4-2019

The Ethics of Cookies: Exploring the Collection of Big Data and Its Ramifications

Sarah Biely
DePauw University

The Ethics of Cookies: Exploring the Collection of Big Data and its Ramifications

Sarah Biely

DePauw University Honor Scholar Program
Class of 2019
Sponsor: Khadija Stewart
Committee: Ted Bitner, Karin Wimbley

Abstract

Technology is taking over the world. In every aspect of human life, technology has been able to provide some sort of help or solution. At the forefront of this revolution is the Internet and with it, the activity of day-to-day life that now takes place online. This rapid takeover pushes technological innovations to develop quickly, pushing boundaries and creating a new way of life.

Today in the United States, websites are allowed to track user data. When a user clicks on a website that intends on documenting the user's actions, the website installs a tracker, otherwise known as "cookie." Websites then use this collected data to create a profile for each and every user that visits their site. This process creates a vast database that has changed the methods of online marketing and increased business revenue. Although websites in the United States are now starting to alert users of cookie collecting, due to the implementation of the European Union's recent General Data Protection Regulation, the alerts are full of lengthy legal jargon, which means that users don't understand what is happening to their data when they browse a website.

The four pillars of ethics, when applied to online data collection, suggest that there are issues with autonomy, nonmaleficence, beneficence, and justice within this data collection field. Due to the fast-paced development of technology, there has not been enough time for regulation to catch up with the major companies that are paving the way for big data practices. Along with the pillars of ethics, privacy and security are at stake, not just for the individual consumer, but for society as a whole.

Keywords: cookies, data collection, ethics, privacy, security

Acknowledgements

I'd first and foremost like to thank my thesis committee. Without Professors Stewart, Bitner, and Wimbley, I would not have made it through this process. Their time, flexibility, and unwavering support have been greatly appreciated. Their guidance and teachings, not only during the time spent on my thesis, but throughout my four years here at DePauw, have been immense. Thank you so much for everything you have done.

I'd also like to thank Kevin Moore, Amy Welch, Tonya Welker, and the rest of the Honor Scholar community. This program was one of the main reasons I chose to attend this university and I am so thankful to have been gifted this incredible opportunity.

Finally, I'd like to thank my family and friends for supporting me throughout the thesis process as well as my four years at this university. I could not be more grateful for my time spent here with the people that have surrounded me.

Sarah Biely

April 2019

Table of Contents

The Ethics of Cookies: Exploring the Collection of Big Data and its Ramifications

"We are not meant to know everything... Did you ever think that perhaps our minds are delicately calibrated between the known and the unknown? That our souls need the mysteries of night and the clarity of day? Young people are creating ever-present daylight, and I think it will burn us all alive. There will be no time to reflect, to sleep, to cool." - Dave Eggers, *The Circle*

Introduction

Today's world is electronic, technological, and internet-driven. The way in which humans interact on a day-to-day basis has changed significantly in the past 20 years and will continue to grow immensely as technological innovations are created. Technology has unarguably improved the quality of life for billions of people. Information and education online are free-flowing and abundant with just one click. "Never, ever in the history of mankind have [humans] had access to so much information so quickly and so easily" (Nunan, Domenico). The way in which technology has been and will be developed is changing the way in which humans live and interact together every day.

The rise in online data is one of the major components of this technological takeover. Data collection on the internet through the collection of data points, user interactions, and online transactions has become a field of study on its own: big data. In this 'information era,' new technologies are created every day, as well as the scandals that follow them. While these technologies make human lives easier on the surface, they raise ethical questions that could pose threats to mankind's morals down the road. Ethical topics regarding privacy, security, and ownership have become the discussion of everyday internet users, technology moguls, and governments alike.

Online data collection has the potential to be discriminatory, violate essential privacies, and give unregulated power to organizations in ways they haven't done before. Privacy scandals like those of Facebook and Google have become national headlines and international causes for concern. However, since the development of technology moves faster than the legislation that monitors it, there are very few laws in the United States that address data privacy and security. While the European Union enacted legislation in 2018 to combat potential privacy violations and other ethical concerns, there are still many issues and technicalities that need to be addressed. The ways in which the United States reacts to, adapts from, and further regulates due to the European Union's legislation will set an example on the world stage of what type and extent of rights humans have on the internet.

With the rising prevalence of big data collection on the internet, government officials and society alike need to question online data collection and its ownership due to their potential ethical ramifications. These ramifications include violations of the four pillars of ethics, including autonomy, justice, nonmaleficence, and beneficence. Additionally, the ideas of power and privacy need to become major concerns when discussions regarding online data collection occur. This questioning and resulting actions need to occur in a timely manner and take into account all parties involved so that data collection and the internet remain positive technological advances for humankind.

What is Big Data?

Data collection, currently and throughout human history, has taken place in many forms. Ranging from oral interviews, surveys, research observations, bubble sheet surveys, or even just observing people or other events are all ways in which data has been collected. One of the first data collection methods that occurred on the internet was through email surveys. An online user would receive an email in text-form and reply with an "x" between two brackets, similar to that of a paper bubble sheet survey. The resulting information would be collected in a database or spreadsheet for further analysis (Topp, Pawloski). A few years later, software applications like *Lasso* and *Tango* became available for researchers to collect data directly from web pages and transfer them to an online database for observation and analysis. By 2001, this data collection methodology was relatively popular, however data collection by email surveys remained most common. This was because online, users "self-selected" themselves to take surveys, whereas email surveys hand-picked the users (Topp, Pawloski). Users that were consciously going out of their way to take a survey were less likely to do so, while users that were asked to do so were more likely to participate.

The Ethics of Cookies

Today, one of the most common forms of online data collection is a “cookie.” In technical terms, a web cookie, also known as a HTTP cookie or browser cookie, is “a small piece of data that a server sends to the user’s web browser. The browser may store it and send it back with the same request to the same server” (“HTTP Cookies”). Cookies are placed through either a web bug (such as a GIF) or the entrance into a specific website. They collect user data for later retrieval, either by the host website or a third party (Miyazaki). The data can be “explicit information provided by the Internet user (e.g., gender, age, zip code, account numbers), behavioral information regarding user movement from one Web page to the next (including time, duration, and sequence of web movement), or the tracking of how many times a particular banner ad has appeared during an online session” (Miyazki). Cookie data is stored by the host website in some sort of database, where the data is later analyzed for specific internal use.

Third parties oftentimes work with external companies on their websites in order to track a user’s online footprint across many platforms. The Federal Trade Commission (FTC) reported in 2000 that “78 percent of the busiest US sites allowed cookie placement by third parties” (Miyazaki). Cookies are placed on websites using a 1x1 pixel “clear GIF”, which means that users most likely do not know that the cookie is there because it is so small. The FTC further described cookies in 2000 as a “nonobvious means of information collection and their undisclosed use as a clear violation of the notice aspect of fair information practices” (Miyazaki).

Today, some web browsers alert users that cookies are being collected, but many times users ignore the alert or do not have a full understanding of what cookie collection actually means.

Cookies persist in many forms. Some cookies are temporary, meaning they only track data while a user is on the given site. Other cookies are designed to last on a user's computer for months or years, no matter if a user is using the original site or not. Although regulations now require companies to alert users of cookie occurrence, length of cookie existence isn't always made clear to the user. Certain web browsers also aren't always coded to delete a web cookie after it has reached its expiration date. Further, even if a user knows that there is a possibility for them to delete a cookie, the cookie is so deeply encoded that it can be tedious and time-consuming to figure out how to delete it (Miyazaki). The ways in which cookies are designed and manufactured online makes it very difficult for users to acknowledge their existence and understand their full impact.

Cookies are just some of many ways online browsers are able to track and store data- from search history to location, purchase history, interests, travel plans, bank preferences… the list is endless. However, big data isn't limited to what a user searches online- today's big data collection reaches to the Internet of Things, which includes data collected from products like smart watches, security systems, Amazon's Alexa, Apple's Siri, and more. Similar to online cookies, the Internet of Things collects data on humans' interactions with their environment in real time. Another rapidly growing field, the Internet of Things faces similar moral questions and ethical repercussions as cookie collection.

The main drive behind online data collection is to create an individual profile of each and every user. Cookies help companies keep customers' purchase history information on file in order to make predictions and suggestions to the customer about what they should purchase next, based on their previous purchases. If a company is successful, they will have a increased likelihood that their customers will find what they're looking for and will therefore purchase more. One specific example of this was used by Target. Target's marketing team mailed out coupons to their customers based on each customer's previous purchases. One of their methods of doing so was to analyze the purchases of pregnant women. "Target ran an algorithm that would score its female customers on the likelihood they were pregnant. If that probability tipped past a certain threshold, the retailer would automatically send out a series of coupons to the woman in question, full of things she might find useful…" (Fry 28). Target predicted that these personalized coupons would increase the likelihood that pregnant mothers would purchase goods from Target again.

In 2012, a father of a teenager contacted Target in outrage because his teenage daughter received coupons in the mail for baby strollers and cribs. She received these coupons due to Target's algorithm- she had been shopping online for items that fit into their profile of a pregnant woman. The father filed a complaint at his local Target store, so the manager of the store "apologized profusely and called the man's home a few days later to reiterate the company's regret about the whole affair" (Fry, 29). A few days later, however, the father followed up with Target customer service and informed them that his daughter indeed was pregnant (Hill).

This incident shows the potential power of algorithms. Even though their customer was upset, Target's executives believed that the company's actions were justified. A spokesperson for the company explained, "We found out as long as a pregnant woman thinks she hasn't been spied on, she'll use the coupons. She just assumes that everyone else on her block got the same mailer for diapers and cribs. As long as we don't spook her, it works" (Fry, 30). The possibility of "spooking" customers with targeted advertising is a concern of customers, however the pros many times outweigh the cons.

Companies consider the reactance to this personalized advertising, or the "motivational state in which consumers resist something they find coercive by behaving in the opposite way to that intended" (Tucker). A negative reactance can occur if a user is worried they are being spied upon or watched too closely, due to their information being collected and re-marketed back to them. Target, in this case, identified and addressed the reaction its customer had when the personalized advertising originally occurred, but its official position, as stated by one of its executives, overlooked the reaction completely.

One of the ways in which big data is used is through targeted advertising. This method of advertising takes data from online user profiles, collected through cookies, for example, and determines what else the user would be interested in. This interest is usually determined by algorithms that a company has either purchased or constructed themselves. The purpose of targeted advertising is to market the best products, ideologies, or whatever else a company might be advertising, to its online users. Online targeted advertising has unlimited possibilities and instant gratification to the host company due to quick analytics tools, which for a marketing company is incredibly powerful. “The Internet provides advertisers with the greatest laboratory ever for consumer research and lead generation. Feedback from each promotion arrives within seconds- a lot faster than the mail” (O’Neill, 75). Having the ability to instantly see what types of advertising positively or negatively affect the popularity of a specific product is changing the way marketing companies operate.

Regulations Throughout History

Regulation of data collection has been a hot topic, especially in the past decade. Even before the internet came into actual existence, data collection has been regulated by the Belmont Report. Established in 1974, it outlines three major principles that were created in response to the unethical and unjust ways the medicinal and psychological field conducted research at the time. The principles emphasized “respect for research participants, beneficence, and justice in participant selection” (Vitak, et al). Resulting legislation created institutional review boards (IRBs) on university campuses, which oversee ethical data collection. One key idea the report brought up was “informed consent,” which is “permission granted in the knowledge of the possible consequences, typically that which is given by a patient to a doctor for treatment with full knowledge of the possible risks and benefits.” Informed consent, whether it be in the medicinal or technological field, outlines the possible consequences to whatever process is occurring. It is particularly difficult to achieve, however, when there are thousands of users participating on one platform, like in the case of online data collection. Along with this idea of informed consent is the idea of transparency. In terms of data collection, websites must determine how transparent they will be with their users in their data collection, specifically in collection, analysis, and use (Vitak, et al).

Another challenge regarding the protection of data is how regulations change from country to country throughout time. Laws change over time, however data collected today is being stored for long periods of time, if there’s even a time limit at all. The ways in which user data is allowed to be shared or processed might change in 50 years, but users won’t know that potential ability when they sign over the use of their data today (Williams, et al). Additionally, in reference to global data practices, the data laws of the United States are constantly playing catch-up with the data laws of the European Union, which provides some challenges to global companies trying to navigate legalities.

In May 2016, the European Union introduced the EU General Data Protection Regulation. This new regulation will replace the EU's current Data Protection Directive, created in 1995, and came into effect as applicable law in May 2018. The reform "aims at modernizing and harmonizing data protection across the EU." It was part of the ambitious Digital Single Market Strategy that the EU launched at the same time in hopes of changing how data collection is practiced and regulated (Burri, Schär). In early 2018, many online users in the United States started to receive emails and notifications about companies' updated privacy policies. This was because the General Data Protection Regulation had gone into effect in the EU. Even though the regulation was legally binding of EU citizens and companies, most online companies have global networks and want to be able to access the European market, therefore making it necessary for their privacy policies to comply with the regulation and thus need to apply their changes worldwide (Fung).

The GDPR is enforced by the European Union's Information Commissioner's Office and requires companies to "be explicit in their efforts to seek consent from consumers before collecting their personal information" (Fung). For users accessing websites, this comes in the form of explicitly clicking a button that says "Accept Cookies." Additionally, the regulation gives users the option to easily access their collected data from companies and further, have it be deleted from a company's database. Article 5 of the GDPR clarifies that data should be

> "processed lawfully, fairly, and in a transparent manner in relation to the data subject (principle of lawfulness, fairness, and transparency); collected for specified, explicit, and legitimate purposes (principle of purpose limitation);

> processing must also be adequate, relevant, and limited to what is necessary (principle of data minimization); as well as accurate and, where necessary, kept up to date (principle of accuracy); data is to be kept in a form that permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed (principle of storage limitation); data processing must be secure (principle of integrity and confidentiality); and the data controller is to be held responsible (principle of accountability)" (Burri, Schär).

Data collection tactics and procedures have become very specific, in order to protect the rights of EU citizens that the EU deems necessary.

According to the GDPR, if a company has a data breach, they are legally required by law to alert the public within 72 hours (Fung). As a result, many companies have opted to appoint a data protection officer, or someone who will regulate how their company collects online data. If a company does not comply with this new regulation, there is the possibility of a large fine, "up to 4 percent of a company's annual global revenue, or €20 million (about $23 million), whichever is higher" (Fung). Whether this fine is substantial enough to deter companies from violating this regulation remains to be seen.

Some companies, however, have neglected to update their privacy policies. In order to avoid fines, they have chosen to simply remove their availability in the EU. For example, if a user in the EU tries to access the Los Angeles Times (as of May 2018), they will be greeted with the following message: "Unfortunately, our website is currently unavailable in most European countries. We are engaged on the issue and committed to looking at options that support our full range of digital offerings to the EU market. We continue to identify technical compliance solutions that will provide all readers with our award-winning journalism" (Fung). Users are unable to further navigate through the website. While this sort of tactic is useful for companies who don't want to update their policies or have not done so quickly enough, the economic disadvantage to losing the European market will take a toll.

While there are high economic stakes for companies affected by the GDPR’s regulation, there are other factors to consider when creating this sort of legislation. This new reform stems from a need to stay up-to-date with technology, which is constantly changing and developing. The EU has been working on this regulation and directive for the past 5 years, however many things in technology can change over 5 years. While it is beneficial to society to create regulations, these regulations could already be outdated because of how quickly technology develops (Burri, Schär). The fast-changing nature of privacy in today’s era already puts the GDPR behind the times.

In recent years, new software has been developed in order to block users' cookies from being collected. So, in order to reach consumers' data, because for many companies it can mean thousands, if not millions, of dollars new software has been developed. Additionally, mobile phones currently do not collect cookies. The new trend in data collection software is known as "fingerprinting." This method allows a website to examine characteristics of a user's computer, such as plugins and software installed, size of the monitor, time zones, fonts, and more. These characteristics are put together to create a sort of "fingerprint," a unique picture of the computer, similar to what a fingerprint is to a human. A study found that 94 percent of browsers that use Java or Flash had unique "fingerprints" (Tanner).

Fingerprinting could potentially be more useful than cookies because the information it collects stays with a user's computer permanently, even if the user erases their cookies. Fingerprinting is also accessible on mobile devices, making it extremely more versatile. While fingerprinting is still relatively new, the companies that both develop and purchase it are reluctant to share this information, due to the potential to "creep out" their customers. Evan Reiser, CEO of a software engineering firm, explains, "At the end of the day, there isn't really a legal case against it, there isn't really a privacy case against it. It's really a PR thing" (Tanner). However, there might not be a legal case against it right now because it is so new.

The ability for a user to delete cookies off of a web browser is possible. There are settings in place on most web browsers that will clear off any cookies that have been installed. By deleting cookies, the process of recording what actions a user takes on a given website will be interrupted, “separating [the user’s] prior online activity from [their] future activity in the tracker’s data. The only noticeable effect this will have for most users is a change in the types of ads that will appear on the websites,” due to the backlog of information being cleared (Parson).

However, just because a user deletes their cookies doesn’t mean that the data sent to the company has been deleted, it only means that the immediate connection between the company’s tracking abilities and the user’s device has been stalled. While this is a good precaution to take for those who don’t want companies tracking their information, new technologies like fingerprinting are circumventing this opt-out possibility.

While the GDPR’s relevance remains questionable, there are controversies within the GDPR that add to its significance and impact. One of the most controversial changes of the GDPR is Article 17, which is "the right to be forgotten." The article expands on an existing right of the Data Protection Directive's "right of erasure," that now allows a data subject to "have their personal data erased and no longer processed, where the data is no longer necessary in relation to the purpose for which it was collected; where a data subject has withdrawn her consent or objects to the processing of personal data concerning her; or where the processing of her personal data is otherwise contrary to the Regulation" (Burri, Schär). This article is controversial due to it being the first time this idea of erasure has been considered. Additionally, the functionality of this article is difficult to comply with, as companies most likely don’t have systems in place to erase user data when requested.

One of the biggest issues the EU is concerned about is the ability for data to move across borders, as well as privacy laws within the international community. According to journalist Mira Burri, the right to privacy is "a key concept in EU law and has been given significant weight that reflects deep cultural values and understandings." The Council of Europe's European Convention on Human Rights protects "the right to private and family life" in Article 8 and separates the right "of respect for private and family life" in Article 7 to the "right to protection of personal data" in Article 8. This shows the EU's commitment to privacy, something that isn't necessarily protected in the United States (Burri, Schär).

The EU General Data Protection Regulation and new Directive were influenced by multiple online data scandals that have taken place within the EU over the past 10 years. In 2014, the Court of Justice of the European Union (CJEU) ruled that "an individual has the right to object to a search engine's linking to personal information and that evaluation of such an objection calls for a balancing of rights and interests, in the context of which account must be taken of the significance of the data subject's rights arising from Articles 7 and 8 of CFREU." This decision was part of a lawsuit filed by a man in Spain who did not want his name, when searched in Google, to come up with any results. This case asserted the rights of individuals against big data (within the EU), but also demonstrated the power and flow of data. The idea of "the right to be forgotten" is a theme that will occur multiple more times before the EU made its most recent data protection regulation (Burri, Schär).

Another event that influenced the data protection regulation was 2014's decision that the Data Retention Directive was invalid, also decided by the CJEU. It arrived after multiple terrorist attempts had occurred in Madrid and London. It "sought to harmonize Member States' laws and required the retention of data from fixed, mobile, or Internet telephony, as well as email communications, for at least six months, and possibly up to two years." The Court found that this directive directly contradicted Articles 7 and 8, even though it had the possibility to help prevent and solve public crime. It ruled that it "did not distinguish between different means of communication, different sorts of data, or different types of users (Burri, Schär). This idea of refusing to hold a person's personal data, whether it be from a cell phone, computer, or other for legal purposes without the user's consent is not unique to the EU- questions and debates about the government's unauthorized access to personal data have occurred in the United States for many years.

Furthermore, there were no defined substantive and procedural conditions with regard to access to the data; nor were objective criteria for the determination of the retention period laid down. This case concerns the ethical implications of big data, that it may "allow very precise conclusions to be drawn concerning the private lives' of individuals- and acknowledged the fact that data retention may have a chilling effect on the right of freedom of expression" ((Burri, Schär). These cases highlight the ever-changing rules and regulations regarding data in the modern age. The stakeholders in these scenarios are everyday citizens, big corporations, and the international governing community (Burri, Schär).

Part of the European Union's 2018 General Data Protection Regulation explicitly states how online data collection should be handled. The Data Directive Act ensures that data is

> "used fairly, lawfully, and transparently, used for specified, explicit purposes, used in a way that is adequate, relevant and limited to only what is necessary, is accurate, and, where necessary, kept up to date, kept no longer than is necessary, and handled in a way that ensures appropriate security, including protection against unlawful or unauthorized processing, access, loss, destruction, or damage" (Government Digital Service).

While the outlines are clear in wording, the implementation of this directive is vague due to its lack of technicalities. How the EU will proceed in legal cases regarding these laws will demonstrate the actual effect and extent of the regulation.

The act also ensures that EU citizens have the right to know what the government and organizations know about them. This includes a citizen's right to "be informed about how [a user's] data is being used, access personal data, have incorrect data updated, have data erased, stop or restrict the processing of [a user's] data, data portability (allowing [a user] to get and reuse their data for different services, and object to how their data is processed in certain circumstances" (Government Digital Service). Additionally, these rights extend to when an organization uses "automated decision-making processed (without human involvement)" (Government Digital Service). This concept opens up communication between the public and the companies in question. No longer are companies able to hide their processes when it comes to how they analyze and measure their user data. Again, in theory this is the case but time will tell how this regulation holds up in practice.

These rights are beneficial to users in theory, but depending on the understanding and extent to which a user is informed greatly affects how these laws will actually affect consumers. Since these laws are broad, the interpretation and implementation of data practices will vary from company to company. While the United States is behind on data protection legislation, any company that works internationally or has clients in the EU must abide by these rules. Whether or not these companies only apply these rules to their EU clients or all of their international clients remains to be seen as well. If a company in the US or EU does not follow these guidelines, the maximum fine is 20 million euros, however it is unlikely that European legislators will begin imposing fines of this measure anytime soon due to the complexity of understanding and interpreting these guidelines (Roberts).

Recent Events

Some of the biggest data collections in the world right now are not governments or big organizations- they are social media websites such as Facebook, Google, Twitter, etc (Nunan, Domenico). There are both pros and cons with this being the case. As lead innovators and researchers, these companies are able to push boundaries when it comes to how data is collected and what happens with the data. These companies have time and resources that the government doesn’t (or can’t allocate to technology), so they are able to move with speed and force when it comes to data collection processing. However, since technologies are created and implemented so quickly, these huge companies, as well as the government, don’t have enough time to regulate them, much less consider their future ethical implications. Researchers Daneil Nunan and Marialaura Domenico ask, “How can consumers trust an organization with information when the organization does not yet know how the information might be used in the future?” (Nunan, Domenico). More, how can the public know the possible negative effects of their data when the company doesn’t even know what they’re going to do with that data?

Facebook, in particular, challenges the separation between the public and private sphere (Debatin, Bernhard, et al). It is an online database in which its users share their personal data with one another. This premise already complicates the idea of privacy, in that a user’s profile is only available to certain users, and public, in that this information is being shared in the first place. Due to its rapid growth and popularity, Facebook has always been ahead of the curve in the realm of technology and thus received the brunt of ethical security and privacy deliberations. Even two years after its inauguration, Facebook users’ passwords were still not encrypted, which meant that any third party could easily hack into a Facebook account. Additionally, simple algorithms available to the public had the ability to download all of the information off of a user’s profile. Additionally, users had to specifically choose to opt out of Facebook’s data collection about them from other sources, meaning that if another company held a portion of a user's data, Facebook could access it and add that information to the same user's Facebook profile. In 2007, that opt-out ability was removed ((Debatin, Bernhard, et al).

Other changes to Facebook's software have inspired distrust and backlash over the years. Facebook's advertising platform has come under fire since its implementation in 2007. Specifically, its "Beacon" online ad system that allows Facebook to track its users' online behavior, such as what they search for and what they shop for online. Originally, Facebook shared this information to users' friends but after many petitions appeared as Facebook Groups that were in opposition to this, Facebook added an option for users to opt-out of the broadcasting. However, Facebook still maintains its tracking on members' activities "on third-party sites that participate in Beacon even if the users are logged off from Facebook and have declined having their activities broadcast to their Facebook friends" (Debatin, Bernhard, et al). Another incident that resulted in backlash occurred in 2006 when a police officer searched a bystander's friend list on Facebook in order to find a suspect who publicly urinated outside of a fraternity house. The police officer found the suspect on the friend list and charged him with two tickets. The Patriot Act additionally allows "state agencies to bypass privacy settings on Facebook in order to look up potential employers" (Debatin, Bernhard, et al). All three of these personal data infringements resulted in the public's growing concern over privacy.

In 2016, Donald Trump's campaign team hired Cambridge Analytica, a political data firm, in order to collect information about the voting American public and their preferences.

During this process, Cambridge Analytica gained access to more than 50 million users' Facebook profiles. The data collected included "details on the users' identities, friend networks, and 'likes.' The idea was to map personality traits based on what people had liked on Facebook, and then use that information to target audiences with digital ads" (Granville). Essentially, if a user was leaning towards one political party, Cambridge Analytica's algorithms could tell based on what posts and pages the user 'liked,' as well as what their Facebook friends did. Then, the company could pick specific ads for that user, either pushing the user in the same direction they were already leaning or posting ads that could potentially convince the user to consider the opposing side.

Mark Zuckerberg, the CEO of Facebook, explained how companies target users through their data during his Congressional hearing in 2018. "The targeting options that are available for advertisers are generally things that are based on what people share… Once an advertiser chooses how they want to target something, Facebook also does its own work to rank and determine which ads are going to be interesting to which people" (Domonoske). A similar process was used with Cambridge Analytica's political advertisements. The way in which politicians spend, create, and market their advertisements and campaigns has always been strictly regulated and supervised. Cambridge Analytica's method of targeting personalized ads on such a large scope is unprecedented.

The data collection at Cambridge Analytica began in 2014. Out of the 50 million users, 270,000 responded to a voluntary survey sent out by the firm, which pulled data from their personal Facebook profiles. This sort of data collection has since been banned by Facebook, although Facebook argued that no confidential information, such as passwords and login information, had been released. In early 2018, Cambridge Analytica denied their collection and use of Facebook data, only to retract their statement a week later, admitting that it had used the data but deleted it two years prior when it realized it had violated Facebook’s privacy rules. Zuckerberg has apologized multiple times regarding this incident, although he puts the blame on the third party that used the data.

Zuckerberg appeared in front of Congress on April 11, 2018, to explain that Facebook users have "complete control over" everything they share on Facebook. However, revealed documents and interviews with past Facebook employees have stated otherwise, that Facebook shares user data without their consent. References were made to the Federal Trade Commission's 2011 consent agreement that "barred the social network from sharing user data without explicit permission" (Dance). One question asked to Zuckerberg was whether or not an “average layperson could look at the terms and conditions and make the evaluation: Is this strong enough protection for me to enter into this arrangement?” (Domonoske)

Terms and conditions are notoriously ignored by users. Their complexity and length, along with their complicated legal style of writing, deters users from reading them, thus creating a space in which users don’t know their legal standing. By agreeing to these terms and conditions, users are giving away pieces of their privacy to companies. Understanding the terms and conditions can help the user determine whether or not they should trust that website and what amount of information to provide it with. Because this collection of data is imperative to online marketers, it is up to them to create this “trusted space” for users to enter and hopefully provide their information to. The content and format of these conditions are the two ways in which marketers are able to create the space. As of the early 2000s, there are no regulations as to what sort of language can and/or should be used, as well as the format of the notices. This has created criticism surrounding companies’ privacy notices (Milne, Culnan).

However, a further step Facebook has taken in this realm is that they also collect information on people who aren’t Facebook users and therefore have not agreed to any data collection. One congressman addressed Zuckerberg during the proceedings, stating “[Facebook] said everyone controls their data, but [Facebook] is collecting data on people who are not even Facebook users, that have never signed a consent, a privacy agreement, and [Facebook is] collecting their data” (Domonoske). Thus, Facebook collects data on online users who have not signed up for Facebook, nor clicked ‘accept’ on the Facebook’s terms and conditions.

In 2017, the New York Times obtained an in-depth report detailing the extensive way in which Facebook shares users' data (Dance). With a whopping 2.2 billion users, Facebook's practices affect almost 30 percent of the world. One of the biggest realizations this report found was that Facebook gave Microsoft's Bing search engine the ability to find any user's name and gave media platforms, Spotify and Netflix, the ability to read users' private messages, all without user consent (Dance).

Facebook’s director of privacy and public policy, Steve Satterfield, said in an interview that “None of the partnerships violated users’ privacy or the FTC agreement” and that “Contracts required the companies to abide by Facebook policies”. Satterfield also mentioned in this interview that “Protecting people’s information requires stronger teams, better technology, and clearer policies, and that’s where [Facebook has] been focused on for most of 2018” (Dance). In total, the Facebook user data, ranging from 2010 to 2017, benefitted over 150 companies. The companies involved were not required to obtain users’ consent before using their data- Facebook argued that these companies were “extensions” of them and therefore were required to follow all of Facebook’s policies.

David Vladeck, the former head of the FTC’s Consumer Protection Bureau made a comment in response to Facebook’s defense, saying that the situation “is just giving third parties permission to harvest data without [the user] being informed or giving consent to it” (Dance). In response to this report, the FTC, Justice Department, and Securities and Exchange Commission opened new investigations in 2018 regarding Facebook’s sharing of data with its partners. Author and data scientist Cathy O’Neil explains that recent events have shown “Facebook’s enormous power to affect what [users] learn, how [users] feel, and whether [users] vote. Its platform is massive, powerful, and opaque. The algorithms are hidden from [users], and [users] see only the results of the experiments researchers choose to publish" (O'Neil, 146). The results from these investigations could potentially be a catalyst for legislation regarding the sharing of user data, since the United States has not seen an event like this before.

Power and Privacy

Two of the biggest considerations that are debated within big data collection are power and privacy. Both are topics discussed extensively outside of the technological realm, from classroom debates to national questions of legislation and further, the given rights of the citizens of the United States. Due to the fast growing nature of big data, power and privacy are pushed towards limits that hadn't been considered before by the American public. Before discussing legislation that would guide or limit either of these subjects within technology, they need to be understood and considered from multiple different perspectives.

Privacy

At the heart of big data collection concerns lies the importance of privacy. Whether or not privacy is a right of the people and at what cost should it be given and taken are questions that need to be answered before big data collection gets exponentially bigger. Privacy is defined in the technological world as "the dilemma about what and how much information a company collects and stores about an individual, who gets access to it, and whether or not individuals get any control or say about what happens to this data" (Rutherfoord, Rutherfoord). Throughout history, philosophers and the general public in the western world have considered privacy a right. Without privacy, it has been argued that individuals will have a lack of freedom and will behave differently when they believe they are being watched, in addition to feeling angry, suspicious, and having lost their spontaneity.

Concerns about privacy are not a new issue. As far back as 1890, scholars worried about putting photographs in newspapers and how that would affect the public's privacy (Nunan, Domenico). Big data has increased in size and scope dramatically over the years, causing further concerns about privacy. In 2011, IBM reported that ninety percent of all the data in the world had been produced between 2009 through 2011(Nunan, Domenico). One way to look at privacy in data collection is through the social contract perspective. In this perspective, "provision of consumer information is expected to yield a certain responsibility (i.e., in the form of an implied social contract) of the receiving organization to collect and care for such information in a responsible manner" (Miyazaki). These "implied social contracts" are essentially social norms that will break consumers' trust in the company if breached. Privacy, then, is part of a trust process that a user has with a company.

Professor James H. Moor of Dartmouth College considers two different perspectives on privacy: one, that it is something vital to human life that needs to be protected, and two, that it is something cultural, “a matter of individual preference, culturally relative, and difficult to justify in general” (Moor). What is considered “privacy” and “private information” in one culture can be very different in another culture. In a world where information is passed globally in a matter of seconds, privacy values and standards change in seconds too. Moor further explains the difference between two important ideas: instrumental value and intrinsic value. Instrumental values are those values “which are good because they lead to something else which is good. Intrinsic values are values which are good in themselves. Instrumental values are good as means; intrinsic values are good as ends” (Moor). Privacy, Moor argues, is an instrumental value that most everyone agrees with. Although it seems that everyone in modern society values privacy, companies that use big data practices are pushing the envelope on what privacy really means and how far it can go.

Dave Eggers paints of a picture of a possible world in which privacy becomes obsolete in, *The Circle*. A young woman enters the workforce of a booming social media company (called the Circle) in the distant future, something similar to Google, Amazon, or Facebook. The social media company is obsessed with sharing information between humans, whether its constant status updates or instant messaging. The young woman, Mae, is at first apprehensive about sharing so much personal information with her colleagues and social circle. However, the CEO of the company explains a new technology during a speech that changes her mind.

The new technology includes a massive amount of mini cameras, intended to be posted in different locations around the world by those who buy them. Due to their abundance, the cameras will be used to help stop crime, locate any person, or show off the world to those who don't have access to it. The cameras can also be worn on a person's body, to show transparency. A fictional congressperson in the novel decides to wear one of the cameras at all times in order to show their commitment to transparency and honesty. Mae begins to buy into this society of sharing every piece of information available. Her friend from home, however, wants to remain a private person and falls to his death after being chased down by cameras. Despite what occurred, Mae believes in the values of the Circle and continues to support the company as it moves to take over the world's banking transactions and voting capabilities. Mae argued that if all of the world's information is in one location, life would be easier, safer, and all around better. Whether this world the Circle has created is a dystopia or utopia, Eggers leaves the reader to decide for themselves.

One revelation Mae has during the book is that: "SECRETS ARE LIES SHARING IS CARING PRIVACY IS THEFT" (Eggers, 201). Privacy as theft, or privacy as a negative part of life, is one of the essential roots of the discussion around data mining. Privacy, then is at the root of this dilemma. Philosopher Charles Fried writes, "Privacy is not simply an absence of information about us in the minds of others, rather it is the control we have over information about ourselves" (Moor). Further, the issue here is not that marketing companies don't have a full picture of their target audience, but that users want to control the picture of themselves. Privacy, even if there are benefits to giving it up, is important to online users, as well as a right that should not be violated.

Power

Power, even within the context of big data collection, can reference many things. In this realm, power can be gained through monetary means, government regulations (or lack of), data ownership, and influence. “The power of big data,” then, can mean many things. The data itself can have power in that it provides information to an observer, but the impact of the data has power in different ways.

Big data provides monetary power to the companies that utilize it. Targeted marketing, like stated earlier, creates a surge in revenue. “[Companies] still make plenty of money in more traditional ways. But the richest companies in the world now generate wealth by putting as many trackers, devices and screens inside [users’] homes and as close to [their] bodies as possible. Accumulated data creates competitive advantage, and money can be made by consolidating everything that is known about an individual” (Wu). Big data monetizes the actions of everyday people, including their interests, preferences, profile, and more. Money, in the capitalistic society of the United States, feeds power.

The Ethics of Cookies

Government regulations, or in the case of the United States, a lack of government regulation, also provides power, specifically to big companies such as Apple, Google, Amazon, etc. These major technology companies don't have strict guidelines or legislation to follow (like those of the EU), so there are no rules when it comes to what is off-limits for data collection. These companies essentially have free will to create the rules as they go due to the fast-paced nature of technology. New uses of data and how it is processed are created faster than the people whose data is being collected, much less legislators, can keep up with. Thus, the potential consequences of data collection occur before the public and government can even begin to consider the ramifications. With the incentive to grow their businesses further, ethics are not at the forefront of concerns they have. Tech companies, at this time in the US, have unchecked power when it comes to regulations.

The ownership of data is powerful, however the idea of data ownership, including what it means to "own" data, when this ownership transaction occurs, and how long this ownership should continue for, are all still being debated in the ethics and technology communities. These questions are not ones that can or will be answered anytime soon- they reach further than just online data collection because they are rooted in deeply philosophical questions about society and the rights of citizens within the United States (and the world) as a whole.

To "own" data is a hotly contested topic that is extremely relevant to today's form of data collection. When a user gives information to a website, which then stores the data for its own benefit, the user is handing over a piece of them that they cannot get back. According to the EU's new GDPR guidelines, users still control their data and have the right to "have it returned to them," or deleted from the collecting company's system. In the United States, this is not the case. Once the data is in the company's system, there are no regulations about the control of data flow. Users do not have the right to request their data be deleted or "returned" to them once they have given it away. Thus, the company at hand has "ownership" over this data. Additionally, there are no regulations as to how long a company is allowed to store data for. While storing basic background information indefinitely doesn't seem that problematic, there are long-term potential issues to consider.

When babies are born in hospitals, doctors take a blood sample to check the baby's state of health. In the United States, the laws regarding what happens to the blood sample after the baby leaves the hospital differ from state to state. While some states require the blood sample to be thrown away, other states choose to keep the blood samples indefinitely in their laboratories, to test for "quality control, assurance monitoring, and public health" ("Your Baby's Screening What Happens to the Blood Sample). Preserving this sort of highly sensitive data, a blood sample, raises many similar ethical questions about "data" ownership and the right to privacy. A baby does not give consent at its birth to have its blood be preserved for life. Further, its parents might not even understand or know that the blood sample is being preserved, because it is explained as a check on the baby's health. The technology we have today can discover thousands of things about a person from a blood sample, but the technology humankind might create in the coming years will be able to discover so much more. This lack of knowing what could be determined from a blood sample or data collected online, is something that a user has no knowledge of when they consent to giving their data. The ownership of data is powerful when it comes to unlimited access because of the potential unknown possibilities there are to process it in the future.

Data ownership additionally has power through influence. Targeted advertising has the power to influence markets and has changed the way citizens in the United States (and all over the world), shop. The Cambridge Analytica scandal demonstrates how users are more likely to click on the posts they already believe in, thus confirming their truths further and reinforcing the power of Facebook in manipulating ideas. An experiment in 2013 ran by Facebook employees manipulated the news feeds of over 690,000 users, without their consent or knowledge, in attempt to change the mood of each user. The experimenters “suppressed any friends’ posts that contained positive words, and then did the same with those containing negative words, and watched to see how the unsuspecting subjects would react in each case” (Fry, 42). Users who saw more positive posts on their news feeds were more likely to post positive posts themselves, while those who saw more negative posts were more likely to post content with negative words. Thus, researchers learned that data can influence the way people are feeling. Data has the ability to influence many more aspects of human life, without the affected users consciously recognizing this change. This ability gives power of unknown boundaries to the companies that control the data.

Ethics Discussion

Data collection happens on many levels, and not just on the internet. When meeting a new person, people make judgements. In order to survive in the world, people must assess their surroundings, taking in data and analyzing it. This data collection, for humans in the real world, is observation and conversation. The way in which humans analyze it, similarly to data collection online, can be problematic, such as the creation of stereotypes and biases. What makes online data collection so dangerous is its automation and lack of third-party analysis. Due to the nature of algorithms, data can be processed incredibly quickly and analyzed using artificial intelligence in order to make predictions and assumptions. Researchers will analyze the data they receive back, but rarely will it be judged based on discriminatory possibilities and bigger-world consequences, because of how quickly it is processed and the degree of positive feedback received. In the current capitalistic society, ethics are rarely at the forefront of big business.

When considering ethics, scholars consider four main pillars: autonomy, beneficence, nonmaleficence, and justice. Dating back to the 4th century BCE, these principles were discussed by the likes of Hippocrates and the ancient Greeks, who applied them to philosophical ideas of what society should be like at the time. While in some ways society has changed, these pillars have not. The Belmont Report included these pillars, citing them as guidelines to the process of responsible research. Through medicine and scientific research, “it is generally held that these principles can be applied, even in unique circumstances, to provide guidance in discovering [researchers’] moral duties within that situation” (McCormick). These four pillars lead the discussion on why today’s big data practices can be so problematic.

Autonomy

Autonomy refers to “a person’s right to make choices, hold views, and to take actions based on personal values and beliefs” (Bishop). Autonomy is at the forefront of the discussion around big data practices. Although today users must agree to the cookies being used on a website in order to proceed using the website, users don’t always know that their data is being tracked. Like stated earlier, many users don’t take the time and effort to read the terms and conditions on a website’s cookie agreement. Every website that uses cookies has one of these agreements, which is unique to how that website uses its user data. Like any contract, the terms and conditions are lengthy and filled with legal jargon, much of which is difficult for a typical user to understand. The time and energy it takes for a user to read and fully comprehend what is happening to their data whilst they use the website is unlikely to be spent on every website a user visits in a day. Thus, an ordinary user lacks knowledge about what happens to their data after they leave the website. This lack of knowledge results in a lack of autonomy of data- if a user doesn't know what is happening to their data, they won't know how to control it.

Another contributing issue to the right of autonomy are algorithms. Algorithms, specifically the machine learning sector, create their own assumptions and conclusions based on the data they process. Once they begin their processes, algorithms can be difficult to understand by even the most informed data scientists. Due to the way they're developed, algorithms make calculations based off the data they receive. Based off of their calculations, patterns begin to form and then predictions can be made. Data scientists sometimes receive the specific predictions made by the algorithms, but many times they are implemented without much oversight. For example, an algorithm that determines that people in the Northwestern USA drink more coffee than other parts of the country might start targeting that Northwestern demographic with coffee ads. Data scientists might only see the results of this through improved sales or might have a general overview of the results gathered by the algorithm, but the intricacies of the decision-making process are hidden.

An algorithm's ability to hide its processes hinders the autonomy of many characters in this scenario. Data scientists lose autonomy in their research capabilities because an algorithm makes decisions that are unreachable to the scientists. On a small scale, this breach in autonomy doesn't seem substantial, however it can have vast consequences the reach much further than the desired audience.

Another issue to consider within this perspective of online data collection today is the ownership of data. In previous years, before data collection occurred online, whoever distributed surveys and collected back the data was the "owner" of the data. With online data collection, however, the idea of the "owner" of the data, as well as the legitimacy of the data is called into question. It is possible that the user could actually be computer automation, or a fake user, thus corrupting the data. This is really only an issue when there is an online form to be completed. The idea of data collection without filling out a form, for example, with the use of online cookies, doesn't relate to this problem. However, the ownership is relatable. The ability to own one’s data is a form of autonomy that should not be violated.

Justice

Justice, the ability to “treat others equitably [and to] distribute benefits/burdens fairly,” is another pillar of ethics that big data encroaches upon (Bishop). The way in which algorithms work make them inherently unjust, although they attempt to do the opposite. Algorithms are fed raw data about a company’s consumers, which is presumed to be as unbiased as possible, then set to work building profiles of the customers. However, algorithms many times aren’t fed the systemic problems that their customers face in the real world, problems that can’t be seen in raw data to an untrained eye. Data scientist Cathy O’Neill explains,

> "[Algorithms are] feeding on each other. Poor people are more likely to have bad credit and live in high-crime neighborhoods, surrounded by poor people. Once the dark universe of WMDs digests that data, it showers them with predatory ads for subprime loans or for-profit schools. It sends more police to arrest them, and when they're convicted it sentences them to longer terms. This data feeds into other [algorithms], which score the same people as high risks or easy targets and proceeds to block them from jobs, while jacking up their rates for mortgages, car loans, and every kind of insurance imaginable" (O'Neil, 199).

Targeted advertising, similar to college admissions, is unfair in its process. At its core, different customers are marketed differently in attempt to best guess their spending habits and interests. Data collected from online cookies reveals patterns in human's online interactions that rank and categorize them. Targeted advertising, the result of this collected data, "establishes a powerful basis for legitimate ad campaigns, but it also fuels their predatory cousins: ads that pinpoint people in great need and sell them false or overpriced promises" (70). One venture capitalist outlined his proposed future of targeted advertising at an advertising company: "The coming avalanche of personalized advertising would be so useful and timely that customers would welcome it. They would beg for more. As he saw it, most people objected to advertisements because they were irrelevant to them. In the future, they wouldn't be" (69).

The Ethics of Cookies

Similar to what the executive from Target said about targeted advertising, its goal is meant to create the most appealing advertisements for each and every internet user.

Google Ads is a huge proponent of this. Google collects cookies across many platforms that a user clicks on, specifically products when a user is online shopping. Then, once the user leaves said product, similar products will appear on Google Ad platforms on another site. For example, if a user is browsing lawn mowers on Amazon for 30 minutes and clicks on 10 different lawn mowers, Google tracks that that user is interested in lawn mowers. Even if that user leaves Amazon without purchasing anything, Google knows that there is a high likelihood that that user is still interested.

Another website, a random blog for example, wants to make money off of advertisements.The owner of the blog installs Google Ads, so that if a user on their website clicks on the Google Ad, the blog owner will receive a small commission. Over time, if many users click on their personalized advertisements, the blog owner will make a significant amount of money, which incentivizes the blog owner to place the advertisements in the first place. If the user interested in lawn mowers clicks on the blog with Google Ads, there is a high chance that lawn mowers, ones that they had been looking at before or similar products, will pop up. Due to their previous interest and continued exposure to the product, Google is betting that the user will return to Amazon and buy the product. This methodology has made online marketing platforms, including Google, millions of dollars. That methodology is created through algorithms, ones that suggest which products users are most likely to click on, products that are specifically targeted to them.

Algorithms are used to come up with concrete answers based on concrete numbers. Due to their nature, there is a widespread belief that algorithms are inherently objective. Although they are written by humans, who are not perfectly objective, the job of an algorithm, especially one that learns on its own, is to create subject matter that is based on real, hard evidence that is unbiasedly true. However, algorithms, whether the fault is at the hands of the human programmer or not, aren’t always perfectly unbiased. Many times algorithms reinforce human bias and discriminatory ideas.

An algorithm is oftentimes referred to as a “black box,” in that programmers don’t know how they work, just that they do. Typically algorithms reflect human behavior and ideas, while trying to remain unbiased. However, since they do reflect humans, they are many times unintentionally biased. For example, in a 2015 study done by the University of Washington, a Google image search for “CEO” found only 11 percent of the photos to include a woman, whereas 27 percent of CEOS during that year were female (Miller). David Oppenheimer, a discrimination law professor at the University of California, Berkeley explains that “Even if [algorithms] are not designed with the intent of discriminating against those groups, if they reproduce social preferences even in a completely rational way, they also reproduce those forms of discrimination" (Miller). Discrimination is illegal, but it can sometimes be hard to tell if an algorithm is doing just that.

One specific incident of clear discrimination is from another 2015 Google study done by Carnegie Mellon University. Researchers found that Google's online advertisement system "showed an ad for higher-income jobs to men much more often that it showed the ad to women" (Miller). While targeted advertising is legal, gender discrimination is not. Researchers weren't able to reach a conclusion on why the results turned out the way they did. Either the advertiser made an effort to target higher-income jobs more often towards men or the algorithm determined that men more often clicked on the ads (Miller). The gender gap between men and women in the workplace is already substantial and this sort of advertising perpetuates this injustice. While it is nearly impossible to figure out why an algorithm makes every decision it does, it is possible for programmers to run simulations against their algorithms in order to check for bias.

Discrimination on the basis of race, ethnicity, religion, national origin, immigration status, sex, gender identity, sexual orientation, disability, and age is illegal. Discrimination, however, happens every day in the United States in many different ways. Algorithms are now contributing to this discrimination. Algorithms, specifically machine learning and artificial intelligence, are programmed to make correlations between different traits in a human's life or online footprint. Any data point, whether it be race, ethnicity, religion, what they purchase online, etc can become a correlation if the algorithm works in a certain way- this is what they're programmed to figure out. Autocorrelations, “when a single aspect of a person’s life is measured repeatedly over time,” and patterns that occur between different people, become part of the algorithm’s way of creating an answer or suggestion to researchers (Williams, et al).

Whether or not these correlations actually serve any purpose or actually reflect the reality of a person’s life is up to the researchers. With a huge amount of data to process, trends become the preferred way to view data, which means that individual discriminations get lost in big numbers but contribute to a bigger problem. Discrimination then can be difficult to combat in an algorithm because of its correlations and lack of context. Algorithms “are designed to find and exploit patterns in big data will pick up on social categories and trace evidence associated with them” (Williams, et al).

Society today has moved towards attempting to fix discrimination by removing the categories themselves. However, removing categories such as gender or race doesn’t help due to systemic discrimination in the real world contributing to correlating data points in a model. For example, take the removal of gender from a success-prediction model of employees. Due to the very real gender wage gap, women will be reported as having a lower salary. However, since gender has been removed from this model, the algorithm will view all employees as the same, therefore concluding the lower-earning employees will be less successful. Women then, are discriminated against in this algorithm due to their gender influencing their data points, even when they have the same chance of being successful as men.

This conclusion is drawn in part from the signaling theory. A concept that explains how "arbitrary biases can arise even when judgments are not explicitly prejudiced and can become self-perpetuating when decision-makers act on these biased judgments" is very common in data research (Williams, et al). Signaling theory takes one data attribute and correlates it to another, making the reference that one attribute implies the existence of another. This theory is very relevant when algorithms and big data sets are used because researchers use any method they can in order to get a beneficial result.

Due to its implicit nature, signaling theory can also be included as a discriminating factor. The former head of the Federal Trade Commission, Edith Ramirez, states, "[at] the very least, companies must ensure that by using big data algorithms they are not accidentally classifying people based on categories that society has decided- by law or ethics- not to use, such as race, ethnic background, gender, and sexual orientation" (Williams, et al). Using signaling theory along with other big data practices need to be assessed for tactics that would cause any sort of discrimination.

Theories and algorithms have the opportunity to be discriminatory by nature. O'Neil defines a Weapon of Math Destruction (WMD) as

> "mathematical models or algorithms that claim to quantify important traits… They have three things in common: opacity, scale, and damage. They are often proprietary or otherwise shielded from prying eyes, so they have the effect of

> being a black box. They affect large numbers of people, increasing the chances that they get it wrong for some of them. And they have a negative effect on people, perhaps by encoding racism or other biases into an algorithm or enabling predatory companies to advertise selectively to vulnerable people…" (O'Neil, 3).

The gender-discrimination salary algorithm or even Facebook's advertising algorithms can all be considered WMDs. WMDs are dangerous in three different ways. First, they have the ability to predict and guess the future. Second, they can imputate, in that they can infer data points based on other known data points. Third, they use proxy variables that result from similar data points outside the current set to create a point inside the set (Williams, et al).

Weapons of Math Destruction and other targeted marketing algorithms threaten justice because they do not treat each and every consumer equally and fairly. Although they attempt to create assumptions and predictions based on "unbiased data," the data they are being fed is inherently biased. Further, because the algorithms create their predictions on their own with little oversight as to how they make the decisions they do, there is little room to identify or fix these injustices.

Nonmaleficence

The third way in which big data conflicts with ethical practices is through nonmaleficence, or the stance to do no harm, is defined as the "obligation not to inflict harm intentionally" (Bishop). Similar to many ideas in justice, algorithms in big data inflict harm without intending to. While algorithms are useful in that they discover patterns and create predictions that humans either wouldn't be able to find or would take a very long time to discover, algorithms also lack a very important quality that humans have- the ability to evolve.

Algorithms form their own stereotypes without checks. They "can't decide guilt. They can't weigh up arguments from the defense and prosecution, or analyze evidence, or decide whether a defendant is truly remorseful" (Fry, 54). On the other hand, humans evolve and are able to make changes in a meaningful manner, not solely due to data and factual evidence, but to changes in philosophy. "As human beings learn and adapt, [humans] change, and so do [their] processes. Automated systems, by contrast, stay stuck in time until engineers dive in to change them" (O'Neil, 203) Automated systems on their own are unable to adapt to social norms and changing cultural knowledge, thus are behind the times in terms of rights and biases.

Velocity and additionally, variety, are big points of discussion in data collection that relate to nonmaleficence. Velocity references the concept of access to data within a timely manner. Data is most useful when it is gathered and analyzed quickly. Although technology has improved, the route of processing data quickly and cheaply is a challenge that researchers face. On the other hand, variety encapsulates the type of data that is being stored. While data used to be stored in a very structured manner, the types of data that are collected today make it difficult and less structured. These data types include data from social media sites, audio, video, organizational messages, internal documents, email, web page data collection, comments from customers, and so many more (Nunan, Domenico). Solutions to these problems include flash-based disk drives and non-relational databases that make unstructured data storage easy. Big data storage units have also populated in massive numbers all over the world. Since variety and velocity are two main aspects in data collection, researchers might be tempted to sidestep security precautions in order to meet their goals.

How data is secured and protected is also incredibly important to the idea of nonmaleficence. Much of the data collected isn’t harmful on its own. It can help researchers market their products towards customers if they see their customer likes a certain trend, but if certain information gets into the wrong hands it can be very harmful. Protecting data, or solving a security breach, can be as easy as wiping out a bank account and resending a credit card, or as difficult as a major company’s credit card breach. Data is additionally being collected on humans in massive amounts, unbeknownst to humans. Anything from an electricity bill to nanotechnology in buildings together is creating a comprehensive picture of each and every human alive.

Due to the rapid pace of big data collection and improved technological advances, privacy is a moving boundary within the field. Privacy can fit into many categories of ethics, however it pertains particularly to nonmaleficence. In order to keep their customers informed and maintain their own privacy policies and standards, tech companies have been known to update their privacy policies frequently. While this helps keep the companies avoid legal trouble, it can confuse users. When a company's privacy policy changes often, it can be difficult for a user to keep track of what sort of data they share and what privacy rights they have for one company, much less the hundreds of different websites users access each month.

The validity of data is also extremely important to nonmaleficence. At a first glance, when it comes to tailoring and personalizing online advertisements, validity may not seem significant. If a user receives an advertisement on their webpage due to an algorithm that has assessed their search history, for a product that isn't interesting to them, they most likely won't click on it and therefore will not buy it. This doesn't seem to be a pressing ethical problem. However, depending on the product and algorithm behind its advertising, this situation can be consequential.

A 2002 journal article discussing the evolution of the internet from its beginnings to what it is today explains that the validity of data, whether online or offline, is extremely important. "Web-based surveys do face a threat in the area of predictive validity in that the populations are usually biased toward those who have access to the Internet and/or the inclination to respond to online surveys" (Topp, Pawloski). In 2002, access to the Internet was a concern and potential bias for data collection. As of October 2018, there are close to 4.176 billion humans or 55 percent of the global population online (Salim). Sixteen years later, Internet access doesn't seem to be as big of a concern and/or bias when it comes to data collection. Now, there are many more factors, algorithms, and biases to take into consideration.

Data analytics and data collection practices are used because companies believe they will add value to their businesses. Through the use of this data, companies trust that the data will be processed and further used according to their standards, not mistreated or leaked. However, due to the velocity of copious amounts of data, it can be hard to ensure that security and trust. Data analytics are usually "driven by a fine blend of 'perceived trustworthiness' and 'evidence of its actual trustworthiness,'" without actually being fully ensured (Shivalker).

While the collection and processing of data raises many ethical questions, the security and safety of this collected data is equally as important. Unlike the UK's GDPR, there is no single document or law in the United States that regulates the collection of data. Instead, the US has a system of federal laws that provide a framework to self-regulation. These frameworks detail "best practices," which essentially are ideals that companies should strive for when it comes to data collection and use, however the companies are not legally binded to follow these ideals (Rottigni). This leaves leeway for companies to do as they please with their collected data.

The "best practices" apply to data security. Data security pertains to the technical aspects behind protecting data. The three main ideals of data security are confidentiality, that "data and information assets must be confined to people authorized to access and not be disclosed to others;" integrity, to keep "the data intact, complete and accurate, and IT systems operational;"

and availability, “an objective indicating that information or system is at disposal of authorized users when needed” (“Key Elements of an Information Security Policy”). Confidentiality, integrity, and availability are all incredibly important aspects of data security, whether it be for the benefit of the organization or its consumers.

Confidentiality renders itself to nonmaleficence. A data breach in any company, big or small, is not only a public relations nightmare, but a risk to the entire well-being of a company. In past years, front-page headlines have covered data breaches at major companies such as Yahoo, eBay, Adobe, and Target. These breaches have included hackers receiving access to user credentials (username and password), credit card information, and profile information, such as date of birth, ethnicity, and more. One data breach from a major company, such as Yahoo, is extremely consequential and concerning to the millions of users it affects. What furthers this concern ten-fold is that any company of any size can have a data breach, hack, or virus in its system at any point in time. While online security systems grow stronger every year with the improvement of technology, there is always the possibility of an attack and a resulting data breach.

Data integrity additionally is incredibly important to users and companies alike, and is directly impacted by security measures. Depending on the company and the type of data being collected, lack of integrity can cause a large amount of damage. On a small scale, if a user’s credit card information on a website’s file system gets attached to another user’s account, there is the opportunity of theft. Continuing on the scale, if medical records are tampered with, life and death could be at stake for a patient. If a patient’s vitals or medical history are changed, doctors could prescribe the wrong treatment or the wrong dosage of medicine. On a worldwide scale, data integrity, as well as data confidentiality, are imperative during war. If a country is planning an attack on a presumably empty warehouse, but has the wrong data, thousands of lives could be at risk. Having the correct information for any scenario is imperative to success and nonmaleficence.

The availability of data is also important to note when it comes to security. Having the right people be able to access stored data and refusing access to the wrong people is the key issue. Similar to a hierarchy in a company, the top executives receive the most sensitive information while the bottom employees receive the least. Since there are typically fewer people at the top, there is a smaller chance data will get leaked. In data collection, storage, and use, the same theory applies. The fewer people with access to add, change, and delete data, the better. The availability of data, who has access and who doesn’t, is crucial to maintaining its security.

All three of these measures, confidentiality, integrity, and availability, can be scaled up or down to apply to any sort of situation pertaining to online data. In regards to cookie collection, the lack of any one of these, much less all three, are why security measures are so crucial for companies to consider and uphold. If any one of these points are violated, the pillar of nonmaleficence is violated as well, harming both the company and its users.

The Ethics of Cookies

Consumers lend their trust to online companies when they enter data into their systems. The idea of “perceived trustworthiness,” could not be more relevant in today’s online security environment. Companies might include pages of legal documents detailing their security promises somewhere on their website, but most users see the image of a lock next to the space where they enter in their credit card number and assume safety. Online “phishing,” or fake websites that entice users to enter in private information while sending it to an external location without a user’s knowledge, are extremely rampant today. The biggest cause of viruses and bugs within a computer system is human error- that a human clicks on a link or site that appears to be something else, therefore installing the detrimental item onto their computer system. While security systems improve, the possibilities for error and intrusion are endless.

Data security is important for many reasons. First and foremost, when a user enters sensitive information into a website, they expect that information to stay unique to that website. Whether it be credit card information, personal profile information, medical history, or other, a user expects and trusts that their information will be safe. Information of this nature can seem non-important to some but life-changing to others, depending on scale. Medical information, for example, is extremely sensitive information. Hospital records then, must be kept incredibly secure. Medical history to healthy individuals might not seem to be a pressing matter, but it can prove to be very stressful to others.

In a focus group led it 2013, regular citizens were asked about the most concerning parts of their medical history. The majority “weren’t that worried about their data being hacked or stolen. They were concerned about having assumptions made about them as a group and then projected onto them as individuals” (Fry, 106). The individuals were worried about how their data might be used against them. The concept of doctor-patient confidentiality is there for a reason- people want their information being kept private. Putting data into automated storage systems online, even with up-to-date protections, increases the likelihood that this trust will be broken.

Beneficence

Beneficence, the obligation to do good, is defined as the ability to “provide benefits to persons and contribute to their welfare. [Beneficence] refers to an action done for the benefit of others” (Bishop). As far as beneficence goes, there is a substantial argument to big data practices being beneficial to today’s modern society, particularly to the United State’s capitalistic economy.

Data scientists use the data they farm from cookies and other customer collection methodologies in order to paint a picture of each and every customer. In past years, marketers created distinct pictures of what they thought small populations of their customers were- based on region, age, etc. Due to the specificity of online data collection, marketers now are able to access increasingly complex profiles of their customer base. Using these profiles, they are able to market their products towards each customer based on their interests. With the implementation of Google Ads, like previously mentioned, user clicks are tracked, collecting information on the types of items they might click on while browsing Amazon or other online shopping websites. Cookies pick up on this information and store it for later use, comparing a clicked on item to what other users clicked on after this item. Here, algorithms are pairing products together that most users tend to click on. Thus, users will be more likely to click on a suggested item because other users did so too.

Algorithms increase the likelihood that a user will click on a suggested item based on past data. This idea of 'suggested' items is something seen across many different types of platforms. Amazon uses it to suggest other, similar products to the know the user is currently looking at, whereas Netflix uses these sort of algorithms to suggest similar movies to the ones users have clicked on. Suggested products increase customer satisfaction because they can help users find something they were already looking for, or more importantly for the company, find something the customer didn't know they were looking for but found and purchased anyways.

Future of Big Data

"Greased data," the concept of data being quick and easy to collect and distribute, is what the current status of data collection strives for. Concerns about privacy occur when users don't know where their data is being sent to or how it is being used. Greased information is “information that moves like lightning and is hard to hold onto” (Moor). Determining the boundaries around specific controls and restricted access is a potential solution to this relevant problem. Restricted access is the idea that “different people may be given different levels of access for different kinds of information at different times” (Moor). For example, not every software developer at an online shopping website will know how to access user credentials, login information, and linked credit card accounts. This is beneficial to both the company and its users because there is a smaller chance that one of the employees will corrupt or steal that sensitive data.

One attempt at data protection is the Never Again Tech Pledge, created and endorsed by many US tech companies. The pledge agrees to “refuse to participate in the creation of [government] databases . . . to target individuals based on race, religion, or national origin and to minimize sensitive data collection” (Williams, et al). This pledge was created after President Trump suggested in 2016 that a database be created to register all Muslim citizens of the United States. As of early 2016, the pledge had more than 90 companies from Silicon Valley as signees, including Google, Indiegogo, Stripe, Giphy, and more (Stone). While this pledge is a step in the right direction, there are many more ways tech companies can change and better their practices to fight discrimination.

Companies still have the opportunity to data mine with algorithms and not be discriminatory. In 2016, Airbnb created a racial discrimination audit program to decrease racial discrimination that users on their website have reported. Additionally, Airbnb made a goal to hire a more diverse staff and especially appoint more diverse leadership. When they made the promise, their staff consisted of 63 percent white employees, 7.1 percent Latino/a employees, and 2.9 percent black employees (Dickey). This range of racial backgrounds is typical of tech companies. The president and CEO of the Leadership Conference on Civil and Human Rights, Wade Henderson, commented that Airbnb's efforts "should be considered by other Silicon Valley companies that have largely failed to reflect the diversity of the nation with their workforces. This report has not addressed every issue of concern but it is an important step in the right direction" (Dickey). If a company has more diversity within their workforce, it is more likely to have differing views. Hiring underrepresented groups, especially to leadership positions that have more power, increase the likelihood that problematic situations can be averted.

Another method to fighting discrimination and biases in data collection is the process of differential item functioning. During a test,

> "typically, a new test item is added, amidst existing test items, and experts reject it if test takers with the same scores everywhere else perform differently on it by demographic subgroup.This process can catch items where wording has additional

> connotations to some groups of test-takers or where other small differences flag social identity rather than the skills the test is intended to measure" (Williams).

In high-stakes studies, researchers will take the time to do this sort of testing in order to rule out any discrimination or bias. All companies that do studies like this, but especially tech companies that use and process cookies, need to make this testing a high priority because there are high stakes when it comes to discrimination.

Transparency within the field of big data is both a concern and potential solution to the ethical dilemmas and considerations. As of now, most big data practices when it comes to the collection of online cookies are transparent. The terms and conditions provided on websites are distinct and prominent when a user first enters the website, but the complexity and length of the legalities and specifications are lost on the user. In an era of quick data and information access through the internet, it is rare that a user takes the time to read each and every cookies collection specification they access on the internet. Thus, users rarely know what they are clicking on, and further, rights and privacy privileges they are signing away when they click 'accept.' Transparency is a method and mindset that big data companies can and should use to improve on the four pillars of ethics.

Conclusion

The Ethics of Cookies

> "And yet, pointing out the flaws in the algorithms risk implying that there is a perfect alternative we're aiming for…. Algorithms *will* make mistakes. Algorithms *will* be unfair. That should in no way distract us from the fight to make them more accurate and less biased wherever we can - but perhaps acknowledging that algorithms aren't perfect, any more than humans are, might just have the effect of diminishing any assumption of their authority" (Fry, 200).

Weapons of Math Destruction, unintentional discriminating algorithms, and unethical data mining practices threaten society today. While massive tech companies forge ahead building new technologies that attempt to better the world and humans' way of living, they risk crossing ethical boundaries. Targeted advertising and data mining practices can be beneficial; figuring out what online users want before they even know they want them, or suggesting a purchase that a user wouldn't normally think of but results in buying can be good and helpful to users. "Mathematical models can sift through data to locate people who are likely to face great challenges, whether from crime, poverty, or educations. It's up to society whether to use that intelligence to reject and punish them- or to reach out to them with the resources they need. We can use the scale and efficiency that make WMDs so pernicious in order to help people. It all depends on the objective we choose" (O'Neil, 97). However, the ways in which these marketing practices work can be extremely discriminatory and can encourage bias. Cookies themselves are not "harmful." The information collected through the use cookies and the profiles that are later created by marketing companies and algorithms are where ethical concerns come into play.

The Ethics of Cookies

In order for market researchers to reach success, their subjects must participate. Without any data, researchers have very little to go off of. When they conduct their studies, market researchers must figure out ethically sound ways to collect and process their data and continually scrutinize it for any sort of discrimination. There can be a balance between business and ethical practices. In the fast-paced environment that is technology and data collection however, the ethical considerations need to be at the forefront of business, whether that comes from regulation or public outcry and pressure. "And yet, pointing out the flaws in the algorithms risk implying that there is a perfect alternative we're aiming for…. Algorithms *will* make mistakes. Algorithms *will* be unfair. That should in no way distract us from the fight to make them more accurate and less biased wherever we can - but perhaps acknowledging that algorithms aren't perfect, any more than humans are, might just have the effect of diminishing any assumption of their authority" (Fry, 200).

Bibliography

Bishop, Laura. *Principles — Respect, Justice, Nonmaleficence, Beneficence*. Kennedy Institute of Ethics. https://www.nwabr.org/sites/default/files/Principles.pdf

Bowie, Norman E., and Karim Jamal. "Privacy Rights on the Internet: Self-Regulation or Government Regulation?" *Business Ethics Quarterly*, vol. 16, no. 3, July 2006, pp. 323–342.

Burri, Mira, and Rahel Schar. "The Reform of the EU Data Protection Framework: Outlining Key Changes and Assessing Their Fitness for a Data-Driven Economy." *JSTOR,* Journal of Information Policy, 4 Nov. 2018, www.jstor.org/stable/pdf/10.5325/jinfopoli.6.2016.0479.pdf?refreqid=excelsior%3Ae05bf02932518a8688d3f2ce3ff20bdf

Coos, Andrada. "EU vs US: How Do Their Data Protection Regulations Square Off?" *Endpoint Protector Blog,* 17 Jan. 2017, www.endpointprotector.com/blog/eu-vs-how-do-their-data-protection-regulations-square-off/.

Costigan, Sean S., and Gustav Lindstrom. "Policy and the Internet of Things." *Connections: The*

Quarterly Journal, vol. 15, no. 2, 2016, pp. 9–18., doi:10.11610/connections.15.2.01.

Cowley, Stacy. "Hold the Phone! My Unsettling Discoveries About How Our Gestures Online Are Tracked." *The New York Times,* The New York Times, 15 Aug. 2018. www.nytimes.com/2018/08/15/business/behavioral-biometrics-data-tracking-html?rref=collection%2Ftimestopic%2FComputer%2BSecurity%2B%28Cybersecurity%29&action=click&contentCollection=timestopics.ion=stream&module=stream_unit&version=latest&contentPlacement=7&pgtype=collection.

Dance, Gabriel JX. "As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants."

The New York Times, The New York Times, 19 Dec. 2018, www.nytimes.com/2018/12/18/technology/facebook-privacy.html.

Debatin, Bernhard, et al. "Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences." *Journal of Computer-Mediated Communication*, vol. 15, no. 1, 2009, pp. 83–108., doi:10.1111/j.1083-6101.2009.01494.x.

Dickey, Megan Rose. "Here's Airbnb's Plan to Fix Its Racism and Discrimination Problem."

TechCrunch, TechCrunch, 8 Sept. 2016, techcrunch.com/2016/09/08/airbnb-plan-fix-racism-discrimination/.

Domonoske, Camila. “Lawmakers Push Zuckerberg On Security, Diversity, Drug Sales On Facebook.” *NPR*, NPR, 11 Apr. 2018, www.npr.org/sections/thetwo-way/2018/04/11/599590470/mark-zuckerberg-is-back-before-congress-for-a-second-day-of-testimony.

Eggers, Dave. *The Circle*. Hamish Hamilton, 2017.

Fry, Hannah. *Hello World: Being Human in the Age of Algorithms*. W. W. Norton & Company, 2018.

Fung, Brian. “Why You're Getting Flooded with Privacy Notifications in Your Email.” *The Washington Post*, WP Company, 25 May 2018, www.washingtonpost.com/news/the-switch/wp/2018/05/25/why-youre-getting-flooded-with-privacy-notifications-in-your-email/?noredirect=on&utm_term=.88e7da274d1e.

Government Digital Service. “Data Protection.” *GOV.UK*, GOV.UK, 16 Sept. 2015, www.gov.uk/data-protection.

Granville, Kevin. “Facebook and Cambridge Analytica: What You Need to Know as Fallout Widens.” *The New York Times*, The New York Times, 19 Mar. 2018, www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html.

Hill, Kashmir. “How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.” Forbes, Forbes Magazine, 31 Mar. 2016, www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6aff5d5d6668.

“HTTP Cookies.” *MDN Web Docs*, Mozilla, 21 Sept. 2018, developer.mozilla.org/en-US/docs/Web/HTTP/Cookies.

Hill, Kashmir. “How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did” *Forbes,* Forbes Magazine, 31 Mar. 2016, www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6aff5d5d6668

"Key Elements of an Information Security Policy." *Infosec Resources*, 12 Feb. 2019, resources.infosecinstitute.com/key-elements-information-security-policy/#gref.

McCormick, Thomas D. "Principles of Bioethics." *Bioethic Tools: Principles of Bioethics*, University of Washington School of Medicine, 2013, depts.washington.edu/bioethx/tools/princpl.html.

Mian, Atif, and Howard Rosenthal. "Big Data in Political Economy." *RSF: The Russell Sage Foundation Journal of the Social Sciences,* vol. 2, no. 7, Nov. 2016, pp. 1–10.

Miller, Claire Cain. "When Algorithms Discriminate." *The New York Times*, The New York Times, 9 July 2015, www.nytimes.com/2015/07/10/upshot/when-algorithms-discriminate.html.

Milne, George R., and Mary J. Culnan. "STRATEGIES FOR REDUCING ONLINE PRIVACY RISKS: WHY CONSUMERS READ (OR DON'T READ) ONLINE PRIVACY NOTICES." *JOURNAL OF INTERACTIVE MARKETING* , vol. 18, no. 3, 2006, pp. 15–29.

Miyazaki, Anthony D. “Online Privacy and the Disclosure of Cookie Use: Effects on Consumer Trust and Anticipated Patronage.” 2008, pp. 19–33.

Moor, James H. “Towards a Theory of Privacy in the Information Age.” *Computers and Society*, Sept. 1997, pp. 27–32.

Nunan, Daniel, and Marialaura Di Domenico. “Market Research and the Ethics of Big Data.” *International Journal of Market Research*, vol. 55, no. 4, 2013, pp. 505–520., doi:10.2501/ijmr-2013-015.

O’Neil, Cathy. “Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.” *Weaponsofmathdesctructionbook.com*, Crown New York, 2016, weaponsofmathdestructionbook.com/.

Parson, Aaron. “What Happens When You Delete Cookies?” *It Still Works*, 10 Jan. 2019, itstillworks.com/happens-delete-cookies-4388.html.

Queiroz, Anderson A. L., and Ruy J.G.B. De Queiroz. “Breach of Internet Privacy through the

Use of Cookies." *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '10*, 2010, doi:10.1145/1839294.1839378.

Reyman, Jessica. "User Data on the Social Web: Authorship, Agency, and Appropriation." *College English*, vol. 75, no. 5, 1 May 2013, pp. 513–533.

Richterich, Annika. "Examining (Big) Data Practices and Ethics." *The Big Data Agenda*, University of Westminster Press, 2018, pp. 15–31.

Roberts, Jeff John. "The GDPR Is in Effect: A Plain English Guide for US Companies." *Fortune*,

25 May 2018,

fortune.com/2018/05/24/the-gdpr-is-in-effect-should-u-s-companies-be-afraid/.

Rottigni, Roberta. "Why IoT Data Protection Has Become More Important than Ever " *ReadWrite*, 22 Nov. 2018, readwrite.com/2018/11/27/why-iot-data-protection-has-become-more-important-than-ever.

Rutherfoord, Rebecca H, and James K Rutherfoord. “SIGITE '10 Proceedings of the 2010 ACM Conference on Information Technology Education.” 7 Oct. 2010.

Salim, Saima. “Global Internet Trends 2018: Almost 4.18 Billion Humans Are Online, 3.4 Billion Are Active Social Media Users.” *Digital Information World*, 19 Oct. 2018, www.digitalinformationworld.com/2018/10/the-state-of-the-digital-world-in-q4-2018.html.

Satariano, Adam. “Europe Worries as Facebook Fights Manipulation Worldwide.” *The New York Times,* The New York Times, 22 Aug. 2018, www.nytimes.com/2018/08/22/business/facebook-Russia-iran-Britain.html

Shivalker, Chirag. “4 Pillars Of Trusted Data Analytics | Articles | Chief Data Officer.” *Articles | Chief Data Officer | Innovation Enterprise*, 14 Nov. 2017, channels.theinnovationenterprise.com/articles/4-pillars-of-trusted-data-analytics.

Stone, Zara. “'Never Again' Is Silicon Valley's Public Pledge To Refuse A Muslim Registry.”

Forbes, Forbes Magazine, 13 Dec. 2016, www.forbes.com/sites/zarastone/2016/12/13/never-again-is-silicon-valleys-public-pledge-to-refuse-a-muslim-registry/#165be56152d7.

Tanner, Adam. “The Web Cookie Is Dying. Here's The Creepier Technology That Comes Next.” *Forbes*, Forbes Magazine, 30 June 2014, www.forbes.com/sites/adamtanner/2013/06/17/the-web-cookie-is-dying-heres-the-creepier-technology-that-comes-next/#bb20abda9bfa.

Topp, Neal W., and Bob Pawloski. “Online Data Collection.” *Journal of Science Education and Technology*, vol. 11, no. 2, June 2002, pp. 173–178., doi:10.1023/a:1014669514367.

Tucker, Catherine E. “Social Networks, Personalized Advertising, and Privacy Controls.” *Journal of Marketing Research*, vol. 51, no. 5, 2013, pp. 546–562., doi:10.1509/jmr.10.0355.

Vitak, Jessica, et al. “Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community.” *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW 16*, 2016, doi:10.1145/2818048.2820078.

Williams, Betsy Anne, et al. “How Algorithms Discriminate Based on Data They Lack: Challenges, Solutions, and Policy Implications.” *Journal of Information Policy*, vol. 8, 2018, pp. 78–115., doi:10.5325/jinfopoli.8.2018.0078.

Wu, Tim. “How Capitalism Betrayed Privacy.” *The New York Times*, The New York Times, 11 Apr. 2019, www.nytimes.com/2019/04/10/opinion/sunday/privacy-capitalism.html.

“Your Baby's Screening What Happens to the Blood Sample.” *Residual Dried Blood Spots | Baby's First Test | Newborn Screening | Baby Health*, Baby's First Test, www.babysfirsttest.org/newborn-screening/what-happens-to-the-blood-sample.