THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF INFORMATION SCIENCES AND TECHNOLOGY

Privacy Policy Readability and Teenagers: A Longitudinal Analysis of Privacy Policy Readability and the Impact on the Privacy Governance Space

KATHRYN FRANKENBERG
SPRING 2022

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree in Security and Risk Analysis
with honors in Security and Risk Analysis

Reviewed and approved* by the following:

Shomir Wilson
Assistant Professor of Information Sciences and Technology
Thesis Supervisor

Nicklaus A. Giacobe
Assistant Teaching Professor of Information Sciences and Technology
Honors Advisor

* Electronic approvals are on file.

ABSTRACT

In recent years, government privacy regulations have been passed with more stringent enforcement to increase transparency for consumers. To be compliant with the new regulations, companies are required to have a privacy policy explaining how the company uses consumer data, what data they collect, what data they store, and 3rd party uses. In addition to the uses of data, three major privacy regulations, the CCPA, COPPA, and the GDPR, have varying provisions relating to the language complexity of the privacy policies. This research assessed if these provisions are met by comparing the readability scores of apps found in a 2018 study to their 2021 counterparts. The hypothesis is that the 2021 privacy policies are more complex than 2018 privacy policies, using readability scores to measure document complexity. This research found that the readability of the 2021 policies was more difficult than the readability of the 2018 policies in all recorded measures (Flesch-Kincaid Grade Level, Flesch-Kincaid Reading Ease, Gunning FOG, SMOG, word count). These results show the increasing complexity of the privacy policies, which is significant because as the policies get more complicated, less readers can be expected to understand the language of these policies. Further, knowing that users cannot be reasonably expected to understand privacy policies changes the purpose of the documents from promoting transparency to predominantly compliance with the laws.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I firstly want to thank Dr. Wilson and his guidance and help with writing this research paper. Being the first official research paper I have written, having the expertise of someone with experience with research and research in privacy was very impactful.

Secondly, I want to thank Dr. Giacobe for his help and guidance towards the end of the project. I am very grateful that he was able to fulfill the role of my honors advisor and secondary reader.

I also would like to to thank my friends for all the support over these last few years The memories we have created together and will create one day makes these next transition years a little less scary.

And finally, I want to thank my family for their continued support and guidance towards me over these last college years. Carrying the Penn State legacy, I'm glad for their consistent advice of making my thesis, a document culminating my experience here, something that I am proud of.

Chapter 1

Introduction

The increase of internet accessibility and popularity among younger audiences has led to new technologies being developed to increase user attention and retention. New updates and features promise increased functionality and personalization based on each user's preferences. However, the increased use of technology has led to the emerging data collection market, and a trend by companies to collect as much data on users as they can. To combat the spread of data collection and increase privacy for assuming and unassuming users, governments have started drafting and passing legislation targeting data collection and increasing user's ability to erase, opt-out, and know what information a company has collected on them. By giving the users more power of their specific data, individuals are given more autonomy and choice in the current internet age.

Parts of the privacy efforts require companies to have privacy policies, which explain "the ways a party gathers, uses, discloses, and manages a customer or client's data" ("Privacy Policy," 2021). The purpose of these policies is to explicitly inform users about the behind-the-scenes by increasing transparency to increase user trust and reduce feelings of being misled. However, these policies often act only as an end for organizations and businesses to be compliant with national privacy laws, instead of them being drafted as an act of good faith between the data collector and those whose data is being collected. Further, there are issues of whether individuals read the policies they are consenting to. This raises concern over whom the policies are written for, which is pertinent to the discussion of policy effectiveness, and the approach of enforcing

privacy through company notices to consumers. This problem becomes more complicated when including the risks posed to at-risk users, specifically teenage users and children.

The three regulations this research focuses on is the CCPA, COPPA, and GDPR. What has been interpreted as the understanding requirements is listed below in Table 1.

Table 1. Regulation Understanding Requirements

	Jurisdiction	Requirement
CCPA	California State Law	"Provide the specific pieces of personal information obtained from the consumer in a format that is easily understandable to the average consumer, and to the extent technically feasible, in a structured, commonly used, machine-readable format that may also be transmitted to another entity at the consumer's request without hindrance." [3BIII] (*California Consumer Privacy Act of 2018*, n.d.)
COPPA	US Federal Law	"It shall be the obligation of the operator to provide notice and obtain verifiable parental consent prior to collecting, using, or disclosing personal information from children. Such notice must be clearly and understandably written, complete, and must contain no unrelated, confusing, or contradictory materials." [312.4] (*Children's Online Privacy Protection Rule ("COPPA")*, 2013)
GDPR	EU Regulation	"The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used. That principle concerns, in particular, information to the data subjects on the identity of the controller and the purposes of the processing and further information to ensure fair and transparent processing in respect of the natural persons

		concerned and their right to obtain confirmation and communication of personal data concerning them which are being processed" [39] (*General Data Protection Regulation (GDPR) – Official Legal Text*, n.d.)

This thesis explores the relationships between the readability scores of privacy policies from 2018 to the readability of these same applications privacy policy readability scores in 2021 performed in the Das et. al. 2018 study "Privacy Policies for Apps Targeted Toward Youth: Descriptive Analysis of Readability". The second chapter of this thesis focuses on related research efforts, specifically the efforts to create systems that can annotate the privacy policies and legal text and legal responses to additional legislation and the subsequent created frameworks and. The third chapter highlights the methodology taken in this research and why the various metrics were recorded. Because this thesis is an extension of a 2018 study, the methodology is a replicated version of the initial study, to compare the 2021 and 2018 results. The fourth chapter mentions the results of this research and shows the comparison between this thesis and the initial studies results. The fifth chapter discusses the results, as well as addresses the motivations for this research, and reiterates the importance of transparency and enforcement between organizations, governments, and the people using these technologies.

Readability is the central variable that will be examined in this thesis. Using the same readability tool and data from the Das. et al. study, this research is a longitudinal analysis of readability scores for applications popular among children and teenagers. The online readability tool, Readability Test Tool, records the Flesch-Kincaid Reading Grade Level, Gunning FOG Index, Simplified Measure of Gobbledygook (SMOG) index (*Readable | Free Readability Test Tool*, n.d.). The Flesch-Kincaid RGL, Gunning FOG index, and SMOG index are three of the

most used readability assessment tools (Walsh & Volsko, 2008). The average reading level comes from averaging the three separate measures. The hypothesis for this paper is that in response to new and thorough enforcement of privacy regulations, the selected policies' readability scores have increased in reading level complexity due to additional legislative provisions on data handling and collection practices. This hypothesis would be reflected in a reduction in the Flesh-Kincaid Reading Ease score, and an increase in the Flesch-Kincaid Reading Grade Level, Gunning FOG Index, and SMOG Index scores.

In the past three years, numerous revisions and lawsuits have come out of businesses who are not compliant with federal and international privacy law and regulations (*Privacy Litigation*, n.d.). Because regulations contain a clause referencing an expectation of understanding and a clarity standard, app developers may have reduced the readability scores (reducing the complexity) for their application's privacy policies (*Children's Online Privacy Protection Rule ("COPPA")*, 2013; *General Data Protection Regulation (GDPR) – Official Legal Text*, n.d.). Their responses to increased scrutiny and public pressure generates a research question regarding the intent of the privacy policies, specifically whether the purpose is for compliance or as a way of promoting transparency.

Chapter 2

Related Work

Children on the internet have posed many safety and privacy risks, both from the perspective of company violations and exploitation from other users. To combat the risk of exploitation of children on the internet, COPPA was drafted to give parents and guardians the power to determine what data is collected from their child's online activity (*Children's Online Privacy Protection Rule ("COPPA")*, 2013). There are two main subsections into this research of privacy policies and their enforcement: efforts to annotate privacy policies and the role of legislation and the frameworks they create. The annotation effort happens at both a small scale, like the OPP-115 Corpus, a collection of 115 privacy policies, and at a large scale effort, like the PrivaSeer Corpus, which is a collection of 1,005,380 privacy policies (Wilson et al., 2016; Srinath et al., 2021). The law focus centers on the laws and regulations that have been passed, the subsequent frameworks these laws created, and the reactions and responses to them and their enforcement. Because this research is centered on the regulatory environment surrounding teenage users younger than 18, this thesis will focus on laws governing the age of digital consent online and children on the internet.

2.1 Annotation Effort

With the increased emphasis on individual privacy placed on organizations, governments, and users, researchers have been turning to ways to figure out how to reduce the complexity of privacy policies while also helping to ensure that the documents are compliant with varying privacy laws. A major issue with monitoring the effectiveness of privacy policies in terms of user

understanding and from the compliance perspective is that there is not a standardized format for policies. Acting as a way to "bridge the gap" between the needs of users, goals of regulators, and the position companies are in, new natural language processing (NLP) technologies and research are arising to help serve the needs of the three bodies (Sadeh et al., 2013). Another highlight of these datasets of privacy policies is their ability to help with scaling the gaps using NLP technology and person annotators, to partly automate privacy policies, predict their structures, and obtain answers to question tasks (Wilson et al., 2016; Srinath et al., 2021).

A complication that has been found in efforts to annotate privacy policies automatically using NLP technologies without incorporating human annotators is identifying the vagueness and complexity of the privacy policies (Liu et al., 2018). Liu et al. attributes the complexity of the documents to the requirement that the documents need to be comprehensive and accurate, and therefore generalize and overly sophisticate the statements (2018). By developing a deep neural network for modeling language, researchers were developing an algorithm to identify phrases that can signal vagueness, with the hope of future research focusing on vagueness prediction in assisting legal counsels to "clarify the privacy text." (Liu et al., 2018).

Though research efforts in classifying vague terms can help reducing the overall complexity of the document, there are also complications when it comes to using NLP technologies to identify data practices within privacy policies without the use of human annotators. To mitigate this, the OPP-115 privacy corpus identified 10 data practice categories in their privacy policy annotation scheme. With these 10 categories, researchers show that they can use partly automated NLP technology to identify what category a data practice belongs in by the categories attributed.

Research efforts, like the OPP-115, that identify regulation-specific annotation schemes can help regulators and organizations make sure that their policies are compliant with the governing privacy regulation for their jurisdiction. Encouraging regulators to identify important areas in the regulation could help with the development of key categories that ensure organizations are meeting the transparency and sharing requirements with their privacy policies.

Though there are many factors into what constitutes compliant privacy, such as the product's industry, user location, user age, as well as the transparency component to how the data is used, these factors can be "mapped" via automated annotations for compliance to the regulation. However, this thesis is focused on the readability expectation of privacy policies. The condition of whether people read the privacy policies is not being tested, but, rather, whether organizations are writing their privacy policies in a way that can be understood by all the products users, from the youngest to the oldest.

2.2 Legislative Effort

In order to protect children from unwarranted data collection, the United States passed the first version of the Children Online Privacy Protection Rule (COPPA) (*Children's Online Privacy Protection Rule ("COPPA")*, 2013). COPPA protects children (defined as people under 13) by requiring parental consent for businesses to collect private information (*Children's Online Privacy Protection Rule ("COPPA")*, 2013). COPPA outlines several key requirements, which passed when it was passed in 1998, were not practices internet sites were engaging in. The practices are:

1. Post a privacy policy on the website

2. Provide notice directly to parents
3. Get parental consent
4. Allow parents to review personal information collected from their children
5. Allow parents to revoke their consent, and delete information collected from their children at the parent's request
6. Establish and maintain reasonable procedures to protect the confidentiality, security, and integrity of children's personal information
7. And, Not condition a child's participation in certain activities on collection of more personal information than is reasonably necessary (Federal Trade Commission Staff Report, 2002).

COPPA was initially draft to combat the targeting of children online for marketing purposes (Jargon, 2019). Because the law outlines protections for children, Congress needed to define the age of a child. Prior to drafting COPPA, governments had used the "Under 12" rule in regards to laws about marketing to children, which was supported by research showing children between 8 and 12 "could distinguish advertising from other content." (Jargon, 2019) When Facebook launched in 2004 and the iPhone was first released in 2007, the age of online consent remained at 13 (Jargon, 2019). 21 years after the implementation of COPPA, websites, apps, and social media platforms set the age of use at 13 in order to not go through the process of seeking parental consent (Jargon, 2019). When the GDPR was implemented, the regulation set the age to 16, the CCPA, a state law, followed suit.

The California Consumer Privacy Act (CCPA) of 2018 states that consumers between the age of 13 and 16 have "the right to opt-out" of data collection, and are a default opt-out state of data collection unless they or their legal guardians explicitly opt-in to data collection, extending

the COPPA protections to this population (*California Consumer Privacy Act of 2018*, n.d.) In 2018, the landmark European Union privacy regulation the General Data Protection Regulation (GDPR) changed the age of online digital consent to privacy policies to 16, adding that member states can reduce the age to a minimum of 13 (*General Data Protection Regulation (GDPR) – Official Legal Text*, n.d.). Though the CCPA, a state law, has extended protections for users between the ages of 13 and 16, there have been several attempts through congressional legislation to raise the federal age of digital consent from 13 to 16 in the United States (*Senators Markey and Hawley Introduce Bipartisan Legislation to Update Children's Online Privacy Rules | U.S. Senator Ed Markey of Massachusetts*, 2019). However, as of September 2021, the federal age of consent for data collection is still 13 in the United States. A further complication is that specific applications have policies governing the age of users, with most setting the age of consent at 13, but some terms and conditions stating that the users need to be 16 or even 18 to use the application.

All these regulations demonstrate that children younger than 13 qualify for increased protections under the law, through requiring parental consent for data collection and restricting the access of data of 3rd parties not integral to the site function. However, there is some controversy as to where that age should be when it comes to minors between the age of 13 and 16, as well as when software services and applications can set their own age requirement too.

Legislators, like Senator Markey, one of the drafters of COPPA, and privacy advocates urge parents to monitor their child's activity and limit the use of social media until the child is at least 13 but encourage children to wait until they are 16 to decide if they want to use certain platforms. Further, the process to verify a minor's age results in more data collection on the minor, furthering the complexity of the issue (Jargon, 2019).

The conflict on the age of consent for data collection in the United States is reflected in the controversy over privacy policies and their readability. These privacy policies are part of the legal requirements for businesses to engage in data collection, as well as are a push to encourage transparency between online business operations and their users. However, these policies are not always written with all users in mind. A study on 1,700 products in EdTech's privacy policies using several readability metrics, specifically Automated Readability Index (ARI), Coleman-Liau Index, Flesch-Kincaid Grade Level, Flesch-Kincaid Reading Ease, Gunning FOG Index, and SMOG Grade, found that the average grade-level of these policies is about that of a 10th grader (*It's Not You; Privacy Policies Are Difficult to Read*, 2018). Considering the implications of this and the notion that 7th and 8th graders are expected to read and consent to these policies regarding their own personal data collection, there is a gap in the accessibility of these policies and the effect of them.

Research repeatedly shows that privacy policies across various industries tend to be too difficult for the average reader to read, with most researchers recommending the reading level for guidelines across industries aim for the 8th grade level (Das et al., 2018; Krumay & Klar, 2020; Milne et al., 2006; Zhang et al., 2020). According to the Center for Plain Language, the average American has the equivalent reading level of a 7th or 8th grader, which is the benchmark the medical community uses for guidelines (Marchand, 2017). A 2003 study done by the National Center for Education Statistics found that the typical American reads between the 7th and 8th grade level, and the United States Department of Health and Human Services (USDHHS) stated that information can only be considered "easy to read" if written at or below the 6th grade level, with material between the 7th and 9th grade level being of "average difficulty", and above 9th grade being considered "difficult" (National Center for Education Statistics, 2003, as cited in

Walsh & Volsko, 2008; United States Dept. of Health and Human Services, 2000, as cited in Walsh & Volsko, 2008)

The Krumay and Klar study analyzed the privacy policies of 15 European countries post GDPR implementation to see if they meet the "clear and understandable language requirement" and "clear and plain language" requirement ((*General Data Protection Regulation (GDPR) – Official Legal Text*, n.d.; Krumay & Klar, 2020). Though only using the section of the privacy policy describing the 'cookie policy', researchers performed both a qualitative and quantitative analysis (Krumay & Klar, 2020). Creating a 5 point liker scale, they used 15 participants in 5 different grade levels identified in the Dale-Chal approach, which was not used in this thesis, but is as follows: 7–8, 9–10, 11–12, 13–15, and above 16 (Krumay & Klar, 2020). The 5 point liker scale used the following measures: length of text (short-long), length of sentences (short-long), complex sentences (some – many), unfamiliar words (some – many), and general readability (very easy – very hard) and documented reading speed (expressed in minutes per 1500 words (Krumay & Klar, 2020). When combining the qualitative results with the quantitative results, the researchers saw the general trend that the "text experienced by all participants as very easy or easy to read (ETR) showed average scores below 12. By contrast, text experienced as hard to read (HTR) by all participants showed average scores above 14."(Krumay & Klar, 2020). However, the researchers did not assess participants comprehension of the text.

This experimental study found that these policies were too complicated for most audience members and recommended companies aim for a 9^{th} or 10^{th} grade reading level. The study recommends an approach using a rules-defined list that incorporates words individuals are familiar with but more obscure or longer in length (the example they provide is email addresses)

(Krumay & Klar, 2020). It is also recommended that this list of words for privacy statements should be updated frequently (Krumay & Klar, 2020).

The Zhang et. al. study also noted readability scores being too high for the intended audience in a study, specifically focusing on the privacy policies of seven COVID-19 contact tracing apps (Zhang et al., 2020). This study included research on the readability of articles on COVID-19, the pandemic, and coronavirus, and it noted that increased text complexity could have an impact on people's confusion and frustration with coronavirus policies. Like the 2018 Das et. al. study's methodology, Zhang et. al. found the polices were too complicated and recommended that app developers should aim for a policy of about the 8th grade level, especially due to the sensitive nature of the health information and geolocation data.

The PrivaSeer Corpus, which contains "1,005,380 privacy policies from 995,475 different web domains", ran a readability analysis of their collected policies (Srinath et al., 2021). These researchers found that the averaged Flesch-Kincaid Grade Level score was 14.87, which translates to an average of 14.87 years of education in the U.S. to understand a privacy policy (Srinath et al., 2021).

A study on the readability of social media applications shared similar findings to the above study, with an average of 12.68 Flesch-Kincaid Grade Level score, with a standard deviation of 2.29 (Meiselwitz, 2013). When put into the context that 82% of teenagers and 50% of adults have social media accounts with many having accounts on multiple platforms, the high scores of readability for privacy policies used by many people daily reflects the issue of balancing transparency with both regulators and consumers (Meiselwitz, 2013).

With the increasing amount of research focused on automatically reading legal documents and sorting the policies into meeting criteria, there has been a push for regulators to

take a bigger stance in enforcing the frameworks and guidelines they created. However, with the prevalence websites and apps have on our daily lives, there is also an expectation that all readers should be able to understand what they are consenting to when they click ‘I agree” on a service’s privacy policy and data practices.

Chapter 3

Methods

This research is an expansion of the 2018 Das et. al. study on apps that targeted toward and used by teenagers. Pulling from the 3.1 million available applications on the Apple App Store and the Google Play Store, the researchers initially narrowed down the list to 1200 apps by pulling the top 300 paid apps on both the App Store and the Play Store, and the top 300 free apps on both stores. From this list of 1200 applications, the researchers reduce the number of apps by using the selective excluding criteria below:

1. encourage the use of money outside in-app purchases (eg, shopping, travel, or real-estate apps)
2. facilitate interaction with unknown people (eg, dating or ride-service apps)
3. are focused on tracking pregnancies or newborn development
4. serve as licensing keys that unlock premium features of other apps (only in the Google Play Store)
5. Shopping apps included apps related to specific stores or corporations (eg, Kohl's, Walmart, or Amazon), buy, and sell apps (eg, letgo or eBay), and coupon or discount apps (eg, Groupon)

This step generated 96 apps that were excluded from the list of 1200. From the list of the remaining 1104 applications, the researchers then chose the top ranked 30 apps in each of the 4 categories: Apple Free, Apple Paid, Google Play Free, and Google Play Paid. From this list of 120 applications, 21 were available on both the Apple store and Google Play store, thus only counted once. Of the remaining 99 apps left, 24 of the applications did not have privacy policies and 11 had the same privacy policy, as they were created by the same developer. This left a total

of 64 potential privacy policies, which is the basis for this 2021 study. Figure 1 outlines this selection criteria process.

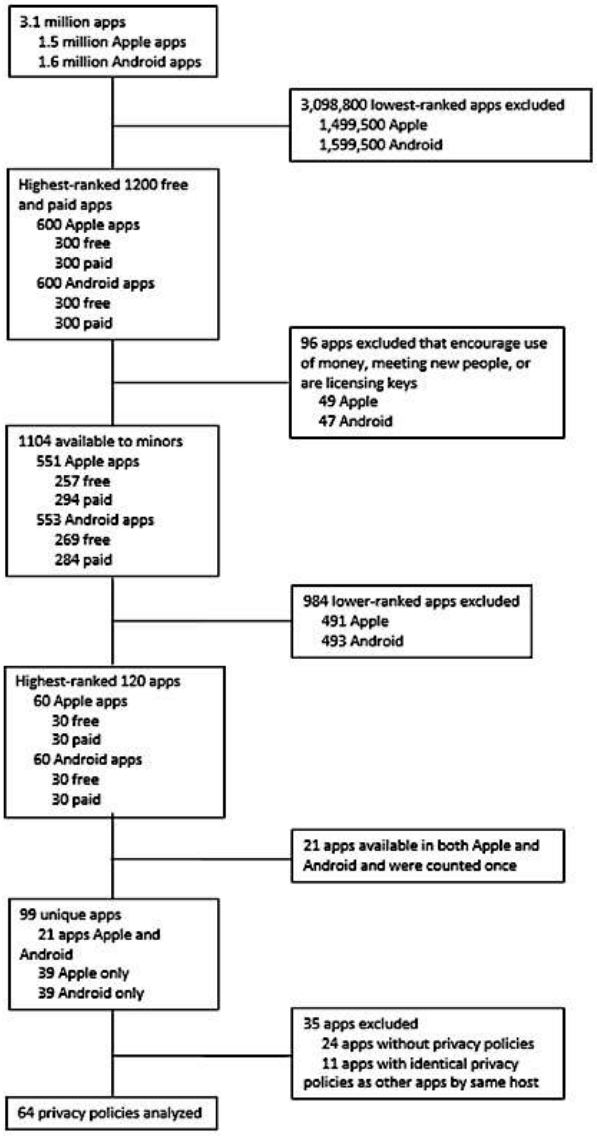

Figure 1. App Selection Criteria (Das et al., 2018)

Once the 64 applications were selected, the research team used the free online readability tool WebFx to generate the readability report after testing several readability tools and finding that the WebFx tools generated consistent results with the simplest user-interface. From there, the research team recorded the word count, Flesch reading ease, Flesch-Kincaid Reading Grade

Level (RGL), Gunning Frequency of Gobbledygook (FOG) RGL, Simplified Measure of Gobbledygook (SMOG) RGL, sentence count, and number of complex words, as well as the average reading score (Das et al., 2018). These readability metrics record the following outlined in Table 2.

Table 2. Readability Approaches, Variables, and Results (Krumay & Klar, 2020)

Approach	Variables	Results	Equation
Flesch-Kincaid Reading Grade Level (RGL)	Average sentence length; average number of syllables per word	Grade levels	FKGL = (0.39 * ASL) + (11.8 * ASW) – 15.59
Gunning FOG RGL	Sentence length; complex words	Grade levels	FOG Index = 0.4 * (ASL + PHW)
SMOG RGL	Number of polysyllable words	Grade levels	SMOG Grading = 2√PSW + 3
Flesch Reading Ease	Average sentence length; average number of syllables per word	Index (reading score)	FRE = 206,835 – (1.015 * ASL) – (84.6 * ASW)
Average Reading Score	FKGL, FOG, SMOG	Grade levels	Average = (FKGL + FOG + SMOG)/3

When the results produce a grade level, this means that the equation computes a value and translates the value to the specific United States grade level. The Flesch Reading Ease test produced a score between 0 to 100, with lower scores representing increasing complexity the more complexity (shown in Table 3) (Das et al., 2018; Flesch, 2016).

Table 3. Score Breakdown of Flesch Reading Ease (Flesch, 2016)

Score	School Level (US)	Notes
100.00 - 90.00	5th grade	Very easy to read. Easily understood by an average 11-year-old student.
90.0 - 80.0	6th grade	Easy to read. Conversational English for consumers.
80.0 – 70.0	7th grade	Fairly easy to read.
70.0 – 60.0	8th & 9th grade	Plain English. Easily understood by 13- to 15-year-old students.
60.0 – 50.0	10th to 12th grade	Fairly difficult to read.
50.0 – 30.0	College	Difficult to read.
30.0 – 10.0	College graduate	Very difficult to read. Best understood by university graduates.
10.0 – 0.0	Professional	Extremely difficult to read. Best understood by university graduates.

This research project used the selected 64 applications. However, in the years between the studies, seven of the apps were discontinued by the developer, changed into a different app, or have a name that is very common making it hard to find the exact app used in the original study. These eight applications and the reason they were not included are listed below in Table 4. This means that the 2021 study is a study of the 57 policies as of September 1st, 2021.

Table 4. Apps Not Included from Original Study

APP NAME	REASON NOT INCLUDED
STICK TEXTING: THE EMOJI KILLER	Could not find the same version from original study
MINECRAFT: STORY MODE	No longer being supported (publisher went out of business)

KIMOJI	Removed due to a copyright violation
SWYPE KEYBOARD	No longer being supported Could not find the same version from original study
POWER CLEAN: OPTIMIZE CLEANER	Could not find the same version from original study
SLEEP CYCLE ALARM CLOCK	Could not find the same version from original study
MUSICAL.LY	No longer being supported – transitioned into TikTok

Once the 57 apps and their privacy policies were generated, the privacy policies were copied and pasted into a word document as unformatted text, and then were pasted into the WebFx "Enter Text" option. This method was chosen over entering the URL because the more formatted privacy policies with UI components, like sections and visuals, reduced the tool's ability to accurately analyze the text. Also, certain policies' URLs were unreadable by the tool, therefore, the copying and pasting approach was more consistent than URL reading.

From there, each privacy policy's Flesch-Kincaid RGL, Gunning FOG RGL, SMOG RGL, Flesch-Kincaid Reading Ease, Word Count, Sentence Count, Number of Complex Words, % of Complex Words, and Recommend Age Range were recorded in an Excel file.

Chapter 4

Results

Of the available 57 privacy policies reviewing in this study, the average word count was 3683 words (standard deviation of 2203.92), with a range of 589 to 10700 words. The privacy policies had a mean RGL of 13.27 (Table 5), which is .5086 higher than the initial 2018 study (n = 57). The mean Flesch reading ease was 41.67, which is 1.06 lower than the 2018 study (N=64).

Table 5. Mean Readability Statistics

Statistic	2018 (64)	2018 Apps (57)	2021 Apps (57)	2021 to 2018 difference
N	64	57	57	--
Mean RGL	12.78	12.76	13.27	0.51
Mean word count	2425	--	3682	1258*
Mean Flesch-Kincaid reading ease	42.73	--	41.67	-1.06*
Mean Flesch-Kincaid Reading level	12.51	12.49	13.04	0.55
Mean Gunning FOG	14.38	14.39	14.96	0.57
Mean SMOG	11.46	11.43	11.81	0.38

Table 5 shows mean readability statistics between the 2018 privacy policies and the 2021 privacy policies. The original study did not report the individual word count and Flesch-Kincaid reading ease results for each individual study but did provide the average for the 64 policies.

Other than those two variables, every other variable in Table 5 was recorded for the modified list (n = 57) of collected policies.

Table 6. Significance Test Results (n = 57)

Statistic	2018 Privacy Polices vs. 2021 Privacy Notice
n	57
Mean RGL	0.11
Mean word count*	--
Mean Flesch-Kincaid Reading Ease*	--
Mean Flesch-Kincaid Reading Level	0.20
Mean Gunning FOG	0.13
Mean SMOG	0.14

Table 6 shows that there was not statistically significant (p = .05) differences between the recorded data in the 2018 study and the 2021 study. Though all the average values reflect an increased complexity, the findings are not statistically significant.

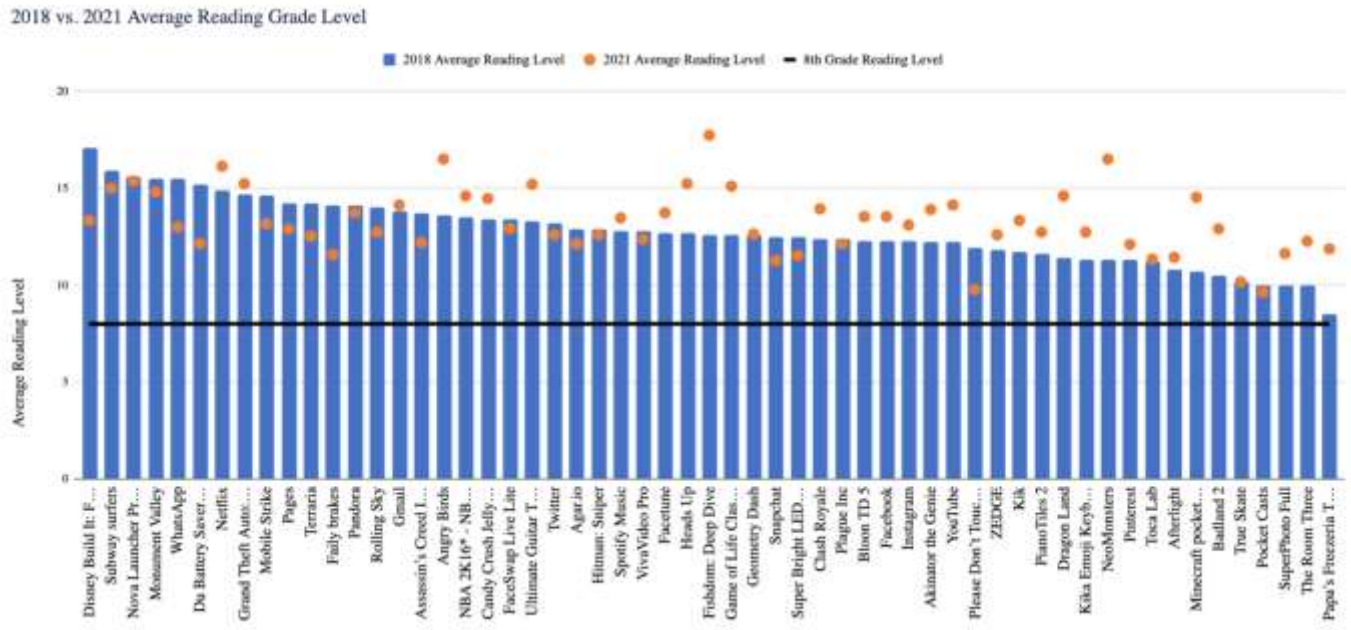

Figure 2. 2018 vs. 2021 Average Reading Level

Figure 2 shows the average reading level, which is the average of the Flesch-Kincaid Grade Level, Gunning FOG, and SMOG scores, for the 2018 and 2021 privacy policies (N=57). The bars represent the 2018 privacy policies and the orange dot represent the 2021 privacy policies. The pink line signifies the 8th grade reading level, the recommended benchmark for privacy policy readability (Das et al., 2018; Krumay & Klar, 2020; Milne et al., 2006;(Zhang et al., 2020)(Das et al., 2018). Looking at the graph, there does seem to be a tend with less variance between the 2021 policies that had higher average readability in 2018 than the policies with a lower average readability in 2018.

Figure 3 and Figure 4 show the upper and lower quartile differences in 2018 compared to 2021. Subtracting the 2018 scores from the 2021 scores for the upper 75% and lower 25%, these charts show in more detail how the policies with the highest average readability in 2018 changed in 2021 and how the policies with the lowest readability scores in 2018 changed in 2021.

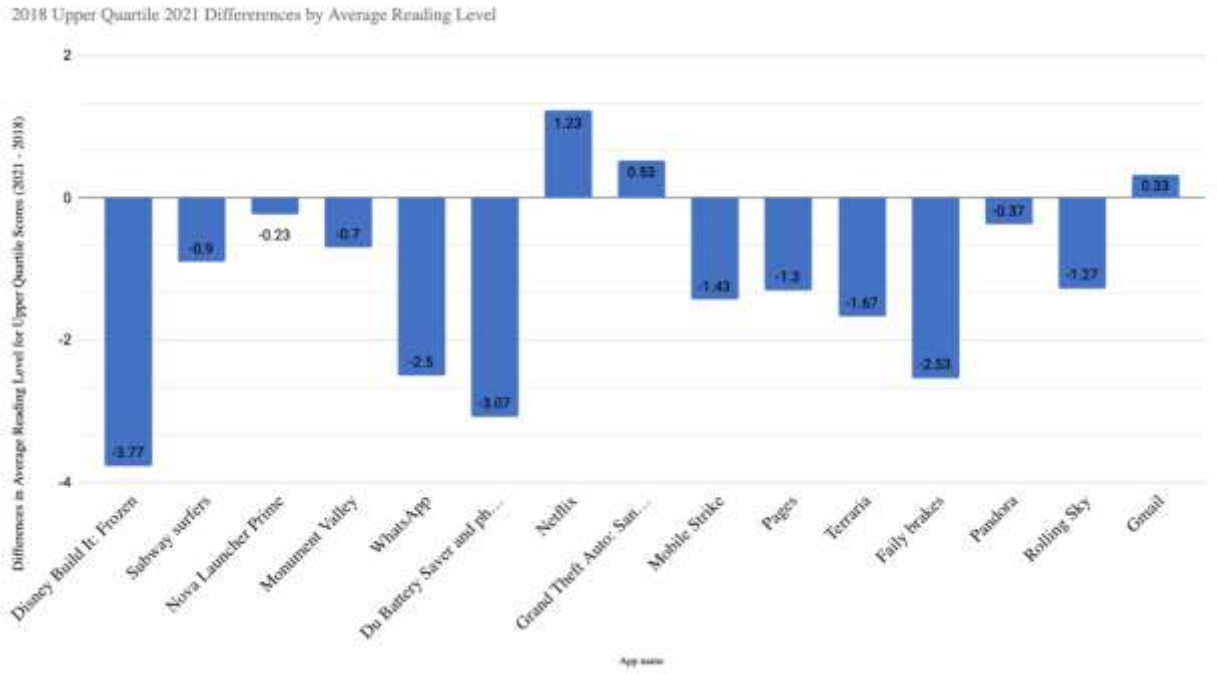

Figure 3. Upper Quartile 2021 and 2018 Difference

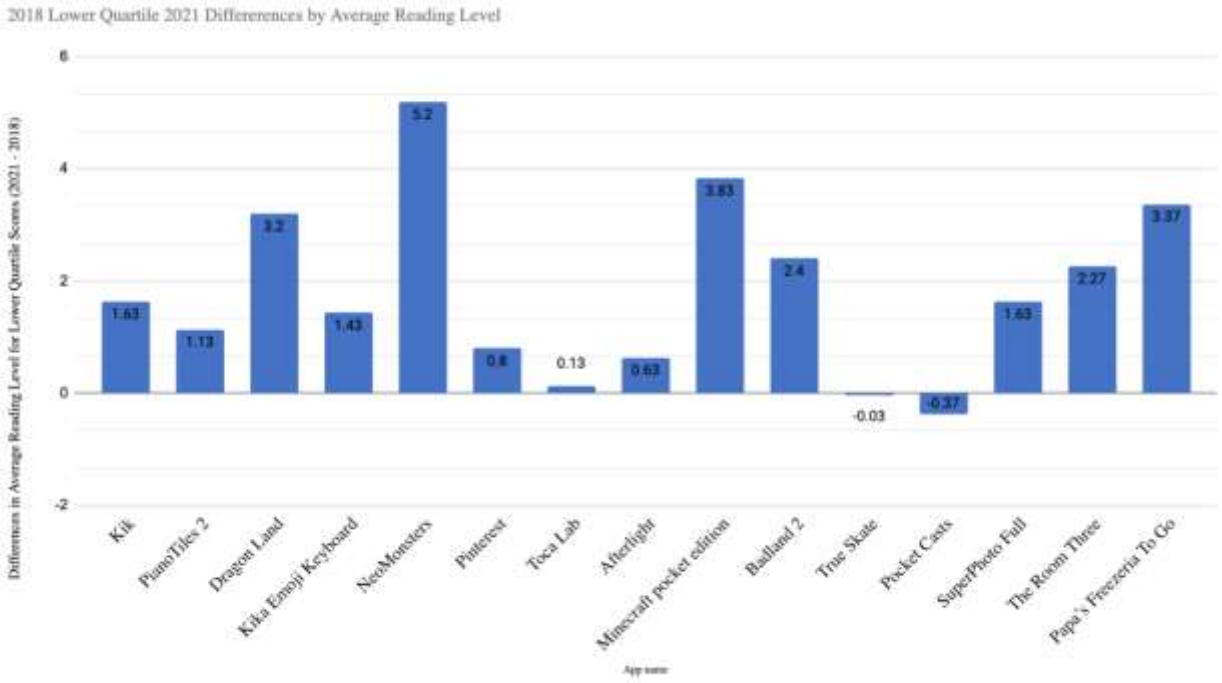

Figure 4. Lower Quartile 2021 and 2018 Difference

Table 7. Upper and Lower Quartile Average Breakdowns

	Upper Quartile (Figure 3)	Lower Quartile (Figure 4)
Average Readability (2018)	14.89	10.70
Average Readability (2021)	13.72	12.52
Average Readability Difference (2021 – 2018)	-1.18	1.82

The Upper Quartile policies saw an average reduction of -1.18 in average readability. The Lower Quartile policies had a 1.82 increase average readability. These are reflected in the overall increase of Mean RGL (.5086 increase) for privacy policies between 2018 and 2021.

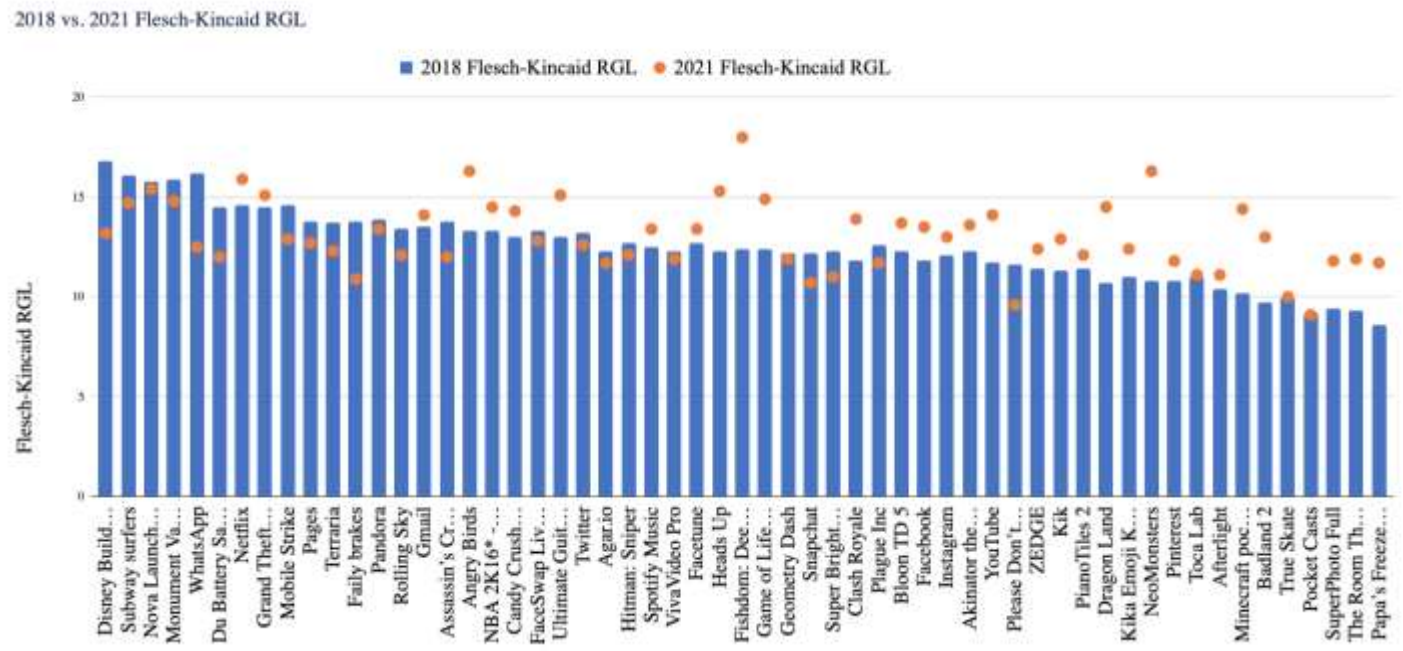

Figure 5. Flesch-Kincaid RGL

Figure 5 shows the changes in privacy policy readability using the Flesch-Kincaid Readability formula.

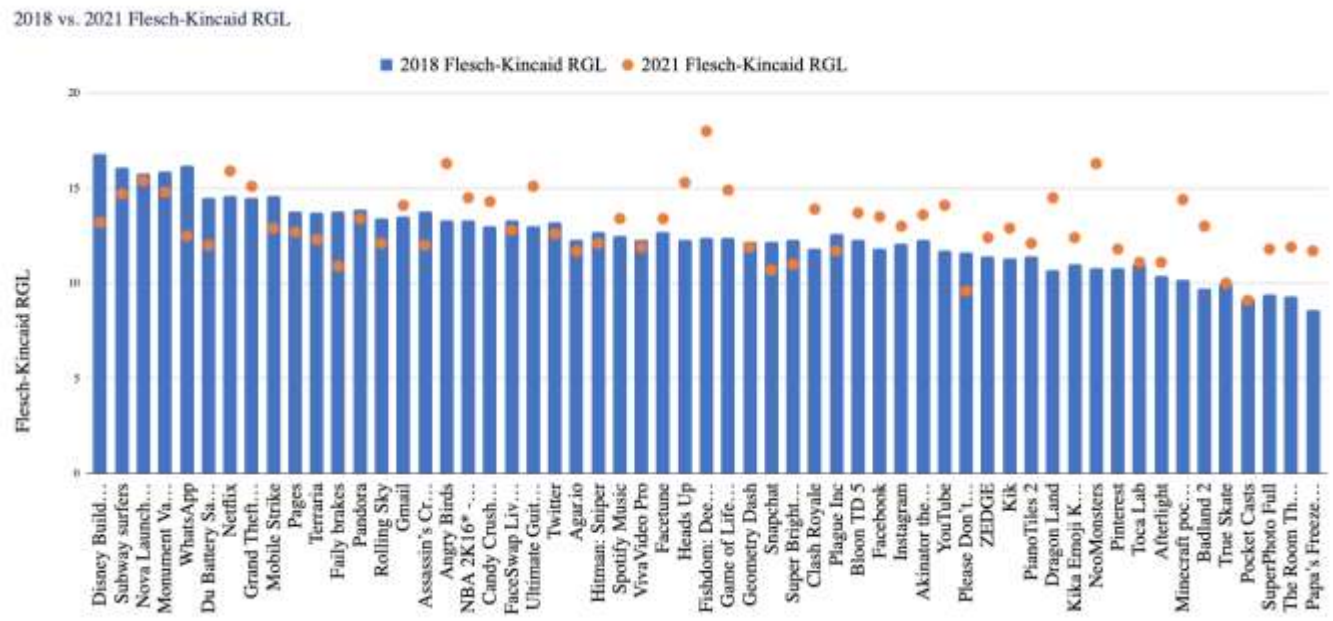

Figure 6. 2018 and 2021 Privacy Policy by Flesch-Kincaid Reading Grade Level (RGL)

Figures 7and Figure 8 show the difference in Gunning FOG (sentence length) and SMOG (number of polysyllable words) between the 2018 and 2021 privacy policies, respectively.

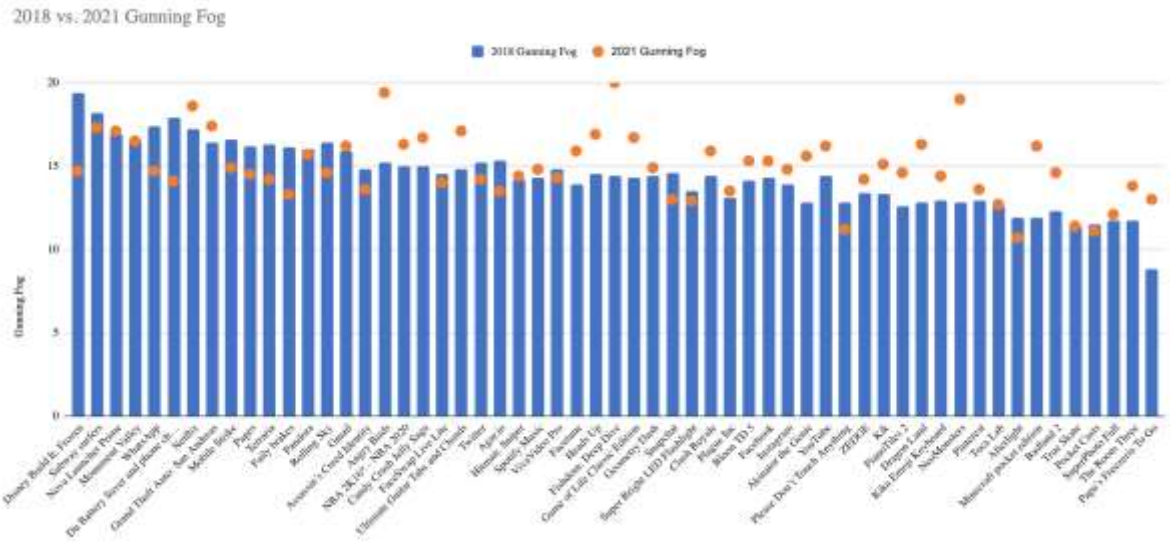

Figure 7. 2018 and 2021 Privacy Policies by Gunning FOG

Gunning FOG, which is a readability metric measuring the average number of words per sentence and the percentage of polysyllable word, is one of the most commonly used readability tool (Using English, 2008 as cited in Walsh & Vosko, 2008; Miles TH, 2008, as cited in Walsh & Vosko, 2008). The 2018 average for Gunning FOG was lower than the 2021 average by .57.

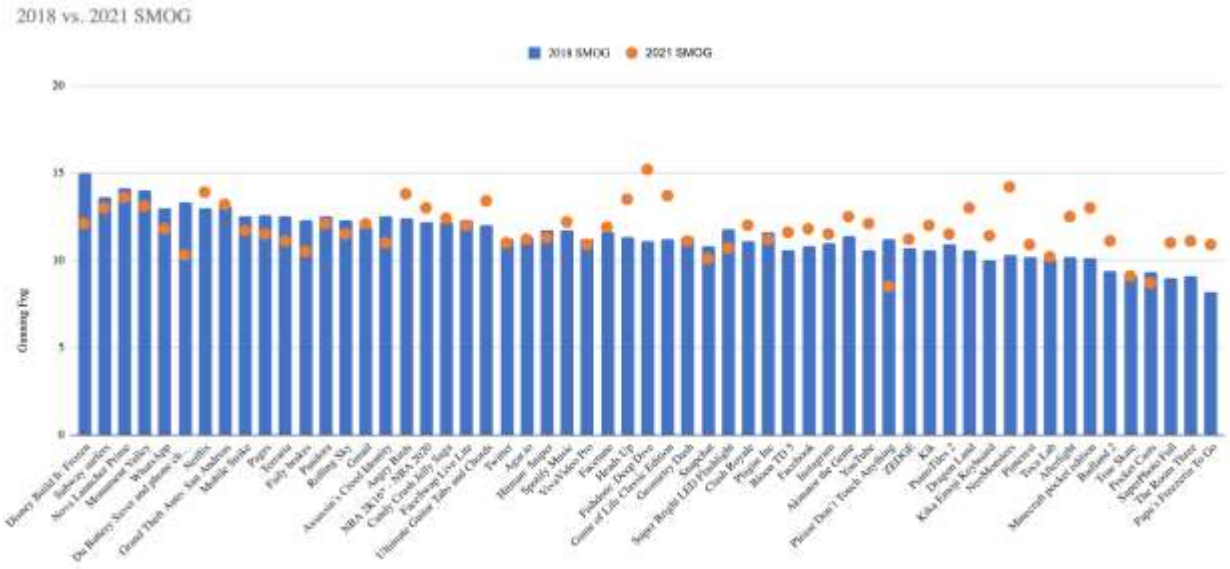

Figure 8. 2018 and 2021 Privacy Policies by SMOG

SMOG (Simplified Measure of Gobbledygook) measures the number of polysyllabic words in at least 30 sentences. The SMOG measure in 2018 was .38 lower than the average for 2021.

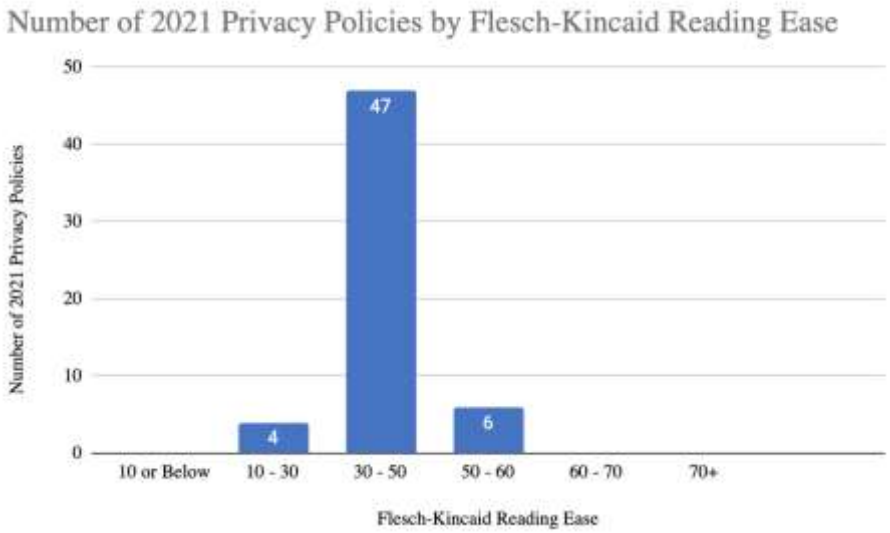

Figure 9. Number of 2021 Privacy Policies by Flesch-Kincaid Reading Ease

Figure 9 shows the 2021 privacy policies by Flesch-Kincaid Reading Ease, which thresholds are outlined in Table 3. The 10-30 range represents the college graduate reading level, the 30-50 range represents the college age level, and the 50-60 range represents the 10th to 12th grade reading level. The 60-70 range equates to the 8th and 9th grade reading level, which is the recommended benchmark for privacy policy readability (Das et al., 2018; Krumay & Klar, 2020; Milne et al., 2006; Zhang et al., 2020).

All 57 of the applications that mention an age have the age of data collection set to 13 or 16 (depending on the user's jurisdiction). However, of the 57 privacy policies analyzed in this study, 18 of them do not mention the minimum age to consent to these privacy policies and 18 do not mention the minimum age of use of the application. Lastly, 15 policies explicitly mention that younger users should consult with their parents or legal guardians before using the application.

Chapter 5

Discussion

The analysis of the 2021 privacy policies shows that the average RGL is 13.27, which is .51 higher than the RGL of the 2018 study (n = 57) and is above the recommended 8.0 reading level for adults in the United States. Another finding is that none of the collected 57 policies in 2021 are at or below in the 7th-9th grade reading level recommendation, which can be seen in Figure 5. Most of the privacy policies are in the 30-50 range, which is the RGL for the college grade level. As of 2021, only 41% of Americans had obtained an associate degree or higher (National Center for Education Statistics, 2021). Adding to the problem is that not only are younger users being affected, but most adults in the United States are as well.

Though the results do not show whether the increase is in response to additional regulation, the results do show that the criteria outlined in the GDPR and CCPA regarding the complexity of the language and a reasonable expectation of understanding are not met, especially regarding younger users consenting to these policies. Due to the contractual nature of the documents, the idea that younger users and their parents may not be able to understand the language in the documents is concerning. Further, explicitly mentioned in the GDPR, the "easily accessible and easy to understand" nature of the document is vital for achieving the principle of transparency and ensuring fairness in respect to individuals' rights to know what and how their information is processed (2013).

While there are no statistically significant differences in the 2021 and the 2018 measures, the 2021 policies reflect the increasing complexity of the documents initially researched in the 2018 study. The trend noted in the 2018 study regarding the incomprehensible nature of the documents to most adults, let alone youth, is still prevalent in today's privacy and compliance

space. However, the 2021 privacy governance space includes more privacy provisions and legislation of data practices and transparency than the initial study's 2018 privacy environment. The new data shows that these provisions were not met when the new legislation was passed, and they are still not being met several years after their passing.

However, the document length might have played an effect as well. Recalling back to Table 2, which highlights the differences in the readability formula, Flesch-Kincaid and Gunning FOG both take the average sentence length (ASL) into consideration, whereas SMOG does not. Because the SMOG score's only variable is the number of polysyllable words (PSW), the differences in the scores among the three could be related to a bigger change in the sentence length than a change in vocabulary. An explanation for the smaller change in the readability scores of the SMOG exam compared to Gunning FOG and Flesch-Kincaid RGL could be that the verbiage describing data practices might not have changed as much as did the length of the documents, which had an average of 2454 (N = 64) in 2018 to 3683 (n = 57) in 2021.

The initial study proposes that stakeholders should play a role in educating youth and guardians about the privacy risk of web-based apps, specifically mentioning pediatricians and other healthcare professionals (Das et al., 2018). The study cited a 2013 Pew Research Center study that showed 70% of teen users on the internet do seek out advice on internet privacy, and that many of these users consult their friends, peers, and their parents (Lenhart et al., 2013). The poll additionally found that children aged 12 and 13 or more likely to talk to and seek help from parents than teenagers (77% versus 67%) (Lenhart et al., 2013). The poll collected data on age, socioeconomic status, and race demographic data, and asked about privacy practices and data hygiene, however, did not mention anything about if the participants read the privacy policy or knew how their data was being used. This reflects the general trend about individuals taking

efforts to protect their privacy from unwarranted intrusion without them being aware about what they are protecting their privacy from.

Further, by reducing the complexity of privacy policies, more individual may be more motivated to read them. Some proposed ways of doing this are by adding 1 sentence summaries to sections that are long in length like the Twitter privacy policy does, and by eliminating some of the legal terminologies that are often what makes the documents illegible to the majority of its audiences, reducing the length of the documents, and by incorporating user feedback, questions, and concerns into the documents with occasional updates and retesting (Das et al., 2018; Zhang et al., 2020). Further, getting more readers to engage with these policies, regardless of age, can help them gain familiarity with privacy practices and afforded protections, which in turn can reduce the complexity of these documents. Proposed ways to reduce complexity as well as encourage readership will be addressed in the next and final chapter of this thesis.

Chapter 6

Recommendations

With the availability of the internet and the diversity of internet users, there are concerns about only using readability scores and formulas for assessing document complexity. Readability can be scored objectively, but there are some complications when it comes to assessing comprehension. A major factor of comprehension and where it can differ from readability is familiarity with the terms and topics addressed in the piece. If a reader has prior experience interacting with the topics, the overall complexity of the document is reduced, but the readability scores would remain the same.

Because all three readability measures used in this study (Flesch-Kincaid, Gunning FOG, and SMOG) use the number of syllables in a word as a factor, polysyllable words can influence readability scores. Words like “anonymization”, “compliance”, “retention”, and other privacy concept words that are increasing in use can be responsible for the increasing readability scores of the documents. Repeating words could be one factor in the increasing readability scores, which can be traced back to government regulation forcing their use for the transparency purpose of the privacy notice. The Krumay and Klar study mentions that only using readability measures that factor in polysyllable words may not be the best way at assessing overall readability due to the nature of the documents (2020). By using human readers in conjunction with readability measures, the Krumay and Klar study was able to assess if the readability scores represented the thoughts and ability of human readers of varying grade levels (2020). However, but not assessing the comprehension of the human readers post-reading, the study only reflects the human readers' ability to read the documents, and not their comprehension and takeaways of the documents.

This thesis recommends additional research be done on assessing comprehension of privacy policies to see if users can understand what the documents are sharing and assessing if users can differentiate data handling practices. This research should include demographic data, like age, socioeconomic status, and citizenship, to see if there is a relationship among these factors. Age and socioeconomic status would be useful for assessing the overall accessibility of these documents, as the Pew Research poll identified that there was a positively increasing relationship between parent household income and the percentage of teen users who seek advice on online privacy management habits (Lenhart et al., 2013). Citizenship would be recorded to see if there is any relationship between citizenship and jurisdictions with privacy regulations. That variable would be assessing if privacy regulation plays any factor in an individual's privacy literacy.

The Krumay and Klar study also outlined the limitations of readability tools, specifically their inability to assess reader capabilities that are not determined by grade level (2020). These factors include elements like content familiarity (a middle schooler very interested in computers could have more technical literacy than someone with a bachelor's degree not interested in computers at all) and individuals with intellectual disabilities (Krumay & Klar, 2020). These tools are also not able to assess privacy policies in other languages (Krumay & Klar, 2020).

Today, readability tools are best used for general trends of the body of text and cannot be used to interpret how a specific reader's impression of the text. By developing a study that furthers studies comprehension trends, researchers could find the most reliable equations for addressing readability in English. Additional research can also focus on how the readability of the text in English changes when translated into another language and vice versa. Further

research would also be beneficial for understanding trends among user familiarity with privacy terms and data handling practices.

It is also recommended to reevaluate how documents promote transparency of organization data handling practices. A potential solution to this problem could be to use the text annotators for text summarization and to have policymakers identify regulation-specific annotation schemes for privacy policies. For example, the OPP-115 privacy corpus developed an annotation scheme based on 10 data categories identified in the 115 analyzed privacy policies (Wilson et al., 2016). The ten data categories derived are below in Table 8.

Table 8. Data Categories outline in OPP-115 Corpus (Wilson et al. 2016)

First Party Collection/Use: how and why a service provider collects user information.	**Data Security**: how user information is protected.
Third Party Sharing/Collection: how user information may be shared with or collected by third parties	**Policy Change:** if and how users will be informed about changes to the privacy policy.
User Choice/Control: choices and control options available to users.	**Do Not Track**: if and how Do Not Track signals[a] for online tracking and advertising are honored.
User Access, Edit, & Deletion: if and how users may access, edit, or delete their information.	**International & Specific Audiences**: practices that pertain only to a specific group of users (e.g., children, Europeans, or California residents).
Data Retention: how long user information is stored.	**Other**: additional sub-labels for introductory or general text, contact information, and practices not covered by the other categories.

[a] references https://www.w3.org/2011/tracking-protection/

In addition to using the recommendation of including one-sentence summaries of the different settings, seen in the Twitter privacy policies, a combination of using an annotation scheme, like what was identified in the OPP-115 corpus, and short summary sentences could improve user familiarity with data handling practices, reduce the complexity of the document,

and promote transparency among the companies, regulating bodies, and users. Further, regulators can identify specific privacy practices they want in privacy policies to create a text annotation schema to help companies identify what information needs to be included in compliant policies and help the enforcement and rectification of policies that are not compliant, in the case of the documents being too complex, lacking a specific section for a data practice, or not available on a certain page of the website, for example.

In addition to efforts that could reduce the complexity of the document, improve, and promote compliance with regulation, and potentially increase the number of people who read privacy policies, there may also need to be an assessment of the age of digital consent in today's internet age. With the increased reliance on the internet in the lives of children and teenagers, promoting good privacy hygiene habits from a young age would be very impactful. However, there is still the need to structurally answer the question of responsibility for the protection of children's privacy on the internet, among organizations, governments, and guardian figures.

There is also some debate about if the age of digital consent should be increased to 16 or 18, or if there should just be increased protections for people 13 to 18 on internet sites (Jargon, 2019). As this thesis found, the language of privacy policies is too complicated for most adults to understand, let alone their children. However, by having regulators promote transparency via accessibility instead of transparency for compliance purposes only, users may learn more about data privacy and can become more aware of their internet habits and personal privacy hygiene. Further, by increasing the accessibility of privacy policies, younger users and teens can more easily identify risky online habits and feel more empowered and embolden to not engage in them.

Though the future of protection of children and teenagers on the internet is not yet determined, it is going to be a collective effort among stakeholders in the child's life, regulators, and organizations to encourage them to protect their own privacy while also navigating the path of regulation and increased protections.

Appendix A
2018 Privacy Policy Data

App Name	2018 ARL	2018 Flesch-Kincaid reading level	2018 Gunning FOG	2018 SMOG[a]
Disney Build It: Frozen	17.1	16.8	19.4	15
Subway surfers	15.9	16.1	18.2	13.6
Nova Launcher Prime	15.6	15.8	16.9	14.1
Monument Valley	15.5	15.9	16.6	14
WhatsApp	15.5	16.2	17.4	13
Du Battery Saver and phone charger	15.2	14.5	17.9	13.3
Netflix	14.9	14.6	17.2	13
Grand Theft Auto: San Andreas	14.7	14.5	16.4	13.1
Mobile Strike	14.6	14.6	16.6	12.5
Pages	14.2	13.8	16.2	12.6
Terraria	14.2	13.7	16.3	12.5
Faily brakes	14.1	13.8	16.1	12.3
Pandora	14.1	13.9	16	12.5
Rolling Sky	14	13.4	16.4	12.3
Gmail	13.8	13.5	15.9	11.9
Assassin's Creed Identity	13.7	13.8	14.8	12.5
Angry Birds	13.6	13.3	15.2	12.4
NBA 2K16	13.5	13.3	15	12.2
Candy Crush Jelly Saga	13.4	13	15	12.2
FaceSwap Live Lite	13.4	13.3	14.5	12.3
Ultimate Guitar Tabs and Chords	13.3	13	14.8	12
Twitter	13.2	13.2	15.2	11.2
Agar.io	12.9	12.3	15.3	11.3
Hitman: Sniper	12.9	12.7	14.2	11.7
Spotify Music	12.8	12.5	14.3	11.7
VivaVideo Pro	12.8	12.3	14.8	11.2
Facetune	12.7	12.7	13.9	11.6
Heads Up	12.7	12.3	14.5	11.3

Fishdom: Deep Dive	12.6	12.4	14.4	11.1
Game of Life Classic Edition	12.6	12.4	14.3	11.2
Geometry Dash	12.6	12.2	14.4	11.3
Snapchat	12.5	12.2	14.6	10.8
Super Bright LED Flashlight	12.5	12.3	13.5	11.8
Clash Royale	12.4	11.8	14.4	11.1
Plague Inc	12.4	12.6	13.1	11.6
Bloon TD 5	12.3	12.3	14.1	10.6
Facebook	12.3	11.8	14.3	10.8
Instagram	12.3	12.1	13.9	11
Akinator the Genie	12.2	12.3	12.8	11.4
YouTube	12.2	11.7	14.4	10.6
Please Don’t Touch Anything	11.9	11.6	12.8	11.2
ZEDGE	11.8	11.4	13.4	10.7
Kik	11.7	11.3	13.3	10.6
PianoTiles 2	11.6	11.4	12.6	10.9
Dragon Land	11.4	10.7	12.8	10.6
Kika Emoji Keyboard	11.3	11	12.9	10
NeoMonsters	11.3	10.8	12.8	10.3
Pinterest	11.3	10.8	12.9	10.2
Toca Lab	11.2	11	12.6	10
Afterlight	10.8	10.4	11.9	10.2
Minecraft pocket edition	10.7	10.2	11.9	10.1
Badland 2	10.5	9.7	12.3	9.4
True Skate	10.2	10	11.4	9.1
Pocket Casts	10	9.2	11.5	9.3
SuperPhoto Full	10	9.4	11.7	9
The Room Three	10	9.3	11.7	9.1
Papa’s Freezeria To Go	8.5	8.6	8.8	8.2

2021 Privacy Policy Data

App Name	2021 Average reading level	2021 Flesch-Kincaid reading level	2021 Gunning Fog	2021 SMOG[a]
Disney Build It: Frozen	13.33333333	13.2	14.7	12.1
Subway surfers	15	14.7	17.3	13
Nova Launcher Prime	15.36666667	15.4	17.1	13.6
Monument Valley	14.8	14.8	16.5	13.1
WhatsApp	13	12.5	14.7	11.8
Du Battery Saver and phone charger	12.13333333	12	14.1	10.3
Netflix	16.13333333	15.9	18.6	13.9
Grand Theft Auto: San Andreas	15.23333333	15.1	17.4	13.2
Mobile Strike	13.16666667	12.9	14.9	11.7
Pages	12.9	12.7	14.5	11.5
Terraria	12.53333333	12.3	14.2	11.1
Faily brakes	11.56666667	10.9	13.3	10.5
Pandora	13.73333333	13.4	15.7	12.1
Rolling Sky	12.73333333	12.1	14.6	11.5
Gmail	14.13333333	14.1	16.2	12.1
Assassin's Creed Identity	12.2	12	13.6	11
Angry Birds	16.5	16.3	19.4	13.8
NBA 2K16	14.6	14.5	16.3	13
Candy Crush Jelly Saga	14.46666667	14.3	16.7	12.4
FaceSwap Live Lite	12.93333333	12.8	14	12
Ultimate Guitar Tabs and Chords	15.2	15.1	17.1	13.4
Twitter	12.6	12.6	14.2	11
Agar.io	12.13333333	11.7	13.5	11.2
Hitman: Sniper	12.6	12.1	14.4	11.3
Spotify Music	13.46666667	13.4	14.8	12.2
VivaVideo Pro	12.36666667	11.9	14.3	10.9
Facetune	13.73333333	13.4	15.9	11.9
Heads Up	15.23333333	15.3	16.9	13.5
Fishdom: Deep Dive	17.73333333	18	20	15.2

Game of Life Classic Edition	15.1	14.9	16.7	13.7
Geometry Dash	12.63333333	11.9	14.9	11.1
Snapchat	11.26666667	10.7	13	10.1
Super Bright LED Flashlight	11.53333333	11	12.9	10.7
Clash Royale	13.93333333	13.9	15.9	12
Plague Inc	12.13333333	11.7	13.5	11.2
Bloon TD 5	13.53333333	13.7	15.3	11.6
Facebook	13.53333333	13.5	15.3	11.8
Instagram	13.1	13	14.8	11.5
Akinator the Genie	13.9	13.6	15.6	12.5
YouTube	14.13333333	14.1	16.2	12.1
Please Don’t Touch Anything	9.766666667	9.6	11.2	8.5
ZEDGE	12.6	12.4	14.2	11.2
Kik	13.33333333	12.9	15.1	12
PianoTiles 2	12.73333333	12.1	14.6	11.5
Dragon Land	14.6	14.5	16.3	13
Kika Emoji Keyboard	12.73333333	12.4	14.4	11.4
NeoMonsters	16.5	16.3	19	14.2
Pinterest	12.1	11.8	13.6	10.9
Toca Lab	11.33333333	11.1	12.7	10.2
Afterlight	11.43333333	11.1	10.7	12.5
Minecraft pocket edition	14.53333333	14.4	16.2	13
Badland 2	12.9	13	14.6	11.1
True Skate	10.16666667	10	11.4	9.1
Pocket Casts	9.633333333	9.1	11.1	8.7
SuperPhoto Full	11.63333333	11.8	12.1	11
The Room Three	12.26666667	11.9	13.8	11.1
Papa’s Freezeria To Go	11.86666667	11.7	13	10.9

References:

California Consumer Privacy Act of 2018. (n.d.). Retrieved April 27, 2021, from https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?division=3.&part=4.&lawCode=CIV&title=1.81.5.

Children's Online Privacy Protection Rule ("COPPA"). (2013, July 25). Federal Trade Commission. https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule

Das, G., Cheung, C., Nebeker, C., Bietz, M., & Bloss, C. (2018). Privacy Policies for Apps Targeted Toward Youth: Descriptive Analysis of Readability. *JMIR MHealth and UHealth*, *6*(1), e7626. https://doi.org/10.2196/mhealth.7626

Flesch, R. (2016, July 12). *Guide to Academic Writing Article—Management—University of Canterbury—New Zealand*. https://web.archive.org/web/20160712094308/http://www.mang.canterbury.ac.nz/writing_guide/writing/flesch.shtml

General Data Protection Regulation (GDPR) – Official Legal Text. (n.d.). General Data Protection Regulation (GDPR). Retrieved April 19, 2021, from https://gdpr-info.eu/

It's Not You; Privacy Policies Are Difficult to Read. (2018, July 17). Common Sense Education. https://www.commonsense.org/education/articles/its-not-you-privacy-policies-are-difficult-to-read

Jargon, J. (2019, June 18). How 13 Became the Internet's Age of Adulthood. *Wall Street Journal*. https://www.wsj.com/articles/how-13-became-the-internets-age-of-adulthood-11560850201

Krumay, B., & Klar, J. (2020). Readability of Privacy Policies. *Data and Applications Security and Privacy XXXIV*, 388–399. https://doi.org/10.1007/978-3-030-49669-2_22

Lenhart, A., Madden, M., Cortesi, S., Gasser, U., & Smith, A. (2013, August 15). Where Teens Seek Online Privacy Advice. *Pew Research Center: Internet, Science & Tech*. https://www.pewresearch.org/internet/2013/08/15/where-teens-seek-online-privacy-advice/

Liu, F., Fella, N. L., & Liao, K. (2018). Modeling Language Vagueness in Privacy Policies using Deep Neural Networks. *ArXiv:1805.10393 [Cs]*. http://arxiv.org/abs/1805.10393

Marchand, L. (2017, March 22). *What is readability and why should content editors care about it?* Center for Plain Language. https://centerforplainlanguage.org/what-is-readability/

Meiselwitz, G. (2013). Readability Assessment of Policies and Procedures of Social Networking Sites. In A. A. Ozok & P. Zaphiris (Eds.), *Online Communities and Social Computing* (Vol. 8029, pp. 67–75). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-39371-6_8

Milne, G. R., Culnan, M. J., & Greene, H. (2006). A Longitudinal Assessment of Online Privacy Notice Readability. *Journal of Public Policy & Marketing*, *25*(2), 238–249.

National Center for Education Statistics. (2021). *Educational Attainment*. National Center for Education Statistics. https://nces.ed.gov/fastfacts/display.asp?id=27

Privacy Litigation. (n.d.). Privacy Law Blog. Retrieved May 1, 2021, from https://privacylaw.proskauer.com/articles/privacy-litigation/

Privacy policy. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Privacy_policy&oldid=1019415362

Protecting Children's Privacy Under COPPA: A Survey on Compliance. (n.d.). Retrieved January 25, 2022, from https://www.ftc.gov/sites/default/files/documents/reports/protecting-childrens-privacy-under-coppa-survey-compliance/coppasurvey.pdf

Readable | Free Readability Test Tool. (n.d.). Retrieved April 27, 2021, from https://www.webfx.com/tools/read-able/

Sadeh, N., Acquisti, A., Breaux, T. D., Cranor, L. F., McDonald, A. M., Reidenberg, J. R., Smith, N. A., Liu, F., Russell, N. C., Schaub, F., & Wilson, S. (2013). *The Usable Privacy Policy Project: Combining Crowdsourcing, Machine Learning and Natural Language Processing to Semi-Automatically Answer Those Privacy Questions Users Care About*. 24.

Senators Markey and Hawley Introduce Bipartisan Legislation to Update Children's Online Privacy Rules | U.S. Senator Ed Markey of Massachusetts. (n.d.). Retrieved April 19, 2021, from https://www.markey.senate.gov/news/press-releases/senators-markey-and-hawley-introduce-bipartisan-legislation-to-update-childrens-online-privacy-rules/

Srinath, M., Wilson, S., & Giles, C. L. (2021). Privacy at Scale: Introducing the PrivaSeer Corpus of Web Privacy Policies. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, 6829–6839. https://doi.org/10.18653/v1/2021.acl-long.532

Walsh, T. M., & Volsko, T. A. (2008). Readability Assessment of Internet-Based Consumer Health Information. *RESPIRATORY CARE*, *53*(10), 6.

Wilson, S., Schaub, F., Dara, A. A., Liu, F., Cherivirala, S., Giovanni Leon, P., Schaarup Andersen, M., Zimmeck, S., Sathyendra, K. M., Russell, N. C., B. Norton, T., Hovy, E., Reidenberg, J., & Sadeh, N. (2016). The Creation and Analysis of a Website Privacy Policy

Corpus. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1330–1340. https://doi.org/10.18653/v1/P16-1126

Zhang, M., Chow, A., & Smith, H. (2020). COVID-19 Contact-Tracing Apps: Analysis of the Readability of Privacy Policies. *Journal of Medical Internet Research*, *22*(12), e21572. https://doi.org/10.2196/21572

ACADEMIC VITA

EDUCATION

The Pennsylvania State University, University Park
Schreyer Honors College
BS in Information Sciences and Technology: Security and Risk Analysis
Minor in Chinese

WORK EXPERIENCE

Privacy Engineering Intern, Google **Summer 2020 and Summer 2021**

- Created the content for an interactive privacy exercise focused on location data centered around encouraging perspective-taking and highlighting privacy tradeoffs
- Audited an internal privacy group review process, identifying redundancies to increase efficiency, with the goal of creating a standardized process
- Examined various teams onboarding information, discovering overlaps and gaps in the trainings
- Engineered and analyzed feedback data into a Differential Privacy training, debuted at DEFCON 2020

Pioneer Internship, Deloitte **Summer 2019**

- Led a team project and inspired a future Talent recruiting campaign, developed a network of professionals and learned how to navigate an office environment
- Experienced working with a Financial Risk Advisory team and how to deliver strong reports

Intelligence Community Centers for Academic Excellence (ICCAE) Scholar

- Selected for and completed a congressionally funded grant program focused on recruitment and hiring to improve diversity in the intelligence community
- Participated as 1 of 66 students selected nationwide to participate in a two-week seminar and simulation focused on briefing skills, structured analytic techniques, and critical thinking through an intelligence-based simulation exercise

Penn State's IAPP Alan Westin Scholarship Winner **Fall 2021**

- Identified by professors as a future leader in the field of privacy and data protection

RESEACH EXPERIENCE

Undergraduate Thesis: Privacy Policy Readability for Apps Used by Teenagers

- Continuing a 2018 readability study analyzing how the readability of the 64 policies changed with increasing regulations and evaluating where the expectation of technical literacy belongs

Pennsylvania Criminal Intelligence Center (PaCIC) Research **Fall 2019 – Fall 2020**

- Identified and mapped social network activity and ideologies, culminating in a research report

SKILLS

SQL, Java and Linux familiarity, R and RStudio familiarity, Tableau, Analyst Notebook

LEADERSHIP/INTERESTS

Schreyer Honors College Mentor and Ambassador
Learning Assistant in the College of IST