTube (2), Facebook (3)) govern and inform their users with regard to using collected data to filter the users' online content? The snapshot is of Google and Facebook's privacy and data policies preserved in July of 2021[1] through the archive on the Wayback Machine and a PDF saved in the researcher's files. Analysis of the policies used Van Dijk (2013) and Proferes' (2015) content analysis rubric for how they inform users of personal data collected that impact their experience on the platforms. The following is a summary of the results of this process. Overall, the personal information collected is nearly identical across the three platforms (Table 4). The analysis of the data and privacy policies provided the framework for learning about user awareness that answers research questions two and three and to learn how companies describe their own practices. The complete analysis spreadsheet of policy clauses is in Appendix C.

[1] Since archived, Google and Facebook have updated their privacy policy three times each.

While there is not a federal law that requires platforms to have a privacy or data policy, there are several other state, federal, and international laws that make it necessary for companies to provide a disclosure for the type of data they collect, how they share it, and how to delete data. For example, the California Consumer Privacy Act (CCPA) requires privacy policies to include information on the Right to Know, the Right to Delete, the Right to Opt-Out of Sale, and the Right to Non-Discrimination (*California Consumer Privacy Act (CCPA)*, 2018). Additionally, the European Union's General Data Protection Regulation (GDPR) which affects any platform conducting business in the EU and requires data collection and processing transparency. As such, privacy and data policies have become similarly formatted and common practice.

Table 6

Information Type by Platform

Information Type	Google	YouTube	Facebook
Your actions on platform	Yes	Yes	Yes
Your actions on other platforms or websites	Yes	Yes	Yes
Details about your online session	Yes	Yes	Yes
Details about your phone usage	Yes	Yes	Yes
Actions taken by others	Yes	Yes	Yes
Actions taken by engineers, editors, curators	Yes	Yes	Yes
Data details between applications	Yes	Yes	Yes
Details about accounts across devices	Yes	Yes	Yes
Interactions on Pages, in Groups, and hashtags	n/a	n/a	Yes

4.1.1 Google and YouTube

Google's Privacy Policy governs all Google products including YouTube (*Privacy & Terms – Google*, 2021). The privacy policy is divided into eight sections: Information Google Collects, Why Google Collects Data, Your Privacy Controls, Sharing Your Information, Keeping Your Information Secure, Exporting & Deleting Your Information, Retaining Your Information,

and Compliance & Cooperation with Regulators. Through these sections Google provides general details about the personal and usage data collected on the user and then provides examples to explain why and how the data is used:

> *We collect information to provide better services to all our users – from figuring out basic stuff like which language you speak, to more complex things like which ads you'll find most useful, the people who matter most to you online, or which YouTube videos you might like.* (*Privacy & Terms – Google*, 2021)

Additionally, Google Search has a website that explains how search works through regular updates highlighting the efforts of engineers while they "organize the world's information and make it universally accessible and useful" (*Google Search – Discover How Google Search Works*, 2022). Through the series of webpages providing an overview, approach, functionality, and features Google uses animated graphics and videos to provide more real-life examples with simple language that give context to the platform. Similarly, YouTube also has a series of webpages that give more context to how they manage their user policies with regard to managing harmful content, child safety, revenue sharing, and copyright (*How YouTube Works – Product Features, Responsibility, & Impact*, 2022). Essentially these websites provide more readable and interactive Terms of Service. They address privacy and collection of data, but direct back to the privacy policy. Since January 2019, Google updated their privacy policy 10 times. These updates happen when state and/or U.S. federal regulations change, a new feature or product becomes available, or more clarity is needed on the topic. Google hosts a page with all previous privacy policies and provides a comparison tool to see what changed within the policy (*Updates: Privacy Policy – Privacy & Terms – Google*, 2022).

Google's Privacy Policy states that they "use the information we collect to customize our services for you, including providing recommendations, personalized content, and customized search results" (*Privacy & Terms – Google*, 2021). This statement indicates that any and all data collected could be used to curate search and video results. Cohen's Kappa for Google's privacy policy analysis is 0.7763.

Table 7

Google – Frequency of Influences	
Data details between applications	9
Details about accounts across devices	8
Your actions on Platform	7
Actions taken by other users	7
Details about your smart phone usage	6
Details about your online session	4
Actions taken by Platform engineers/editors/curators	4
Your actions on other websites	2
n/a	2

Table 8

Google – Frequency of Personal Data Collection	
Browser, application, or device in use	12
Details about your online session	10
Details about your smart phone usage	10
Interactions with Platform (clicks, likes, comments, subscribes)	8
Information from advertisers, app developers, and games from when you are not using Platform	6
Usage of Platform products and features	6
Location	5
Purchases or financial transactions through Platform products	5
Address book contacts	4
Information from non-Platform apps and advertisers	4
Voice and audio information when you use audio features	4
n/a	1

4.1.2 Facebook

Facebook's (Meta's) Data Policy governs all Facebook (Meta) products including Facebook, Instagram, and Messenger for United States users (*Facebook Data Policy*, 2021). The data policy is broken into nine sections: What kinds of information do we collect, How do we use this information, How is this information shared, How do the Facebook (Meta) companies work together, How can I manage or delete information about me, How do we respond to legal requests or prevent harm, How do we operate and transfer data as part of our global services, How will we notify you of change to this policy, and privacy notice for California residents. Facebook's Data Policy is slightly more explicit with the types of information they collect on the user. Similar to Google, they provide examples of why and how they use the collected data:

> *We collect information about how you use our Products, such as the types of content you view or engage with; the features you use; the actions you take; the people or accounts you interact with; and the time, frequency and duration of your activities. For example, we log when you're using and have last used our Products, and what posts, videos and other content you view on our Products.* (*Facebook Data Policy*, 2021)

Throughout Facebook's Data Policy they refer to *partners* without providing a definition or list of who they are, however, examples include advertisers, app developers, and publishers. Facebook did end the program with third-party data brokers to receive information for market segmentation. Now those third-party companies use Meta for Business to segment the Facebook market for advertisements. Facebook is not as transparent as Google regarding their data policy history. They host a page with a previous data policy from 2016 that then links to their current

policy, but nothing in between (*Facebook: Previous Data Policy*, 2016). The versions could be identified through using the Internet Archive and Microsoft Word to find the differences in text.

Facebook's Data Policy states, "We use the information we have to deliver our Products, including to personalize features and content (including your ads, Facebook News Feed, Instagram Feed, and Instagram Stories) and make suggestions for you (such as groups or events you may be interested in or topics you may want to follow) on and off our Products" (*Facebook Data Policy*, 2021). This statement provides the assumption that any and all data collected can be used to curate a newsfeed. Cohen's Kappa for Facebook's data policy analysis is 0.7821.

Table 9

Facebook – Frequency of Influences	
Details about your smart phone usage	12
Details about accounts across devices	10
Your actions on Platform	9
Details about your online session	8
Data details btw applications	8
Your actions on other websites	7
Actions taken by other users	6
Actions taken by Platform engineers/editors/curators	2
n/a	0

Table 10

Facebook – Frequency of Personal Data Collection	
Details about your smart phone usage	13
Browser, application, or device in use	10
Details about your online session	8
Interactions with Platform (clicks, likes, comments, subscribes)	8
Information from advertisers, app developers, and games from when you are not using Platform	8
Usage of Platform products and features	7
Information from non-Platform apps and advertisers	6
Location	5

Purchases or financial transactions through Platform products	5
Address book contacts	5
Voice and audio information when you use audio features	5
n/a	0

All three platforms collect nearly identical sets of personal information from the user (Table 4). As such, these are the general classifications of the privacy and data policies developed from the content analysis of the policies. These classifications supplied options in the survey questions presented to participants. After analysis of the privacy/data policies, the frequencies of the codes were tabulated. There are differences in the explicitness between how Google (and YouTube) and Facebook describe how the information influences the results and newsfeeds (Table 5 and 7). For example, Google's Privacy Policy provides six examples of how "details about smartphone usage" impacts the curation of results whereas Facebook's Data Policy mentions it 12 times. However, the frequency tabulations of the types of personal information collected are similar, which is to be expected (Table 6 and 8).

4.2 Survey Instrument and Population Characteristics

To answer research question two, three, and four, I developed a 60-question survey and distributed it through Qualtrics from January 7 to 24, 2022. The survey asks questions about use of relevant platforms, knowledge of personal information collected and used, and trustworthiness of multiple platforms. Descriptive statistical analysis and crosstabs are used to explore differences among respondents based on demographics, reported platform use, and reported reading of privacy policies from the survey data (Connaway & Radford, 2017). All chi-square tests use an alpha level of .05. The results are not necessarily in the order of the survey instrument, but rather in an order that illustrates user knowledge and how variables in usage,

policy reading habits, and demographics may vary their knowledge. The full survey instrument is in Appendix B.

To ensure a representative sample of the United States population using the 2020 U.S. Census percentages, I purchased a sample through Qualtrics with partial aid from the University of Wisconsin-Milwaukee, School of Information Studies Doctoral Research Funds. The goal percentages were set prior to the recent report by the U.S. Census which concluded that the 2020 census undercounted the Black or African-American population by 3.3%, the Hispanic or Latino population by 4.99%, and the American Indian or Alaska Native populations living on reservations by 5.64% (U.S. Census Bureau, 2022).

4.2.1 Age, Race, Gender, Political Affiliation, and Education

Table 11

Age

What is your age?	N = 394
18-24	9.11%
25-34	18.97%
35-44	16.75%
45-54	17.24%
55-59	8.87%
60-64	8.87%
65-74	13.55%
75+	6.65%

Table 12

Race

What is your race?	N = 401 (includes those selecting multiple races)
African-American or Black	13.08%
American Indian or Alaskan Native	2.66%
Asian or Pacific Islander	5.81%
Other	2.91%
White	75.54%

Table 13

Gender Identity

What is your gender identity?	N = 395
Man	47.17%
Woman	51.84%
Non-binary / non-conforming	0.74%
Prefer not to say	0.25%
Prefer to self-describe	0.00%

Table 14

Political Affiliation

What is your political affiliation?	N = 391
Democrat	38.61%
Republican	31.71%
Independent	24.29%
Prefer not to disclose	5.37%

Table 15

Education

What is the highest level of education completed?	N = 396
Less than high school	1.52%
High school or equivalent	6.57%
Some college	28.79%
Bachelor's degree	30.56%
Master's degree	24.24%
Professional degree	6.06%
Doctoral degree	1.26%

4.3 Understanding Users' Awareness of Algorithmically Delivered Content

This section answers research question two which asks, what do United States-based adult users of these platforms know about the personal and usage data being collected and their awareness about how the platforms use this data to moderate and serve online content? This is the most high-level research question to gather baseline awareness prior to further investigation of usage characteristics and demographics from research questions three and four.

4.3.1 Awareness of Influence on Search Results and Newsfeeds

Table 16

Overall Customization

Does ____ use personal information to customize search results/video results/newsfeed?						
	Google		YouTube		Facebook	
Yes	236	60%	226	58%	245	63%
No	60	15%	69	18%	47	12%
Unsure	99	25%	98	25%	99	25%
	395		393		391	

Drawing from the general classifications, developed in section 4.1, of the privacy and data policies, respondents answered questions about what influences search results, video results, and newsfeeds. The following tables provide data on all three platforms. The consistencies in the user awareness includes "their actions on the platform" and "actions taken by other users", with Facebook 5% and 9% higher, respectively. The notable commonality between the three platforms is the users' unawareness that actions by other users influence their information feeds. Of the scenarios given, all of them influence search results, video results and autoplay, and newsfeeds based on the language in the privacy/data policy.

Table 17

Actions Influencing Search Results

Do any of the following actions influence your search results?	Yes	No	Unsure	No/Unsure
Your actions on Google	47%	37%	16%	53%
Your actions on websites other than Google	44%	38%	18%	56%
Details about your online session	44%	36%	20%	56%
Details about your smart phone usage	32%	45%	23%	68%
Actions taken by other users	32%	45%	23%	68%
Actions taken by Google engineers/editors/curators	40%	35%	25%	60%
Data details between applications	36%	37%	27%	64%
Details about accounts across devices	34%	38%	28%	66%

Table 18

Actions Influencing YouTube

Do any of the following actions influence your video results?	Yes	No	Unsure	No/Unsure
Your actions on YouTube	47%	36%	18%	53%
Your actions on websites other than YouTube	37%	39%	25%	63%
Details about your online session	41%	35%	23%	59%
Details about your smart phone usage	38%	36%	26%	62%
Actions taken by other users	34%	38%	29%	66%
Actions taken by YouTube engineers/editors/curators	38%	35%	28%	62%
Data details between applications	37%	33%	30%	63%
Details about accounts across devices	37%	32%	31%	63%

Table 19

Actions Influencing Facebook

Do any of the following actions influence your newsfeed?	Yes	No	Unsure	No/Unsure
Your actions on Facebook	53%	28%	19%	47%
Your actions on websites other than Facebook	43%	35%	22%	57%
Details about your online session	43%	32%	25%	57%
Details about your smart phone usage	38%	35%	27%	62%
Actions taken by other users	41%	32%	27%	59%
Actions taken by Facebook engineers/editors/curators	44%	30%	26%	56%
Data details between applications	42%	31%	27%	58%
Details about accounts across devices	41%	30%	29%	59%

As reported in Table 18, Facebook users are more aware that the actions taken by other users influence their platform experience than YouTube or Google. This is likely the nature of social media and a notification at the top of the post that a friend liked or commented on the post. The right column combines the “no” and “unsure” responses to identify the percentage of respondents who have not answered a firm “yes.” This column indicates over half of platform users are not confident in their awareness of what actions influence their algorithmically curated experience.

4.3.5 Awareness of Personal Data Collected About the User

The previous section addressed the survey questions that asked respondents what influences their platform experience, this section provides the results to the survey questions asking about personal data collection. The survey questions used more specific language taken from the privacy and data policies and because the prompts reflect the policies there is a slight difference in language between Google/YouTube and Facebook. This section also provides a breakout of the statistically interesting relationships between platform usage and knowledge of data collection and influences. The survey asked how frequently participants used a search engine, the number of times a day they checked their YouTube timeline and how much times they watched videos, and the number of times a day the participants checked their Facebook newsfeed and how much time the spent interacting with their newsfeed.

4.3.5.1 Google

Overall, Google users are most aware of the collection of information pertaining to "location" and "usage of Google products" (Table 19). These are to be expected as location-based search results are some of the more obvious curated results. The two categories at odds are the "use of Google products" and "address book contacts," however, there is an assumption that someone who uses Drive or Maps would also use a Gmail account.

Table 20

Personal Data Collected by Google

Is the following personal data collected about you while using Google?	Yes	No	Unsure	No/ Unsure
Interactions with search results	54%	20%	26%	46%
Browser, application, or device in use	59%	17%	24%	41%
Location	68%	17%	15%	32%
Address book contacts	38%	29%	33%	62%

Usage of Google products (YouTube, Maps, Drive) and features	60%	18%	22%	40%
Purchases or financial transactions through Google products	49%	24%	26%	51%
Details about your online session	53%	22%	25%	47%
Details about your smart phone usage	45%	27%	28%	55%
Information from advertisers, app developers, and games from when you are not using Google	48%	25%	27%	52%
Information from non-Google apps and advertisers	41%	26%	33%	59%
Voice and audio information when you use audio features	44%	26%	30%	56%

4.3.5.2 YouTube

YouTube usage and knowledge of personal information collected faired different than that of Google, even though they share a privacy policy.

Table 21

Personal Data Collected by YouTube

Is the following personal data collected about you while using YouTube?	Yes	No	Unsure	No/ Unsure
Interactions with videos (likes, comments, subscribes)	57%	19%	24%	43%
Browser, application, or device in use	52%	24%	24%	48%
Location	52%	24%	24%	48%
Address book contacts	33%	35%	32%	67%
Usage of Google products (YouTube, Maps, Drive) and features	49%	25%	26%	51%
Purchases or financial transactions through Google products	39%	30%	31%	61%
Details about your online session	48%	24%	27%	52%
Details about your smart phone usage	39%	28%	33%	61%
Information from advertisers, app developers, and games from when you are not using YouTube	45%	23%	32%	55%
Information from non-YouTube apps and advertisers	41%	25%	34%	59%
Voice and audio information when you use audio features	41%	26%	33%	59%

4.3.5.3 Facebook

Facebook users are more aware of the personal data collected by the platform to influence their newsfeeds (Table 22). However, there is a statistical difference between those aware of that smart phone usage influences the newsfeed versus those aware smartphone usage is collected while using Facebook.

Table 22

Personal Data Collected by Facebook

Is the following personal data collected about you while using Facebook?	Yes	No	Unsure	No/ Unsure
Interactions on Pages, in Groups, and hashtags	58%	19%	23%	42%
Frequency of communication to people and groups	53%	22%	25%	47%
Address book contacts	45%	25%	30%	55%
Usage of Facebook products (Instagram, Messenger, WhatsApp) and features	56%	18%	26%	44%
Purchases or financial transactions from Facebook products (Instagram, Messenger)	48%	22%	30%	52%
Details about your online session	47%	21%	32%	53%
Details about your smart phone usage	42%	26%	33%	58%
Information from advertisers, app developers, and games from when you are not using Facebook	50%	20%	30%	50%
Information from non-Facebook apps and advertisers	43%	24%	33%	57%

Table 23

What is the purpose of your search? (check all that apply)

Look up news on a topic or event	239	20%
Navigation to sites	209	17%
Answer medical or health questions	179	15%
Find entertaining content	175	14%
Check accuracy of news or information	165	14%
Learn about politics or current events	134	11%
Complete a work-related task	94	8%
Other	20	2%
Keep up with family and friends	0	0%
Share pictures and videos	0	0%
Share links to content	0	0%

Table 24

What is the purpose of your use of YouTube? (check all that apply)		
Keep up with family and friends	199	23%
Find entertaining content	138	16%
Share pictures and videos	132	15%
Look up news on a topic or event	96	11%
Share links to content	72	8%
Learn about politics or current events	60	7%
Check accuracy of news or information	58	7%
Navigation to sites	43	5%
Answer medical or health questions	38	4%
Complete a work-related task	26	3%
Other	9	1%

Table 25

What is the purpose of your use of Facebook? (check all that apply)		
Keep up with family and friends	199	23%
Find entertaining content	138	16%
Share pictures and videos	132	15%
Look up news on a topic or event	96	11%
Share links to content	72	8%
Learn about politics or current events	60	7%
Check accuracy of news or information	58	7%
Navigation to sites	43	5%
Answer medical or health questions	38	4%
Complete a work-related task	26	3%
Other	9	1%

4.4 Usage Characteristics

This section answers the third research question (RQ3), which asks are there any associations between usage characteristics of United States-based adult users of these platforms and users' awareness of algorithmic curation using personal information? While some of these findings are what we might expect and confirm other qualitative studies, others allow for a

deeper investigation into how reading (even skimming) privacy polices impacts user awareness and how time spent on a platform affects user awareness. The following research hypotheses are tested throughout this section using chi-square tests with a 95% confidence level.

H_1: There is an association between time spent on the platform and the users' awareness of algorithmic curation using [personal information].

H_2: There is an association between reading or skimming the privacy policy and the users' awareness of algorithmic curation using [personal information].

H_3: There is an association between those privacy-focused search users and their awareness of algorithmic curation using [personal information].

Table 5 provides a summary of personal information by platform. Some variables are uniform across the platforms while others are specific.

4.4.1 Usage Levels

Usage level data provides an opportunity to explore if spending time on a platform increases the users' knowledge about that platform.

4.4.1.1 Google

Unsurprisingly, over half of the participants use search engines once or more a day. Those in the group of 18-24 (traditional age of college students) are among the lowest in frequency of searchers. Of the 15 participants who claim to never use search engines, 7 have a high school diploma or equivalent, 5 have some high school or less, and 3 have some college or university. The discussion chapter further discusses this group.

Table 26

Crosstabulation of Search Engine Use Frequency by Awareness of Personal Information Influencing Search Results

		Does Google use personal information to customize search results?			
		Yes	No	Unsure	Total
How often do you use a search engine?	Never	4	5	6	15
	Less than monthly	7	5	10	22
	Monthly	21	10	6	37
	Weekly	34	9	23	66
	Daily	84	20	38	142
	More than once a day	86	11	16	113
Total		236	60	99	395

A chi-square test was conducted to examine H_1, the relationship between search frequency and knowledge of personal information to customize search results. The relationship between these variables was significant, X^2 (10, N = 395) = 27.327, p .002. Those who search more frequently are more likely to be aware of how personal information is used to customize results (Table 26).

4.4.1.2 YouTube

Only 8.88% of participants claim to either not have a YouTube account or not use the platform. YouTube users are checking the website either daily (31%) or weekly (31%) and spending 30 minutes to two hours watching videos (67%). The length of time daily users spent on the platform watching videos is evenly distributed between the ages of 25-54. A chi-square test was conducted to examine H_1, both the relationship between how often and for how long YouTube users were on the platform and overall knowledge about if YouTube uses personal information to customize video results. There was a statistically insignificant relationship

between the frequency of checking YouTube and the knowledge that YouTube uses personal information to customize video results, X^2 (10, N = 358) = 14.768, p .141. Users who check their video feed more than once a day are less likely to be aware of how personal information is used to customize video results. There is a statistically significant relationship between the length of time spent watching videos and the awareness that YouTube uses personal information to customize video results, X^2 (6, N = 357) = 23.687, p .001. Those users watching more than 30 minutes of videos at a time are more likely to be aware of how personal information is used to customize search results.

4.4.1.3 Facebook

Over a quarter of total participants did not have or do not actively use their Facebook account. Overall, those actively using their Facebook account checked their timeline daily (40%) for an average of less than 30 minutes to two hours (64%). Those aged 25 to 55 years old checked their newsfeeds and spent time on the account similarly. Unlike Google and YouTube, Facebook users have a different relationship with the amount of time spent on the platform and their awareness of algorithmic curation based on personal information. Two chi-square tests were performed to examine H_1, the relationship between frequency of checking their Facebook newsfeed and the length of time spent on the platform, and their awareness of whether personal information is used to customize their newsfeed. The relationship between frequency of checking, X^2 (6, N = 288) = 6.493, p .370; and length of time spent on the newsfeed, X^2 (6, N = 286) = 9.071, p .170 are both insignificant. The users who check their newsfeed regularly and users who spend more time on the platform are less likely to be aware if personal information is used to customize their newsfeed.

4.4.1.4 Internet Usage

The majority of the participants connected to the internet through broadband/DSL/cable (38%) or through unlimited mobile data (28%), while only 4% identified dial-up as their main internet connection. Other participants access the internet through satellite connection (8%), LTE hotspots (8%), and public access (2%). Unsurprisingly, Google was the most used search engine followed by Yahoo! And Bing respectively (Table 27). The *other* responses provided were browser names, not search engines. Twenty-four participants chose only privacy-focused search engines.

Table 27

Search Engine Use

What search engine do you use?	%	Count
Bing	16.58%	66
Brave	6.03%	24
DuckDuckGo	12.56%	50
Google	88.69%	353
StartPage	3.27%	13
Yahoo!	18.34%	73
Other	2.76%	13
Total		592

For the purposes of search engines and the investigation into what searchers know about how their personal data is impacting the curation of their search results, how they access the search engine (Table 28) matters as evidenced with the 15 people who claim not to use search engines, but then go on to report they search in mobile browsers or URL bar input. Further research is needed to learn about what potential implications exist for how mobile search might be obfuscating the search function.

Table 28

Search Engine Access

How do you access the search engine?	%	Count
Mobile browser	63.31%	252
Website.com from a web browser	41.20%	164
URL bar input	21.35%	85
Voice activated search (Siri, Alexa, Google Home, etc.)	15.07%	60
Other (click on button, app, laptop, phone, tablet, manually)	2.26%	9
Total		570

4.4.2 Policy Reading Habits

Another check of user awareness of how their personal information impacted their platform experience is to ask about their policy reading habits. The respondents of all three platforms claimed to have read the privacy policies at a much higher rate than reported in a 2019 Pew Research study (Auxier et al., 2019). At the time of the Pew study, only 22% of participants claimed to have read the privacy policy all the way through, 35% read part of the way through, and 43% glanced the policy over. Like this dissertation research, the Pew study used a web survey to capture participant data but with a much larger sample size. As an average across all three platforms, of the respondents of this dissertation study, 69%, claim to have read or skimmed the privacy policy. In addition to policy reading habits, participants were also asked if users agree to a Terms of Service upon using the platform and if the platforms have the option to turn off algorithmic filtering. It is well known that users either over report reading policies or do not fully read them as evidenced in research by Obar and Oeldorf-Hirsh (2020) that launched a fake social media site and added "gotcha" clauses to the privacy policy that included "providing first-born as payment" and "providing the NSA with user data." A common limitation and risk of self-reported survey data is a misrepresentation of one's actions and correspondence to what they think the researcher wants to hear.

4.4.2.1 Google

Table 29

Google Policy Reading Habits

	n	Yes	No	Unsure
Do you agree to terms of service when using platform?	395	61.77%	18.73%	19.49%
Does Google have a privacy policy?	396	59.34%	11.87%	28.79%
Of those saying yes, have you read/skimmed the privacy policy?	234	71.79%	28.21%	n/a

Of the 60% who believe Google has a privacy policy, 72% claim to have read the policy (Table 29). A further investigation of the 168 respondents who answered yes to reading or skimming Google's privacy policy compared their knowledge of what influences search results and the personal data collected while using the platform, based on the survey questions developed from the open coding of the policy analysis. On average, this subset of respondents answered the influencing and data collection questions at 75% correct. Alternatively, 42% of the respondents (*n*=168) who read or skimmed the privacy policy scored worse than the average. Table 30 presents a breakdown between those who reported reading the privacy policy and the user knowledge of actions that influence search results.

Table 30

Crosstabulation of Self-Reported Reading of Privacy Policy and Actions that Influence Search Results

		Have you read or skimmed Google's privacy policy?				
Do any of the following actions influence your search results?		Yes		No		Total
Your actions on Google	Yes	112	81%	26	19%	138
	No	42	55%	34	45%	76
	Unsure	14	70%	6	30%	20
Total		168	72%	66	28%	234

		Yes		No		Total
Your actions on websites other than Google	Yes	100	79%	26	21%	126
	No	50	63%	29	37%	79
	Unsure	18	62%	11	38%	29
Total		168	72%	66	28%	234

		Yes		No		Total
Details about your online session	Yes	105	81%	25	19%	130
	No	45	61%	29	39%	74
	Unsure	18	60%	12	40%	30
Total		168	72%	66	28%	234

		Yes		No		Total
Details about your smart phone usage	Yes	91	83%	18	17%	109
	No	59	66%	31	34%	90
	Unsure	17	50%	17	50%	34
Total		167	72%	66	28%	233

		Yes		No		Total
Actions taken by other users	Yes	79	83%	16	17%	95
	No	62	63%	37	37%	99
	Unsure	27	68%	13	33%	40
Total		168	72%	66	28%	234

		Yes		No		Total
Actions taken by Google engineers/editors/curators	Yes	87	78%	24	22%	111
	No	48	63%	28	37%	76
	Unsure	33	70%	14	30%	47
Total		168	72%	66	28%	234

		Yes		No		Total
Data details between applications	Yes	89	82%	20	18%	109
	No	53	64%	30	36%	83
	Unsure	26	62%	16	38%	42

Total		168	72%	66	28%	234
		Yes		No		Total
Details about accounts across devices	Yes	84	81%	20	19%	104
	No	55	66%	28	27%	83
	Unsure	27	60%	18	17%	45
Total		166	72%	66	63%	232

A chi-square test was conducted to test H_2, the relationship between reading Google's privacy policy and user awareness of the actions that influence their search results. The test revealed one insignificant relationship between those who read the policy and being aware of the actions taken by Google's engineers, editors, or curators on their search results, X^2 (2, N = 234) = 5.234, p .073. Of the remaining categories, users who read the privacy policy were more likely to be aware of the actions that influenced their search results.

Table 31

Crosstabulation of Self-Report Reading of Privacy Policy and Personal Data Collected About the User – Google

		Have you read or skimmed Google's privacy policy?				
Is the following personal data collected about you while using Google?		Yes		No		Total
Interactions with search results	Yes	119	76%	37	24%	156
	No	26	76%	8	24%	34
	Unsure	23	53%	20	47%	43
Total		168	72%	65	28%	233
		Yes		No		Total
Browser, application, or device in use	Yes	119	74%	42	26%	161
	No	26	84%	5	16%	31
	Unsure	23	55%	19	45%	42
Total		168	72%	66	28%	234

		Yes		No		Total
Location	Yes	127	79%	46	21%	173
	No	26	75%	8	25%	34
	Unsure	14	55%	11	45%	25
Total		167	72%	65	28%	232

		Yes		No		Total
Address book contacts	Yes	88	79%	23	21%	111
	No	45	73%	15	27%	60
	Unsure	34	55%	28	45%	62
Total		167	72%	66	28%	233

		Yes		No		Total
Usage of Google products (YouTube, Maps, Drive) and features	Yes	121	75%	40	25%	161
	No	27	77%	8	23%	35
	Unsure	20	53%	18	47%	38
Total		168	72%	66	28%	234

		Yes		No		Total
Purchases or financial transactions through Google products	Yes	103	74%	36	26%	139
	No	40	77%	12	23%	52
	Unsure	24	59%	17	41%	41
Total		168	72%	65	28%	233

		Yes		No		Total
Details about your online session	Yes	111	75%	37	25%	148
	No	32	76%	10	24%	42
	Unsure	25	58%	18	42%	43
Total		168	72%	65	28%	233

		Yes		No		Total
Details about your smart phone usage	Yes	98	78%	28	22%	126
	No	43	78%	12	22%	55
	Unsure	27	51%	26	49%	53
Total		168	72%	66	28%	234

		Yes		No		Total
Information from advertisers, app developers, and games from when you are not using Google	Yes	108	78%	30	22%	138
	No	36	72%	14	28%	50

	Unsure	24	52%	22	48%	46
Total		168	72%	66	28%	234

		Yes		No		Total
Information from non-Google apps and advertisers	Yes	93	78%	27	23%	120
	No	40	77%	12	23%	52
	Unsure	34	56%	27	44%	61
Total		167	72%	66	28%	233

		Yes		No		Total
Voice and audio information when you use audio features	Yes	95	76%	30	24%	125
	No	42	76%	13	24%	55
	Unsure	30	57%	23	43%	53
Total		167	72%	66	28%	233

A chi-square test was conducted to test H_2, the relationship between reading Google's privacy policy and user awareness of the actions collected by Google. The test revealed three insignificant relationships between those who read the policy and being aware location data collected, X^2 (2, N = 232) = 3.681, p .159, financial transactions, X^2 (2, N = 232) = 4.614, p .100, and details about their online session, X^2 (2, N = 233) = 5.135, p .077. For all other data points, those who read the privacy policy were more likely to be aware of data collected.

4.4.2.2 YouTube

Table 32

YouTube Policy Reading Habits

	n	Yes	No	Unsure
Do you agree to terms of service when using platform?	396	69.91%	15.15%	18.94%
Does YouTube have a privacy policy?	395	61.77%	11.65%	26.58%
Of those saying yes, have you read/skimmed the policy?	242	66.12%	33.88%	n/a

When compared with Google, more YouTube users (62%) believe YouTube has a privacy policy and 66% claim to have read the policy. On average, the subset of responders who claim to have read the privacy policy answered the influencing and data collection questions at 75% correct. Almost half (43%) of the respondents scored worse than the overall average. The YouTube and Google privacy policies are the same. Of the 168 respondents who claim to have read the Google privacy policy, 15 respondents claim to have not read the YouTube privacy policy. Again, of those 168 for Google, 23 respondents claimed they did not know if YouTube had a privacy policy. Table 33 presents a breakdown between those who reported reading the privacy policy and the user knowledge of actions that influence search results.

Table 33

Crosstabulation of Self-Reported Reading of Privacy Policy and Actions that Influence Video Results

		Have you read or skimmed YouTube's privacy policy?				
Do any of the following actions influence your video results?		Yes		No		Total
Your actions on YouTube	Yes	107	73%	39	27%	146
	No	41	58%	30	42%	71
	Unsure	11	46%	13	54%	24
Total		159	66%	82	34%	241
		Yes		No		Total
Your actions on websites other than YouTube	Yes	93	78%	27	23%	120
	No	48	60%	32	40%	80
	Unsure	18	44%	23	56%	41
Total		159	66%	82	34%	241
		Yes		No		Total
Details about your online session	Yes	98	74%	34	26%	132
	No	40	60%	27	40%	67
	Unsure	20	49%	21	51%	41
Total		158	66%	82	34%	240

		Yes		No		Total
Details about your smart phone usage	Yes	95	78%	27	22%	122
	No	44	63%	26	37%	70
	Unsure	19	40%	28	60%	47
Total		158	66%	81	34%	239

		Yes		No		Total
Actions taken by other users	Yes	79	73%	29	27%	108
	No	51	65%	28	35%	79
	Unsure	29	54%	25	46%	54
Total		159	66%	82	34%	241

		Yes		No		Total
Actions taken by YouTube engineers/editors/curators	Yes	88	76%	28	24%	116
	No	44	61%	28	39%	72
	Unsure	24	48%	26	52%	50
Total		156	66%	82	34%	238

		Yes		No		Total
Data details between applications	Yes	89	75%	30	25%	119
	No	42	64%	24	36%	66
	Unsure	27	49%	28	51%	55
Total		158	66%	82	34%	240

		Yes		No		Total
Details about accounts across devices	Yes	84	75%	28	25%	112
	No	43	63%	25	37%	68
	Unsure	29	50%	29	50%	58
Total		156	66%	82	34%	238

A chi-square test was conducted to test H_2, the relationship between reading YouTube's privacy policy and user awareness of the actions that influence their video results or autoplay.

The test revealed only significant relationships between those who read the policy and being aware of the actions of other users on their video results. Users who read the privacy policy were more likely to be aware of the actions that influenced their video results or autoplay.

Table 34

Crosstabulation of Self-Reported Reading of Privacy Policy and Personal Data Collected About the User – YouTube

		Have you read or skimmed YouTube's privacy policy?				
Is the following personal data collected about you while using YouTube?		Yes		No		Total
Interactions with videos (likes, comments, subscribes)	Yes	122	68%	57	32%	179
	No	10	33%	20	77%	30
	Unsure	17	53%	15	47%	32
Total		159	66%	82	34%	241
		Yes		No		Total
Browser, application, or device in use	Yes	108	69%	48	31%	156
	No	33	66%	17	34%	50
	Unsure	17	52%	16	48%	33
Total		158	66%	81	34%	239
		Yes		No		Total
Location	Yes	112	70%	48	30%	160
	No	29	67%	14	33%	43
	Unsure	16	44%	20	56%	36
Total		157	66%	82	34%	239
		Yes		No		Total
Address book contacts	Yes	87	80%	22	20%	109
	No	42	62%	26	38%	68
	Unsure	30	47%	34	53%	64
Total		159	66%	82	34%	241
		Yes		No		Total
Usage of Google products (YouTube, Maps, Drive) and features	Yes	107	72%	41	28%	148
	No	31	61%	20	39%	51

	Unsure	21	50%	21	50%	42
Total		159	66%	82	34%	241

		Yes		No		Total
Purchases or financial transactions through Google products	Yes	92	72%	36	28%	128
	No	39	68%	18	32%	57
	Unsure	27	50%	28	50%	55
Total		158	66%	82	34%	240

		Yes		No		Total
Details about your online session	Yes	98	69%	44	31%	142
	No	36	69%	16	31%	52
	Unsure	23	51%	22	49%	45
Total		157	66%	82	34%	239

		Yes		No		Total
Details about your smart phone usage	Yes	95	76%	30	24%	125
	No	38	67%	19	33%	57
	Unsure	26	44%	33	56%	59
Total		159	66%	82	34%	241

		Yes		No		Total
Information from advertisers, app developers, and games from when you are not using YouTube	Yes	105	76%	34	24%	139
	No	28	68%	13	32%	41
	Unsure	25	42%	35	58%	60
Total		158	66%	82	34%	240

		Yes		No		Total
Information from non-YouTube apps and advertisers	Yes	97	78%	28	22%	125
	No	36	71%	15	29%	51
	Unsure	25	39%	39	61%	64
Total		158	66%	82	34%	240

		Yes		No		Total
Voice and audio information when you use audio features	Yes	100	76%	31	24%	131
	No	34	62%	21	38%	55
	Unsure	22	42%	30	58%	52
Total		156	66%	82	34%	238

A chi-square test was conducted to test H_2, the relationship between reading YouTube's privacy policy and user awareness of the actions collected by YouTube. The test revealed three insignificant relationships between those who read the policy and being aware of interactions with videos, X^2 (2, N = 241) = 2.740, p .254, browser, application, or device in use, X^2 (2, N = 239) = 3.817, p .148, and details about their online session, X^2 (2, N = 239) = 5.229, p .073 data points collected. For all other data points, those who read the privacy policy were more likely to be aware of data collected.

4.4.2.3 Facebook

Table 35

Facebook Policy Reading Habits

	n	Yes	No	Unsure
Agree to terms of service when using platform	396	76.01%	8.84%	15.15%
Do they have a privacy policy	394	69.29%	11.68%	19.04%
Of those saying yes, have you read/skimmed the policy	270	67.78%	32.22%	n/a

Of the three platforms, Facebook users were the most aware (69%) of the platform having a privacy policy and 68% claim to have read the policy. Of the 183 respondents who stated they have read or skimmed Facebook's data policy they scored, on average, an 80% on the influencing newsfeed and personal data collection questions correctly while 43% respondents scored lower than the overall average.

Table 36

Crosstabulation of Self-Reported Reading of Privacy Policy and Actions that Influence Facebook Newsfeeds Results

Do any of the following actions influence your newsfeed?		Have you read or skimmed Facebook's privacy policy?				
		Yes		No		Total
Your actions on Facebook	Yes	119	70%	52	30%	171

	No	44	62%	27	38%	71
	Unsure	18	69%	8	31%	26
Total		181	68%	87	32%	268

		Yes		No		Total
Your actions on websites other than Facebook	Yes	102	73%	38	27%	140
	No	52	63%	31	37%	83
	Unsure	26	59%	18	41%	44
Total		180	67%	87	33%	267

		Yes		No		Total
Details about your online session	Yes	104	73%	38	27%	142
	No	53	65%	29	35%	82
	Unsure	24	56%	19	44%	43
Total		181	68%	86	32%	267

		Yes		No		Total
Details about your smart phone usage	Yes	95	75%	32	25%	127
	No	55	64%	31	36%	86
	Unsure	31	56%	24	44%	55
Total		181	68%	87	32%	268

		Yes		No		Total
Actions taken by other users	Yes	104	74%	37	26%	141
	No	47	62%	29	38%	76
	Unsure	30	59%	21	41%	51
Total		181	68%	87	32%	268

		Yes		No		Total
Actions taken by Facebook engineers/editors/curators	Yes	102	73%	37	27%	139
	No	45	62%	28	38%	73
	Unsure	33	61%	21	39%	54
Total		180	68%	86	32%	266

		Yes		No		Total
Data details between applications	Yes	98	73%	37	27%	135
	No	51	65%	27	35%	78
	Unsure	32	58%	23	42%	55
Total		181	68%	87	32%	268

		Yes		No		
Details about accounts across devices	Yes	107	77%	32	23%	139
	No	41	59%	28	41%	69
	Unsure	33	55%	27	45%	60
Total		181	68%	87	32%	268

A chi-square test was conducted to investigate H_2, the relationship between reading Facebook's privacy policy and user awareness of the actions that influence their newsfeed. The test revealed one significant relationship, X^2 (2, N = 268) = 12.026, p .002. Based on the crosstab in Table 36, those users who have read the privacy policy are more aware of how their actions across devices influence their newsfeed than users who did not read it. The remaining variables do not have a statistically significant relationship, meaning that reading the privacy policy does not improve their awareness of how personal information influences their newsfeed.

Table 37

Crosstabulation of Self-Reported Reading of Privacy Policy and Personal Data Collected About the User – Facebook

		Have you read or skimmed Facebook's privacy policy?				
Is the following personal data collected about you while using Facebook?		Yes		No		Total
Interactions on posts, Pages, in Groups, and hashtags	Yes	134	69%	60	31%	194
	No	27	69%	12	31%	39
	Unsure	21	58%	15	42%	36
Total		182	68%	87	32%	269
		Yes		No		Total
Frequency of communication to people and groups	Yes	126	71%	52	29%	178
	No	30	63%	18	37%	48
	Unsure	26	60%	17	40%	43
Total		182	68%	87	32%	269
		Yes		No		Total
Address book contacts	Yes	110	72%	43	28%	153
	No	44	70%	19	30%	63
	Unsure	28	53%	25	47%	53
Total		182	68%	87	32%	269
		Yes		No		Total
	Yes	132	72%	51	28%	183

Usages of Facebook products (Instagram, Messenger, WhatsApp) and features	No	27	68%	13	32%	40
	Unsure	22	51%	21	49%	43
Total		181	68%	85	32%	266

		Yes		No		Total
Purchases or financial transactions through Facebook products	Yes	112	70%	47	30%	159
	No	36	71%	15	29%	51
	Unsure	34	58%	25	42%	59
Total		182	68%	87	32%	269

		Yes		No		Total
Details about your online session	Yes	111	70%	47	30%	158
	No	35	71%	14	29%	49
	Unsure	34	57%	26	43%	60
Total		180	67%	87	33%	267

		Yes		No		Total
Details about your smart phone usage	Yes	104	73%	38	27%	142
	No	41	71%	17	29%	58
	Unsure	36	54%	31	46%	67
Total		181	67%	86	33%	269

		Yes		No		Total
Information from advertisers, app developers, and games from when you are not using Facebook	Yes	119	78%	46	22%	165
	No	31	78%	14	22%	45
	Unsure	31	51%	27	49%	58
Total		181	72%	87	28%	268

		Yes		No		Total
Information from non-Facebook apps and advertisers	Yes	111	77%	34	23%	145
	No	33	65%	18	35%	51
	Unsure	36	52%	33	48%	69
Total		180	68%	85	32%	265

A chi-square test was conducted to test H_2, the relationship between reading Facebook's privacy policy and user awareness of the actions collected by Facebook. Opposite of Google and YouTube, the test revealed three significant relationships between those who read the policy and

details about their smart phone usage, X^2 (2, N = 267) = 8.219, p .016, information from advertisers, app developer, and games from when they are not using Facebook, X^2 (2, N = 268) = 6.870, p .032, and information from non-Facebook apps, X^2 (2, N = 265) = 13.053, p .001 data points collected. For all other data points, reading the privacy policy did not have a statistical impact on awareness of data collected.

4.4.2.4 Reliability of Information Sources

Three sections of the survey questions were to gain a better understanding of user knowledge of how their personal information influences their platform experience, it is also important to know how reliable the users believe the platforms are. As shown in previous results, the knowledge of personal data collection and influence is fairly consistent across platforms, which leads to questions about reliability. The survey asked a series of questions about search engine bias, if search results are the same for different people, and asked to rank the reliability of several sources of information. The question about search engine bias (Table 38) is taken from a 2012 Pew Research study on American's perceptions of information found on the internet. The results of my dissertation survey show a healthy amount of skepticism with regard to search engines as a source of fair and unbiased information.

Table 38

Search Engine Bias

In general, do you think internet search engines are a fair and unbiased source of information?		
Yes	144	36%
No	82	21%
Depends	126	32%
Don't know	43	11%

The results in Table 39 demonstrate a different approach to asking about how Google displays information. Participants were asked "do two people get the same search results if they entered the same search terms at the same time on Google?" The answers were nearly split equally between the three answers, further demonstrating the uncertainty of how search results are algorithmically curated.

Table 39

Do two people get the same search results if they entered the same search results at the same time on Google?

Same	117	29.55%
Different	125	31.57%
Unsure	136	34.34%

As a follow-up, participants were then asked why they chose their answer. While this was not a required question, 146 respondents provided an explanation. Table 40 provides direct quotes from each answer category that are illustrative of the responses. What is notable about the responses is the lack of trust in the platform in the different and unsure categories and the desire to test the scenario prior to answering definitively.

Table 40

Why?

Same	Different	Unsure
Same information as given to everyone	Their history	I'm not 100% sure of any things on the internet
Because it's the same question should be the same answer	Depends on data sourced	Don't know
Information same for everyone	I really don't know why but I see it happens	Never tried it so I have no way of knowing

	Common sense	Because they spell out different opinions to keep us confused about what is really going on with world views	I have no idea I have never heard of this before
	That's a good search engine	Because I heard it was based on their profile or do something	I don't know and I don't trust any of these platforms. They are always making tons of money somehow.

As with exploratory research, a pattern began to emerge when analyzing the textual data. There are four main categories of belief with regard to the construction of search results. With 309 textual replies, 16% of the respondents indicate some awareness of something influencing their search results. While 10% of participants believe search results should be the same for everyone and 9% believe since they have not tested the scenario themselves, they cannot answer. The largest category, 26%, do not know or do not have the words to explain why search results may or may not be the same for two different people.

Table 41

Search Result Belief Frequency

Beliefs	*Responses*	*Summary*
"It's tailored that way"	48	Respondents convey awareness of personalization in one way or another. Indicating knowledge of something influencing search results.
"Because it's the same question should be the same answer"	31	Respondents explicitly state searchers should receive the same information if searching the same thing. This belief showed up in the "same" and "unsure" category.
"Because I haven't tested it out to see if two people get the same search results."	28	Respondents expressed the desire to test the scenario to be able to answer effectively. This belief appeared entirely in the "unsure" category.

"I don't know"	81	Respondents either do not feel comfortable explaining or do not have the words to explain their choice.

Given the amount of information available to the average internet user, the survey asked participants to rate the following types of information platforms or sources by reliability as represented in Table 42. It is notable that respondents rated the choice "family, friends, and colleagues" as the most reliable of all the sources while they rated Facebook as the least reliable source.

Table 42

Source Reliability

How reliable are different sources of information?	Very reliable	Mostly reliable	Somewhat reliable	Unreliable	N/A
Family, friends, colleagues	23%	30%	30%	9%	8%
Television news	20%	30%	29%	14%	7%
Newspapers	19%	27%	33%	12%	8%
Cable television news	19%	29%	31%	14%	7%
Facebook	18%	16%	30%	31%	6%
YouTube	18%	21%	35%	18%	8%
News radio	17%	28%	37%	10%	8%
Talk radio	17%	21%	35%	17%	10%
Wikipedia	16%	20%	29%	19%	17%
Online news	15%	25%	39%	13%	8%
Twitter	13%	19%	21%	29%	18%
Parler	10%	14%	15%	22%	39%

4.4.4.5 Privacy-Focused Users

As reported earlier, 24 participants indicated they only use privacy-focused search engines, Brave, DuckDuckGo, StartPage, or a combination of the three. The demographics of this group include participants who self-identify as men (50%), women (42%), non-binary (8%). They are aged 25 to 64 years old (74%) and have either a high school diploma (29%) or attended

some college or university (38%). The participants report their political affiliation as Republican (46%), Democrat (33%), and Independent (13%).

A chi-square test was conducted to learn the relationship between privacy-focused search engine users and their knowledge of platforms' use personal data to customize results or newsfeed. The only statistically significant relationship existed between these users and Facebook, X^2 (2, N = 391) = 9.168, p .010. Privacy-focused search engine users are more likely to believe Facebook does not use personal information to customize their newsfeed than those users of other types of search engines.

Further investigation into the specific types of personal information collected using chi-square tests found several statistically significant relationships. Privacy-focused users are more likely to believe personal information is not collected from other Google products, X^2 (2, N = 393) = 11.113, p .004. They are also more likely to believe personal information is not collected about their online session X^2 (2, N = 395) = 8.358, p .015, or smart phone X^2 (2, N = 395) = 12.916, p .002, or non-Google applications, X^2 (2, N = 394) = 14.064, p .001. For YouTube and Facebook, there are no statistically significant associations for any specific information that might influence the search results or autoplay, or newsfeed on the platforms.

Chi-square tests were performed about how personal information influences search results or newsfeeds. A statistically significant relationship between search engine users and Google using smartphone information to influence search results, X^2 (2, N = 395) = 12.916, p .002 where privacy-focused search engine users are less likely to believe that smart phone details influence search results. For YouTube, there is no statistically significant relationship for any specific information that might influence the search results or autoplay on the platform. Facebook users demonstrated one statistically significant relationship. Privacy-focused search

engine users are less likely to believe that usage of other Facebook products (Messenger, Instagram, WhatsApp) might influence their newsfeed, X^2 (2, N = 391) = 8.007 p .018.

Looking further into this group, a chi-square test identified a statistically significant relationship between privacy-focused search engine users and the reliability of Parler, X^2 (4, N = 392) = 22.500, $p < .000$. Those using privacy-focused browsers were more likely to say Parler is mostly reliable. This was the only statistically significant relationship among the information sources and privacy-focused search engine users.

4.5 Demographic Variables and Platform Knowledge

This section answers research question four which asks, how does user knowledge about algorithmically delivered content vary based on demographic characteristics, including age, gender, race, education level, and political affiliation? The following research hypotheses are explored through chi-square tests run at a 95% confidence level.

H_{4A}: There is an association between age and users' awareness of algorithmic curation using [personal information].

H_{4B}: There is an association between gender and users' awareness of algorithmic curation using [personal information].

H_{4C}: There is an association between race and users' awareness of algorithmic curation using [personal information].

H_{4D}: There is an association between education and users' awareness of algorithmic curation using [personal information].

H_{4E}: There is an association between political affiliation and users' awareness of algorithmic curation using [personal information].

Table 5 provides a summary of personal information by platform. Some variables are uniform across the platforms while others are specific. Due to the validation requirements of crosstabs some demographic data was collapsed to ensure data validity during the tests. Age was consolidated into four groups: 18-24, 25-44, 45-59, and over 60 years of age. Education levels were simplified into college graduate and not a college graduate. Political affiliation was simplified by removing those participants who identified as Democrat, Republican, or Independent. Gender was consolidated into the binary of man and woman.

4.5.1 Usage Levels

As platform usage had an impact on the users' overall awareness of algorithmically curated content, this section investigates whether there are any demographic associations with platform usage.

4.5.1.1 Google

Chi-square tests were conducted to examine the relationship between demographic variables and search engine use frequency. Education and search were the only statistically significant relationship, X^2 (5, N = 394) = 18.020, p .003. Users who identified as those without at least an undergraduate degree demonstrated lower usage of search engines.

4.5.1.2 YouTube

Chi-square tests were performed to examine the relationship between demographic variables and YouTube usage. There were no statistically significant relationships between demographics and YouTube usage. Demographic variables did not show associations with how frequently or how long YouTube users visited the platform.

4.5.1.3 Facebook

The relationship between political affiliation and the frequency of checking their Facebook timeline was significant, X^2 (6, N = 273) = 16.281, p .012. Republicans demonstrated more frequent checking of their Facebook timeline than Democrats or Independents. However, there is a not a statistically significant relationship between political affiliation and time spent on Facebook, X^2 (6, N = 271) = 4.051, p .670.

4.5.2 Policy Reading Habits

Age is the only demographic characteristic with an association between the relationship of the respondents' awareness of the existence of a privacy policy and reported reading of the privacy policy. No other demographic characteristic had statistically significant results.

Google users under 59 years of age are less aware of a privacy policy in place, X^2 (6, N = 394) = 28.673, p <.000. Comparatively, YouTube users under 60 years of age are more aware of their privacy policy, X^2 (6, N = 393) = 30.410, p <.000. However, for Facebook the younger users were less aware of the existence of a privacy policy, X^2 (6, N = 392) = 39.280, p <.000. Facebook users 45-59 were aware of the privacy policy, while users 18 to 44 were more likely to believe one did not exist as opposed to being unsure of one existing.

Of those who answered 'yes' to the existence of a privacy policy, they were asked if they have read or skimmed the policy. Again, age was the only statistically significant demographic characteristic. For Google users 18 to 44, they claim to have read or skimmed the privacy policy more so than those older than 45 years of age, X^2 (3, N = 241) = 7.898, p .048. YouTube users 25-59 claim to have read or skimmed the policy, X^2 (3, N = 241) = 9.582, p .022. Facebook users had no significant relationship between age and reading of the privacy policy.

4.6.3 Personal Information Used, Influenced, and Collected

Three questions were asked of the participants to gauge their level of awareness of personal information used to customize their online experience, what facets of personal data collected might be used to influence that experience, and what facets of personal information are collected while using the platform. The following works to bring conclusions to $H_{4A\text{-}4E}$.

4.5.3.1 Google

For Google users, significant relationships appear between age, gender, and education. The age relationship appears between those 25-59 being more aware that personal information is used and those over 60 being unsure of whether personal information is used to during their search experience, X^2 (6, N = 393) = 33.506, p <.000. Users under 24 years old do not believe personal information is used. The chi-square test between Google and gender also demonstrates a statistically significant relationship, X^2 (2, N = 389) = 15.559, p <.000. Women were less likely to believe personal information is used to customize their search experience and were more likely to feel unsure about the use of personal information. The chi-square test for education demonstrated a statistically significant relationship, X^2 (2, N = 394) = 14.614, p .001. Those users with at least an undergraduate degree are more likely to know that personal information is used to construct search results.

During the analysis of specific personal data collection points determined by the privacy policy analysis, a chi-square test revealed a statistically significant relationship between political affiliation and awareness of data collected by other Google products, X^2 (4, N = 369) = 16.198, p .003. Independents are more likely to be aware while Democrats are less likely to be aware and Republicans are more likely to be unsure of this data collection practice. Another statistically significant relationship appeared between political affiliation and the awareness of voice and

audio data collected. X^2 (4, N = 369) = 13.731, p .008. Those users identifying as Democrat are less likely to believe this information is collected while Independents and Republicans are more likely to be unsure.

Gender held three statistically significant relationships with specific data collection points or potential influences, browser, app device in use collected, X^2 (2, N = 392) = 8.391, p .015, location collected, X^2 (2, N = 389) = 8.605, p .014, and details about online session collected, X^2 (2, N = 391) = 11.493, p .003. For all data points, those users identifying as men are more likely to be aware of the influence or data collection whereas those users identifying as women are more likely to be unsure. However, this is only true for Google. YouTube and Facebook chi-square tests revealed no statistical significance between gender and either data collection or potential data influences.

4.5.3.2 YouTube

The chi-square test conducted to examine the relationship between YouTube and demographic characteristics revealed a statistically significant relationship to age, X^2 (6, N = 391) = 34.466, p <.000. The younger users (18-59) are more aware of their personal information being used while the older users (60+) are more likely to be unsure of personal information is used by the platform. These results mimic those of Google.

Age continues to be the significant factor when examining the relationship between demographic variables and personal data points that may influence users' online experience. These data points are actions on websites other than YouTube, X^2 (6, N = 393) = 41.922, p < .000; details of the online session, X^2 (6, N = 392) = 32.419, p <.000; smart phone usage, X^2 (6, N = 394) = 30.043, p <.000; actions taken by other users, X^2 (24, N = 396) = 44.967, p .006; actions of engineers, X^2 (6, N = 394) = 22.002, p .001; between applications, X^2 (6, N = 394) =

27.244, p <.000; across devices, X^2 (6, N = 392) = 29.226, p <.000. For all variables listed, users aged 18-54 are more likely to be aware of the personal data points that may be used to influence their online experience.

The following personal data points have statistically significant relationships when examining demographic variables and personal data points collected while using the platform. These data points are interactions with videos, X^2 (6, N = 393) = 26.803, p <.000; browser or device in use, X^2 (6, N = 390) = 27.633, p <.000; location, X^2 (6, N = 391) = 26.765, p <.000; address book, X^2 (6, N = 393) = 35.407, p < .000; usage of Google products, X^2 (6, N = 393) = 24.005, p .001; purchases, X^2 (6, N = 391) = 24.346, p <.000; other applications, X^2 (6, N = 391) = 32.214, p < .000; and voice and audio, X^2 (6, N = 388) = 32.042, p <.000.

Again, age is the only significant relationship. Users aged 18-24 are more likely to believe that the listed personal data points are not collected, users aged 25-54 are more likely to be aware of the data collected, and users over 55 are unsure. There is a conflict between those respondents who are younger having an awareness that personal information is used to influence their online experience, but more unaware of the personal data collected.

4.5.3.3 Facebook

A chi-square test was conducted to examine the relationship between demographic characteristics and whether or not the users believed the platform uses their personal information to customize their online experience. There was a statistically significant relationship between age and personal data that influences the newsfeed, X^2 (6, N = 388) = 12.699, p .048. Users 18-24 and 45-59 are more likely to be aware of actions taken by engineers, whereas users 25-44 are unaware, and users aged 60 and over are unsure.

There were statistically significant relationships between age and personal data collected, purchase information, X^2 (6, N = 391) = 29.937, p <.000 and smart phone usage, X^2 (6, N = 389) = 21.730, p .001. For both relationships, users aged 18-24 are more likely to be unaware of data collected, users aged 25-59 are more likely to be aware, and those over 60 are unsure.

4.5.4 Reliability of Information Sources

This section provides further investigation into demographic variables through the lens of source reliability. This question combines the three studied platforms with other sources of information retrieval. A high-level hypothesis was developed to learn more about these associations.

H_5: There is an association between specific demographic variables and specific sources of information reliability.

Table 43 indicates the statistically significant associations between demographic variables and information sources.

Table 43

Chi-square test results examining the relationship between demographic variable and source of information reliability

How reliable are different sources of information?	Age	Gender	Education	Political affiliation
Search engine	X^2 (12, N = 391) = 27.962, p .006			X^2 (8, N = 364) = 16.373, p .037
News radio	X^2 (12, N = 390) = 26.292, p .010		X^2 (4, N = 392) = 11.688, p .020	X^2 (8, N = 364) = 18.30, p .019
Talk radio	X^2 (12, N = 390) = 45.692, p .000			

Family, friends, colleagues	X^2 (12, N = 390) = 35.533, p .000			X^2 (8, N = 364) = 19.607, p .012
Online news	X^2 (12, N = 390) = 23.560, p .023			X^2 (8, N = 364) = 19.563, p .012
Newspaper	X^2 (12, N = 390) = 26.720, p .008			X^2 (8, N = 362) = 42.716, p .000
Television news	X^2 (12, N = 388) = 25.747, p .012			X^2 (8, N = 362) = 39.336, p .000
Cable news		X^2 (4, N = 388) = 11.079, p .026		X^2 (8, N = 362) = 48.572, p .000
Facebook	X^2 (12, N = 391) = 47.871, p .000	X^2 (4, N = 386) = 21.237, p .000		X^2 (8, N = 360) = 18.079, p .021
YouTube	X^2 (12, N = 389) = 38.689, p .000	X^2 (4, N = 384) = 25.781, p .000		
Twitter	X^2 (12, N = 389) = 50.964, p .000		X^2 (4, N = 387) = 12.535, p .014	X^2 (8, N = 361) = 15.718, p .047
Parler	X^2 (12, N = 387) = 47.459, p .000	X^2 (4, N = 382) = 14.577, p .006	X^2 (4, N = 385) = 13.338, p .010	
Wikipedia	X^2 (12, N = 387) = 37.519, p .000		X^2 (4, N = 385) = 17.630, p .001	

Age continues to be the demographic variable with the most significant relationships. Those aged 35-54 are likely to rate YouTube are very reliable, while those aged 55-75 and older rated newspapers unreliable. Similarly, those aged 25-54 rated Facebook very reliable and those older than 55 rated the platform unreliable. This pattern continues for all information sources, even more traditional media like newspapers and television news. The chi-square tests between race and reliability of information source did not reveal any significant relationships.

Gender has significant relationships with four information sources. Men are more likely to rate cable news as mostly reliable and women are more likely to rate it as somewhat reliable. Women are more likely to rate Facebook as mostly and somewhat reliable whereas men are more likely to rate it unreliable. Men are more likely to rate YouTube as unreliable whereas women are more likely to rate it as somewhat reliable. Men are more likely to rate Parler as very and mostly reliable whereas women are more likely to rate it as not applicable, or they are unaware of the platform.

Education levels have significant relationships with four information sources. Participants with a bachelor's degree are more likely to rate search engines as very reliable. Participants with a high school diploma or less are more likely to rate Facebook as very reliable. Participants with at least a bachelor's degree are more likely to rate Wikipedia as a mostly reliable source.

There is an overall skepticism of Republican participants when rating the reliability of information sources. Democrats are more likely to rate search engines as mostly or very reliable. Democrats are more likely to rate news and talk radio as mostly or very reliable. Republicans are more likely to rate family, friends, and colleagues are somewhat or mostly reliable. Republicans are more likely to rate online news as unreliable. Democrats are more likely to rate newspapers are mostly or very reliable. Democrats are more likely to rate television news are mostly or very reliable. Republicans are more likely to rate cable news as somewhat reliable or unreliable. Democrats are more likely to rate Facebook as mostly or very reliable. Republicans are more likely to rate Twitter as unreliable.

4.6 Conclusions

This chapter worked to answer four research questions:

RQ1: How do platforms (Google (1), YouTube (2), Facebook (3)) govern and inform their users with regard to using collected data to filter the users' online content?

RQ2: What do United States-based adult users of Google, YouTube, and Facebook know about the personal and usage data being collected and their awareness about how the platforms use this data to moderate and serve online content?

RQ3: Are there any associations between usage characteristics of United States-based adult users of these platforms and users' awareness of algorithmic curation using personal information?

RQ4: How does user knowledge about algorithmically delivered content vary based on demographic characteristics, including age, race, education level, and political affiliation?

I analyzed the privacy/data policies for Google, YouTube, and Facebook to identify how the platforms informed their users about the personal data collected that filters their search results and newsfeeds. Within this analysis I identified eight categories of influence and eleven categories of personal data collected. These categories create the framework for the survey questions to identify the knowledge and awareness of Google, YouTube, and Facebook users. Overall, the privacy/data policies use situational examples to explain the types of data they collect and how they use that data. The platforms do not provide an explicit list of all data points collected or how the platform could use or segment the data.

With a sample representative of the United States population, 396 US based adults completed a 60-survey question survey that asked about the knowledge of personal data

collected while using these platforms, and also how often they used the platform and if they read the privacy policy.

Regarding the awareness of how platforms use personal data to moderate and serve online content, participants of the survey were mostly unsure or did not know personal information was used in the construction of their results or newsfeeds. There is one exception to this, Facebook users are more aware of how their actions on Facebook influenced their experience. As a social media site, one where interaction is assumed and encouraged, this finding makes sense.

There are several discrepancies between the knowledge of personal data collected and if that personal data influences or filters the results or news feed. For example, 57% of respondents answered yes when asked if interactions with videos are collected while using YouTube; while only 47% said those interactions influenced video results. Another example, 53% of respondents answered yes when asked if details about their online session are collected while using Google; while only 44% said that details about their online session influence search results. Finally, 42% of respondents answered yes when asked if details about their smart phone usage are collected while using Facebook; while only 38% said that details about smart phone usage influenced newsfeeds. These discrepancies are consistent between three platforms.

Age is the most consistent demographic variable with an association to algorithmic awareness with those users younger than 55 years old generally being more aware of how personal information is used to curate their online experience, and those over 55 years old being unsure of if or how personal information is used.

4.6.1 Limitations

This study utilized a cross-section sampling method to gather a representative sample without researcher bias. The utilization of Qualtrics for a web-based survey distribution and the panel limits participation to those people who use the internet and participate in their research compensation program. The 60-question survey is on the longer end of survey design and as such did not allow for many follow-up questions to learn the full extent of the participants' understanding. As with any survey, participants are providing a self-assessment of their understanding which cannot be verified. Terms of Service and Privacy Policies change frequently, and this analysis is a snapshot of those policies. The framework for analysis is replicable on newer policies and other informational methods of platform governance.

Chapter 5: Discussion

The previous chapter presented the qualitative and quantitative findings from the data/privacy policy analysis and 60-question user survey. This chapter tells a story informed by the results, following the outline of the literature review, first of users' information practices through the context of information and algorithmic literacy. It then discusses how user beliefs and awareness of algorithmic personalization in this study compares to related work on algorithmic understanding. To further the discussion about information policy, an analysis was conducted of users' habits surrounding the reading of privacy policies to compare how recent proposed legislation might change how users view platform policies. The following section explores the theoretical and practical implications to this research, by unpacking the black box, suggesting new information literacy outcomes, responses to the proposed legislation, and recommendations for platforms. The first two sections provide more context while continuing to address research questions two and three, while the third section addresses research question one.

> **RQ1:** How do platforms (Google (1), YouTube (2), Facebook (3)) govern and inform their users with regard to using collected data to filter the users' online content?
>
> **RQ2:** What do United States-based adult users of Google, YouTube, and Facebook know about the personal and usage data being collected and their awareness about how the platforms use this data to moderate and serve online content?
>
> **RQ3:** Are there any associations between usage characteristics of United States-based adult users of these platforms and users' awareness of algorithmic curation using personal information?

RQ4: How does user knowledge about algorithmically delivered content vary based on demographic characteristics, including age, race, education level, and political affiliation?

5.1 User Information Practices

Users choose platforms for a variety of information types and sources. There are similarities within the participants of this study. Users of search engines are mostly using them to look up news on a topic or event, navigate to sites, answer medical or health questions, find entertaining content, and check the accuracy of news or information. Users of YouTube mostly use the app to keep up with family and friends, find entertaining content, share pictures and videos, and look up news on a topic or event. Users of Facebook mostly use it to keep up with family and friends, find entertaining content, share pictures and videos, and look up news on a topic or event. The respondents indicated they use all three platforms to look up news on a topic or event and find entertaining content.

The available data of the reasons for platform use and source reliability begin to provide insight into meaning making and the cognitive processes within information seeking. While only 36% of participants believe that search engines are a fair and unbiased source of information, they use them to find information on news, health questions, and fact checking. Even though users indicated they frequent Facebook and YouTube for the same reasons, participants found Facebook somewhat reliable to unreliable, and YouTube mostly reliable to somewhat reliable. It appears Facebook has a public stigma of being a questionable information source.

Although the most cited reason for using Facebook was to keep up with family and friends, users do not seem to equate the people and the platform when rating the reliability of Facebook. Facebook reliability is low, but reliability of friends and family is high. This indicates a departure from Wilson's (1981) model of information behavior wherein the system and human

information source were not well delineated (Knight & Spink, 2008; Xie, 2008). Presumably the users' Facebook account contains those family, friends, and colleagues. There is a disconnect iin the survey results between the people and the platform. This could be due to a heavy critique of Facebook in the news and watching executives testifying in Congress (2018, 2021) seeing as 42% of those who ranked Facebook as unreliable "feel like Facebook is discussed in the news" and 29% "keep up with news about Facebook."

There were 15 users who answered "never" when asked how often they use a search engine. This group identifies as a 35-44 (34%) or over 60 years old (34%) white (67%) woman (73%) with a high school education (67%). Her political identity could either be democrat, independent, or republican as they are equally represented. When asked what search engine they use, their responses were either mobile browser or URL bar input. Facets of information poverty appear within this subgroup through Chatman's (1996) first proposition, people perceive themselves to be devoid of information sources since they believe they never search even while using search functionality. Additionally, they do not seem to have the capability to use and understand the information source as they are not aware that they are using a search engine even as they claim to search in a browser (Britz et al., 2012).

These 15 users suggest that searchers are not explicitly aware that when searching via a smartphone or URL bar that they are accessing a search engine. It would require more research into how mobile users assess their information practices, but with nearly 60% of global internet traffic being mobile, this finding leads to a hypothesis that mobile search is obfuscating search awareness and evaluation techniques ("What percentage", 2022).

There is a subset of Facebook users who spend less time on the platform, both frequency of checking their newsfeed and minutes on the platform that are more aware of data collection

practices. I regret not asking for how long users had a Facebook account because I would surmise that the users who currently spend less time on the platform but are more aware of data collection practices are users who frequented the site 10-15 years ago and now use it sparingly.

5.2 User Awareness

Algorithmic awareness is not ubiquitous and is impacted by several factors (Cotter & Reisdorf, 2020; Klawitter & Hargittai, 2018). The presence of awareness is important as algorithms are acting as gatekeepers to information retrieval as we know they promote and repress information based on the user's profile (Gillespie, 2014, 2018; Halavais, 2018; Noble, 2018). Eslami et al. (2015) found that less than half (37.5%) of their participants were aware of algorithmic filtering on Facebook, whereas, nine years later this study asked if personal information is used to customize search results and newsfeeds and found 60% of Google, 58% of YouTube, and 63% of Facebook users are aware that algorithmic curation is happening. The increase of awareness could be due to the phrasing of the question, I did not use the phrase "algorithmic curation" but rather one of the facets of algorithmic curation, personal information. Another possibility is the social discourse surrounding Facebook during and after the 2016 presidential election. The Cambridge Analytica leak drew broad news coverage and outlined how personal data was used to create voter profiles with the intent of selling them to campaigns (Confessore, 2018).

5.2.1 Personal Information

A surprising disconnect in user awareness takes place between what they believe influences search results, autoplay, or newsfeeds and what data points they believe the platform collects about the user. For all three platforms, user awareness of personal information collected is higher than their awareness of how personal information influences their online experience.

For example, 13% more users believe Google collects details about smart phone use than believe that Google uses that data to influence search results. Even "online session" details were reported by participants to be collected more than used in algorithmic curation.

Cotter (2022) posits that the users' practical knowledge "captures how people locate and configure algorithms with their social world" (p. 15). Many factors impact users' social worlds, one of which is education. This dissertation study found that those with at least a bachelor's degree were more aware of algorithms using personal information to customize their online experience than those without a college education. Similarly, Cotter and Reisdorf (2020) found those with higher socioeconomic status were also more aware of algorithmic filtering.

For years the term "digital divide" referred to the access of digital services and while there is still a gap in access to high-speed internet in the United States, a new digital divide is growing – algorithmic awareness. Gran et al. (2021) argue that the awareness of how algorithmic systems impact access to information requires more attention. Their findings in the highly digitized country of Norway found that 62% of their participants (n=1624) had no or low awareness of algorithms. In comparison to the results of this dissertation study, 60% of participants have no or low awareness of influences on search results, autoplay, or newsfeeds. In a study about Facebook's newsfeed, researchers found that 44% of participants acknowledged an entity modifying their experience based on the users' interactions with the content (Rader & Gray, 2015). The concept of an entity or recognition of a system in place was also found in this study which is discussed further below.

In 2012 the Pew Research Institute conducted a survey of internet users learning about their demographic characteristics, socioeconomic statuses, how they used the internet, and their attitudes toward information found utilizing online searches (Purcell et al., 2012). One of the

questions repeated in this dissertation study showed a large change in trust toward search engines returning fair and unbiased information. In the near decade since the Pew study appeared, the belief that search engines provide fair and unbiased information has decreased by 30%. A more skeptical search user appeared as survey takers selecting "depends" increased by 29%. There are sampling differences between this study (*n*=396), an English web-based survey and the Pew study (*n*=1,729), landline and mobile telephone interviews conducted in Spanish and English. Comparing the results is meant to be illustrative about the how the discourse over the last decade has increased the awareness the term and practice of "misinformation" and "fake news," creating an online environment where some people are more skeptical of the information they receive. A Google Trends search for "fake news" shows a spike in October 2016 and "misinformation" in January 2020. The dissertation survey data was collected in January 2022.

One of the research hypotheses evaluated algorithmic awareness through the usage of the platform. There is overlap between participants who use search engines, YouTube, and Facebook and the results are reported as users of the specific platform. While those searchers who reported frequent use of search engines were more likely to be aware of personal information being used to customize results, the same result was found for users who check their YouTube feed more than twice a day and/or watch more than 30 minutes of video a day. The length of time spent on YouTube does not necessarily dictate the awareness of algorithmic curation. This could be due to the social nature of the platform where creators encourage engagement with their content through asking viewers to like, subscribe, and comment on their videos. This request of the creator indicates the value of these actions to both the creator and viewer. Facebook users who check their newsfeed more regularly were more likely to be aware of the use of personal

information to customize results. However, Facebook users who spend more time on the platform are less likely to be aware of the customization from personal information.

5.2.2 Algorithmic Curation of Search Results

The last questions of the survey asked, "do two people get the same search results if they entered the same search terms at the same time on Google" and "why." The data collected begins to attend to algorithmic folk theories surrounding search engine functionality and algorithmic curation.

Only 69 people (55%) who responded with "different" explained why the search results would be different. The majority (68%) of these explanations mentioned some form of personalization from usage history to location and several mentioned "the algorithm" specifically. A few of the respondents replied with "I don't know". The idea of "the algorithm" as an actor is a common thread in the comments, this external or nebulous thing that performs the work to provide the information needed.

Of the 80 people (68%) who responded with "same" and provided a follow-up explanation there was a strong feeling that the same question should elicit the same response as if there is only one correct answer per query. Almost as strongly, people responded with "I don't know" indicating that their initial choice was a gut feeling, or they could not articulate how or why the results might be the same. In this category, a few respondents indicated the results would be the same because they platform is "trustworthy".

Sixty people (44%) provided explanations for their response of "unsure" and from those explanations, three themes developed. First, they truly did not know and didn't want to overestimate their awareness. Second, they have not tested it and therefore could not prove it one

way or another. And finally, again, a few responses focused on not trusting the platform or its motives.

Table 44

Search Result Belief Frequency

Beliefs	*Responses*	*Summary*
"It's tailored that way"	48	Respondents convey awareness of personalization in one way or another. Indicating knowledge of something influencing search results.
"Because the same question should be the same answer"	42	Respondents explicitly state searchers should receive the same information if searching for the same thing. This belief showed up in the "same" and "unsure" category.
"Because I haven't tested it out to see if two people get the same search results."	35	Respondents expressed the desire to test the scenario to be able to answer effectively. This belief appeared entirely in the "unsure" category.
"I don't know"	55	Respondents either do not feel comfortable explaining or do not have the words to explain their choice.

A reminder of the four most common beliefs surrounding the question, "if search engines provide the same search results at the same time" appears above. Forty-eight of the responses indicated the awareness of personalization through the use of words and phrases like "tailoring" and "based on search history," again echoing similar findings in Power's (2017) work where only 1 participant (n=147) mentioned algorithms, most participants believed the display of information was due to popularity. A large portion of the respondents also felt that the same question should return the same answers, which is a logical, factual way to perceive search engines. And while this sentiment is largely true for questions with a specific answer; 15% of the respondents are using search engines to look up medical or health information which does not often have a straightforward answer and depends on the user's skill.

Location provides the most obvious influence to search results for the searcher while others are less obvious. A limitation of this study and future research opportunity is the inability to follow-up with the participants to dive deeper into their understanding of why search results

should be the same. Of the responses, 48 people specifically mentioned personalization of search results indicating either awareness or, since this question appeared at the end of the survey, learned something during other sections of the survey. The recognition of something impacting the users' platform experience is a common finding where awareness exists but the user does not have the vocabulary to name it (Rader & Gray, 2015; Eslami et al., 2016; Powers, 2017; Cotter, 2022). Respondents who answered "unsure" reported a desire (11%) to test the search engine to learn if the same results are delivered to different people at the same time. However, 57% of this group answered "yes" when asked if Google uses personal information to customize search results conflicting perhaps with their "unsure" assessment. The demographic composition of this group is a high school educated (74%) woman (66%) over the age of 45 (77%) with Republican political affiliation (49%). The skepticism among Republicans in this dissertation study has been consistent throughout as they've been more likely to answer "unsure" to awareness questions. The same is true for women who are not as confident as men in their algorithmic awareness.

Ytre-Arne & Moe (2021) developed five thorough folk theories through their interviews that lead to identifying how users find algorithms confining, practical, reductive, intangible, and exploitative. Within the responses from this dissertation study, there are phrases that indicate similar feelings like "I think they want everybody to be on the same wavelength" (confining), "I don't trust any of these platforms; they are always making money somehow" (exploitative), or "it is already programmed [that way]" (reductive).

In addition to participants explaining their awareness about how search engines return results, we see a group who want to test to know for certain whether two people can or do receive the same search results. Oeldorf-Hirsch and Neubaum (2022) identify three categories of human-algorithm interaction: knowing, feeling, and doing. The participants in this dissertation

study are motivated to “do” algorithmic interaction through the research of algorithmic functionality. The participants already know about their existence and have felt that the results vary based on input.

5.2.3 Privacy-Focused Users

The data suggests a compelling argument about the 24 users who claim to only use privacy-focused search engines. The usage of privacy-focused search engines is a form of identity-signaling more than a desire for privacy. In the last year, and while this data was collected, far-right publications like Breitbart and groups like QAnon promoted the use of DuckDuckGo, specifically, as an anti-censorship search engine. *The New York Times* reported a rise in popularity of DuckDuckGo when people were not finding Covid vaccine information that supported their viewpoints on Google (Thompson, 2022). Sites containing misinformation were more readily available on DuckDuckGo than Google, which aligns with some recent studies about misinformation ranking higher on Bing than Google as DuckDuckGo sources from Bing’s search results (Bush & Zaheer, 2019; Results Sources, 2022; Urman et al., 2022). With 14% of participants indicating they use search engines to check the accuracy of news or information, the absence of factual information in search results is problematic. The increase of misinformation on some search engines and not others is due to data voids. Data voids occur when a query does not have enough associated results allowing for manipulation of those results to specific search terms (Golebiewski & boyd, 2019).

Further investigation of the privacy-focused users into the reliability of information sources revealed a positive association between these users and the belief that Parler was mostly reliable as an information source. Parler is a social media platform popular with far-right, conspiracy theorists where posts are often racist, misogynistic, and antisemitic. Even though

Republicans are more likely to choose one of these search engines, Republicans represent only 46% of the privacy-focused users, with Democrats making up 33%, Independents 12%, and the remaining participants preferred not to disclose their political affiliations.

There is a lack of existing literature indicating a specific reason for why someone might choose a privacy-focused search engine. The evidence presented in this dissertation suggests that these users did not choose the privacy-focused search engines due to a desire for privacy. Rather, the evidence indicates the use of privacy-focused search engines are not about the value proposition of privacy, but due to an ideological power shift which leads to a group norm. "Authority is increasingly expressed algorithmically," in a black box society and when the algorithm is not providing the information desired, there is a shift in information systems (Pasquale, 2016, p. 8).

5.2.4 Demographic Association with Algorithmic Awareness

There is a consensus in the field of education and technology studies that digital natives are those who grew up interacting with technology from a young age. Depending on socioeconomic status, this could be as early as 1980 (Prensky, 2001). There is less of a consensus surrounding whether age and, therefore, exposure impacts the user's knowledge or awareness of the technology. An analysis of digital native discourse found too many assumptions are made about a generation of people. Age is only one of many factors to consider about a person like access to the internet and personal or political reasons for being informationally aware or choosing one technology over another (Selwyn, 2009). It is difficult to say definitively that young people and those now entering middle age are innately technologically aware. However, in this dissertation study, age is the primary demographic variable with statistical significance in algorithmic awareness tests.

For both Google and YouTube, users under the age of 59 (birth year of 1963 at the time of data collection) were more aware of personal data being used to curate their online experience than those 60 years of age and older. For Facebook users the data were more nuanced; their awareness differed based on data points collected and data points that influenced where users were more aware of information collected rather than if the information was used to curate their newsfeed. Overall, for personal information collected users 18-24 were unaware, users 25-54 were aware, and those over 60 were unsure of personal data points collected while using platforms. For data points that influence algorithmic curation users 18-24 were aware, users 23-35 were unaware, users 36-54 were aware, and those over 60 were unsure. Again, the disparity exists between the awareness of data collected and how it may be used.

Of the demographic characteristics tested, an investigation of gender and algorithmic awareness reported that women were more likely to be unsure of their awareness than men. The study conducted by Gran, Booth, & Bucher (2020) also found that men perceived their level of algorithmic awareness at a higher rate than women.

5.3 Platform Privacy and Data Policies

5.3.1 Platform Policy Analysis

The analysis of the privacy and data policies of Google and Facebook identified eight ways they tell their users what information they use to personalize their search results, autoplay, and newsfeed. The analysis also found 12 specific points of data collection outlined by Google and Facebook. Winkler & Zeadally (2016) identified seven methods of data collection of Facebook, Google, Twitter, and LinkedIn: personally identifiable information, user-generated content, device information, location information, payment information, off-platform activities, and how the user interacts with and uses the web platform. Their findings indicated that

Facebook collected data across all seven categories whereas Google's privacy policy at the time did not indicate the collection of user-generated content, payment information, and off-platform activities. Google has updated their privacy 12 times since the publication of the Winkler & Zeadally article and the policy collection for this dissertation. The privacy policy analysis conducted for this dissertation found that Google now collects user-generated content, payment information, and off-platform activities. The privacy policies analyzed for this dissertation study found that Google and Facebook's privacy policies are similar in their layout and content.

Privacy policy standardization likely stems from the European Union's General Data Protection Regulation (GDPR) which requires businesses collecting data from any EU citizen to disclose what information they collect, how they use it, and how to remove their data. Additionally, websites and platforms had to show a pop-up so that users could agree or decline the use of tracking cookies. For United States users, this EU regulation made tracking cookies more obvious as platforms applied the pop-up globally. States are beginning to pass legislation that requires platforms to disclose their personal data collection policies and provide users a way to request their data and delete it. A leader in this area is California, which passed their privacy legislation in 2018 (California Consumer Privacy Act, 2018).

Digital advertising is a large portion of Google and Facebook's revenue, and the advertisements appear in search results, during videos, and in newsfeeds indicating that the data collected is used ("Google Third Quarter", 2022; "Meta Third Quarter", 2022). Mager's (2012) work investigating how "capitalist ideology gets inscribed in search algorithms by way of social practices" extrapolates on this by determining that Google profits from user data (p. 770). What is more than that is how the social practices of the platforms are described through capitalistic practices in their privacy policies. Overtime there has become a cyclical practice of first

prescribing the data collection practices, and then recognizing user behavior to identify additional data points.

5.3.2 Policy Reading Habits

Consumers are inundated with usage policies from apps and platforms, and they overstate how thoroughly they read the usage policies. The documents are lengthy, often repeat similar verbiage, and are seen as a nuisance (Obar & Oeldorf-Hirsch, 2020). One of the current proposed bills, Terms-of-service Labeling, Design, and Readability (TLDR) Act, works to shorten or simplify Terms of Service to help the consumer not be overwhelmed with information (S.3501, 2022). But the consumer still needs to be informed of how the platform company is collecting and using personal data to curate results and to serve other algorithmically determined content. Additionally, with more time spent on the platform the community norms of the platform often drive what is known about how the platform operates. The policy reading habits and awareness vary slightly by platform and this could be due to how the technology is conceptually understood by the user (Martin, 2012). Those claiming to have read the privacy policy of Google (72%) and YouTube (66%) are statistically more aware of the relationship regarding actions that influence search and video results pages. Ibdah et al. (2021) reported that 77% of their participants read or attempted to read privacy policies at least once and found a positive correlation between policy reading habits and knowledge of data collection practices. In comparison, Pew Research (2019) found that "one in five Americans say they always or often read privacy policies before agreeing to them." The participants in my survey reported reading the privacy policies at higher rates, likely because that is what they think I want to hear, or they should be doing.

5.3.1.1 Google and YouTube

If the participants answered "yes" when asked if the platform has a privacy policy, they were then asked a follow-up question to see if they had either read or skimmed the policy. Only 59% of the participants reported that Google has a privacy policy, but 72% of that population stated they read or skimmed the policy. There was a statistically significant relationship between those who read or skimmed the privacy policy and the users' awareness of actions that influence their search results, except for one – "actions taken by Google engineers, editors, or curators" influence the search results. This category of action is not as explicit in the privacy policy as it relates to how the platform is maintained, but the search engine or any platform would not exist without these actions. It reinforces the idea that users are not aware of how other people impact their platform experience and conceptualize the platform as an individual actor. YouTube users' policy reading habits were slightly different but the relationship between those who read or skimmed the privacy policy and their awareness of actions that influence their video results were all statistically significant. Sixty-two percent of users believe YouTube has a privacy policy and 66% of those have read or skimmed the privacy policy.

5.3.3 Facebook

While comparing Facebook users claiming to have read the privacy policy with the awareness of influences on their newsfeed, most of the relationships were found to be statistically insignificant. This is likely due to the explicitly social nature of Facebook. The two statistically significant relationships between reading the privacy policy and what influences the newsfeed are "details about your smartphone usage" and "details about accounts across devices." As stated in the previous chapter, users are more aware of personal data collected rather than the personal data used to influence their platform experience. More thorough and specific

explanation of how their interests and interactions are mirrored in search results and newsfeeds in the platforms' data/privacy policies would improve user awareness.

5.4 Implications

5.4.1 Theoretical Implications

The theoretical framework for this dissertation employed the three methodological tactics when researching black boxed algorithms (Bucher, 2018). First, I outlined what is known about platform algorithmic curation through the analysis of privacy policies. I then used algorithmic folk theories and studies about algorithmic knowledge gaps which provided the focus to study the beliefs and perceptions of algorithmically delivered content. And finally, I conducted a survey to interrogate those (un)known configurations to learn where or if the cracks in the black box exist. The following theoretical implications are what I find to be the biggest takeaways from this research project.

5.4.1.1 Programmed Sociality

Bucher (2012) coined the term programmed sociality, "sociality in social networking sites is algorithmically and dynamically shaped around the pursuit of participation" (p. 10) and she situated the term around unseen networks while using social media. Additionally, search engines are understood by researchers to be a part of a social search system that Halavais (2018) defines as the act of interacting with others while searching. However, both actions are invisible to the user. The respondents of this dissertation survey are mostly unsure or do not believe that actions taken by other users influence their search results or newsfeeds.

Table 45

Platform Comparison of "Actions taken by other users"

Do any of the following actions influence your newsfeed?	Yes	No	Unsure	No/Unsure
Google: Actions taken by other users	32%	45%	23%	68%
YouTube: Actions taken by other users	34%	38%	29%	66%

Facebook: Actions taken by other users	41%	32%	27%	59%

The findings in this section reinforce Bucher's (2018) theory of programmed sociality wherein "actors are articulated in and through computational means of assembling and organizing" (p. 4). Google's mission is to "organize the world's information and make it universally accessible." To do this, Google has many computational means, or algorithmic systems, in order to make information easier to find. Google's How Search Works page explains that the company uses natural language processing to learn and determine contextual word meaning. Because of this we can assume that search terms and links clicked are associated together and after time, the search terms and links can change based on social awareness. The more obvious use of social tagging is through the use of hashtags on YouTube and Facebook. Hashtags influence how a post is found or viewed by the user and with what information is trending at the time. However, the use, or participation, of hashtags is user driven and explicitly tagged as such. The use of a search engine is implicit participation in the social tagging of information on the internet.

There are multiple practices that occur where other people influence search results. Google employs human Quality Raters who help categorize information to improve search results (*How Our Quality Raters Make Search Results Better - Google Search Help*, n.d.). The Quality Raters undergo extensive training in hopes of standardizing how information is categorized, however information categorization is highly subjective and cultural. Additionally, Google's algorithmic systems receive feedback from users through the clicking of links after a search query. This feedback loop provides additional categorization detail to both search terms and information links. As such, programmed sociality does not only apply to social media sites but to other forms of information retrieval as well.

5.4.1.2 Information Poverty

Information poverty has traditionally been about equitable access to information. However, as Chatman (1996) demonstrates there are facets of information poverty that include personal perceptions to information seeking practices like perceiving themselves as devoid of sources even when having access to them, like the prior example of the 15 users who claim they do not use search engines. However, other conditions of information poverty are influenced by outsiders who withhold access to information. This condition includes algorithmically curated information as some information is withheld or prioritized based on the personal information used to curate search results and newsfeeds. As such, the definition of information poverty is evolving to include not only access to information but also understanding how you come to interact with that information, how you use and understand that information (Britz et at., 2012). The results of this study highlight the fact that just because users are aware data are collected, does not mean they are aware of how it is used. This is only one facet of information poverty. Others could be addressed with additional research around the readability of privacy policies.

5.4.1.3 Peeking Inside of the Black Box

Historically, maintaining the black box heuristic is beneficial for the platforms as their proprietary algorithms are profitable for advertisements and the efficiency of results or experience make the product appealing. What I have found is that users are paying attention to the behaviors of platforms and how they are interacting with them, 63% of participants report that they keep up with news about Google, 54% report that they keep up with news about YouTube, and 53% report that they keep up with the news about Facebook. The study results indicate that users are aware that platforms collect personal data about them when they use the services, however, they are less aware of the type of data collected and how it might be used.

As users start to look inside of the black box, they are reading the privacy policies to better understand the platforms. Users are less sure about what information is collected and used, but this correlates with the level of specificity in the privacy policies. As demonstrated in the findings, users of YouTube know that the platform collects interactions with the videos, but they are less aware that the interactions impact their video feeds. Similarly, users of Google are aware of details of their online session (browser, other websites visited, other search terms used) being collected while using the platform but less aware that the other online data impact search results.

These findings relate to Bucher's (2018) approach to black box theory through the recognition of "specific contexts and situations" where algorithmic curation is at play (p. 40). Facebook users are more aware of how Facebook uses their personal information to influence their timeline. This is partly due to the social nature of the platform, and because of the visual representation of advertisements. When asked about search engine functionality, many users identified how keenly unaware they were of how search algorithms worked while at the same time knowing more than just the search terms influenced the results. And for Bucher, opening the black box is not knowing exactly what the algorithm is but identifying its impacts as they are happening.

The privacy-focused users of this study complicate the black box heuristic. Even though their choice of search engine does not use personal data to customize results, they were more likely to believe platforms do not use personal information. These users may have chosen privacy-focused search engines because they believed the companies have less censorship than Google. The claim of censorship in this example is due to a difference in point of view about Covid-19 vaccination information. Further research around motivations of choosing privacy-

focused search engines is needed, however recent studies have shown that DuckDuckGo saw an increase in activity when Breitbart and other right-wing pundits claimed to find more "factual" information on the site (Urman et al., 2022).

Pasquale (2016) ends *The Black Box Society* with a hopeful move toward an intelligible society, one where financial and communication systems are made transparent and understandable. Recognizing a black box society creates opportunity for citizens to be skeptical about the platforms with which they interact, to be more aware of how those platforms function within social and political arenas and encourages the continued poking when a crack is found in the box. Users of these platforms have already created their own ways of knowing or understanding platform, demonstrated through the user beliefs identified in this dissertation study and others. Some users are aware that their search history impacts future search results, while others believe everyone does and should receive the same information. Ultimately, it does not necessarily matter if users know how it technically works, but rather how users perceive platforms work. The users are breaking down one black box and creating their own in the process. What does matter is how and if they recognize the curation of their information ecosystem and what impacts that has on their evaluation of information received.

5.4.2 Practical Implications

5.4.2.1 Information Literacy

The findings of this dissertation can be used to inform information, media, and/or algorithmic literacy instructors. From a library and information science perspective, so much of the curriculum for information literacy is directed toward a traditionally college aged person with an assumed level of technological literacy based on their age. What the findings show is that age does not have a consistent impact on how much a user knows about a platform. A greater

understanding of how users interpret the information which they interact with allows for more specific and in-depth instruction. There is an assumption that people who use platforms have a greater knowledge of how they operate. The term "digital native" has been used in education and technology studies to define the phenomenon; however, the findings of this dissertation are not conclusive enough to say there is a correlation between platform use and knowledge. This is an important factor to consider even as future information professionals enter graduate school.

The Association of College and Research Libraries (ACRL) provides information literacy guidance, among other professional support for academic library workers. When the ACRL updated their Framework for Information Literacy for Higher Education in 2014, it received praise from librarians who wanted to move away from the check-list type approach to information literacy instruction. It also received criticism as being less prescriptive and more open for interpretation than the previous Standards. The Framework consists of six frames through which to view information literacy – authority is constructed and contextual, information creation as a process, information has value, research as inquiry, scholarship as conversation, and searching as strategic exploration. Of these six frames, two address information retrieval systems – information has value and searching as strategic exploration. However, neither of the frames are explicit in how the system impacts searching activities. The results of this dissertation study demonstrate the lack of understanding about how the users' identity shapes their experience online. There is a disconnect between what data the users believe is collected and how that data is used to construct their online experience. Information literacy is not only about how to use and understand information, but where and how information is delivered. It is because of that I propose the following, an additional frame for the ACRL Framework, Systemic Issues within Information Retrieval Platforms, in the appropriate format.

Systemic Issues within Information Retrieval Platforms

Information dissemination requires infrastructure. Infrastructure changes how the person accesses, consumes, and interacts with the information. There are social, political, and economic factors influencing information retrieval platforms through advertising, algorithmic curation, business model, information creators, and location.

Access to information creates social, political, and economic barriers to information seeking. After the creation process, information becomes accessible through a variety of venues, both physical and digital. The information seeker's access to that information changes based on location, affiliation with higher education, and ability, both physical and digital, to access, open, and consume the information. Experts understand these pressures on platforms to modify and commodify how information is accessed. It is through these systemic issues that information seekers must navigate to fully understand how their information is populated and curated with searching, especially on digital platforms.

Knowledge practices

Learners who are developing their information literate abilities

- understand the social, political, and economic influences on information retrieval platforms;
- recognize differences in information retrieval platforms;
- assess the content provided through the platform and match it with their information needs;
- develop strategies for identifying personalized information;
- recognize their worldview can be mirrored back to them;

- modify information seeking behaviors when necessary.

Dispositions

Learners who are developing their information literate abilities

- are willing to experiment with multiple information retrieval platforms;
- resist complacency when information seeking is difficult;
- demonstrate awareness of the social, political, and economic impact of platforms to the access of information;
- value the differences in information platforms.

5.4.2.2 Response to Proposed Federal Information Policies

Over the last three years, Congress has been drafting and proposing federal legislation that aims to regulate how personal information is used to produce platform experiences, like search results. None of these bills went to the floor for a vote. They were read and then sent to relevant committees for more work. The bills indicate either what Congress believes is needed for their constituents, or some sort of virtue signaling for their political platform. The following bills were identified through Congress.gov by searching for "algorithm," "platform," "personal information," or a combination of these words. They were limited to those bills that directly discuss algorithmic curation using personal information. While this analysis is outside the scope of my research questions, I wanted to see if any of these bills might have the potential to increase user awareness based on the results reported earlier.

Table 46

Proposed Legislation 2019-2022

Bill Number	Year	Title	Summary from Congress.gov

H.R.6416 S.3520	2022	Banning Surveillance Advertising Act	This bill restricts online advertising that targets an individual, internet-connected device, or group of individuals or devices based on personal information. Personal information includes information that is reasonably linkable to an individual or connected device such as internet browsing history or the content of communications. The bill generally prohibits (1) online advertisers from using personal information, including personal information that identifies an individual as a member of a specified protected class, to target advertising; and (2) advertising facilitators (i.e., entities that receive compensation for disseminating online advertisements) from using personal information to disseminate targeted advertising or knowingly enabling online advertisers to do so.
H.R.6407 S.3501	2022	Terms-of-service Labeling, Design, and Readability (TLDR) Act	Require the FTC to create rules for commercial websites and mobile apps to create summary terms-of-service statements. Summary statements must be concise, easy to understand, machine-readable, and located at the top of the existing ToS page. Summary statements must include: the categories of sensitive consumer information collected and whether that data is necessary for basic functioning, legal liabilities of a consumer using the service, a change log, and a list of reported data breaches from the last three years. https://trahan.house.gov/uploadedfiles/tldr_act_one-pager.pdf
H.R.2154 S.3029	2021	Protecting Americans from Dangerous Algorithms Act	Amend section 230(c) of the Communications Act of 1934 to prevent immunity for interactive computer services. Specifically, the bill removes this immunity from a social media company with more than 10 million monthly users if it utilizes an algorithm, model, or other computational process to amplify or recommend content to a user that is directly relevant to a claim involving (1) interference with civil rights, (2) neglect to prevent interference with civil rights, or (3) acts of international terrorism.
H.R.5596	2021	Justice Against Malicious Algorithms Act	This bill limits federal liability protection that applies to a provider of an interactive computer service (e.g., a social media company) for claims related to content provided by a third party if the provider makes personalized recommendations of online content that cause physical or emotional injury. Specifically, the liability protection (sometimes referred to as Section 230 protection) shall not apply to a service provider that • has more than 5 million monthly visitors for more than 3 of the preceding 12 months, • uses an algorithm or similar computational process to make personalized recommendations based on information specific to an individual, and • knowingly or recklessly makes a personalized recommendation that materially contributes to a physical or severe emotional injury to a person. However, the liability protection shall continue to apply to (1) recommendations made directly in response to a user's search; and (2) service providers of web hosting, domain registration, data storage, and related services that are used by another service

			provider in the management, control, or operation of that provider's services.
H.R.3611 S.1896	2021	Algorithmic Justice and Online Platform Transparency Act	This bill establishes requirements for certain commercial online platforms (e.g., social media sites) that withhold or promote content through algorithms and related computational processes that use personal information. The platforms must • make disclosures about their collection and use of personal information and their content moderation practices; • retain specified records that describe how the algorithms use personal information and assess whether the algorithms produce disparate outcomes based on race and other demographic factors in terms of access to housing, employment, financial services, and related matters; • employ algorithms safely and effectively; and • allow users to access and transfer their personal information. If a platform uses algorithms to publish or sell advertising, it must maintain a library of the advertisements. The Federal Trade Commission must also adopt rules concerning deceptive advertising. A platform's chief executive officer or other senior officer must certify compliance with disclosure requirements. Additionally, platforms may not (1) employ algorithms or other design features that result in discrimination or similar harms based on demographic or biometric factors, or (2) process information such that it impairs voting rights. Further, users of a platform may not violate civil rights laws using the platform's algorithms. The bill prohibits waivers or other methods that limit rights under the bill; provides whistleblower protections for individuals who report violations; and authorizes enforcement by specified federal agencies, states, and private individuals. The bill also provides funding for an interagency task force to study the discriminatory use of personal information by platforms' algorithms.
H.R.5921 S.2024	2021	Filter Bubble Transparency Act	This bill establishes requirements for large online platforms that use algorithms applying artificial intelligence or machine learning to user-specific data to determine the manner in which content is displayed to users. Specifically, if an online platform applies such techniques to user-specific data that is not expressly provided by the user, the platform must (1) notify users that the platform uses such data, and (2) make a version of the platform available that uses only user-specific data that has been expressly provided by the user and which enables users to switch between the two platforms.
S.2763	2019		
H.R.7012	2020	Protecting Democracy from Disinformation Act	Amends the Federal Campaign Act of 1971. This bill restricts the ability of online platforms and advertisers to target political advertisements to a specific group of individuals based on online behavioral data.

The regulations proposed would not be obvious to the end user, aside from the TLDR Act if the user reads the Terms of Service and the Filter Bubble Transparency Act, which would change the design of the platform. Due to most of the participants either answering "no" or "unsure" to questions about data use and collection, some form of regulation is needed to make the functionality of the platforms more obvious to the user. However, users might not care to know how the platform works, even if the information or experience is filtered (Ibdah et al., 2021).

Of the seven proposed bills, the TLDR Act and Algorithmic Justice and Online Platform Transparency (AJOTP) Act make efforts to further user awareness through the simplification of policies and making specific requirements for data collection and use disclosures. It is unfortunate that Congress is assuming that data collected are listed in the terms of service. The TLDR Act does not specify a summary of privacy policies in addition to terms of service, since in my analysis platforms link to the privacy policy instead of outlining data collection and use practices in the terms of service. The AJOTP Act works similarly to the General Data Protection Regulation in that it requires platforms to more specifically outline their data collection, use, and retention practices. The AJOTP Act goes further by requiring companies to explain the algorithms that "use personal information and assess whether the algorithms produce disparate outcomes based on race and other demographic factors" in order to determine if discrimination or harm occurs through design based on demographic information (S.1896, 2022; H.R. 3611, 2022).

The Filter Bubble Transparency Act outlines the most obvious platform change to the user because of the required design feature (ex.: toggle button) to see non-personalized search

results or newsfeeds. If passed, the participants who felt they wanted to test whether search results would be the same for different people would have the opportunity to experience non-personalized results. Additionally, the ability to toggle between their personalized results and non-personalized results would allow people to see their filter bubble and/or be exposed to more and different information.

Senators Cassidy and Lujan and Representative Trahan introduced the "Terms-of-service Labeling, Design, and Readability Act" or the "TLDR Act" in Congress in 2022 "as an attempt to simplify and provide more transparency to long and technical terms of service. The TLDR Act would require the Federal Trade Commission (FTC) to provide guidelines for short-form terms of service. The proposed Act requires three elements for a short-form terms of service; a summary statement that provides the amount of effort required to read the entire terms of service, a graphic data flow diagram, and an interactive data format (XML). With the authority of the Unfair or Deceptive Acts or Practices regulation the FTC would fine platforms that do not comply or state attorney generals who believe that at least 1,000 residents of the state have been adversely impacted can file a lawsuit against the platforms that do not comply.

Many of the bills listed in Table 46 seek to regulate the data collection practices of platforms regulating advertisements likely because Google and Facebook dominate three-quarters of the United States' digital ad market (Accountable Tech, 2021). The high ad revenue provides Congress with the leverage they need to enforce the law, if passed. For example, the "Banning Surveillance Advertising Act of 2022," introduced in both the House and Senate, seeks to restrict online advertising that targets individuals through the collection of personal data including browsing history and connected devices (S.3520 & H.R.6416, 2022). "Protecting Democracy from Disinformation Act" is more specific about online advertising as it relates to

political advertisements and how they target individuals on behavioral data collected (H.R.7012, 2020).

The "Algorithmic Justice and Online Platform Transparency Act" (AJOPTA) (S.1896 & H.R.3611, 2022) seeks to require platforms to make disclosures about how and why they collect personal data, detailed records on how personal data is used to display information and allow users to access and transfer their data. My analysis of Google and Facebook's privacy/data policies found they do disclose generalizations of the type of information they collect and how they use it. If a bill like AJOPTA were to be enacted, these platforms would need to make their data collection and use practices more specific, transparent, and provide users with access to their data.

An investigation by the Associated Press into Google's location tracking practices began in 2018 with 40 states who have consumer protection laws in place. Google settled the investigation in November 2022 with a $391.5 million payment to be split between the 40 states (Collins & Gordon, 2022). It appears before federal regulation is passed by Congress, states will work together to hold platforms accountable through upholding their current laws and/or writing additional legislation to protect their citizens.

An effective policy would be a combination of the proposed legislation. To increase algorithmic awareness, both the Terms of Service and Privacy/Data Policies should be provided as a short summary in easy-to-understand language, in addition to the full policy. Visuals, such as infographics or charts, should provide more clarity and simplification as well. Additionally, personalization should be made more obvious to the user, either through the toggle suggested by the Filter Bubble Transparency Act or through notation under or next to results indicating if, how, and why the result is personalized. Facebook does something similar now with their

advertisements. The user can click on the three dots in the upper right of an ad and select "why am I seeing this ad." The Federal Trade Commission would likely be responsible for the oversight and enforcement of proposed legislation as much of it is written for consumer protection. Specifically, the Banning Surveillance Advertising Act prohibits the use of protected class information (race, age, gender) in personalizing or targeting advertisements.

5.4.2.3 Recommendations for Google, YouTube, and Facebook

The spread of mis/disinformation and extreme polarization of society are partly due to how people find and interact with information. Even though filter bubbles have been in public discourse for over 10 years, people find themselves stuck in echo chambers created by the platforms. Without the skills, knowledge, or desire to learn new techniques these problems will continue. However, platforms have the opportunity to build transparency into their platforms that could indicate if, how, and why search results, videos, or social media posts are presented through the utilization of personal information. At the end of Noble's (2018) book, she makes the recommendation for a new type of search engine, where you opt-in to the biases you want to see or impact the search results. The visual aspect of her proposal creates transparency for the user – both from a platform perspective but also how the user recognizes their own information seeking practices.

Platforms should work to better separate content and advertising. The FTC requires that advertisements be fair and not use deceptive practices. Google and YouTube's advertising is clearly marked with "ad" in the search results, and as a commercial in YouTube. While Facebook does denote advertisements, some appear in the user's newsfeeds like user generated posts. Not all advertising has malicious intent, but Facebook acknowledges that their political digital advertisements created issues during U.S. elections. In 2020 they implemented a ban on

new social issue, political, and election advertisements a week prior to election day (Shrimaker, 2022).

Even though these are billion-dollar companies with fiduciary responsibilities to their shareholders, they should also feel a sense of responsibility to their users to provide transparency and accountability when supplying millions of people, the information they need to live. An informed public is necessary for a functioning democracy, and there is a significantly low awareness of how and why information is delivered the way it is.

Chapter 6: Conclusions

This dissertation study sought to learn about the awareness users have about their algorithmically constructed online experiences of using search engines, newsfeeds, and video autoplay. The reason for such a study is to better learn if people are aware of the filter bubbles in which they receive their information through learning user awareness of what personal information is collected and how it might be used to curate their online experience. Previous research on algorithmic awareness has been largely qualitative with interviews or focus groups to talk with users about why they believe some friends' Facebook posts appear more than others (Eslami et al., 2016) or how preference is given to news publications in Google News (Powers, 2017). To provide additional quantitative data and tie user awareness directly to how they are made aware of algorithmic filtering, the privacy policy analysis provided the basis of the data collection. While research has demonstrated that the reading of privacy policies is low, it is one of the standardized and common forms of communication between a platform and the user (Obar & Oeldorf-Hirsch, 2020). The importance of this research is found in identifying where users require more information regarding algorithmic curation to be able to develop and execute accurate information and algorithmic literacy instruction and to support legislation that would require platforms to be more transparent with their users.

Mutual construction and black box theory provide the theoretical framework for this dissertation. Mutual construction demonstrates the symbiotic relationship between the user and the technology where there is a dependency on user interaction with the technology for optimal functionality. This relationship is what is black boxed from the user. In most personal relationships a person can see or feel what is being given and taken, however this process is opaque between a user and algorithmic system. Users begin to see the outcomes of the

relationship, like surveillance advertising, but do not know the full scope of what and how that data is used. For a more information- and algorithmic-literate society, there is a need to make more transparent the inner workings of algorithmic systems. The rampant spread of mis- and disinformation is a symptom of users not understanding how their online experience is curated based on their personal data. This chapter summarizes overall and key findings from an exploratory mixed methods dissertation study, contributions to the discipline, and directions for future research.

6.1 Overall Findings

> **Research Question 1 (RQ1):** How do platforms (Google (1), YouTube (2), Facebook (3)) govern and inform their users with regard to using collected data to filter the users' online content?

I analyzed the privacy/data policies for Google, YouTube, and Facebook to identify how the platforms informed their users about the personal data collected that filters their search results and newsfeeds. Within this analysis I identified eight categories of influence and eleven categories of personal data collected. These categories create the framework for the survey questions to identify the knowledge and awareness of Google, YouTube, and Facebook users. While privacy policies are not written to educate the user, they are one common touchpoint across all three platforms studied. Overall, the privacy/data policies use situational examples to explain the types of data they collect and how they use that data. Both companies' privacy/data policy includes a statement that provides the assumption that data collected could be used to provide recommendations, personalize content, customize search results, or personalize features and content.

Research Question 2 (RQ2): What do United States-based adult users of these platforms know about the personal and usage data being collected and their awareness about how the platforms use this data to moderate and serve online content?

With a sample representative of the United States population, 396 U.S.-based adults completed a 60-survey question survey that asked about the knowledge of personal data collected while using these platforms, and also how often they used the platform and if they read the privacy policy. Participants indicated their primary reasons for using Google were to look up news, navigate to specific sites, answer health questions, find entertaining content, and check the accuracy of information. Participants reported they use YouTube and Facebook for the same reasons to keep up with family and friends, find entertaining content, share pictures or videos, and look up news.

When asked generally if personal information is used to moderate and serve online content, participants of the survey were mostly unsure or did not know personal information was used in the construction of their results or newsfeeds. There is one exception to this, Facebook users are more aware of how their actions on Facebook influenced their experience. Even though more than half of all participants are aware of privacy/data polices for the platform, the specifics of data collection and use demonstrate where there are gaps in the knowledge.

There are several discrepancies between the knowledge of personal data collected and if that personal data influences or filters the results or news feed. For example, 57% of respondents answered yes when asked if interactions with videos are collected while using YouTube, while only 47% said those interactions influenced video results. Another example, 53% of respondents answered yes when asked if details about their online session are collected while using Google, while only 44% said that details about their online session influence search results. Finally, 42%

of respondents answered yes when asked if details about their smart phone usage are collected while using Facebook, while only 38% said that details about smart phone usage influenced newsfeeds. These discrepancies are consistent between three platforms.

> **Research Question 3 (RQ3):** Are there any associations between usage characteristics of United States-based adult users of these platforms and users' awareness of algorithmic curation using personal information?

The frequency of conducting searches, checking platforms, or time spent on platforms did not have a significant association with awareness of algorithmic curation. Alternatively, users who claim to have read or skimmed the privacy policy did have a significant association with their awareness of algorithmic curation.

> **Research Question 4 (RQ4):** How does user awareness about algorithmically delivered content vary based on demographic characteristics, including age, race, education level, and political affiliation?

Age is the most consistent demographic variable with an association to algorithmic awareness with those users younger than 55 years old generally being more aware of how personal information is used to curate their online experience, and those over 55 years old being unsure of if or how personal information is used.

6.2 Key Findings

Users of search and social media platforms are aware that personal information is collected about them while using the platforms. They are less aware of how that information is used to impact their online experience. The contribution section outlines how explicit

information literacy instruction can address this gap in user knowledge, but more work needs to be done to address the systemic issues of information retrieval.

6.2.1 Subgroups within the Participants

The exploratory research revealed two key subgroups of platform users. First users who believe they do not use search engines, but then name either a search engine or the browser where they search. This finding highlights people's dependency on smart phones for information retrieval and how search functionality is obscured by in-browser search. One of the pillars of information literacy is being aware of where information is coming from, specifically when the information is found online in less than obvious transactions.

The second subgroup are the respondents who report they only use privacy-focused search engines. Privacy-focused search engines (PFSE) market themselves as an ethical, non-tracking alternative to Google. After deeper analysis into the demographics of this group, we learned the users to be mostly Republican males without a college education. Furthermore, this group is more likely to believe Parler is a mostly reliable information source. An investigation into intrinsic motivations of PFSE users found a recent push by Breitbart to use DuckDuckGo due to claims of Google suppressing information about Covid-19 vaccination deaths.

6.2.2 Demographics that Matter

The analysis of survey participants revealed no significant results when using race as a variable. Age, gender, education, and political affiliation all demonstrated a variety of associations with user awareness of algorithmically delivered content.

With regard to age, the cusp between being using a platform frequently, being aware of privacy policies, or being aware of personal information collection is in the mid 50s. Where

those older than their mid-50s are less frequent users or less aware and those younger are more frequent users and more aware of personal data collection. Typically speaking, the "digital native" concept was constructed for Millennials (born 1982-1997) as those who grew up with technology and should be aware of how it is used. As Millennials are in their 30s and 40s, the age group of "digital natives" is now Gen Z (born 1997-2013). The data from this dissertation does find that users in their mid-20s and younger have less awareness than older users about how personal information is used to algorithmically curate information.

The association gender has with algorithmic awareness is the lack of certainty that women demonstrate in their answers about personal data collection and use. Gender had associations with reliability with information sources where men were more likely to find family and friends reliable and women were more likely to find them very reliable. Participants with exposure to higher education were more likely to be aware of how personal data is collected and used for algorithmic curation. Political affiliation demonstrated that Republicans are more likely to be unsure of personal data collection and use by an online platform for algorithmic curation. Democrats and Independents demonstrated more certainty in their answers, and were often incorrect.

6.2.3 Beliefs of Search Results

The beliefs that stood out from the free text responses were the recognition of personalization techniques, results should be the same for the same queries, the desire to prove difference or sameness of search results, mistrust of online platforms or companies, and people not knowing or not having the words to explain their beliefs. Of these beliefs, all except those recognizing personalization should be refined and examined further using think-aloud methodology. The search beliefs documented in this dissertation support findings found in

studies researching folk theories of algorithmic systems. These findings include where respondents identified an algorithm or entity that is determining the search results (Rader & Gray, 2015), the recognition of their search history impacting the results (Eslami et al., 2016), and the users identified how the companies can be exploitative and confining to a specific set of search results (Ytre-Arni & Moe, 2021).

6.3 Contributions

Users are trying to look into the black box of information retrieval platforms. More than half of all users in this study keep up with news about Google, YouTube, and Facebook. This is one of the more general ways to learn about the platforms and understand their function in our world. The users surveyed are also curious about how search engines function with their desire to test search outcomes between people. While the proprietary black box remains unharmed, the social commentary around platforms and the delivery of online content is beginning to open for the interested users. The privacy-focused search engine users in this study complicate the findings as it could be assumed they choose privacy-focused search engines for the lack of personal information used. But data suggests that the increased use of these search engines is due to encouragement from conservative news outlets.

The findings of this research contribute to an information literacy curriculum that meets users at their level of knowledge and expands upon their understanding of information seeking behaviors. The curriculum has implications for both the Association for College and Research Libraries (ACRL) and the American Association for School Librarians (AASL) who both have frameworks for information literacy instruction, although neither of them specifically address algorithms, search engines, or algorithmically delivered content. In January 2023, the state of New Jersey became the first state to require information literacy instruction as a part of K-12

education (Sitrin, 2023; NJ S.B. 588 & NJ A.B. 4169, 2023). This is hopes to be the start of a trend of state education boards adding information literacy as a part of the standard curriculum, and not as something a few teachers and librarians implement through their course work.

The tangible contribution to this curriculum comes in the form of an additional frame to the ACRL Framework for Information Literacy. The proposed frame centers how information infrastructure impacts a person's access to, consumption of, and interactions with information. Specifically, how algorithmic curation and business models deliver content based on a person's data. The results of this dissertation study show a disconnect between users' awareness of personal data collected when compared to their awareness of personal data used to algorithmically deliver content.

For most Americans, the use of information seeking platforms is a necessary part of life, as is agreeing to the Terms of Service and Privacy/Data Policies. As reported earlier, it would take nearly 76 business days for the average American to read the policies of the websites they interact with and congresspeople are working to increase that transparency to the American people (*Cassidy, Lujan, Trahan Introduce Bill to Inform Consumers, Increase Online Transparency*, 2022). Platform governance, as well as functionality, is another black boxed process. Legislators proposed a bill that would require platforms to simplify Terms of Service documents to increase the legibility and comprehension of the policies (TLDR, 2020). The findings from this dissertation study show that most users are aware that privacy policies exist, and more than half of the respondents claim to have read or skimmed the policy. However, the users are not comprehending what the policy is saying because their knowledge about personal information used does not align with personal information collected or they did not read the policy.

6.4 Limitations

The limitations of this study include those related to historical snapshots of data (privacy policies) and self-reported survey data. The policies analyzed in this study were accurate and timely when collected but have since updated several times since. Even with best efforts in place, textual analysis is subjective to the interpretation of the researcher. While the survey data is representative of the United States population, some demographic variables needed to be collapse to for data analysis validity reasons. There is no way to verify the information provided in self-reported surveys and they contain the risk that participants misrepresent their knowledge. The 60-question survey is on the longer end of survey design and did not allow for follow-up questions.

6.5 Future Research

Throughout the writing of this dissertation, I have wished I could speak to subsets of the respondents of the survey, like the 15 who "do not use search engines" and the 24 who use only privacy-focused search engines, to ask follow-up questions and learn more about how they understand their interactions with platforms. Future research would use the findings of this dissertation to develop questions for either interviews or focus groups to further investigate perceptions of algorithmic curation and how inferred identity impacts search results and newsfeeds. Work has already begun to conduct focus groups with searchers who use privacy-focused search engines. The methodology outlined in DeVos et al. (2022) which conducts user-driven algorithmic audits to identify and name algorithmic behavior in real time would provide an excellent framework for future studies built off of the learned data of users wanting to test search results.

Think-aloud interviews would expand the work and explore deeper into specific areas of algorithmic awareness, like the disconnect between the trust of family, friends, and colleagues and the mistrust of Facebook as information sources to potentially expand the corpus of folk theories around Facebook newsfeeds. The survey instrument asked questions regarding internet use and access as well as habits of the users keeping up with the platforms in the news. These questions were not immediately impactful in the analysis for the research questions. Further analysis to segment on these data points might reveal additional details about the user to better identify a point of intervention for information, media, or algorithmic literacy instruction.

Results of the study showed that mobile phones are obfuscating search functionality. Recently while helping someone search for an address on their phone, I asked where their browser (search) app was located. They responded they did not have one on their phone. I searched for "Safari," found it, and they had 150 tabs open. This is an example of a well-educated person unaware of how they were locating information on their smart phone. Mobile search in concert with the rise of artificial intelligence chatbots could lead to homogeneous information landscapes. I am curious about the supposed objectivity these two technologies create.

A project is underway with a colleague where we are utilizing the methodology from this dissertation to learn about student and instructor awareness of Canvas' data collection and privacy features. With the rise of student surveillance tools and built-in analytics, we are interested to learn from not only students but the instructors who rely on learning management systems to conduct class.

Finally, I have questions remaining around how federal and state legislators work to regulate platform speech and their use of surveillance advertising. As public policy tends to be

reactionary, I am curious to speak to state legislators about how they learn about technical issues and what they see as future problems to address. I realize most of this legislation does not move forward and is seen as grandstanding, but at this point the platforms are regulating themselves, creating policy that is outside of the control of citizens.

References

Accountable Tech. (2021, December 3). Petition for rulemaking to prohibit surveillance advertising. *Federal Trade Commission*.https://www.ftc.gov/system/files/attachments/other-applications-petitions-requests/r207005_-_petition_for_rule_to_prohibit_surveillance_advertising_0.pdf

ACRL. (2000, January 1). *Information Literacy Competency Standards for Higher Education*. https://alair.ala.org/handle/11213/7668

ACRL. (2015, February 9). *Framework for Information Literacy for Higher Education* [Text]. Association of College & Research Libraries (ACRL). http://www.ala.org/acrl/standards/ilframework

ALA Public Programs Office. (2017). Coming soon: Media literacy training for public libraries. *Programming Librarian*. https://programminglibrarian.org/articles/coming-soon-media-literacy-training-public-libraries

Alghowinem, S. (2019). A safer YouTube Kids: An extra layer of content filtering using automated multimodal analysis. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), *Intelligent Systems and Applications* (pp. 294–308). Springer International Publishing. https://doi.org/10.1007/978-3-030-01054-6_21

Anderson, M., Perrin, A., Jiang, J., & Kumar, M. (2019, April 22). 10% of Americans don't use the internet. Who are they? *Pew Research Center: Fact Tank*. https://www.pewresearch.org/fact-tank/2019/04/22/some-americans-dont-use-the-internet-who-are-they/

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias: There's software used across the country to predict future criminals. And it's biased against

blacks. *ProPublica*. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Ashby, W. R. (1956). *An introduction to cybernetics*. Chapman & Hall. http://pespmc1.vub.ac.be/books/IntroCyb.pdf

Auxier, B., Rainie, L., Anderson, M., Perrin, A., Kumar, M., & Turner, E. (2019, November 15). 4. Americans' attitudes and experiences with privacy policies and laws. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/

Bakke, A. (2020). Everyday Googling: Results of an observational study and applications for teaching algorithmic literacy. *Computers and Composition*, *57*, 102577. https://doi.org/10.1016/j.compcom.2020.102577

Balkin, J. M. (2018). The First Amendment in the second Gilded Age. *Buffalo Law Review*, *66*(5), 35.

Ballanco, M. J. (2013). Searching for the First Amendment: An inquisitive free speech approach to search engine rankings. *George Mason University Civil Rights Law Journal*, *24*(89).

Ballotpedia. (n.d). *Georgia's 14th Congressional District*. Retrieved September 28, 2022, from https://ballotpedia.org/Georgia%27s_14th_Congressional_District.

Baltar, F., & Brunet, I. (2012). Social research 2.0: Virtual snowball sampling method using Facebook. *Internet Research*, *22*(1), 57–74. https://doi.org/10.1108/10662241211199960

Banning Surveillance Advertising Act of 2022, H.R.6416, House of Representatives, 117th Congress (2022). https://www.congress.gov/bill/117th-congress/house-bill/6416

Banning Surveillance Advertising Act of 2022, S.3520, Senate, 117th Congress (2022). https://www.congress.gov/bill/117th-congress/senate-bill/3520

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *On-Line Review*, *13*, 407–431.

Bauman, A., & Bachmann, R. (2017). Online consumer trust: Trends in research. *Journal of Technology Management and Innovation*, *12*(2), 68–79. https://doi.org/10.4067/S0718-27242017000200008

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society*, *11*(6): 985-1002.

Beer, D. (2017). The social power of algorithms. *Information, Communication & Society*, *20*(1), 1–13. https://doi.org/10.1080/1369118X.2016.1216147

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation*, *38*(2), 61–71. https://doi.org/10.1108/eb026722

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim Code*. Polity Press.

Bijker, W. (2001). Understanding technological culture through a constructivist view of science, technology, and society. In S. H. Cutcliffe & C. Mitcham (Eds.), *Visions of STS: Counterpoints in science, technology and society studies* (pp. 19–34). SUNY Press.

Bowker, G. C., & Star, S. L. (2000). *Sorting things out*. MIT Press.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. *Ethics and Information Technology*, *15*(3), 209–227. https://doi.org/10.1007/s10676-013-9321-6

BrandVerity. (2020). *Online Consumer Search Trends Report*. https://www.brandverity.com/blog/three-ways-search-can-boost-your-consumer-brand-in-2020

Brenneur-Garel, I. (2021). Artificial intelligence in legal practice: Exploring theoretical frameworks for algorithmic literacy in the legal information profession. *Legal Information Management, 21*: 97-117.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, *30*, 107–117.

Britz, J. J. (2004). To Know or not to know: A moral reflection on information poverty. *Journal of Information Science*, *30*(3), 192–204. https://doi.org/10.1177/0165551504044666

Britz, J. J., Hoffman, A., Ponelis, S., Zimmer, M., & Lor, P. (2012). On considering the application of Amartya Sen's capability approach to an information-based rights framework. *Information Development*, *29*(2), 106–113.

Brock, K., & Shepherd, D. (2016). Understanding how algorithms work persuasively through the procedural enthymeme. *Computers and Composition*, *42*, 17–27. https://doi.org/10.1016/j.compcom.2016.08.007

Buchanan, T., & Benson, V. (2019). Spreading disinformation on Facebook: Do trust in message source, risk propensity, or personality affect the organic reach of "fake news"? *Social Media + Society*, *5*(4), 2056305119888654. https://doi.org/10.1177/2056305119888654

Bucher, T. (2012). *Programmed sociality: A software studies perspective on social networking sites* [University of Oslo]. http://tainabucher.com/wp-content/uploads/2009/08/Bucher_Ph.D.diss_.pdf

Bucher, T. (2018). *If...then: Algorithmic power and politics*. Oxford University Press.

Bucher, T. (2021). *Facebook* (1st edition). Polity.

Bush, D. & Zaheer, A. (2019). Bing's top search results contain an alarming amount of disinformation. *Stanford: Freeman Spogli Institution of International Studies*, December 17. https://fsi.stanford.edu/news/bing-search-disinformation

California Consumer Privacy Act (CCPA). (2018, October 15). State of California - Department of Justice - Office of the Attorney General. https://oag.ca.gov/privacy/ccpa

Cassidy, Lujan, Trahan Introduce Bill to Inform Consumers, Increase Online Transparency. (2022, January 13). Bill Cassidy | Press Releases. https://www.cassidy.senate.gov/newsroom/press-releases/cassidy-lujan-trahan-introduce-bill-to-inform-consumers-increase-online-transparency

Chatman, E. A. (1996). The impoverished life-world of outsiders. *Journal of the American Society for Information Science, 47*(3), 193–206. https://doi.org/10.1002/(SICI)1097-4571(199603)47:3<193::AID-ASI3>3.0.CO;2-T

Childers, T. (1975). *The information-poor in America*. The Scarecrow Press, Inc.

Collins, D. & Gordon, M. (2022). 40 states settle Google location-tracking charges for $392M. *Associated Press News*, November 14. https://apnews.com/article/google-privacy-settlement-location-data-57da4f0d3ae5d69b14f4b284dd084cca

Complaint, United States v. Google LLC, (October 20, 2020). https://www.justice.gov/opa/pr/justice-department-sues-monopolist-google-violating-antitrust-laws

Confessore, N. (2018, April 4). Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times*. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html

Connaway, L. S., & Radford, M. L. (2017). *Research Methods in Library and Information Science* (6th ed.). Libraries Unlimited.

Cotter, K. (2022). Practical knowledge of algorithms: The case of BreadTube. *New Media & Society*. https://doi.org/10.1177/14614448221081802

Cotter, K., DeCook, J., Kanthawala, S., & Foyle, K. (2022). In FYP we trust: The divine force of algorithmic conspirituality. *International Journal of Communication*, *16*, 1-23. https://doi.org/1932–8036/20220005

Cotter, K. & Reisdorf, B. (2020). Algorithmic knowledge gaps: A new dimension of (digital) inequality. *International Journal of Communication, 14*: 745-765. https://ijoc.org/index.php/ijoc/article/view/12450/2952

Couldry, N. (2019). Capabilities for what? Developing Sen's moral theory for communications research. *Journal of Information Policy*, *9*(1), 43–55. https://doi.org/10.5325/jinfopoli.9.1.0043

Crane, D. A. (2014). After search neutrality: Drawing a line between promotion and demotion. *I/S: A Journal of Law and Policy for the Information Society, 9,* 397-406.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE Publications, Inc.

DataReportal. (2022, April). *Facebook statistics and trends*. Facebook Statistics and Trends. https://datareportal.com/essential-facebook-stats

Demographics of internet and home broadband usage in the United States. (2019, June 12). *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/fact-sheet/internet-broadband/

Dencik, L., Jansen, F., & Metcalfe, P. (2018). *A conceptual framework for approaching social justice in an age of datafication* (p. 9) [Working paper]. DATAJUSTICE. https://datajusticeproject.net/wp-content/uploads/sites/30/2018/11/wp-conceptual-framework-datajustice.pdf

DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). "Algorithms ruin everything": #RIPTwitter, folk theories, and resistance to algorithmic change in social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3163–3174. https://doi.org/10.1145/3025453.3025659

DeVos, A., Dhabalia, A., Shen, H., Holstein, K., & Eslami, M. (2022). Toward user-driven algorithm auditing: Investigating users' strategies for uncovering harmful algorithmic behavior. *CHI Conference on Human Factors in Computing Systems*, 1–19. https://doi.org/10.1145/3491102.3517441

Diakopoulos, N. (2015). Algorithmic accountability: Journalistic investigation of computational power structures. *Digital Journalism*, *3*(3), 398–415. https://doi.org/10.1080/21670811.2014.976411

Dimock, M. (2019, January 17). Defining generations: Where Millennials end and Generation Z begins. *Pew Research Center*. https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/

Discover How Google Search Works. (2020, October 21). Google Search. https://web.archive.org/web/20201021092214/https://www.google.com/search/howsearchworks/

Dogruel, L. (2021). Folk theories of algorithmic operations during Internet use: A mixed methods study. *The Information Society*, *37*(5): 287-297.

Dutton, W. H., Reisdorf, B. C., Blank, G., Dubois, E., & Fernandez, L. (2019). The internet and access to information about politics: Searching through filter bubbles, echo chambers, and disinformation. In M. Graham & W. H. Dutton (Eds.), *Society and the Internet* (2nd ed., pp. 228–247). Oxford University Press.

Dutton, W. H., Reisdorf, B., Dubois, E., & Blank, G. (2017). Search and politics: The uses and impacts of search in Britain, France, Germany, Italy, Poland, Spain, and the United States (SSRN Scholarly Paper ID 2960697). *Social Science Research Network*. https://doi.org/10.2139/ssrn.2960697

Edelson, L., Nguyen, M.-K., Goldstein, I., Goga, O., McCoy, D., & Lauinger, T. (2021). Understanding engagement with U.S. (mis)information news sources on Facebook. *IMC '21: ACM Internet Measurement Conference*, 444–463. https://doi.org/10.1145/3487552.3487859

Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Hamilton, K., & Sandvig, C. (2015, April). "I always assumed that I wasn't really that close to [her]": Reasoning about invisible algorithms in News Feeds. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 153-162. https://doi.org/10.1145/2702123.2702556

Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016). First I "like" it, then I hide it: Folk theories of social feeds. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2371–2382. https://doi.org/10.1145/2858036.2858494

Facebook Data Policy. (2021, May). https://web.archive.org/web/20210526204336if_/https://www.facebook.com/policy.php

Facebook: Previous Data Policy. (2016). https://www.facebook.com/about/privacy/previous

Facebook, social media privacy, and the abuse of data (S.Hrg. 115-683). (2018). https://www.congress.gov/event/115th-congress/senate-event/LC64510/text?s=1&r=59

Fiesler, C., & Hallinan, B. (2018, April). "We are the product": Public reactions to online data sharing and privacy controversies in the media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1-13.

Filter Bubble Transparency Act, S. 2763, Senate, 116 (2019). https://www.congress.gov/bill/116th-congress/senate-bill/2763

Garber, M. (2013, August 23). *How Google's autocomplete was ... Created / invented / born*. The Atlantic. https://www.theatlantic.com/technology/archive/2013/08/how-googles-autocomplete-was-created-invented-born/278991/

Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies* (pp. 167–194). The MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Gillespie, T. (2016). Algorithms. In B. Peters (Ed.), *Digital keywords: A vocabulary of information, society, and culture* (pp. 18–30). Princeton University Press. https://press.princeton.edu/books/hardcover/9780691167336/digital-keywords

Gillespie, T. (2017). Algorithmically recognizable: Santorum's Google problem, and Google's Santorum problem. *Information, Communication & Society*, *20*(1), 63–80. https://doi.org/10.1080/1369118X.2016.1199721

Gillespie, T. (2018). *Custodians of the internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

GMI Blogger. (2022, April 18). YouTube user statistics 2022. *Official GMI Blog*. https://www.globalmediainsight.com/blog/youtube-users-statistics/

Golebiewski, M. & boyd, d. (2019). Data voids: Where missing data can be easily exploited. Data & Society. https://datasociety.net/wp-content/uploads/2019/11/Data-Voids-2.0-Final.pdf

Google. (2021, February 4). *Privacy Policy – Privacy & Terms*. https://policies.google.com/privacy?hl=en-US#whycollect

Google announces third quarter 2022 results. (2022, October 5). Alphabet Investor Relations. https://abc.xyz/investor/static/pdf/2022Q3_alphabet_earnings_release.pdf

Google Search—Discover How Google Search Works. (2022, March 18). Google Search - Discover How Google Search Works. https://web.archive.org/web/20220318143347/https://www.google.com/search/howsearchworks/

Grimmelmann, J. (2013). Speech engines. *Minnesota Law Review*, *98*(3), 868–952.

Haas, A., & Unkel, J. (2017). Ranking versus reputation: Perception and effects of search result credibility. *Behaviour & Information Technology*, *36*(12), 1285–1298. https://doi.org/10.1080/0144929X.2017.1381166

Halavais, A. (2018). *Search Engine Society* (2nd ed.). Polity Press. https://www.wiley.com/en-us/Search+Engine+Society%2C+2nd+Edition-p-9781509516834

Hargittai, E. (2009). An update on survey measures of web-oriented digital literacy. *Social Science Computer Review*, *27*(1), 130–137. https://doi.org/10.1177/0894439308318213

Hargittai, E. (2016). "What can I really do?" Explaining the privacy paradox with online apathy. *International Journal of Communication, 10*: 3737-3757. https://ijoc.org/index.php/ijoc/article/viewFile/4655/1738

Herring, S. C. (2010). Web content analysis: Expanding the paradigm. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (pp. 233–249). Springer Netherlands. https://doi.org/10.1007/978-1-4020-9789-8_14

Hillis, K., Petit, M., & Jarrett, K. (2012). *Google and the Culture of Search*. Routledge.

Hosseinmardi, H., Ghasemian, A., Clauset, A., Mobius, M., Rothschild, D. M., & Watts, D. J. (2021). Examining the consumption of radical content on YouTube. *Proceedings of the National Academy of Sciences, 118*(32), e2101967118. https://doi.org/10.1073/pnas.2101967118

How our Quality Raters make Search results better—Google Search Help. (n.d.). Retrieved May 7, 2022, from https://support.google.com/websearch/answer/9281931?hl=en

How YouTube Works—Product Features, Responsibility, & Impact. (2022, March 19). https://web.archive.org/web/20220319200423/https://www.youtube.com/howyoutubeworks/

Hunt, R., & McKelvey, F. (2019). Algorithmic regulation in media and cultural policy: A framework to evaluate barriers to accountability. *Journal of Information Policy, 9,* 307–335. https://doi.org/10.5325/jinfopoli.9.2019.0307

Ibdah, D., Lachtar, N., Raparthi, S. M., & Bacha, A. (2021). "Why should I read the privacy policy, I just need the service": A study on attitudes and perceptions toward privacy policies. *IEEE Access, 9,* 166465–166487. https://doi.org/10.1109/ACCESS.2021.3130086

Introna, L. D., & Nissenbaum, H. (2000). Shaping the web: Why the politics of search engines matters. *The Information Society*, *16*(3), 169–185. https://doi.org/10.1080/01972240050133634

Jenner & Brock. (2001, September). *CIPA and Schools*. American Librarian Association. https://www.ala.org/advocacy/advleg/federallegislation/cipa/schoolsandcipa

Jungherr, A., & Schroeder, R. (2021). Disinformation and the structural transformations of the public arena: Addressing the actual challenges to democracy. *Social Media + Society*, *7*(1), 2056305121988928. https://doi.org/10.1177/2056305121988928

Kantayya, S. (Director). (2020). *Coded Bias* [Film]. Netflix.

Karizat, N., Delmonaco, D., Eslami, M., & Andalibi, N. (2021). Algorithmic folk theories and identity: How TikTok users co-produce knowledge of identity and engage in algorithmic resistance. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1-44.

Kim, S.-Y. (2017). Intercoder reliability techniques: Cohen's Kappa. In M. Allen (Ed.), *The SAGE Encyclopedia of Communication Research Methods* (pp. 735–738). SAGE Publications, Inc. https://methods.sagepub.com/reference/thesage- encyclopedia-of-communicationresearch- methods/i6767.xml

Klawitter, E. & Hargittai, E. (2018). "It's like learning a whole other language": The role of algorithmic skills in the curation of creative good. *International Journal of Communication, 12*(2018): 3490-3510.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). Discrimination in the age of algorithms. *Journal of Legal Analysis*, *10*, 113–174. https://doi.org/doi:10.1093/jla/laz001

Knight, S. A., & Spink, A. (2008). Toward a web search information behavior model. In A. Spink & M. Zimmer (Eds.), *Web Search: Multidisciplinary Perspectives* (pp. 209–234). Springer. https://link.springer.com/chapter/10.1007/978-3-540-75829-7_12

Koenig, A. (2020). The algorithmics know me and I know them: Using student journals to uncover algorithmic literacy awareness. *Computer and Composition*, *58*(102611).

Kozak, N. I. (2018). Fighting for the Internet: Online Blackout Protests and Internet Legislation in the United States, 1996-2018. *M/C Journal*, *21*(3), Article 3. https://doi.org/10.5204/mcj.1415

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, *42*(5), 361–371. https://www.proquest.com/scholarly-journals/inside-search-process-information-seeking-users/docview/216895301/se-2

Lanel, G. H. J., & Jayawardena, H. (2020). A study on graph theory properties of on-line social networks. *International Journal of Scientific and Research Publications (IJSRP)*, *10*(3), 267-274. https://doi.org/10.29322/IJSRP.10.03.2020.p9929

Latour, B. (2005). *Reassembling the social: An introduction to Actor-Network Theory*. Oxford University Press.

Lawrence, S. (2000). Context in web search. *Bulletin of the Technical Committee on Data Engineering*, *23*(3), 25–32.

Ledwich, M., & Zaitsev, A. (2020). Algorithmic extremism: Examining YouTube's rabbit hole of radicalization. *First Monday*. https://doi.org/10.5210/fm.v25i3.10419

Leonardi, P. M. (2009). Crossing the implementation line: The mutual constitution of technology and organizing across development and use activities. *Communication Theory*, *19*(3), 278-310.

Lessig, L. (2000). Code is law: On liberty in cyberspace. *Harvard Magazine*. https://cartorios.org/wp-content/uploads/2020/11/LESSIG._Lawrence_Code_is_law.pdf

Lessig, L. (2003). Law Regulating Code Regulating Law. *Loyola University Chicago Law Journal*, *35*(1), 1–14.

Libert, T. (2015). Privacy implications of health information seeking on the web. *Communications of the ACM*, *58*(3), 68–77. https://doi.org/10.1145/2658983

Loveland v. Facebook, 20-cv-6260-JMY (United States District Court, E.D. Pennsylvania. May 3, 2021). https://casetext.com/case/loveland-v-facebook

Mager, A. (2010, October 1). *The social construction of search engines*. Humlab, Umeå University. https://www.umu.se/en/research/projects/the-social-construction-of-search-engines/

Mager, A. (2012). Algorithmic ideology: How capitalist society shapes search engines. *Information, Communication & Society*, *15*(5), 769–787. https://doi.org/10.1080/1369118X.2012.676056

Marcella, R., & Chowdhury, G. (2018). Eradicating information poverty: An agenda for research. *Journal of Librarianship and Information Science*. https://doi.org/10.1177/0961000618804589

Martin, K. (2012). Information technology and privacy: Conceptual muddles or privacy vacuums? *Ethics and Information Technology*, *14*(4), 267–284. https://doi.org/10.1007/s10676-012-9300-3

Meta—Q1 2022 Earnings. (2022, April 27). https://investor.fb.com/investor-events/event-details/2022/Q1-2022-Earnings/default.aspx

Meta Reports Third Quarter 2022 Results. (2022, October 26). Meta Investor Relations. https://investor.fb.com/investor-news/press-release-details/2022/Meta-Reports-Third-Quarter-2022-Results/default.aspx

Miller, C.C. (2013, May 15). Google introduces new search tools to try to read our minds. *New York Times*. https://archive.nytimes.com/bits.blogs.nytimes.com/2013/05/15/google-introduces-new-search-tools-to-try-to-read-our-minds/

Nakamura, S., Konishi, S., Jatowt, A., Ohshima, H., Kondo, H., Tezuka, T., Oyama, S., & Tanaka, K. (2007). Trustworthiness analysis of web search results. *Research and Advanced Technology for Digital Libraries*, 38–49. https://doi.org/10.1007/978-3-540-74851-9_4

Nawaz, A. (2022, January 5). *How disinformation around Jan. 6 riot has divided Americans*. PBS NewsHour. https://pbswisconsin.org/watch/newshour/misinformation-1641423149/

Noble, S. U. (2012, Spring). Missed connections: What search engines say about women. *Bitch*, *12*, 6.

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, *23*(1), 128–147. https://doi.org/10.1080/1369118X.2018.1486870

Oeldorf-Hirsch, A., & Neubaum, G. (2021). What do we know about algorithmic literacy? The status quo and a research agenda for a growing field. *SocArXiv*. [Preprint, updated June 2022]. https://osf.io/preprints/socarxiv/2fd4j/

Olson, H. A. (2001). The power to name: Representation in library catalogs. *Signs*, *26*(3), 639–668. JSTOR.

Olson, H. A. (2002). *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Kluwer Academic.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (1st edition). Crown.

Orlowski, J. (Director). (2020). *Social Dilemma* [Film]. Netflix.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*(3), 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin.

Pasquale, F. (2016). *The black box society: The secret algorithms that control money and information*. Harvard University Press. https://www.hup.harvard.edu/catalog.php?isbn=9780674970847

Pinch, T., & Bijker, W. E. (2012). The social construction of facts and artifacts: Or how the sociology of science and the sociology of technology might benefit each other. In W. E. Bijker, T. P. Hughes, & T. Pinch (Eds.), *The social construction of technological systems: New directions in the sociology and history of technology* (pp. 11–44). MIT Press.

Platform Accountability and Consumer Transparency Act, S. 4066, Senate, 116 (2020). https://www.congress.gov/bill/116th-congress/senate-bill/4066/

Powers, E. (2017). My News Feed is Filtered? *Digital Journalism*, *5*(10), 1315–1335. https://doi.org/10.1080/21670811.2017.1286943

Privacy & Terms – Google. (2021, May 22). https://web.archive.org/web/20210522012956/https://policies.google.com/?hl=en-US

Proferes, N. (2015). *Informational Power on Twitter: A Mixed-Methods Exploration of User Knowledge and Technological Discourse about Information Flows*. University of Wisconsin-Milwaukee.

Protecting Democracy from Disinformation Act, H.R. 7012, House of Representatives, 116 (2020). https://www.congress.gov/bill/116th-congress/house-bill/7012

Public Library Association. (n.d.) *Digital literacy*. Accessed on January 12, 2021. https://www.ala.org/pla/initiatives/digitalliteracy

Purcell, K., Brenner, J., & Raine, L. (2012). *Main findings: Search engine use 2012*. Pew Research Center. https://www.pewresearch.org/internet/2012/03/09/main-findings-11/

Rader, E., & Gray, R. (2015). Understanding user beliefs about algorithmic curation in the Facebook news feed. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 173–182. https://doi.org/10.1145/2702123.2702174

Rainie, L. & Anderson, J. (2017, February 8). Code-dependent: Pros and cons of the Algorithmic Age. *Pew Research Center*. https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/

Results Sources. (2022). https://help.duckduckgo.com/duckduckgo-help-pages/results/sources/, retrieved October 17, 2022.

Roberts, S. T., & Noble, S. U. (2016). Empowered to name, inspired to act: Social responsibility and diversity as calls to action in the LIS context. *Library Trends*. http://ir.lib.uwo.ca/fimspub/34/?utm_source=ir.lib.uwo.ca%2Ffimspub%2F34&utm_medium=PDF&utm_campaign=PDFCoverPages

Robertson, R. E., Lazer, D., & Wilson, C. (2018). Auditing the personalization and composition of politically-related search engine results pages. *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, 955–965. https://doi.org/10.1145/3178876.3186143

S.Hrg. 115-683—*Facebook, social media privacy and the use and abuse of data*. (2018). https://www.congress.gov/event/115th-congress/senate-event/LC64510/text?s=1&r=59https://www.congress.gov/event/115th-congress/senate-event/LC64510/text?s=1&r=59

Sawant, S., Nair, A., Sultana, S. A., Rajendran, A., & Madathil, K. C. (2021). Evaluation of the Content Quality of YouTube Videos on Alternate Therapies for Prevention or Treatment of COVID-19. *Journal of Consumer Health on the Internet*, *25*(2), 133–155. https://doi.org/10.1080/15398285.2021.1902226

Sawyer, & Jarrahi, M. H. (2014). Sociotechnical approaches to the study of information systems. In *Computing Handbook, Third Edition: Information Systems and Information Technology* (pp. 5-1–5-27). https://doi.org/10.1201/b16768

Schultheiß, S., Sünkler, S., & Lewandowski, D. (2018). We still trust in Google, but less than 10 years ago: An eye-tracking study. *Information Research: An International Electronic Journal*, *23*(3). https://eric.ed.gov/?id=EJ1196314

Search Engine Market Share United States of America. (2020, December). StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share/all/united-states-of-america

Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, *4*(2), 2053951717738104. https://doi.org/10.1177/2053951717738104

Seaver, N. (2019). Knowing Algorithms. In J. Vertesi & D. Ribes (Eds.), *DigitalSTS: A Field Guide for Science & Technology Studies* (pp. 412–422). Princeton University Press,. https://digitalsts.net/wp-content/uploads/2019/03/26_Knowing-Algorithms.pdf

Seidenberg, S. (2009). Copyright in the age of YouTube. *ABA Journal*, *95*(2), 46–51.

Self-Identification of LGBTQ Employees. (n.d.). Human Rights Council Foundation. Retrieved January 23, 2021, from https://www.thehrcfoundation.org/professional-resources/self-identification-of-lgbtq-employees

Selwyn, N. (2009). The digital native – myth and reality. *Aslib Proceedings: New Information Perspectives*, *61*(4): 364-379.

Senate, No. 588, NJ S.B. 588 (22R), New Jersey State Legislature, 2022–2023 (2023). https://www.njleg.state.nj.us/bill-search/2022/S588

Serra, R. (1973). Television delivers people. https://www.moma.org/collection/works/118185

Seymour, T., Frantsvog, D., & Kumar, S. (2011). History of search engines. *International Journal of Management & Information Systems*, *15*(4), 47–58.

Shane, T. (2020, August 10). People are using Facebook and Instagram as search engines. During a pandemic, that's dangerous. *Nieman Lab*. https://www.niemanlab.org/2020/08/people-are-using-facebook-and-instagram-as-search-engines-during-a-pandemic-thats-dangerous/

Shrimaker, N. (2022, September 28). Upcoming restriction period for U.S. ads about social issues, elections, or politics. *Meta*. https://developers.facebook.com/blog/post/2022/09/28/upcoming-restriction-period-for-us-ads/

Sismondo, S. (2010). *An Introduction to Science And Technology Studies* (2nd edition). Wiley-Blackwell.

Sitrin, C. (2023, January 1). New Jersey becomes first state to mandate K-12 students learn information literacy. *POLITICO*. https://www.politico.com/news/2023/01/05/new-jersey-is-the-first-state-to-mandate-k-12-students-learn-information-literacy-00076352

Solove, D. J. (2007). I've got nothing to hide and other misunderstandings of privacy. *San Diego Law. Review*, *44*, 745.

Soukup, P. A. (2014). Looking at, through, and with YouTube. *Communication Research Trends*, *33*(3), 3–34.

Spink, A. (2002). A user-centered approach to evaluating human interaction with Web search engines: An exploratory study. *Information Processing & Management*, *38*(3), 401–426. https://doi.org/10.1016/S0306-4573(01)00036-X

Statista. (2021). *U.S. search engines: Number of core searches 2020*. Statista. https://www.statista.com/statistics/265796/us-search-engines-ranked-by-number-of-core-searches/

Statista. (2022, May 25). *Global time spent on YouTube.com 2021*. Statista. https://www.statista.com/statistics/1257254/youtubecom-time-spent-per-visit/

Storms, E. (2019). Exploring actor-network theory in the investigation of algorithms. In *ACM CHI Conference on Human Factors in Computing Systems*. 1-6. https://www.eliasstorms.net/publication/storms-2019b/storms-2019b.pdf

Sweney, M. (2022, December 13). Twitter 'to lose 32m users in two years after Elon Musk takeover.' *The Guardian*. https://www.theguardian.com/technology/2022/dec/13/twitter-lose-users-elon-musk-takeover-hate-speech

Thompson, S.A. (2022, February 23). Fed up with Google, conspiracy theorists turn to DuckDuckGo. *New York Times*. https://www.nytimes.com/2022/02/23/technology/duckduckgo-conspiracy-theories.html

Toepoel, V. (2017). Online Survey Design. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (2nd ed.). (pp. 184–202). SAGE Publications Inc.

Top 100: The Most Visited Websites in the US. (2021, June). Semrush. https://www.semrush.com/blog/most-visited-websites/

TLDR Act, H.R.6407, House of Representatives, 117th Congress (2022). https://www.congress.gov/bill/117th-congress/house-bill/6407

TLDR Act, S.3501, Senate, 117th Congress (2022). https://www.congress.gov/bill/117th-congress/senate-bill/3501

Tripodi, F. (2018). *Searching for alternative facts: Analyzing scriptural inference in conservative news practices* (p. 64). Data & Society. https://datasociety.net/library/searching-for-alternative-facts/

Updates: Privacy Policy – Privacy & Terms – Google. (2022). https://policies.google.com/privacy/archive?hl=en-US

Urman, A., Makhortykh, M., Ulloa, R., & Kulshrestha, J. (2022). Where the earth is flat and 9/11 is an inside job: A comparative algorithm audit of conspiratorial information in web search results. *Telematics and Informatics, 72*(101860).

U.S. Census Bureau. (2021, July). *U.S. Census Bureau QuickFacts: United States*. U.S. Census Bureau QuickFacts. https://www.census.gov/quickfacts/US

U.S. Census Bureau. (2022, March 10). *Census Bureau Releases Estimates of Undercount and Overcount in the 2020 Census*. Census.Gov. https://www.census.gov/newsroom/press-releases/2022/2020-census-estimates-of-undercount-and-overcount.html

Vaidhyanathan, S. (2011). *The Googlization of everything (and why we should worry)*. University of California Press. https://www.ucpress.edu/book/9780520272897/the-googlization-of-everything

van Dijck, J., & Poell, T. (2013). Understanding social media logic. *Media and Communication*, *1*(1), 2–14. https://doi.org/10.17645/mac.v1i1.70

Von Hilgers, P. (2011). The history of the black box: The clash of a thing and its concept. Cultural Politics, 7(1), 41-58.

Walsh, D. (2018, August 1). *How do online bots shift opinions?* MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/how-do-online-bots-shift-opinions

Warner, J. (2007). Selection power and selection labor for information retrieval. *Journal of the American Society for Information Science & Technology*, *58*(7), 915–923. https://doi.org/10.1002/asi.20576

Wilson, T. D. (1981). On user studies and information needs. *Journal of Documentation*, *37*(1), 3–15. https://doi.org/10.1108/eb026702

Winner, L. (1980). Do artifacts have politics? *Daedalus*, *109*(1), 121–136.

Winner, L. (1993). Upon opening the black box and finding it empty: Social constructivism and the philosophy of technology. *Science, Technology, & Human Values*, *18*(3), 362–378. https://doi.org/10.1177/016224399301800306

What percentage of internet traffic is mobile? (2022). *Oberlo*. https://www.oberlo.com/statistics/mobile-internet-traffic

Wolfram, D., & Xie, H. I. (2002). Traditional IR for web users: A context for general audience digital libraries. *Information Processing & Management*, *38*(5), 627–648. https://doi.org/10.1016/S0306-4573(01)00056-5

Wu, Z., Sanderson, M., Cambazoglu, B. B., Croft, W. B., & Scholer, F. (2020). Providing Direct Answers in Search Results: A Study of User Behavior. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 1635–1644. https://doi.org/10.1145/3340531.3412017

Xie, I. (2008). *Interactive information retrieval in digital environments*. IGI Publishing.

Xie, I. (2010). Information Searching and Search Models. In M. J. Bates & M. N. Maack (Eds.), *Encyclopedia of Library and Information Sciences* (3rd ed.). (pp. 2592–2604). CRC Press. https://doi.org/10.1081/E-ELIS3

Yangyuen, W., Phetkaew, T., & Nuntapichai, S. (2020). A Conceptual Framework for an Information Behavior Model Based on the Collaboration Perspective between User and System for Information Retrieval. *Journal of Information Science Theory and Practice*, *8*(3), 30–46. https://doi.org/10.1633/JISTAP.2020.8.3.3

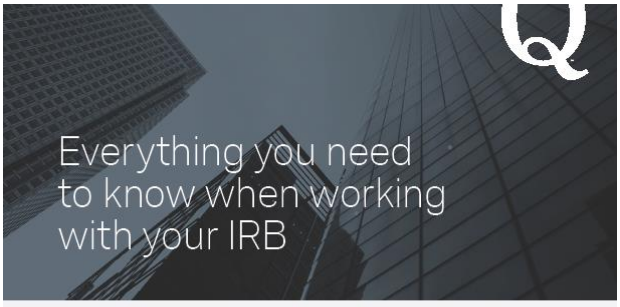
Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society*, *43*(5), 807–824. https://doi.org/10.1177/0163443720972314

Winkler, S. & Zeadally, S. (2016). Privacy policy analysis of popular web platforms. *IEEE Technology and Society Magazine, 35*(2): 75-85. https://doi.org/10.1109/MTS.2016.2554419

Zimmer, M. (2008). Privacy on planet Google: Using the theory of "contextual integrity" to clarify the privacy threats of Google's quest for the perfect search engine. *Journal of Business & Technology Law*, *3*(1), 109–126.

Zimmer, M., Kumar, P., Vitak, J., Liao, Y., & Chamberlain Kritikos, K. (2020). 'There's nothing really they can do with this information': Unpacking how users manage privacy boundaries for personal fitness information. *Information, Communication & Society*, *23*(7), 1020–1037. https://doi.org/10.1080/1369118X.2018.1543442

Appendix A: Qualtrics Documentation



HOW ARE PARTICIPANTS RECRUITED AND VERIFIED?

Participants are recruited from various sources, including website intercept recruitment, member referrals, targeted email lists, gaming sites, customer loyalty web portals, permission-based networks, and social media, etc.

Consumer panel members' names, addresses, and dates of birth are typically validated via third-party verification measures prior to their joining a panel.

B2B participants are subject to additional quality control measures such as LinkedIn matching, phone calls to the participant's place of business, and other third-party verification methods (TrueSample, RelevantID, Verity, etc.).

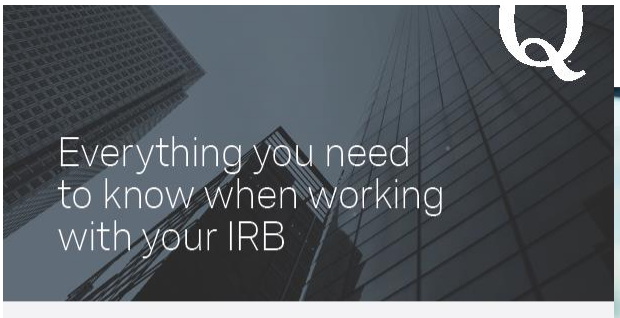
HOW MUCH IS EACH PARTICIPANT INCENTIVIZED?

Our panelists join from a variety of sources. They may be airline customers who chose to join in reward for SkyMiles, retail customers who opted in to get points at their favorite retail outlet, or general consumers who participate for cash or gift cards, etc.

When participants are invited to take a survey, they are informed what they will be compensated.

Because each respondent is compensated differently, it would be inappropriate to inform how much they will be compensated in the cover letter. Instead, we recommend including verbiage along the lines of: "You will be compensated the amount you agreed upon before you entered into the survey."

qualtrics. To find out more, call 02 8310 8031.



HOW ARE PARTICIPANTS RECRUITED AND VERIFIED?

Participants are recruited from various sources, including website intercept recruitment, member referrals, targeted email lists, gaming sites, customer loyalty web portals, permission-based networks, and social media, etc.

Consumer panel members' names, addresses, and dates of birth are typically validated via third-party verification measures prior to their joining a panel.

B2B participants are subject to additional quality control measures such as LinkedIn matching, phone calls to the participant's place of business, and other third-party verification methods (TrueSample, RelevantID, Verity, etc.).

HOW MUCH IS EACH PARTICIPANT INCENTIVIZED?

Our panelists join from a variety of sources. They may be airline customers who chose to join in reward for SkyMiles, retail customers who opted in to get points at their favorite retail outlet, or general consumers who participate for cash or gift cards, etc.

When participants are invited to take a survey, they are informed what they will be compensated.

Because each respondent is compensated differently, it would be inappropriate to inform how much they will be compensated in the cover letter. Instead, we recommend including verbiage along the lines of: "You will be compensated the amount you agreed upon before you entered into the survey."

qualtrics. To find out more, call 02 8310 8031.

qualtrics.

Updated: 4/30/19

ESOMAR 28

28 Questions to Help Buyers of Online Samples

Appendix B: Survey Instrument

Algorithmic Awareness of Online Platforms

C1

University of Wisconsin-Milwaukee Informed Consent to Participate in Research

Study title: Algorithmic Awareness of Online Platforms

Researcher: Meghan Dowell, Doctoral Candidate, School of Information Studies

I'm inviting you to take a survey for research. This survey is completely voluntary. There are no negative consequences if you don't want to take it. If you start the survey, you can always change your mind and stop at any time.

What is the purpose of this study?

I want to understand the ways users of major platforms understand how the search results and news feeds are algorithmically filtered.

What will I do?

This survey will ask questions about your internet use, specifically on search engines, Facebook, and YouTube. It includes questions about the frequency of use, general knowledge on digital terminology, and questions about platform functionality. The survey will take about 10-15 minutes.

Risks

Some questions may be personal or upsetting. You can skip them or quit the survey at any time. Online data being hacked or intercepted: Anytime you share information online there are risks. We're using a secure system to collect this data, but we can't completely eliminate this risk.

Breach of confidentiality: There is a chance your data could be seen by someone who shouldn't have access to it. We're minimizing this risk in the following ways: Data is anonymous. I'll store all electronic data on a password-protected, encrypted computer.

Possible benefits:

Participation in this study benefits society in helping understand how platform users approach the use of their personal data and the knowledge of how their information is algorithmically filtered.

Estimated number of participants: 400

How long will it take? 10-15 minutes

Costs: None

Compensation: You will be compensated according to the amount you agreed upon before entering the survey, set forth by Qualtrics Panel.

Future research: De-identified data (all identifying information removed) may be shared with other researchers. You won't be told specific details about these future research studies.

Where will data be stored? On the researcher's computers, and the servers for the online survey software (Qualtrics).

How long will it be kept? Four years.

Who can see my data? I (the researcher) will have access to de-identified (no names, birthdate, address, etc.) data. This is so we can analyze the data and conduct the study. Qualtrics could link your Panel ID and associated personal information with your survey responses. Make sure you have read Qualtrics Panel's participant and privacy agreements to understand how your personal

information may be used or disclosed. Agencies that enforce legal and ethical guidelines, such as The Institutional Review Board (IRB) at UWM, The Office for Human Research Protections (OHRP). I may share my findings in publications or presentations. If I do, the results will be aggregate (grouped) data, with no individual results. If I quote you, I'll use pseudonyms (fake names).

Questions about the research, complaints, or problems:

Contact Meghan Dowell, mldowell@uwm.edu

Questions about your rights as a research participant, complaints, or problems:

Contact the UWM IRB (Institutional Review Board) at 414-662-3544 / irbinfo@uwm.edu. Please print or save this screen if you want to be able to access the information later.

IRB #: 22.140

IRB Approval Date: January 4, 2022

Agreement to Participate

Your participation is completely voluntary, and you can withdraw at any time. To take this survey, you must be: At least 18 years old, and a US resident. If you meet these criteria and would like to take the survey, please answer yes to the following question.

C2 I am at least 18 years old, a US resident, and I agree to participate in this survey.

- ○ Yes (1)
- ○ No (2)

Skip To: End of Block If I am at least 18 years old, a US resident, and I agree to participate in this survey. = No

End of Block: Consent

Start of Block: Survey

S1 Demographics

Q1 What is your age?

- ○ under 18 (6)
- ○ 18-24 (1)
- ○ 25-34 (2)
- ○ 35-44 (3)
- ○ 45-54 (4)
- ○ 55-59 (5)
- ○ 60-64 (7)
- ○ 65-74 (8)
- ○ 75+ (9)

Skip To: End of Block If What is your age? = under 18

Q1.5 What is your race?

- [ ] African-American or Black (1)
- [ ] American Indian or Alaskan Native (3)
- [ ] Asian or Pacific Islander (2)
- [ ] White (4)
- [ ] Other (5) _____

Q1.5b What is your ethnicity?

- ○ Hispanic (1)
- ○ Non-hispanic (2)

Q2 What is your gender identity?

- ○ Man (1)
- ○ Woman (2)
- ○ Non-binary / non-conforming (3)
- ○ Prefer not to say (4)
- ○ Prefer to self-describe (5) _____

Q3 What is the highest level of education completed?

- ○ Less than high school (1)
- ○ Some high school (2)
- ○ High school graduate or equivalent (3)
- ○ Some college/university (4)
- ○ Bachelor's degree (5)
- ○ Master's degree (6)
- ○ Professional degree (7)
- ○ Doctorate degree (8)

Q4 Please indicate your occupation:

- ○ Management, professional, and related (1)
- ○ Service (2)
- ○ Education (10)
- ○ Sales and office (3)
- ○ Farming, fishing, and forestry (4)
- ○ Construction, extraction, and maintenance (5)
- ○ Production, transportation, and material moving (6)
- ○ Government (7)
- ○ Retired (8)
- ○ Unemployed (9)

Q5 What is your political affiliation?

- ○ Democrat (1)
- ○ Republican (2)
- ○ Independent (3)
- ○ Prefer not to disclose (4)
- ○ Prefer to self-describe (5) _____

Page Break

S2 Internet Literacy and Usage

Q6 What type of internet connection do you have? (check all that apply)

- [ ] Dial-up (1)
- [ ] Broadband/DSL/Cable (2)
- [ ] Satellite (3)
- [ ] Mobile (limited data) (4)
- [ ] Mobile (unlimited data) (5)
- [ ] LTE hotspot (6)
- [ ] Public access (7)
- [ ] Don't know (8)
- [ ] Other (9) _____

Q7 Where do you access the internet? (check all that apply)

- [ ] Home (1)
- [ ] Work (2)
- [ ] Phone (3)
- [ ] Library (4)
- [ ] Family/friend home (5)
- [ ] Another internet connected place (6)

Q9 What is your familiarity with the following terms?

	Not at all familiar (1)	A little familiar (2)	Mostly familiar (3)	Very familiar (4)
Advanced search (1)	○	○	○	○
PDF (2)	○	○	○	○
Wiki (3)	○	○	○	○
Blog (4)	○	○	○	○
Cache (5)	○	○	○	○
Malware (6)	○	○	○	○
Phishing (7)	○	○	○	○
FYP (8)	○	○	○	○
News feed (9)	○	○	○	○
Trending (10)	○	○	○	○
Cookies (11)	○	○	○	○
Pixel tags (12)	○	○	○	○
Autoplay (13)	○	○	○	○

Page Break

S2a Internet Literacy and Usage: Search Engines

Q10 How often do you use a search engine?

- Never (1)
- Less than monthly (2)
- Monthly (3)
- Weekly (4)
- Daily (5)
- More than once a day (6)

Q11 Which search engine do you use? (check all that apply)

- [ ] Bing (1)
- [ ] Brave (2)
- [ ] DuckDuckGo (3)
- [ ] Google (4)
- [ ] StartPage (5)
- [ ] Yahoo! (6)
- [ ] Other (7) _____

Q12 How do you access the search engine? (check all that apply)

- [ ] Website.com from a web browser (1)
- [ ] URL bar input (2)
- [ ] Mobile browser (3)
- [ ] Voice activated search (Siri, Alexa, Google Home, etc) (4)
- [ ] Other (5) _____

Q14 What is the purpose of your search? (check all that apply)

- ☐ Navigation to sites (2)
- ☐ Look up news on a topic or event (3)
- ☐ Check accuracy of news or information (4)
- ☐ Answer medical or health questions (5)
- ☐ Learn about politics and current events (6)
- ☐ Find entertaining content (7)
- ☐ Complete a work-related task (8)
- ☐ Other (9) _____

Q15 I feel like Google is discussed in the news.

- ○ Yes (1)
- ○ Somewhat (2)
- ○ No (3)
- ○ Unsure (4)

Q16 I keep up with news about Google.

- ○ Yes (1)
- ○ Somewhat (2)
- ○ No (3)
- ○ Unsure (4)

Page Break

S2b Internet Literacy and Usage: YouTube

Q17 Do you use YouTube?

- Yes (1)
- No (2)

Skip To: S2c If Do you use YouTube? = No

Q18 How often do you check YouTube to watch videos?

- Monthly (1)
- Weekly (2)
- Daily (3)
- Less than 5 times per day (4)
- 5 - 10 times per day (5)
- 10+ times per day (6)

Q19 For how long do you watch videos per day?

- ○ Less than 30 minutes (1)
- ○ 30 minutes - 1 hour (2)
- ○ 1 - 2 hours (3)
- ○ 2+ hours (4)

Q20 How do you access YouTube? (check all that apply)

- ☐ YouTube.com from web-browser (1)
- ☐ Mobile app (2)
- ☐ Smart TV app (3)
- ☐ Other (4) _____

Q21 What is the purpose of your use of YouTube? (check all that apply)

- [ ] Navigation to sites (2)
- [ ] Look up news on a topic or event (3)
- [ ] Check accuracy of news or information (4)
- [ ] Answer medical or health questions (5)
- [ ] Learn about politics or current events (6)
- [ ] Find entertaining content (7)
- [ ] Complete a work-related task (8)
- [ ] Learn a new skill (9)
- [ ] Create and upload content (10)
- [ ] Other (11) _____

Q22 I feel like YouTube is discussed in the news.

- Yes (1)
- Somewhat (2)
- No (3)
- Unsure (4)

Q23 I keep up with news about YouTube.

- Yes (1)
- Somewhat (2)
- No (3)
- Unsure (4)

Page Break

S2c Internet Literacy and Usage: Facebook

Q24 Do you have a Facebook account?

- ○ Yes (1)
- ○ No (2)

Skip To: S3 If Do you have a Facebook account? = No

Q24b If you have a Facebook account, do you actively use it?

- ○ Yes (1)
- ○ No (2)

Skip To: S3 If If you have a Facebook account, do you actively use it? = No

Q25 How often do you check your timeline in Facebook?

- Monthly (1)
- Weekly (2)
- Daily (3)
- Less than 5 times per day (4)
- 5 - 10 times per day (5)
- 10+ times per day (6)

Q26 For how long do you use Facebook per day?

- Less than 30 minutes (1)
- 30 minutes - 1 hour (2)
- 1 - 2 hours (3)
- 2+ hours (4)

Q27 How do you access Facebook? (check all that apply)

- ☐ Facebook.com from web-browser (1)
- ☐ Mobile app (2)
- ☐ Other (3) _____

Q28 What is the purpose of your use of Facebook? (check all that apply)

- ☐ Navigation to sites (2)
- ☐ Look up news on a topic or event (3)
- ☐ Check accuracy of news or information (4)
- ☐ Answer medical or health questions (5)
- ☐ Learn about politics or current events (6)
- ☐ Find entertaining content (7)
- ☐ Complete a work-related task (8)
- ☐ Keep up with family and friends (9)
- ☐ Share pictures and videos (10)
- ☐ Share links to content (11)
- ☐ Other (12) _____

Q29 I feel like Facebook is discussed in the news.

- Yes (1)
- Somewhat (2)
- No (3)
- Unsure (4)

Q30 I keep up with news about Facebook.

- Yes (1)
- Somewhat (2)
- No (3)
- Unsure (4)

Page Break

S3 Platform Knowledge: Google

Answer the following questions to the best of your ability using your current knowledge.

Q31 Do users of Google agree to a Terms of Service when conducting a search?

- ○ Yes (1)
- ○ No (2)
- ○ Unsure (3)

Display This Question:

If Do users of Google agree to a Terms and Conditions when conducting a search? = Yes

Q31a Have you read or skimmed Google's Terms of Service?

- ○ Yes (1)
- ○ No (2)

Q32 Does Google have a privacy policy regarding search users?

- Yes (1)
- No (2)
- Unsure (3)

Display This Question:

If Does Google have a privacy policy regarding search users? = Yes

Q32a Have you read or skimmed Google's privacy policy?

- Yes (1)
- No (2)

Q33 Does Google use personal information to customize search results?

- Yes (1)
- No (2)
- Unsure (3)

Q34 Does Google give the option to turn off algorithmic filtering?

- Yes (1)
- No (2)
- Unsure (3)

Q35 Does Google give the option to delete user history of personal information?

- Yes (1)
- No (2)
- Unsure (3)

Q36 Does Google charge users to use their search engine?

- Yes (1)
- No (2)
- Unsure (3)

Q37 Please identify which of the following ways Google generates revenue. (check all that apply)

- [ ] Advertising (1)
- [ ] Search Engine Optimization (SEO) (2)
- [ ] Selling user data (3)
- [ ] Email services (4)
- [ ] Cloud storage services (5)
- [ ] Pay for higher result rankings (6)

Q38 Do any of the following actions influence your search results?

	Yes (1)	No (2)	Unsure (3)
Your actions on Google (1)	○	○	○
Your actions on websites other than Google (2)	○	○	○
Details about your online session (3)	○	○	○
Details about your smart phone usage (4)	○	○	○
Actions taken by other users (5)	○	○	○
Actions taken by Google engineers/editors/curators (6)	○	○	○
Data details between applications (7)	○	○	○
Details about accounts across devices (8)	○	○	○

Q39 Is the following personal data collected about you while using Google?

	Yes (1)	No (2)	Unsure (3)
Interactions with search results (1)	○	○	○
Browser, application, or device in use (2)	○	○	○
Location (3)	○	○	○
Address book contacts (4)	○	○	○
Usage of Google products (YouTube, Maps, Drive) and features (5)	○	○	○
Purchases or financial transactions through Google products (6)	○	○	○
Details about your online session (7)	○	○	○
Details about your smart phone usage (8)	○	○	○

Information from advertisers, app developers, and games from when you are not using Google (9)	○	○	○
Information from non-Google apps and advertisers (10)	○	○	○
Voice and audio information when you use audio features (11)	○	○	○

Page Break

S3b Platform Knowledge: YouTube

Answer the following questions to the best of your ability using your current knowledge.

Q40 Do users of YouTube agree to a Terms of Serice when watching and/or posting a video?

- ○ Yes (1)
- ○ No (2)
- ○ Unsure (3)

Display This Question:

If Do users of YouTube agree to a Terms and Conditions when watching and/or posting a video? = Yes

Q40a Have you read or skimmed YouTube's Terms of Service?

- ○ Yes (1)
- ○ No (2)

Q41 Does YouTube have a privacy policy regarding its platform users?

- Yes (1)
- No (2)
- Unsure (3)

Display This Question:

If Does YouTube have a privacy policy regarding its platform users? = Yes

Q41a Have you read or skimmed YouTube's privacy policy?

- Yes (1)
- No (2)

Q42 Does YouTube use personal information to customize video results?

- Yes (1)
- No (2)
- Unsure (3)

Q43 Does YouTube give the option to turn off algorithmic filtering?

- Yes (1)
- No (2)
- Unsure (3)

Q44 Does YouTube give the option to delete user history of personal information?

- Yes (1)
- No (2)
- Unsure (3)

Q45 Does YouTube charge users to use their search engine?

- Yes (1)
- No (2)
- Unsure (3)
- Other (4) _____

Q46 Please identify which of the following ways YouTube generates revenue. (check all that apply)

- [ ] Advertising (1)
- [ ] Search Engine Optimization (SEO) (2)
- [ ] Selling user data (3)
- [ ] Video production services (4)
- [ ] Cloud storage services (5)
- [ ] Pay for higher result rankings (6)

Q47 Do any of the following actions influence your video results?

	Yes (1)	No (2)	Unsure (3)
Your actions on YouTube (1)	○	○	○
Your actions on websites other than YouTube (2)	○	○	○
Details about your online session (3)	○	○	○
Details about your smart phone usage (4)	○	○	○
Actions taken by other users (5)	○	○	○
Actions taken by YouTube engineers/editors/curators (6)	○	○	○
Data details between applications (7)	○	○	○
Details about accounts across devices (8)	○	○	○

Q48 Is the following personal data collected about you while using YouTube?

	Yes (1)	No (2)	Unsure (3)
Interactions with videos (likes, comments, subscribes) (1)	○	○	○
Browser, application, or device in use (2)	○	○	○
Location (3)	○	○	○
Address book contacts (4)	○	○	○
Usage of Google products (YouTube, Maps, Drive) and features (5)	○	○	○
Purchases or financial transactions through Google products (6)	○	○	○
Details about your online session (7)	○	○	○

Details about your smart phone usage (8)	○	○	○
Information from advertisers, app developers, and games from when you are not using YouTube (9)	○	○	○
Information from non-YouTube apps and advertisers (10)	○	○	○
Voice and audio information when you use audio features (11)	○	○	○

Page Break

S3c Platform Knowledge: Facebook Answer the following questions to the best of your ability using your current knowledge.

Q49 Do users of Facebook agree to a Terms of Service when creating an account?

- Yes (1)
- No (2)
- Unsure (3)

Display This Question:

If Do users of Facebook agree to a Terms and Conditions when creating an account? = Yes

Q49a Have you read or skimmed Facebook's Terms of Service?

- Yes (1)
- No (2)

Q50 Does Facebook have a privacy policy regarding active users of its platform?

- Yes (1)
- No (2)
- Unsure (3)

Display This Question:

If Does Facebook have a privacy policy regarding active users of its platform? = Yes

Q50a Have you read or skimmed Facebook's privacy policy?

- Yes (1)
- No (2)

Q51 Does Facebook use personal information to customize the newsfeed?

- Yes (1)
- No (2)
- Unsure (3)

Q52 Does Facebook give the option to turn off algorithmic filtering?

- Yes (1)
- No (2)
- Unsure (3)

Q53 Does Facebook give the option to delete user history of personal information?

- Yes (1)
- No (2)
- Unsure (3)

Q54 Does Facebook charge users to use their platform?

- Yes (1)
- No (2)
- Unsure (3)
- Other (4) _____

Q55 Please identify which of the following ways Facebook generates revenue. (check all that apply)

- [ ] Advertising (1)
- [ ] Search Engine Optimization (SEO) (2)
- [ ] Selling user data (3)
- [ ] Video production services (4)
- [ ] Cloud storage services (5)
- [ ] Pay for higher result rankings (6)

Q56 Do any of the following actions influence your newsfeed?

	Yes (1)	No (2)	Unsure (3)
Your actions on Facebook (1)	○	○	○
Your actions on websites other than Facebook (2)	○	○	○
Details about your online session (3)	○	○	○
Details about your smart phone usage (4)	○	○	○
Actions taken by other users (5)	○	○	○
Actions taken by Facebook engineers/editors/curators (6)	○	○	○
Data details between applications (7)	○	○	○
Details about accounts across devices (8)	○	○	○

Q57 Is the following personal data collected about you while using Facebook?

	Yes (1)	No (2)	Unsure (3)
Interactions on Pages, in Groups, and hashtags (1)	○	○	○
Frequency of communication to people and groups (2)	○	○	○
Address book contacts (3)	○	○	○
Usage of Facebook products (Instagram, Messenger, WhatsApp) and features (4)	○	○	○
Purchases or financial transactions from Facebook products (Instagram, Messenger) (5)	○	○	○

Details about your online session (6)	○	○	○
Details about your smart phone usage (7)	○	○	○
Information from advertisers, app developers, and games from when you are not using Facebook (8)	○	○	○
Information from non-Facebook apps and advertisers (9)	○	○	○

Page Break

S4 Accuracy and Trust

Q58 In general, do you think internet search engines are a fair and unbiased source of information?

- Yes (1)
- No (2)
- Depends (3)
- Don't know (4)

Q59 How reliable are different sources of information?

	Very Reliable (1)	Mostly Reliable (2)	Somewhat Reliable (3)	Unreliable (4)	N/A (5)
Search engine results (1)	○	○	○	○	○
News radio (2)	○	○	○	○	○
Talk radio (3)	○	○	○	○	○
Family, friends, colleagues (4)	○	○	○	○	○
Online news (5)	○	○	○	○	○
Newspapers (6)	○	○	○	○	○
Television news (7)	○	○	○	○	○

Cable television news (8)	○	○	○	○	○
Facebook (9)	○	○	○	○	○
YouTube (10)	○	○	○	○	○
Twitter (11)	○	○	○	○	○
Parler (12)	○	○	○	○	○
Wikipedia (13)	○	○	○	○	○

Q60 Do two people get the same search results if they entered the same search terms at the same time on Google?

- ○ Same (1)
- ○ Different (2)
- ○ Unsure (3)

Display This Question:

If Do two people get the same research results if they entered the same search terms at the same tim... = Different

Q60a If different, why?

_____

Display This Question:

If Do two people get the same research results if they entered the same search terms at the same tim... = Same

Q60b If the same, why?

_____

Display This Question:

If Do two people get the same research results if they entered the same search terms at the same tim... = Unsure

Q60c If unsure, why?

_____

End of Block: Survey

Appendix C: Privacy Policy Analysis

Table 47: Google & YouTube Policy Analysis

Platform	Date of Policy	Policy
Google and YouTube	22-May-21	Privacy Policy

Policy Clause	Influences Results/Autoplay/Newsfeed (Author Generated)	Type of Data Collected (Author Generated)
When you're not signed into a Google Account, we store the information we collect with unique identifiers tied to the browser, application, or device you're using.	Details about your online session, Details about accounts across devices, Data details btw applications	Details about your smart phone usage, Browser, application, or device in use, Information from non-Platform apps and advertisers
We also collect the content you create, upload, or receive from others when using our services.	Your actions on platform, Actions taken by other users	Interactions with Platform (clicks, likes, comments, subscribes)
We collect information about the apps, browsers, and devices you use to access Google services, which helps us provide features like automatic product updates and dimming your screen if your battery runs lows.	Your actions on platform, Details about your smart phone usage, Details about accounts across devices	Details about your smart phone usage, Details about your online session, Browser, application, or device in use
The information we collect includes unique identifiers, browser type and settings, device type and settings, operating system, mobile network information including carrier name and phone number, and application version number.	Details about your online session, Details about accounts across devices	Browser, application, or device in use, Details about your smart phone usage

The activity information we collect may include: terms you search for, videos you watch, views and interactions with content and ads, voice and audio information when you use audio features, purchase activity, people with whom you communicate or share content, activity on third-party sites and apps that use our services, chrome browsing history you've synced with your Google Account	Your actions on platform, Your actions on other websites, Details about accounts across devices, Data details btw applications, Actions taken by other users, Data details btw applications	Interactions with Platform (clicks, likes, comments, subscribes), Usage of Platform products and features, Purchases or financial transactions through Platform products, Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Purchases or financial transactions through Platform products, Usage of Platform products and features
Your location can be determined with varying degrees of accuracy by: GPS, IP address, sensor data from your device, information about things near your device, such as Wi-Fi access points, cell towers, and Bluetooth enabled devices	Details about your online session, Details about accounts across devices, Details about your smart phone usage	Location, Browser, application, or device in use, Details about your online session, Details about your smart phone usage
In some circumstances, Google also collects information about you from publicly accessible sources.	Your actions on other websites	Information from non-Platform apps and advertisers
We use various technologies to collect and store information, including cookies, pixel tags, local storage, such as browser web storage or application data caches, databases, and server logs.	Details about accounts across devices, Data details btw applications, Actions taken by platform engineers/editors/curators, Details about your smart phone usage, Details about your online session, Your actions on other websites, Your actions on platform, Data details btw applications	Details about your online session, Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Usage of Platform products and features, Purchases or financial transactions through Platform products, Details about your online session, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers

We also allow specific partners to collect information from your browser or device for advertising and measurement purposes using their own cookies or similar technologies.	n/a	Information from advertisers, app developers, and games from when you are not using Platform, Browser, application, or device in use, Details about your smart phone usage
We share information publicly to show trends about the general use of our services.	Actions taken by other users	n/a
We collect information to provide better services to all our users - from figuring out basic stuff like which language you speak, to more complex things like which ads you'll find most useful, the people who matter most to you online, or which YouTube videos you might like.	Details about your online session, Data details btw applications, Actions taken by other users, Data details btw applications	Voice and audio information when you use audio features, Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Details about your online session, Details about your smart phone usage, Usage of Platform products and features
We use your information to deliver our services, like processing the terms you search for in order to return results or helping you share content by suggesting recipients from your contacts.	Your actions on platform, Actions taken by other users, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Address book contacts, Details about your online session
We also use your information to ensure our services are working as intended, such as tracking outages or troubleshooting issues that you report to us.	n/a	Browser, application, or device in use, Usage of Platform products and features, Details about your online session, Information from advertisers, app developers, and games from when you are not using Platform

We use the information we collect to customize our services for you, including providing recommendations, personalized content, and customized results.	Your actions on platform, Your actions on other websites, Details about your online session, Details about your smart phone usage, Actions taken by other users, Actions taken by platform engineers/editors/curators, Actions taken by platform engineers/editors/curators, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Location, Browser, application, or device in use, Address book contacts, Usage of Platform products and features, Purchases or financial transactions through Platform produ cts, Details about your online session, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers, Voice and audio information when you use audio features
Depending on your settings, we may also show you personalized ads based on your interests.	Details about accounts across devices, Data details btw applications, Actions taken by Platform engineers/editors/curators, Actions taken by other users, Details about your smart phone usage, Details about your online session, Your actions on other websites, Your actions on Platform	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Location, Address book contacts, Usage of Platform products and features, Purchases or financial transactions through Platform produ cts, Details about your online session, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers, Voice and audio information when you use audio features

We use automated systems that analyze your content to provide you with things like customized search results, personalized ads, or other features tailored to how you use our services.	Your actions on Platform, Your actions on other websites, Details about your online session, Details about your smart phone usage, Actions taken by other users, Actions taken by Platform engineers/editors/curators, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Location, Address book contacts, Usage of Platform products and features, Purchases or financial transactions through Platform produ cts, Details about your online session, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers, Voice and audio information when you use audio features
We collect information about your location when you use our services, which helps us offer features like driving directions for your weekend getaway or showtimes for movies playing near you.	Details about your smart phone usage	Location

Table 48: Facebook Policy Analysis

Platform	**Date of Policy**	**Policy**
Facebook	26-May-21	Data Policy

Policy Clause	**Influences Results/Autoplay/Newsfeed (Author Generated)**	**Type of Data Collected (Author Generated)**

We collect the content, communications and other information you provide when you use our Products, including when you sign up for an account, create or share content, and message or communicate with others. This can include information in or about the content you provide (like metadata), such as the location of a photo or the date a file was created.	Your actions on platform, Actions taken by other users, Your actions on Platform, Details about accounts across devices, Details about your smart phone usage, Details about your online session, Your actions on other websites, Actions taken by Platform engineers/editors/curators, Data details btw applications	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Location, Address book contacts, Usage of Platform products and features, Purchases or financial transactions through Platform produ cts, Details about your online session, Details about your smart phone usage, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers, Voice and audio information when you use audio features

It can also include what you see through features we provide, such as our camera, so we can do things like suggest masks and filters that you might like, or give you tips on using camera formats. Our systems automatically process content and communications you and others provide to analyze context and what's in them for the purposes described below.	Your actions on Platform, Your actions on other websites, Details about your online session, Details about your smart phone usage, Actions taken by other users, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Location, Address book contacts, Usage of Platform products and features, Purchases or financial transactions through Platform products, Details about your online session, Details about your smart phone usage, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers, Voice and audio information when you use audio features
We collect information about the people, Pages, accounts, hashtags and groups you are connected to and how you interact with them across our Products, such as people you communicate with the most or groups you are part of.	Your actions on Platform, Your actions on other websites, Actions taken by other users, Data details btw applications	Interactions with Platform (clicks, likes, comments, subscribes), Address book contacts, Details about your online session, Information from advertisers, app developers, and games from when you are not using Platform
We also collect contact information if you choose to upload, sync or import it from a device (such as an address book or call log or SMS log history), which we use for things like helping you and others find people you may know and for the other purposes listed below.	Your actions on Platform, Details about your online session, Details about your smart phone usage, Actions taken by other users, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Address book contacts, Details about your online session, Details about your smart phone usage

We collect information about how you use our Products, such as the types of content you view or engage with; the features you use; the actions you take; the people or accounts you interact with; and the time, frequency and duration of your activities. For example, we log when you're using and have last used our Products, and what posts, videos and other content you view on our Products. We also collect information about how you use features like our camera.	Your actions on Platform, Your actions on other websites, Details about your online session, Actions taken by other users, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Usage of Platform products and features, Details about your online session, Information from advertisers, app developers, and games from when you are not using Platform, Voice and audio information when you use audio features
If you use our Products for purchases or other financial transactions (such as when you make a purchase in a game or make a donation), we collect information about the purchase or transaction. This includes payment information, such as your credit or debit card number and other card information; other account and authentication information; and billing, shipping and contact details.	Your actions on Platform, Details about your smart phone usage	Interactions with Platform (clicks, likes, comments, subscribes), Details about your smart phone usage, Purchases or financial transactions through Platform products
We also receive and analyze content, communications and information that other people provide when they use our Products. This can include information about you, such as when others share or comment on a photo of you, send a message to you, or upload, sync or import your contact information.	Actions taken by other users	Address book contacts
Information such as the operating system, hardware and software versions, battery level, signal strength, available storage space, browser type, app and file names and types, and plugins.	Details about your online session, Details about your smart phone usage, Details about accounts across devices	Browser, application, or device in use, Details about your online session, Details about your smart phone usage

Information about operations and behaviors performed on the device, such as whether a window is foregrounded or backgrounded, or mouse movements (which can help distinguish humans from bots).	Your actions on Platform, Your actions on other websites, Details about your online session, Details about your smart phone usage, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Usage of Platform products and features, Details about your online session, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers, Voice and audio information when you use audio features
Unique identifiers, device IDs, and other identifiers, such as from games, apps or accounts you use, and Family Device IDs (or other identifiers unique to Facebook Company Products associated with the same device or account).	Your actions on other websites, Details about your smart phone usage, Data details btw applications, Details about accounts across devices	Usage of Platform products and features, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers
Bluetooth signals, and information about nearby Wi-Fi access points, beacons, and cell towers.	Details about your smart phone usage, Details about your online session	Browser, application, or device in use, Location, Details about your smart phone usage
Information you allow us to receive through device settings you turn on, such as access to your GPS location, camera or photos.	Details about your smart phone usage, Details about accounts across devices	Browser, application, or device in use, Location, Usage of Platform products and features, Details about your smart phone usage, Voice and audio information when you use audio features

Information such as the name of your mobile operator or ISP, language, time zone, mobile phone number, IP address, connection speed and, in some cases, information about other devices that are nearby or on your network, so we can do things like help you stream a video from your phone to your TV.	Details about your smart phone usage	Details about your smart phone usage, Location, Browser, application, or device in use
Data from cookies stored on your device, including cookie IDs and settings. Learn more about how we use cookies in the Facebook Cookies Policy and Instagram Cookies Policy.	Your actions on Platform, Your actions on other websites, Details about your online session, Details about your smart phone usage, Details about your smart phone usage, Actions taken by other users, Data details btw applications, Details about accounts across devices	Interactions with Platform (clicks, likes, comments, subscribes), Browser, application, or device in use, Usage of Platform products and features, Purchases or financial transactions through Platform products, Details about your online session, Details about your smart phone usage, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers
Advertisers, app developers, and publishers can send us information through Facebook Business Tools they use, including our social plug-ins (such as the Like button), Facebook Login, our APIs and SDKs, or the Facebook pixel. These partners provide information about your activities off Facebook—including information about your device, websites you visit, purchases you make, the ads you see, and how you use their services—whether or not you have a Facebook account or are logged into Facebook.	Actions taken by Platform engineers/editors/curators	Purchases or financial transactions through Platform products, Information from advertisers, app developers, and games from when you are not using Platform, Information from non-Platform apps and advertisers