5-2020

Understanding Personal Data in the World of Social Media

Nicholas Scott Rodgers
Utah State University

UNDERSTANDING PERSONAL DATA IN THE WORLD OF SOCIAL MEDIA

by

Nicholas Scott Rodgers

Capstone submitted in partial fulfillment of the requirements for graduation with

University Honors

with a major in

Computer Science
in the Department of Computer Science

Capstone Mentor Dr. Chad Mano	**Departmental Honors Advisor** Dr. Dan Watson
Committee Member N/A	**University Honors Program Director** Dr. Kristine Miller

UTAH STATE UNIVERSITY
Logan, UT

Spring 2020

Abstract

Personal data is behind many of the online interactions that people have through social media and other online sites and services. This data allows sites to understand their users, which in turn allows them to provide better content for their users. This data is also used to determine user interests, which these online services use to target more relevant advertising to their users, and share the information that they collect about their users with third parties. It is only recently that this personal data is being regulated by lawmakers, the businesses running these sites are held accountable for managing the data that they collect on users, and that the users of these sites are given a greater degree of control over how their data is used, however these regulations are not universal. Additionally, the data that online services have collected on their users has been used by third parties for malicious purposes, such as the Russian interference in the 2016 U.S. presidential election. The collection of personal data allows for the development of new technologies such as facial recognition and genealogy through DNA, but without regulation, moral dilemmas surround the collection and use of such personal data.

For my capstone I built a website structured like a social media application. On this site I have written a number of posts that communicate the findings of my research to the users of this site. Additionally, on this website visitors are able to see some of the type of personal data that social media sites can collect about them while they use the service.

For my mom and dad. I wouldn't be here without your support.

Acknowledgments

I would like to thank Dr. Dan Watson, who has been willing to work with me and provide validation on my decisions as I worked on this project. Going into this project I do not believe either of us had any idea of the scope of the project or everything that it would entail. I am thankful that he stuck with me long enough to see it through. I would also like to thank Dr. Vladimir Kulyukin, who pointed me in the right direction as I was beginning my research, and introduced different aspects of the issues surrounding personal data that I may not have considered. Finally I would like to thank Dr. Chad Mano for his willingness to serve as my faculty mentor.

I would also like to thank the USU Honors program for their flexibility and support throughout my undergraduate career and especially during the COVID-19 pandemic. Without the support of their advisors, especially Amanda Adison, I do not think that I would have completed this project.

Table of Contents

Abstract.....i
Acknowledgments.....iii
Final Product.....1
Introduction.....1
What is Personal Data?.....1
How Personal Data is Collected and Used.....3
Risks Associated with Collecting Personal Data.....4
Consumer Protection over Personal Data.....6
Present and Future Issues Relating to Personal Data.....7
My Website.....9
I. Homepage - Newsfeed.....10
II. Security.....12
III. Profile.....13
IV. Citations.....14
Conclusion.....14
Reflection.....16
Works Cited.....18
Biography.....21

Table of Figures

Figure 1: Homepage 10
Figure 2: Security Page 14
Figure 3: User Profile Page.....15
Figure 4: Citations page.....16

Vocabulary

Data – Information that can be processed or stored by computers such as text, audio, and images.

Deepfake - A video generated by artifical intelligence and machine learning to intentionally deceive viewers by giving the appearance that the person in the video is doing or saying something that they have not.

California Consumer Protection Act (CCPA) – The California law that regulates the collection, use, storage, and sale of the personal data of California citizens.

General Data Protection Regulation (GDPR) – The set of European Union (E.U.) laws that regulate the collection, use, storage, and sale of the personal data of E.U. citizens by any companies that use or collect this data.

Personal Data - A smaller subset of computer data, information relating to an identifiable living person, including health, financial, and behavioral data.

Personal Health Information (PHI) – A subset of Personal Data, PHI is data that is related to an individual's health or their medical records, such as health history and insurance information. In the U.S. PHI is protected by the Health Insurance Portability and Accountability Act of 1996 (HIPAA).

Personally Identifiable Information (PII) - A subset of Personal Data, PII is data that can directly identify a specific individual such as a social security number or email address.

Terms of Service (ToS) – A legal document that online businesses and services make the user agree to in exchange for providing them service which include legal information about how the site can be used and how personal data can be collected.

Final Product

Word Count: 5,592

Introduction

During my time as an undergraduate student in the Computer Science program at Utah State University, I have had a few jobs in web development, and will continue to work in this field once I graduate. This field of rapidly evolving technology is powered by data, and as the world moves increasingly online, it becomes more and more important to have a basic understanding of the data that powers our online interactions, what it is used for, and what powers individual users have over their own data. I myself did not know much about what this data consisted of when I started this project, and I am sure that there are many others that share this same gap in knowledge. As someone that will most likely be interacting with this data frequently in my career I believe it is important that I know more about it. As someone that also uses social media frequently and generates this data, I believe it is also important that I understand how this data gets used, and share this information with others who would like to take more responsibility for their online presence.

Additionally, through the classes I have taken for my Spanish minor as well as the honors program, I have come to understand both the power and importance of communication with others. For my capstone project, I wanted to do something that would be representative of the type of work that I have done and what I have learned in my undergraduate career. For my project I chose to build a website that showcases what I have learned in my research as well as well as communicate my findings to readers in an accessible and familiar setting. I wanted to answer a handful of questions regarding personal data to form a basic understanding of what it is and why it is important. These questions consisted of defining what personal data is, how it is allowed to be collected, how it is used, potential risks, how to protect personal data, and how our understanding of and relationship with personal data may change in the future.

What is Personal Data?

In Computer Science, data simply refers to information that can be processed or stored by a computer such as text, images, audio and video. Data is what drives every online interaction and is transmitted globally at rates of 92,653 gigabytes per second [1]. Personal data is a smaller subset of this data, and is defined by the European Commission as, “any information that relates to an identified or identifiable living individual,” meaning if the data can be related to a person, it is personal data [2]. This definition encompasses quite a lot of information. Personal data can be data directly linked to a person, such as their name, email address, financial information, or medical information. Data that can be used to directly identify a person is referred to as

Personally Identifiable Information (PII). This definition also encompasses data that is linked to a person in an indirect manner, such as the ID of a tracking tag in their browser, a user's IP address, or behavior data about a user on a site, including the user's list of "liked" posts or a user's "friends" list. This data is still deemed personal data even if it has been encrypted or obfuscated in such a way that it does not directly tie to a user but can still be used to re-identify a person. Data is deemed no longer personal after it has been rendered anonymous in such a way that it is impossible to revers and link back to a person.

Using the European Commission's definition of personal data is useful as it provides a broad but well-defined legal interpretation of personal data, which provides a basis for understanding the different issues that arise when discussing its use. The European Commission has also developed a list of regulations regarding the use of personal data in the European Union (EU), called the General Data Protection Regulation (GDPR), which provides legal protection to users over their personal data and regulates how businesses can use personal data. This research project is focused more on understanding the use of personal data in a very broad perspective, as it is often more difficult to understand why data that is not directly tied to a person is important. As such, the main focus of this research project is on the use of personal data other than PII, although there is a significant overlap between PII and non-PII personal data. For example, a photo uploaded by a user may have their face within it, but another picture may not have any distinguishable characteristics.

Additionally, it is useful to consider personal data from a legal perspective as online service providers generally have users agree to their Terms of Service when they sign up for an account. The Terms of Service (ToS) is a legally binding document which lists the legal guidelines for using the service, limitations of liabilities, privacy policies detailing the use of personal data, and more. However, these documents are often very long and have to use legal language that the average user may not understand. According to the Norwegian Consumer Council, in order to read all of the Terms of Service from 33 apps on an "average" smartphone, it would take over 24 hours [3]. Given that these businesses may also change their terms of service, this becomes a large task for the average consumer, and as such many are disinclined to read or understand what they are agreeing to by signing up. Additionally, this research was unable to find any legal regulations for clarifying to users what is contained in a ToS document and as such many businesses do not provide such clarification. Facebook, among a handful of other sites, provides a page that outlines their data policy, providing a more user-friendly outline of how their data is stored, used, and shared [4]. There are also initiatives like Terms of Service; Didn't Read, which aim to classify websites based on the user protections listed in their ToS, as well as give a succinct overview of the agreements made by signing up for the service [5].

How Personal Data is Collected and Used

To use the services provided by a site like a social media application, a person generally has to create an account to use the site and become a site "user." User information such as their name, email address, date of birth, profile photo, and possibly even financial information are all tied to this account. Additionally this account identifies interactions the user has while using the service, such as the type of content that they find engaging, the type of content that the user shares, the other users that the user interacts with, and the advertisements that the user interacts with and responds to [4]. Based on a combination of the data that a user provides, as well as the user interactions, the company that owns the site can build up the information that they have about the user to have a better understanding of them. This is commonly done through the use of machine learning and proprietary algorithms.

Technology companies such as Facebook and Google make a large portion of their revenue through the use and sale of advertising space to other companies. As such social media sites often are designed to hold a user's attention for as long as possible, and to bring the user back to the application multiple times throughout the day. This allows the sites to show more advertisements to their users, and thus generate more revenue. Most users are intend to use these sites as distractions and are happy to give their attention to the services, and 71% of Snapchat users, 55% of Instagram users, and 51% of Facebook users report using the service multiple times throughout the day, and 51% of users ages 18 to 24 say that they would find it difficult to give up using social media. With roughly seventy percent of adults in the United States using Facebook and YouTube, these sites are able to collect a lot of information about their users [6].

Businesses are able to leverage this user information for a number of different purposes. One such purpose is to recommend content to the user. Based on a user's interactions, such as posts liked and commented on, as well as other user profiles followed, the site can develop suggestions of other posts that the user may be interested in out of the vast number of other posts on the site. This recommendation algorithm is typically handled by a Machine Learning algorithm, which collects information about the posts a user interacts with, suggests new posts, and based on how a user responds can tune its recommendations. Instagram uses this type of machine learning algorithm to suggest content to users in the "Explore" tab of their application [7]. This type of recommendation system adds value to the user experience by suggesting more relevant content to what a user is interested in, and to the site as well by increasing user engagement and thus advertisement viewing. Some companies also use the data that they collect to perform research in fields such as disaster relief, economic patterns and demographic data [8].

When third-party businesses use companies like Facebook and Google to show advertisements, they often get access to more than just ad space on the website. Advertisers are given tools by these companies that allow them to target their ad at specific types of viewers according to generalized demographics of users. Advertisers are often able to target their ads at people based on age, gender, location, language, and interests,

among other features [9, 10]. This allows the advertisements to deliver their message more effectively to the user, as well as allows the advertisers to target their desired audience. These companies hosting these ads, like Facebook and Google, then collect information about the users that interact with the advertisements. This allows the advertisers to understand how effective their advertising is and if it is reaching and connecting with their desired audiences. Exact personal data pertaining to individual users is aggregated, so instead of sharing with advertisers the information of the people that the ads reached, the advertisers only see anonymous statistics about the type of people seeing the ads and how the ads are performing, and if the ads are leading users to make purchases or not [4]. Additionally sites like Facebook and Google use tracking cookies to follow their users around the web. This allows the advertising hosts to track users as they see ads on third-party sites and applications, determine which advertisements to show them, and report their behaviors from these off-platform sites [11].

This system of advertising has become common, but because algorithms develop the demographic groups for users, there have been instances in which the advertising system has been taken advantage of. Facebook recently removed their “conspiracy theory” and “pseudoscience” categories after the tech news site The Markup found out it was possible to target advertisements to users that believed in false information during the COVID-19 pandemic [12]. Additionally, advertisements have been used to spread misinformation and sow discord about political candidates, leading to the United States Congress asking social media sites to fact-check or limit political advertising on their sites. Twitter responded by removing all political advertising, Google limited the political advertisements on their properties, and Facebook made no such concessions, with Facebook’s director of product management Rob Leathern citing the “absence of regulation,” by U.S. lawmakers [13].

As advertising is one of the main methods sites like Facebook and Google make money, there is a financial incentive to choose not to limit political advertising. In the fourth quarter of 2019, Facebook reported a total revenue of $21 billion, with $20.7 billion coming from advertising [14]. In this same time period, Alphabet, the parent company of Google, reported a total revenue of $46 billion, with $36 billion coming from advertising [15]. Facebook also reported that during the three months of the fourth quarter, they had 2.5 billion monthly active users, meaning that the average user only generates $2.76 in advertising revenue for the site.

Risks Associated with Collecting Personal Data

There are a number of risks associated with collecting personal data, especially Personally Identifiable Information. The main concern with having so much information about the users of a site is the risk of the site being hacked or leaking data, and thus exposing the personal information about its users. When data like financial information is breached, there are legal precedents for how companies are to respond, such as after the

2017 Equifax data breach of 147 million people, in which the company agreed to a global settlement of up to $425 million to help those affected by the breach [16]. There are less instances of other types of personal data being leaked, partially because the field is so new. However, in 2016 the analytics company Cambridge Analytica was able to gain access to the psychological profiles of more than 87 million Facebook users without their permission, and used this information to build up voter profiles in order to influence their behavior. This company worked with the Trump campaign during the 2016 election cycle, as well as with the campaign of former national security advisor John Bolton, an provided them with voter data and analytics for their campaigns. This breach and misuse of personal information was considered by some lawmakers to be harmful to democracy, and lawmakers demanded that Facebook's chief executive Mark Zuckerburg testify before Congress. Facebook made promises to be more transparent about the use of its data and to attempt to prevent another incident of this type from happening again but the 87 million users that had their data leaked never received any form of compensation [17].

There are also ways in which groups have influenced U.S. elections without illegally gaining access to the personal data of users. The 2016 U.S. presidential race was notable for Russian interference. In a report released by the U.S. Department of Justice detailing the extent of Russian interference, it was found that a Russian group called the Internet Research Agency, or IRA, ran a campaign aimed to sow discord in U.S. politics through the use of social media and hacking. The IRA purchased ads on social media platforms in which they shared false political information and deliberately targeted users in order to cause discourse and distrust in the U.S. electoral system. Through what the U.S. Department of Justice calls "information warfare" the IRA also placed illegitimate users on social media platforms in order to engage and stimulate other users. The Russian government also was involved in the hacking and leaking of the email accounts of presidential candidate Hillary Clinton's campaign volunteers and employees [18]. The spread of misinformation during this election was a big issue and continues to be a problem during everyday life and into the COVID-19 pandemic, that social media platforms are still working on addressing.

False information is able to spread online much further and faster than factual information for a number of reasons related to how humans take in information. A study published in the journal *Science* found that because false information is more novel than factual information, it inspires fear, disgust, and surprise in the recipients of this information. In turn, the recipients of such information are more likely to share the novel information and continue the spread of the misinformation. The study also found that the effects of spreading such misinformation was especially more pronounced for political news than for news about terrorism, natural disasters, or financial information, spreading nearly three times faster than the other categories because users were more willing to believe false political information. This stands in contrast to the spread of factual information and truthful news stories, which are not as novel, and inspire sadness, joy, and trust, but does not cause users to share the same information. Additionally, the study found that robots and algorithms were more

likely to pick up on false information and spread it than they were for truthful news stories [19]. This means that in addition to those that are intentionally spreading false information online, there are others that are more likely to spread this information unintentionally, due to how the social media sites are structured and how the algorithms decide to suggest content to users.

Consumer Protection over Personal Data

Personal data is still a new field of information and it is only recently that there have been moves regulate the use of personal data and provide users and consumers of these services with legal protections in the event that their personal data is misused or mishandled. As of 2020, there are very few federal laws governing the use of personal data. The U.S. Privacy act of 1974 limits the type of personal information that the federal agencies can collect about the citizens and outlines how to notify citizens about this data. This law was written before the increase in personal devices that we see today, but ammended in 2015 to include more provisions for digital information [20]. There is also the 1996 Health Insurance Portability and Accountability Act (HIPAA) that regulates data privacy in regard to private health information (PHI), meaning individual citizen's medical records, insurance information, and billing information, and limits who can see and share this information [21]. There is the Gramm-Leach-Bliley Act (GLBA) which regulates how financial institutions explain how they share information with their clients and requires these institutions to safeguard sensitive data about their clients [22]. There is also the Children's Online Privacy protection Act (COPPA) which regulates what information online companies are allowed to collect about minors, specifically preventing companies from asking for the PII of children under 12 years old. COPPA has been updated and amended several times since it was first written, and was updated in 2013 to amend the definitions of key terms such as "personal information," and "Web site or online service directed to children," [23]. Otherwise, there are no laws in terms of federal-level regulation that detail how online companies are allowed to collect, store, share, and sell other types of personal data.

There are a number of U.S. states that have passed or are currently working on passing laws regarding the privacy regulations of their citizens. As of May 2020, Massachusetts, New York, Maryland, and North Dakota all have bills in progress that are attempting to address the personal data of their citizens [24, 25, 26. 27]. California is currently the only U.S. State that has passed and implemented extensive legislation regarding the use of personal data by online companies in the California Consumer Protection Act (CCPA). Under the CCPA, California consumers are given the right to know what personal data online companies collect about them, the right to delete personal information held by businesses, the right to opt out from the sale of their personal information, and the right to not be discriminated against for enacting any of those rights. Additionally, the CCPA designates what Businesses are subject to these regulations as well as what businesses have to do in order

to be compliant with the regulations. This includes how businesses are to respond to requests to access or delete data and fines for breaching these regulations [28].

Another prominent legal framework regarding the collection, use, and sale of personal data was implemented by the European Union (EU) in 2018. The General Data Protection Regulation (GDPR) is one of the most extensive and strict privacy and security laws in the world, and applies to all companies that collect data related to EU citizens. Under the GDPR, consumers are granted more rights over their personal data, such as the right to have their data permanently deleted, the right to be informed about the collection, use, and sale of their personal data, the right to correct their information, the right to restrict the processing of their personal information, and the right to object to the collection of personal data. The GDPR also holds businesses that collect personal information accountable for the proper processing, storage, and security and access to personal data, as well as mandates that these companies that collect personal data obtain informed consent from the users over the collection of personal information [29].

Present and Future Issues Relating to Personal Data

In my research I identified a few different issues related to personal data that may become more prominent in the future as the field continues to grow and develop. These issues include the use and regulation of facial recognition in areas such as law enforcement, the use of DeepFake technology to spread misinformation, and the collection of DNA and genetic information for genealogy and medical uses. These are issues that are still being figured out as they continue to further our understanding of privacy as well as what falls under the definition of personal data.

Facial recognition is a tool to match faces to people in photos. This is generally done through the use of a machine learning algorithm, where a computer learns to distinguish different faces from each other. Facial recognition was used on Facebook to recognize users in photos posted by other users, but made the setting opt-in rather than on by default in 2019 [30]. Facebook also uses machine learning image recognition algorithms to determine what is in photos that users upload and provide captions of these photos for users that may be visually impaired [31]. In order to develop these machine learning algorithms, a large dataset of images and photos is needed. When users agree to the Terms of Service of social media apps, they often allow the company that owns the application or service to use the content that they upload to develop and refine their machine learning algorithms [4]. However, images uploaded to these social media apps are often viewable by the public as well. This has allowed other parties, such as Clearview AI to scrape, or go to individual user profiles and download, the images from sites and develop their own algorithms using this data that users did not intend to give to these third-party companies.

Clearview AI makes a software product that allows law enforcement to use facial recognition to find criminals. Their product has been used to stop shoplifting, identity theft, murder and more, but the images that it uses has been scraped from YouTube, Facebook, Instagram, and other websites [32]. This tool has been proven useful for stopping criminals but it has been a taboo practice for larger tech companies that have had the capabilities to do this, as companies such as Google believed that it would erode and, according to Google CEO Eric Schmidt in 2011, be used both for good and, "in a very bad way," [33]. Citing privacy concerns, some legislators have taken steps to ban the use of facial recognition, such as the city of San Francisco did in 2019, but the use and misuse of facial recognition is still a developing issue [34].

Another issue that continues to evolve is is a tool that has the potential to be used to spread misinformation. Another product of machine learning and artificial intelligence, Deepfake videos are videos that have been doctored by artificial intelligence to make it look and sound like someone has said or done something that they have not by manipulating or generating misleading video content. Especially prevalent on the social news site Reddit, the technology has been used to create pornographic videos of celebrities, including former first lady Michelle Obama and Olivia Wilde [35]. This technology has caught the attention of federal lawmakers, who are already working on combating fake news on social media sites. Senator Mark Warner of the Senate Intelligence Committee was quoted saying, "I'm much more worried about what could come next – could bad actors target kids with fake videos from the people they trust?" [36]. This technology is still evolving but because it currently requires a lot of video and audio data, as well as a lot of computer processing power to create convincing videos, most Deepfakes are only of public figures such as celebrities and potentially politicians, where there is an abundance of video and audio footage of these people.

Finally there is the topic of the collection of genetic information from companies such as 23andMe and AncestryDNA that collect DNA samples from their clients, analyze their genes, and give them information on their genes and health information. These companies claim that the genetic data that they collect and research is useful in the medical field, to help understand and treat genetic diseases [37]. However, genetic information is closely shared with relatives, meaning that if one family member consents to use one of these services, it may be possible to connect this information to other family members. The infamous Golden State Killer, a serial killer from California, was found and apprehended after his genetic information matched with a relative in the database GEDmatch, which has raised privacy concerns from researchers in genealogy and forensic science [38]. As DNA is shared with others, but also unique to individuals, it becomes difficult to categorize, and other potential uses of this technology in the future are hard to predict.

My Website

The intention for my capstone was to create a website on which I could share the findings of my research with users of social media. It was my intention that the website look and feel like a social media website so that the target audience would relate to it. This website was written in React and used the Material-UI style library to style the individual components. I chose to name the site MyFace, as a portmanteau of MySpace and Facebook, with a nod to the fact that the contents of the site deal with the use of personal information. On the site I wrote a number of different posts that explained the different areas of my research and how it related to how they may use social media sites. I also created a page which lists all of my references and citations so the user is able to learn more on their own. Additionally, I created a page which outlines what sort of legal protections and controls over personal data that a user may have, depending on the laws and regulations of where they live. Finally, I made a page that showed some of the information that sites are able to track from their visitors, without them even having to log in. This website is viewable at: https://nickrodgers42.github.io/capstone-react/. For best results, the use of the Google Chrome browser is recommended.

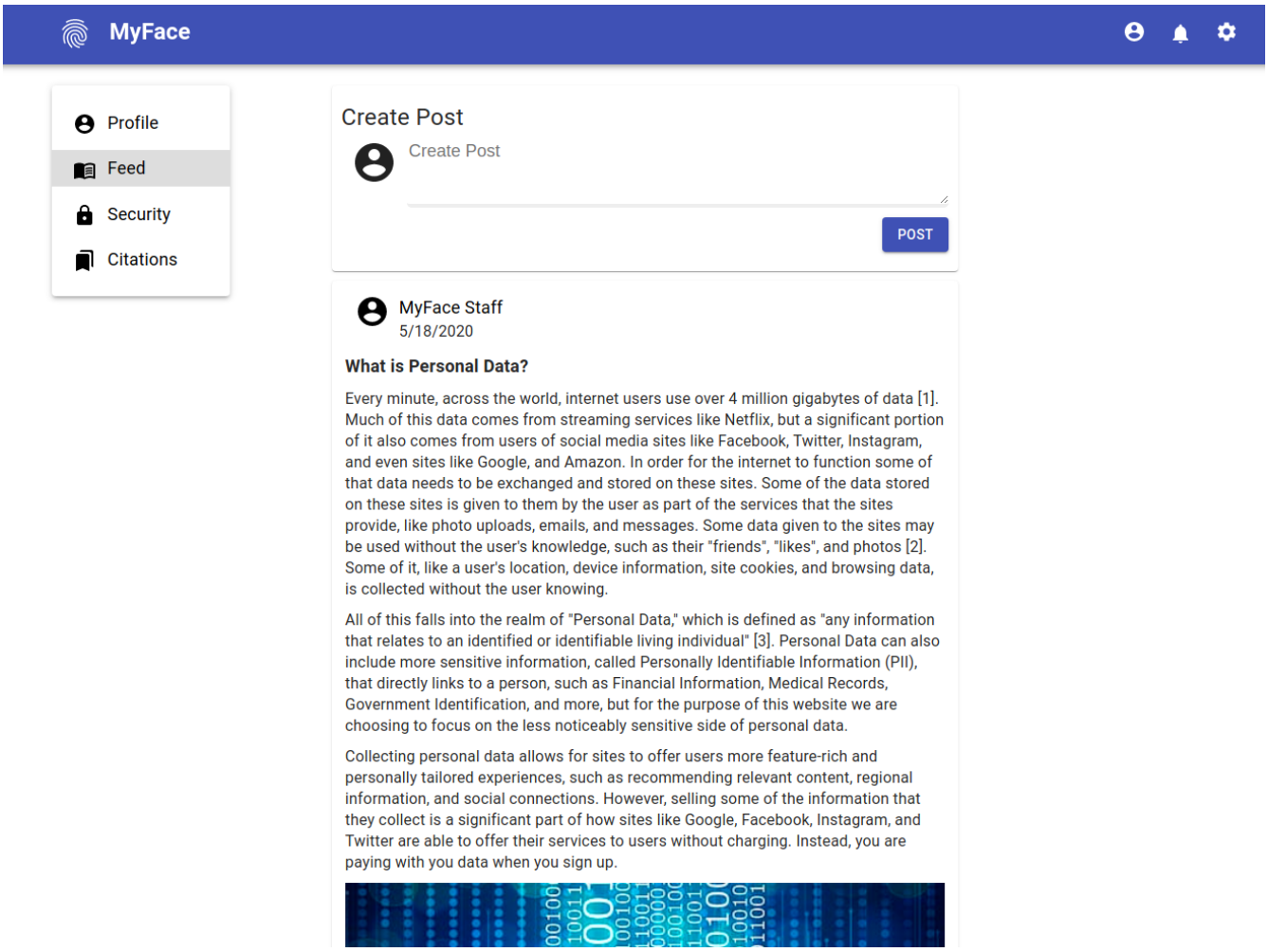

Figure 1: The homepage that the user sees when opening the site. This page contains posts written about different topics relating to personal data

I. Homepage - Newsfeed

The first page that a user sees on this website is the home page. This page contains nine different “posts” that I have written which detail the different areas of my research. These posts appear to be written by the “MyFace Staff” in order to to make it clear that the content of the site is coming from the creator of the site itself. See Figure 1 for a screenshot of this page. When a user opens the site for the first time, a dialogue box pops open, prompting the user to enable the use of cookies on the site. No data is actually transmitted to another device, everything is just stored locally on the user’s machine. The different posts that I have written on this page are as follows:

- What is Personal Data? - This post gives a definition of what personal data is and introduces the user to the different types of personal data that exist. Additionally it gives a bit of information to the user of why it is useful for social media sites and other tech companies to collect this type of data.

- I Agree to the Terms and Conditions – This post explains what the Terms of Service on websites are and why they are important. Additionally, it gives information on why the Terms of Service are usually very complicated and how it is useful for users to be more aware of what they are signing up for by agreeing to the Terms of Service.
- Why you can't put your phone down – This post explains how social media uses machine learning algorithms to generate profiles of users. Additionally, this post discusses how these algorithms suggest content to users and learn from user behavior to suggest better and more relevant content.
- You're being followed, even off of social media – This post discusses how social media sites use tracking cookies to log user behavior on third-party applications and websites in order to suggest more relevant advertising and collect more personal data. This post also briefly introduces some of the rights that users have with respect to the use of cookies.
- What the advertiser sees and how much your data is worth – This post discusses how ad providers such as Facebook and Google aggregate the data that they collect about their users before it is shown to the advertisers, and what sort of benefits advertisers gain from the analytic tools these platforms provide. This post additionally discusses how targeted advertising works and how this is beneficial for both the advertisers and the users, such as by showing ads based on interest and location. Finally, this post discusses how much money these sites make off of user data.
- A leak in the Data Stream: Personal Data and #FakeNews – This post discusses the various ways that personal information can be misused by bad actors. This section includes information on why false information spreads quicker than truthful information, and notable examples of when this has happened in the past, such as during the 2016 presidential election.
- Is there any upside to all this? - This post discusses the various positive benefits gained from sites collecting user data. These benefits include examples of sites creating better experiences for their users and developing better technology from this data.
- Can I get some Privacy? - This post explains actions that users can take to limit the personal data that businesses are able to collect about them. Additionally this post goes into more details about what legal protections exist for users from laws like the CCPA and the GDPR.
- Looking towards the Future – This post discusses the different personal data issues that are currently developing, and discusses how they are shaping modern understanding of personal data. These issues include using facial recognition in law enforcement, Deepfake videos, and the collection of genetic information.

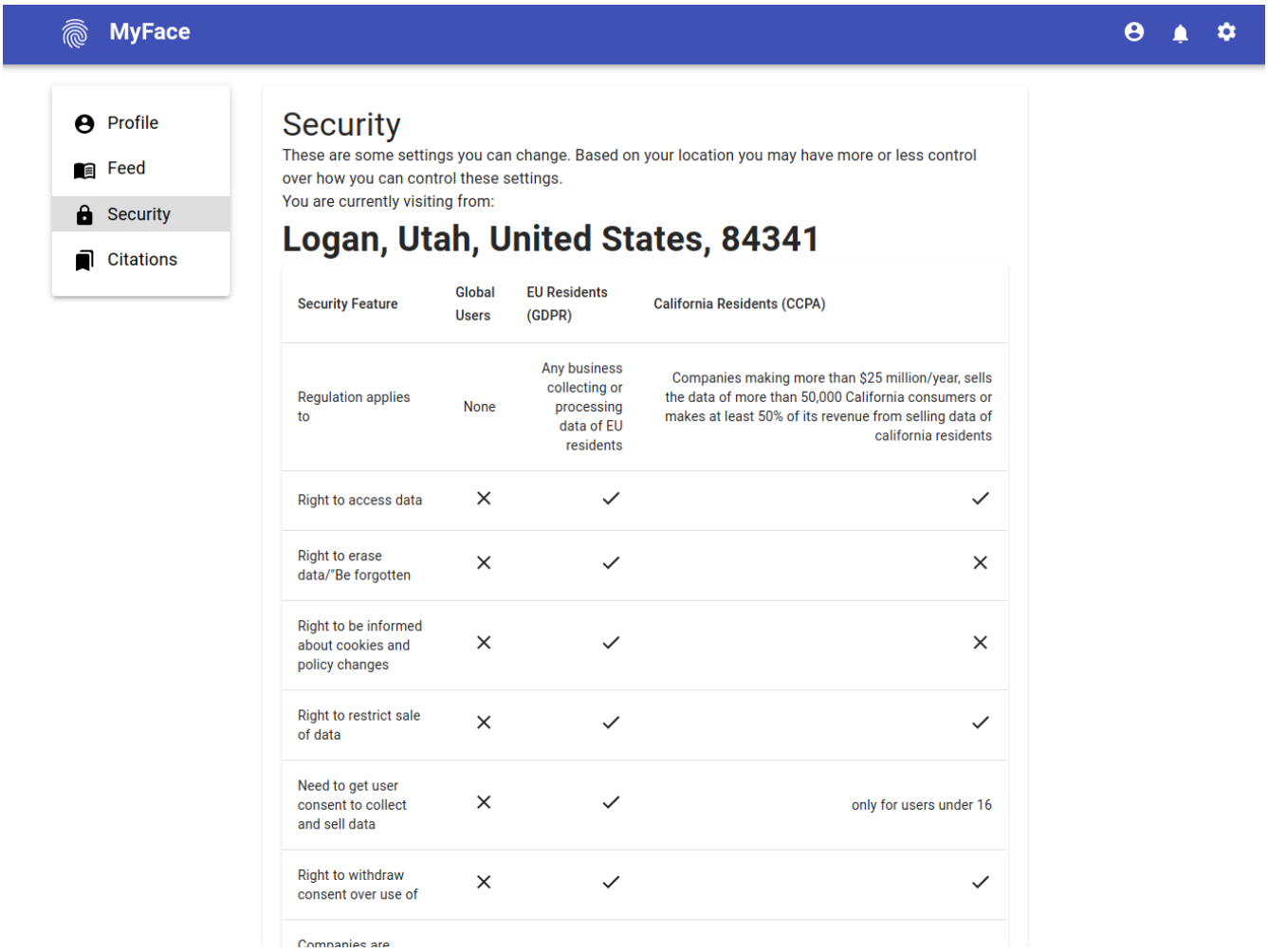

Figure 2: The security page shows the different rights that users have in relation to their personal data in both the EU and California

II. Security

This page shows a comparison between global data protection regulation, the European Union's GDPR, and California's CCPA. When a user visits this page, an API call is made that uses the IP address of the user to determine an approximate location of where the user is visiting from, and display this information to the user, letting the user know that their rights to personal data are based on their location. To see a screenshot of this page, see Figure 2.

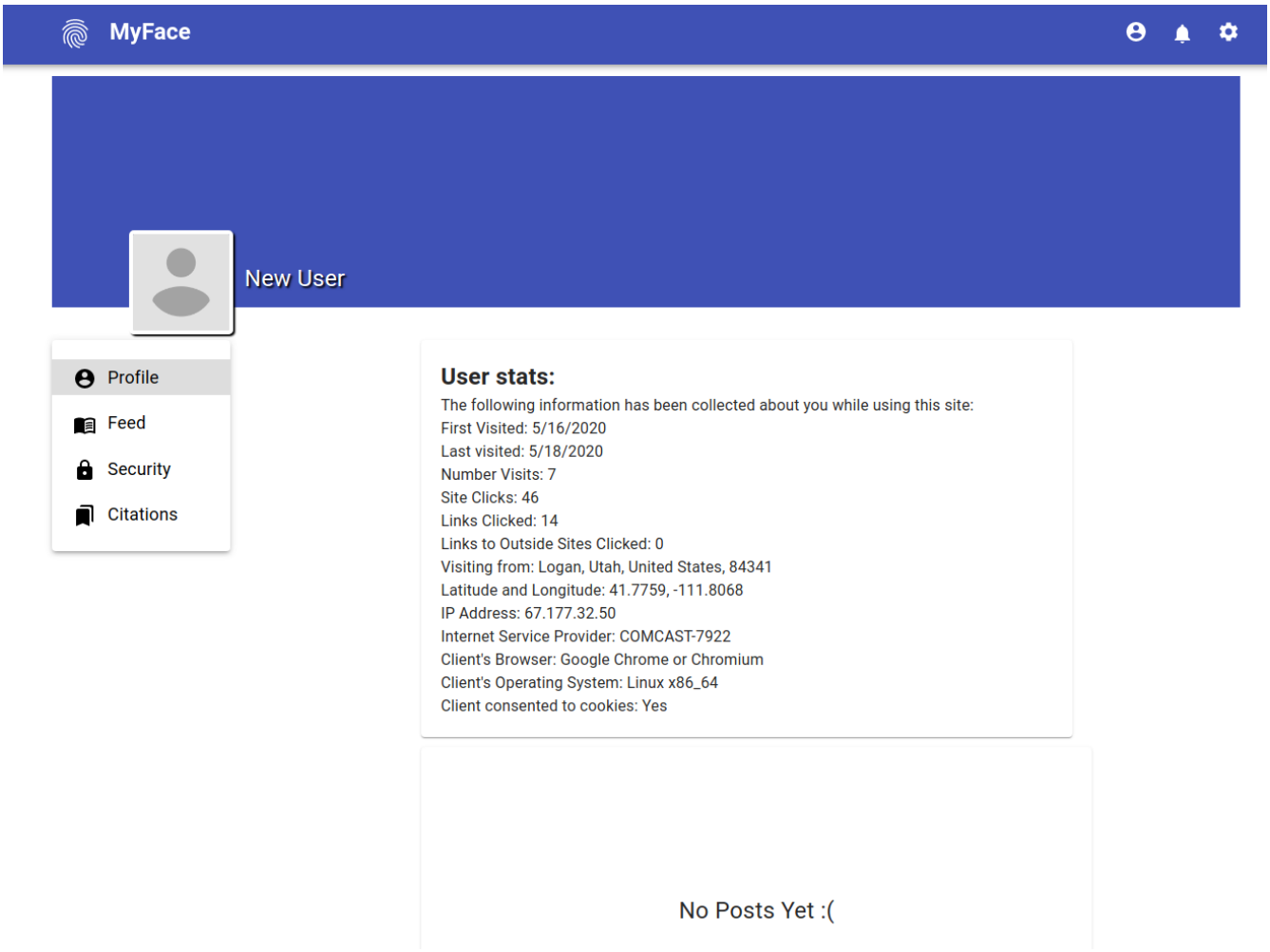

Figure 3: The user profile page shows the user some of the information that sites can learn about their visitors, even without explicit permission

III. Profile

This page is navigable through either clicking on the username on one of the posts, the “Profile” on the sidebar, or the account icon in the navigation toolbar. Here a user can see different statistics that have been collected about them, such as the first time the user visited the site, the last time the user visited the site, information obtained from the user’s IP address, what type of web browser and operating system the user is running on, and how many links the user has clicked on. The goal of this page is to show the user some of the information that a site can collect about them without their knowledge. To see a screenshot of the page see Figure 3.

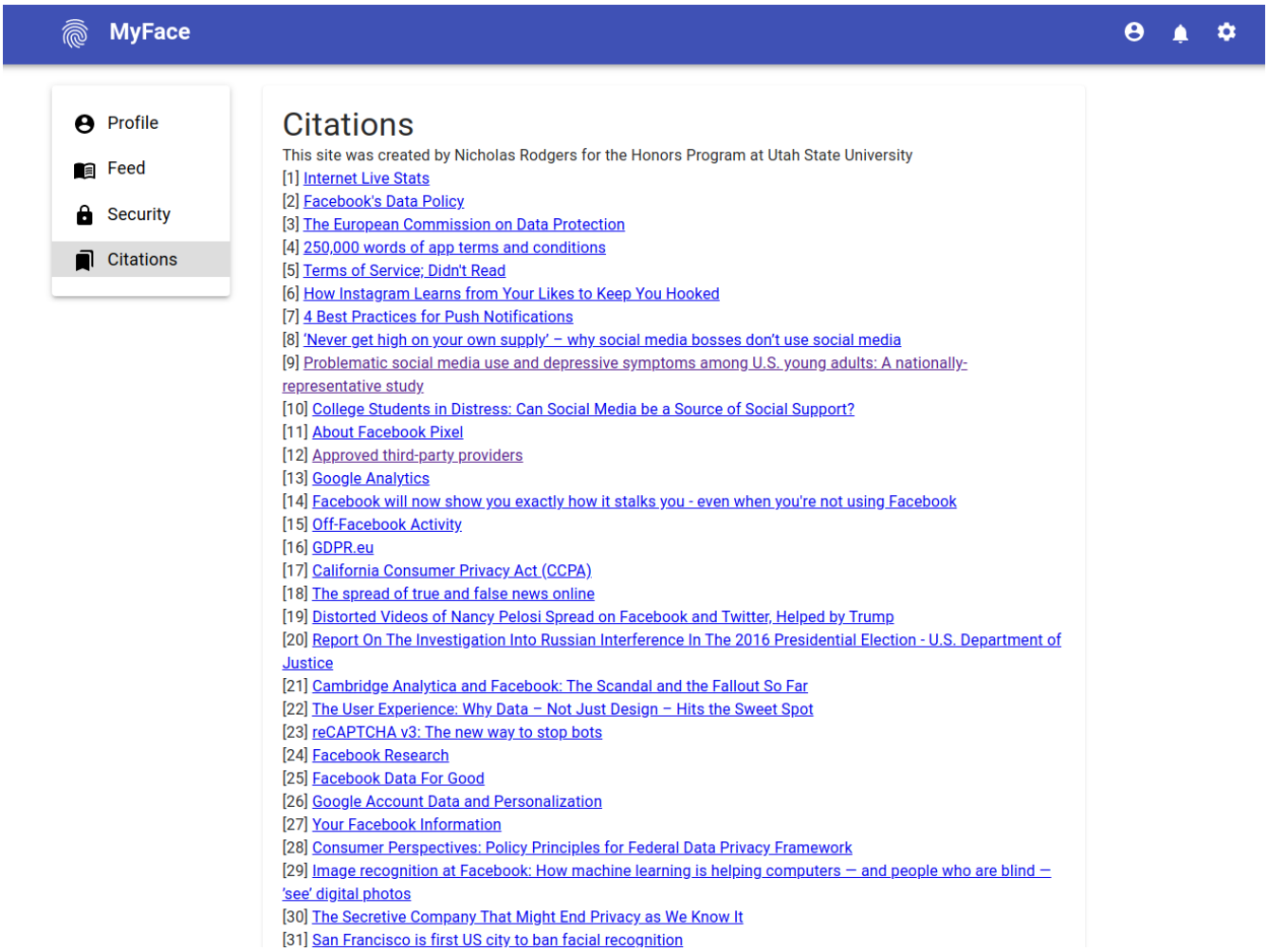

Figure 4: The citations page provides links to resources I used to write the posts allowing site visitors to continue to learn about topics they are interested in

IV. Citations

This page includes a list of citations that are marked throughout the home page, with links to the sources in case a user would like to learn more about any particular issue. To see a screenshot of this page, see Figure 4.

Conclusion

The field of personal data is quite extensive and there are many different factors that influence the creation and use of personal data. Personal data is necessary to power online interactions and is one of the factors that contributed to the wealth of technology that we have today. At the same time there are risks associated with the collection of personal data, and often consumers are not aware of what personal data is, how it gets collected, or how it is used. Personal data powers online connections, but it also has been used to create systems which are

able to spread false information very quickly. For many social media users there are no rules regulating the use of their personal data. It is only recently that that some U.S. states and the European Union have moved to create laws giving users more rights towards managing this data and held the businesses that collect this data responsible for ensuring that it is used properly. As technology continues to advance, there continue to be new issues related to personal data, and there only seem to be more issues on the horizon.

The site that I have created for my capstone project demonstrates both what I have learned through this research project, but additionally what I have learned during my undergraduate career in the Computer Science program at Utah State University. This site showcases my programming, my research, and my communication.

Reflection

Word Count: 1,090

Competing this capstone project for the honors program tied all of my experiences from my undergraduate career together in a way that I believe few other projects could. My two main areas of study, my major in Computer Science, and my minor in Spanish, are very separate from one another and it is not very often that they overlap. While this project did not include the use of the Spanish language, I believe that the skills that I have developed in both fields of study helped contribute to the final product. Additionally, taking on this capstone project gave me a broader perspective of a not often discussed are of the career field that I am going to be working in after my time at Utah State, allowing me to consider this topic of personal data not just from a programmer's point of view, but from multiple angles and identities. Trying to understand and be able to communicate what I have learned about such a broad topic proved challenging but worth the effort.

Throughout my time in the Computer Science program at Utah State, I have been aware that I have been developing my programming skills and gaining knowledge about this field that will be useful to me in my future career. I have had a job and an internship both in the web development field that have allowed me to learn how to apply the sort of programing skills that I have been taught outside of the classroom. These types of skills are readily apparent and easy to measure progress in this area. Another set of skills that were essential for this project but that I less aware that I was developing came from what I learned through the classes that I took for my Spanish minor and Honors courses. More than just learning about Spanish language and culture in my minor, by taking these classes I was developing both my communication skills as well as developing a more global perspective from which to consider issues. Both of these skills were only amplified by the Honors courses that I took at Utah State, where I learned how to interact with and discuss many other topics from outside of my major. If the Honors program has taught me anything during my time at Utah State University, it is how to develop my understanding of the little details while keeping a bigger picture in mind.

Deciding on a topic for my capstone project proved difficult, as it was only into the fall semester of my senior year that I felt as though I had the technical skills and understanding of the research being done in computer science to have any sort of understanding on how to structure a research project in this field. However, I am very pleased that this challenge led me to the topic that I chose to focus on for my capstone project, understanding how personal data is used. This topic required admittedly less high-level technical knowledge about a field of study within computer science, but instead a more open minded and creative perspective. This topic is important for me because as I start to work in Computer Science, I know that it is important that I keep in mind the larger context and consequences of the work that I do. Modern technology has been extremely beneficial for many, producing many of the tools necessary to make a better world, but at the same time it is not without drawbacks, risks, and moral dilemmas. It is important that those that are developing this technology take time to consider both the positive and potential negative impacts that their creations can have, and how to respond to new developments as they occur. Additionally, I believe that it is important that the general users of

these platforms that collect and use personal data are provided the opportunity to learn about these issues, as it will lead to both the companies that manage the sites, as well as the users, to make more informed decisions about their personal data, such as leading to more protections for users through legislation or more transparency in how this data is collected and handled.

One side effect of completing this project that I was unaware of was how I would grow and develop my understanding of myself as a professional. I am not going to pretend that I was great about meeting deadlines or that the final product of this capstone had all of the features that I wanted to incorporate, but I learned a lot about what I am capable of. I am thankful for Dr. Watson's support in this project, and his encouragement of the decisions that I made as the project progressed. Learning how to manage my time between programming the website, completing my schoolwork for my senior year, doing the necessary research, and just managing daily life in the face of the COVID-19 pandemic turned out to be challenging. I was forced to make decisions about what sort of features and content I wanted to include in the final product while maintaining its integrity in order to be able to deliver a final project that I was still satisfied with. Additionally during this project, my research led me to learn quite a bit from many different areas that I was unaware of going into this project. Doing so made me a better researcher and more knowledgeable about the topic, as I had to continually figure out how to keep the research relevant to the project as well as ensure that the information that I was collecting was accurate, and that it did not stray too far away from my main goals.

The Computer Science program at Utah State University has given me the knowledge and tools that I need to be successful in my future career as a programmer. The classes that I have taken for my Spanish minor have taught me a lot about the Spanish language and the culture and history of Spanish-speaking countries, but it has also developed my communication skills and taught me how to think with a global perspective. The Honors program and especially this capstone have been able to tie together my learning throughout my undergraduate career and make me a more knowledgeable and open-minded person. The topic of personal data that I chose for my capstone project is an important area of study to keep in mind for my future career, but I believe the skills that it has taught me will be even more useful.

Works Cited

[1] InternetLiveStats. “One Second” Internet: https://www.internetlivestats.com/one-second/, 2020 [May 17, 2020].

[2] European Commission. “What is personal data” Internet: https://ec.europa.eu/info/law/law-topic/data-protection/reform/what-personal-data_en April 27, 2016 [May 17, 2020]

[3] Ø. Kalestad “250,000 words of app terms and conditions” Internet: https://www.forbrukerradet.no/side/250000-words-of-app-terms-and-conditions/ May 24, 2016 [May 17, 2020]

[4] Facebook “Data Policy” Internet: https://www.facebook.com/policy.php April 19, 2018 [May 17, 2020]

[5] Terms of Service; Didn’t Read, “About” Internet: https://tosdr.org/about.html 2020 [May 17, 2020]

[6] Pew Research Center, “Social Media Use in 2018” Internet: https://www.pewresearch.org/internet/2018/03/01/social-media-use-in-2018/ March 1, 2018 [May 17, 2020]

[7] Facebook Artificial Intelligence. “Powered by AI: Instagram’s Explore recommender system” Internet: https://ai.facebook.com/blog/powered-by-ai-instagrams-explore-recommender-system/ Nov. 25, 2019 [May 17, 2020]

[8] Facebook Research, “Research Areas” Internet: https://research.fb.com/research-areas/ 2020 [May 18, 2020]

[9] Facebook for Business. “Facebook Ads” Internet: https://www.facebook.com/business/ads 2020 [May 17, 2020]

[10] Google Ads “Customize your YouTube Media Strategy” Internet: https://ads.google.com/home/tools/reach-planner/ 2020 [May 17, 2020]

[11] Facebook, “About Facebook Pixel.” Internet: https://www.facebook.com/business/help/742478679120153?id=1205376682832142&helpref=faq_content 2020 [May 17, 2020]

[12] E. Culliford. “Facebook gets rid fo ‘pesudoscience’ ad-targeting category” Internet: https://www.reuters.com/article/us-health-coronavirus-facebook-ads/facebook-gets-rid-of-pseudoscience-ad-targeting-category-idUSKCN2253CC April 23, 2020 [May 17, 2020]

[13] M. Isaac, C. Kang, “Facebook Says it Won’t Back Down from Allowing Lies in Political ads.” Internet: https://www.nytimes.com/2020/01/09/technology/facebook-political-ads-lies.html Jan. 9, 2020 [May 17, 2020]

[14] Facebook, “Facebook Reports Fourth Quarter and Full Year 2019 Results.” Internet: https://investor.fb.com/investor-news/press-release-details/2020/Facebook-Reports-Fourth-Quarter-and-Full-Year-2019-Results/default.aspx Jan. 29, 2020 [May 17, 2020]

[15] Alphabet, “Alphabet Announces Fourth Quarter and Fiscal year 2019 Results.” Internet: https://abc.xyz/investor/static/pdf/2019Q4_alphabet_earnings_release.pdf?cache=05bd9fe Feb. 3 2020, [May 17, 2020]

[16] Federal Trade Commission, “Equifax Data Breach Settlement.” Internet: https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement Jan. 2020 [May 17, 2020]

[17] N. Confessore, “Cambridge Analytica and Facebook: The Scandal and the Fallout So Far.” Internet: https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html April 4, 2018 [May 17, 2020]

[18] R. Mueller, III (March 2019) “Report On the Investigation Into Russian Itnerference in the 2016 Presidential Election” [On-line]. I. Available: https://www.justice.gov/storage/report.pdf [May 18, 2020]

[19] S. Vosoughi, D. Roy, S. Arai. (2018, March) “The spread of true and false news online” *Science*. [On-line]. 359(6380), pp. 1146-1151. Available: https://science.sciencemag.org/content/359/6380/1146 [May 18, 2020]

[20] The United States Department of Justice, “Privacy Act of 1974,” Internet: https://www.justice.gov/opcl/privacy-act-1974 Jan. 15, 2020 [May 18, 2020]

[21] United States District Court for the District of Columbia, “Ciox Health, LLC, v Alax Azar, et al.,” [On-line] Case 18-CV-0040, Available: https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2018cv0040-51 [May 18, 2020]

[22] Federal Trade Commission, “Gramm-Leach-Bliley Act,” Internet: https://www.ftc.gov/tips-advice/business-center/privacy-and-security/gramm-leach-bliley-act Sept. 4 2015, [May 18, 2020]

[23] Federal Trade Commission, (Jan. 17, 2013) “Children’s Online Privacy Protection Rule (“COPPA”),” [On-line] 78(12) Available: https://www.ftc.gov/system/files/2012-31341.pdf [May 18, 2020]

[24] Commonwealth of Massachusetts, “Bill S.120,” Internet: https://www.legis.nd.gov/assembly/66-2019/bill-actions/ba1485.html Feb. 13, 2020 [May 18, 2020]

[25] The New York State Senate (May 9, 2019) “Senate Bill S5642” Internet: https://www.nysenate.gov/legislation/bills/2019/s5642 [May 18, 2020]

[26] Maryland General Assembly, “Online Consumer Protection Act,” Internet: http://mgaleg.maryland.gov/mgawebsite/legislation/details/sb0613?ys=2019rs 2019 [May 18, 2020]

[27] North Dakota Legislative Branch, “Bill Actions for HB 1485,” Internet: “https://www.legis.nd.gov/assembly/66-2019/bill-actions/ba1485.html” April 1, 2019 [May 18, 2020]

[28] California Legislative Information, “California Consumer Privacy Act of 2018,” [On-line] Title 1.81.5(1798.100-1798.199). Available: http://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=CIV&division=3.&title=1.81.5.&part=4.&chapter=&article= [May 18, 2020]

[29] Official Journal of the European Union, (April 27, 2016). “Regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)” [On-line] I(119/1). Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [May 18, 2020]

[30] Facebook, “An Update About Face Recognition on Facebook.” Internet: https://about.fb.com/news/2019/09/update-face-recognition/ Sept. 3, 2019 [May 18, 2020]

[31] Facebook Engineering, “Under the hood: building accessibility tools for the visually impaired on Facebook,” Internet: https://engineering.fb.com/ios/under-the-hood-building-accessibility-tools-for-the-visually-impaired-on-facebook/ April 4, 2016 [May 18, 2020]

[32] K. Hill, “The Secretive Company That Might End Privacy as We Know It.” Internet: https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html Feb. 10, 2020 [May 18, 2020]

[33] B. Bosker, “Facial Recognition: The One Technology Google is Holding Back.” Internet https://www.huffpost.com/entry/facial-recognition-google_n_869583 June 1, 2011 [May 18, 2020]

[34] D. Lee, “San Francisco is first US city to ban facial recognition.” Internet: https://www.bbc.com/news/technology-48276660 May 15, 2019 [May 18, 2020]

[35] K. Roose, “Here Come the Fake Videos, Too.” Internet: https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html March 4, 2018 [May 18, 2020]

[36] A. Breland, “Lawmakers worry about the rise of fake video technology.” Internet: https://thehill.com/policy/technology/374320-lawmakers-worry-about-rise-of-fake-video-technology Feb. 19, 2018 [May 18, 2020]

[37] 23andMe, “Research,” Internet: https://www.23andme.com/research/ 2020, [May 18, 2020]

[38] G. Kolata, H. Murphy, “The Golden State Killer Is Tracked Through a thicket of DNA, and Experts Shudder,” Internet: https://www.nytimes.com/2018/04/27/health/dna-privacy-golden-state-killer-genealogy.html April 27, 2018, [May 18, 2020]

Biography

Nicholas Scott Rodgers was born in West Jordan, Utah and raised in Logan, Utah. After attending high school at Oak Ridge High School in El Dorado Hills, California, Nicholas returned to Logan to pursue a Bachelor's Degree in Computer Science at Utah State University. Fascinated by languages, both human and computer-based, Nicholas has also pursued and completed minors in Spanish and Mathematics during his time at USU. Additionally, Nicholas has held positions working as a web developer's assistant at the Space Dynamics Lab in North Logan, Utah in addition to working as a software development intern for Pluralsight in South Jordan, Utah. He has also studied abroad in Logroño, Spain. Following graduation, Nicholas will be moving to New York City, New York to work as a software development engineer for Amazon.