The University of Texas at Austin
Center for Identity

Human and Privacy Rights

Razieh Nokhbeh Zaeem
K. Suzanne Barber

UTCID Report #21-09

November 2021

Human and Privacy Rights

Razieh Nokhbeh Zaeem and K. Suzanne Barber
Center for Identity, The University of Texas at Austin
razieh,sbarber@identity.utexas.edu

November 12, 2021

1 Privacy Is a Human Right

Privacy is important. Individuals should have the right to control the disclosure of personal data that describes them, identifies them and reveals information about them. Individuals should not be subjected to invasions of their privacy, family, home or correspondence, nor to assaults upon their reputation. These rights to privacy must be protected by law to guard against such interference, invasions or attacks.

Many, many different types of personal data serve to describe, identify, and reveal information about a person, their devices and their organizations. This information can be organized into four categories:

- What you KNOW—memorized assets such as name, address, or mother's maiden name.
- What you HAVE—held assets such as driving license and employee credential badges.
- What you ARE—biometric assets describing physical characteristics.
- What you DO—assets describing behavioral patterns such as Internet browsing or location patterns.

This personal data serves as identifiers, the "Little i," that describe and imprint the "Big I"—individuals. Consequently, this "Little i" describes, imprints, points to and ultimately offers benefits or threatens the "Big I" in the physical world.

Because the "Little i," is directly related to "Big I," theft, fraud and abuse of personal data can be used to expose, exploit, threaten or attack an individual.

In a world of increasing physical and digital surveillance where an enormous amount of information is collected about individuals, privacy is challenged in every aspect of an individual's life. Privacy achieved though anonymity is extremely difficult if not impossible. Consequently, privacy rights must encompass

the right to control one's personal data, "Little i" to best protect the "Big I" in the physical world.

Agency and control are prerequisites for controllable and measurable privacy. For example, the U.S. constitution gives individuals the expectation for certain rights and protections. These constitutional rights empower individuals with agency (power and control) and give individuals well-defined authority over their assets and promise consequences when violated. For example, individuals own assets in the physical world such as a home, car, or filing cabinet and have expectations based on their constitutional rights that protect ownership of those assets and guard against unauthorized forfeiture, search or theft.

An individual's personal data, their "identity assets," are constantly being collected and, in fact, individuals knowingly and unknowingly share their identity assets in exchange for goods, services and access on a daily basis. This personal data is often referred to by the authors as "identity assets" due to the importance of this information in constructing identities and the value of this information as it serves as a means to access a wide range of goods, services, and systems. In fact, identity has often been called "the new currency" or "the new oil" of the Internet.

Whether used to grant benefits, gain access or pose threats, these identity assets are inseparable from us as human beings. Thus, the human rights offered to us as individuals must be extended to the rights of privacy and control of information that identifies us. Without the rights to control one's identity assets, our privacy rights and thereby our human rights are at risk.

2 Defining Privacy

"Privacy is a plurality of different things." [66]. "Currently, privacy is a sweeping concept, encompassing (among other things) freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations." [66] Privacy has been connected to the creation of knowledge, to dignity, and to freedom [56]. Philosophers, legal theorists, jurists, psychologists, and computer scientist have endeavored to understand what it means and how it should be protected.

The Merriam-Webster dictionary defines privacy as "the quality or state of being apart from company or observation", "freedom from unauthorized intrusion", and "secrecy"[1]. The European General Data Protection Regulation (GDPR) defines data protection as "keeping data safe from unauthorized access" and data privacy as "empowering users to make their own decisions about who can process their data and for what purpose"[2]. According to the GDPR, all "natural persons" have a right to data privacy under the European Union

[1] https://www.merriam-webster.com/dictionary/privacy?utm_campaign=sd&utm_medium=serp&utm_source=jsonld

[2] https://gdpr.eu/data-privacy

Copyright 2021. The Center for Identity. All rights reserved.

law. The California Consumer Privacy Act (CCPA) secures new privacy rights for California consumers[3].

3 Privacy in the Digital Society

Towards the end of the twentieth century, the development of new information technologies–in particular, the rise of the computer and the Internet–has made privacy erupt into the front-line [35, 52].

3.1 Privacy Policies: the Reality of Privacy Online

Many websites collect, share, and use their users' Personally Identifiable Information (PII)—"any information relating to an identified or identifiable natural person" [75]. Online privacy policies are legal documents that explain how an organization collects, handles, shares, uses, and discloses user data. Privacy policies have grown into the de facto method of communicating such data practices for organizations, and particularly their websites.

The ever-growing use of the Internet and the collection of PII over it has raised concerns for over two decades [27, 19]. In particular, the problem of how companies handle users' PII collected over the Internet involves three main players: companies, regulators, and users.

Companies, across industries, currently are faced with tough decisions when constructing their privacy policies. On the one hand, many business models are built on collecting, using, sharing, and selling personal information. Such information can be profitable for the company, and can be leveraged to improve their product offerings and consumer-facing services. On the other hand, collecting and storing personal information about consumers carries considerable risk, as evidenced by the financial and public relations fallout from high-profile hacks. Data breaches are occurring at alarming rates, and the fallout from such hacks can be massive. Companies must assess the balance of these risk/value propositions as they construct their privacy policies.

In response to high profile data breaches, *regulators* and policy makers—the second important player—have employed two lines of strategy: (1) holding corporations liable for breaches, imposing fines and sanctions on organizations that handled consumer data inappropriately and (2) attempting to increase the transparency of privacy and data management practices in privacy policies. The regulators, however, must constantly assess the current state of privacy policies across industries and evaluate the effects of the regulations they establish [62]. Many regulatory bodies around the globe have long enforced requirements on posting privacy policies online. Over the past two decades, as the concerns regarding PII use and misuse were growing, newer laws have gone into effect to protect user privacy. Prominent examples of such laws are the General Data

[3] `https://oag.ca.gov/privacy/ccpa`

Copyright 2021. The Center for Identity. All rights reserved.

Protection Regulation (GDPR)[4] in the European Union and the California Consumer Privacy Act (CCPA)[5] in the United States.

Finally, in the face of companies' carefully constructed privacy policies and regulators' endeavors to encourage transparency in privacy policies, *users* have neither the time [49, 50, 39, 48] nor the inclination [30, 51] to read privacy policies thoroughly, choosing instead to agree absentmindedly to the various privacy policies. More than ever, consumers need information to help them compare what a privacy policy offers with the status quo (e.g., average privacy practices among the companies that provide similar services).

3.1.1 Prevalence of Privacy Policies

The Federal Trade Commission (FTC) has provided several reports on online privacy practices since 1995. Its 1998 report [27] on US commercial websites' privacy disclosures revealed that while 92% of websites were collecting PII, only 14% disclosed any privacy policies. In its 2000 report [28], the FTC investigated a group of 335 websites chosen randomly and another group of 100 most busiest websites, both groups from the US market. The FTC noted that a vast majority of the websites studied collected some PII, e.g., 97% of the random sample and 99% of the busiest websites asked for email addresses. The same study found that, in year 2000, 88% of the random group and all of the 100 busiest websites disclosed some form of their privacy policy. In the same time frame, a survey of 100 most heavily used websites [19] focused on Notice, Choice, Access, and Security—the four Fair Information Practice Principles (FIPPs).

Since the first round of studies performed by the FTC, many researchers have analyzed the content of privacy policies in various ways. Many have investigated privacy policies with respect to a set of factors. Prior to the GDPR, the majority of such investigations were focused on the FTC's four Fair Information Privacy Practices. These investigations have considered different sample sets of privacy policies [41, 63, 92, 65, 57, 14].

It appears that privacy policies are becoming more and more common. In 2002 [47] Liu et al. examined web sites of the Global 500 and showed that only 61% of companies in the US had posted privacy policies. They extended their search effort for companies' privacy policies by emailing the companies and asking for their policies, when one could not be found online. Even with that extra effort, they showed that only 24% of the websites without posted privacy policy that they contacted indeed did have a policy elsewhere. In 2017 [83] an evaluation of 10% of all listings on NYSE, Nasdaq, and AMEX stock markets revealed that 69% had their privacy policies posted online.

3.1.2 Who Reads Privacy Policies?!

The topic of user interest, or lack thereof, in privacy policies has been on the minds of many researchers. McDonald and Cranor [48] noted that if users were

[4] `https://gdpr-info.eu`
[5] `https://oag.ca.gov/privacy/ccpa`

Copyright 2021. The Center for Identity. All rights reserved.

to read–just once a year–the privacy policy for each site they visit, they would need to spend over 200 hours doing so. In fact, less than half of website users claim to have *ever* read a privacy policy [49]. Studies that used self-reported data from users found that only 4.5% claim to always read them [50] and more reliable server side observation of websites revealed that only 1% or less of users click on a website's privacy policy [39]. Studies using advanced eye tracking techniques come to the same conclusion: users barely take the effort to read privacy policies thoroughly [68]. More recent work [54] demonstrated that three out of four users completely ignore privacy policies. Other users skim through policies that take 29 to 32 minutes to read in less than two minutes.

Researchers have also long criticized [30, 51, 23, 25] poor readability of online privacy policies. An average privacy policy was previously measured to need 12 to 14 years of schooling to comprehend [30, 51, 67, 2]. Many researchers have investigated privacy policy readability in general, and inside one sector in particular [20, 76, 30, 24, 69, 10, 5]. Prior work has also considered privacy policy readability across market sectors [42, 51, 25]. In addition, multiple studies of online privacy policies found that privacy policies are getting longer and harder to read, with their readability score decreasing over time [51, 33].

Therefore, it should not be surprising that researchers have attributed [23] the fact that users rarely read privacy policies to the lack of readability in these lengthy documents. After two decades of research, readability of privacy policies is still an important and relevant issue. First, the investigation of privacy policy readability across categories enlightens users and policy makers about how users read (or ignore) privacy policies. Second, the poor readability of these documents makes the case for the thriving field of automatic analysis of privacy policies (e.g., Polisis [31], Pribots [32], PrivacyCheck [88, 53, 82, 84], MAPS [94], PolicyLint [3], and PrivacyGuide [71]), as we cover in Section 4. Informing users, for example by clearly displaying privacy policies [74, 55], motivates them to incorporate privacy into their online decisions. Therefore, it is vital to educate users through tools, information, and statistics about privacy policies.

4 Privacy Enhancing Technologies (PET) to Digest Privacy Policies

Privacy policies are lengthy and hard to read, yet are profoundly important as they communicate the practices of an organization pertaining to user data privacy. To address the poor readability of privacy policies, an emerging field of research focuses on Privacy Enhancing Technologies (PETs) that summarize and visualize online privacy policies (e.g., Polisis [31], Pribots [32], PolicyLint [3], Privee [93], PrivacyGuide [71], tools from the Usable Privacy project [64], other similar tools and research [26, 94, 11], and our own publicly available PrivacyCheck [88, 84, 53]). Rich PET tools not only inform users about details of privacy policies, but also empower them to understand privacy policies at a higher level, make informed decisions, and even select competitors with better

Copyright 2021. The Center for Identity. All rights reserved.

privacy policies.

The flourishing field of PET development has resulted in research and (sometimes publicly accessible) tools that digest long privacy policies and automatically answer questions about them. These tools utilize Machine Learning (ML), Natural Language Processing (NLP), crowd-sourcing, etc. In this section, we review the most related PET tools and research.

Privee [93] was the first automatic privacy policy analysis tool to utilize machine learning. Building on the crowd sourcing privacy analysis framework ToS;DR [72], Privee combines crowd sourcing with rule and machine learning classifiers to classify privacy policies that are not already rated in the crowd sourcing repository. Privee, however, does not go beyond this basic analysis.

Polisis, available as a web page[6] and a browser extension, utilizes deep learning to summarize what user data privacy policies collect and share. At its core, Polisis is a neural network classifier trained on privacy policies retrieved from the Google Play store. In addition to providing the summary, Polisis visualizes user data collection/sharing, mapping types of data the policy collects/shares to the collection/sharing reasons outlined therein. Furthermore, Polisis displays user choices, security, data retention, etc. as graphs, making it easier for the user to comprehend what is covered in the *current* privacy policy. Notably, Polisis particularly extracts statements about how the policy claims to handle changes in its content. None of these capabilities, nonetheless, go beyond the analysis of the current policy at hand. Even the "policy change" is limited to information extraction from the current privacy policy. Pribots [32] is from the authors of Polisis and is a chat bot that answers free form questions about a given privacy policy.

The Usable Privacy Project[7] [64] takes advantage of machine learning and crowd sourcing to semi-automatically annotate privacy policies. This project annotates [79, 78] a corpus of 115 policies with attributes and data practices, the same corpus that Polisis and Pribots use to extract coarse- and fine-grained classes.

PolicyLint [3] is a natural language processing tool that identifies potential contradictions that may arise inside the same privacy policy. PrivacyGuide [71] (not publicly available) is a machine learning and natural language processing tool inspired by the GDPR. PolicyLint, PrivacyGuide, and many other recently developed tools [13, 11] are solely focused on automatic extraction of information from *one* privacy policy.

Researchers have also investigated the consistency, or lack thereof, between privacy policies of mobile applications and how their actual code treats user data [94, 95]. As one prominent example, MAPS [94] analyzes privacy policies of more than one million mobile applications.

At the Center for Identity at the University of Texas at Austin[8] we target many aspects of identity management and privacy [90, 87, 89, 91, 59, 89]. We

[6] `https://pribot.org/polisis`
[7] `https://usableprivacy.org`
[8] `https://identity.utexas.edu`

Copyright 2021. The Center for Identity. All rights reserved.

developed PrivacyCheck v1 [88], v2 [84, 53], and v3 [81] as detailed in the next section.

4.1 PrivacyCheck

PrivacyCheck is a publicly available browser extension that summarizes privacy policies with machine learning. It automatically answers twenty questions, rooted in the FIPPs (Fair Information Practice Principles) and GDPR (European General Data Protection Regulation). Our previous work covered how we chose these questions and trained LightGBM machine learning models for them (FIPPs questions [88] and GDPR questions [84, 53]). We have also applied PrivacyCheck in a variety of applications: e.g., to study the effect of the GDPR on the landscape of privacy policies [84], to compare privacy policies in the public and private sectors [80], to study privacy policies across industries [83], and to study PET usage patterns [82].

To use PrivacyCheck, the user navigates to a web page using the Chrome browser and then opens and runs the PrivacyCheck Chrome extension. PrivacyCheck's machine learning models digest the privacy policy and assign two scores to it, one for the (FIPPs-based) User Control and one for the GDPR standards. Clicking on each of the scores takes the user to score breakdowns explaining why the privacy policy received this score, based on the questions and their corresponding answers according to this privacy policy.

PrivacyCheck is free and publicly available online[9] and can also be found by searching for "PrivacyCheck" on the Google Chrome Web Store[10]. PrivacyCheck currently has over 800 users around the globe.

The first version of PrivacyCheck [88] used to summarize privacy policies based on ten *User Control* privacy questions that were rooted in the work of the Organization for Economic Cooperation and Development [61], and the Federal Trade Commission Fair Information Practices [28]. The second version of PrivacyCheck added ten new *GDPR* questions [84, 53]. PrivacyCheck v2 also added (1) a consumer-facing tool with higher performance, (2) new interface, and (3) the ability to find the top three competitors with better privacy policies (according to the User Control or GDPR standards) in the same market sector as of the privacy policy under evaluation. Finally, new capabilities were introduced in PrivacyCheck v3. In particular, the third version added the ability to (1) find the competitors of an organization with Alexa traffic analysis and compare policies across them, (2) follow privacy policies the user has agreed to and notify the user when policies change, (3) track policies over time and report how often policies change and their trends, (4) automatically find privacy policies in domains, and (5) provide a bird's-eye view of privacy policies the user has agreed to.

[9] `https://tinyurl.com/ydf7h7dr`
[10] `https://chrome.google.com/webstore`

Copyright 2021. The Center for Identity. All rights reserved.

4.2 Corpora of Privacy Policies

The studies that investigate privacy policies (e.g., their readability, or the effect of regulation on them) and the more recent body of PET that utilizes machine learning and natural language processing to automatically summarize privacy policies greatly benefit, if not rely on, corpora of privacy policies collected from the web or mobile app stores. Such corpora of privacy policies are valuable tools at the researchers' disposal to investigate privacy policies. For example, they facilitates comparison among different methods of privacy policy summarization by providing benchmarks, and can be used in unsupervised machine learning to summarize privacy policies. Any tool or research technique that addresses the length and poor readability of these policies, such as those that apply machine learning, natural language processing, and crowd-sourcing to automatically summarize privacy policies (e.g., [8, 70, 31, 88, 94, 4]), would require (or benefit from) large corpora of privacy policies.

Many studies have privately gathered corpora of privacy policies and some have publicly shared them with the research community. Some researchers have dedicated their attention to manually annotating privacy policy corpora. The fact that these corpora are manually annotated by researchers or crowd-sourced workers is prohibiting them from including more than a couple hundred privacy policies. Two of the most widely used privacy policy corpora are OPP-115 [78] and APP-350 [94], containing 115 and 350 privacy policies respectively. There exist other manually labeled corpora containing less than 1K policies (e.g., 236 policies [8], 400 policies [88, 53, 83, 84], 45 policies [71], and 64 policies [16]). While valuable for supervised machine learning, these corpora fall short when it comes to unsupervised machine learning, natural language processing, and testing/validation because of their limited size.

Seeking to put together larger corpora of privacy policies, one could collect privacy policies from (1) the web or (2) mobile app stores. Interestingly, there are multiple corpora of *mobile app* (particularly Google Play) privacy policies that are available. For example, Kumar et al. used 150K [40] policies from Google Play and Sunyaev et al. [69] considered the privacy policies of 183 health iOS and Android apps. Notable is the MAPS framework [94], which evaluated the privacy policies of over one million Android apps and released 441,626 app privacy policies with their app categories. Mobile app privacy policies have received a lot of attention, arguably, among other reasons, because of the research that analyzes a mobile app's code alongside its privacy policy [22, 34, 6, 3, 4]. Privacy policies of websites, nonetheless, are equally important. There are meaningful differences between the contents of web privacy policies and mobile app privacy policies. For instance, the use of cookies is more applicable to web privacy policies or, generally speaking, mobile apps can obtain finer grade location information when compared to websites and should address how they deal with this location information in their privacy policies.

There have been recent efforts to curate very large corpora of *web* privacy policies [67, 2, 86]. Among corpora gathered from the web, some are larger but not available to the public—for instance, corpora of 9,295 policies [96] and 130K

Copyright 2021. The Center for Identity. All rights reserved.

policies [31]. Some smaller corpora of web privacy policies are made publicly available, for example, a corpus of 1,010 policies [58]. Srinath and his colleagues [67] created a corpus of one million privacy policies. They crawled the web for links with the words "privacy" or "data protection" in the URL. We [86] recently published a dataset of over 100K English website privacy policies across 15 categories. These categories come from the DMOZ (now known as Curlie[11]) project—a community of volunteers who created, manually categorized, and maintained a collection of over 1.5 million links according to a hierarchical ontology. Our use of Curlie/DMOZ is particularly advantageous—we are reusing a manual categorization already done by volunteers.

4.3 PET Usage Patterns

There are a variety of PET tools to summarize privacy policies, but how do actual users take advantage of these tools? We were the first to monitor the usage patterns of about a thousand actual PrivacyCheck users, the first work to track the usage and traffic of an ML-based privacy analysis tool. Results show: (1) there is a good number of privacy policy URLs checked repeatedly by the user base; (2) the users are particularly interested in privacy policies of software services; and (3) PrivacyCheck increased the number of times a user consults privacy policies by 80%. Our work demonstrates the potential of ML-based privacy analysis tools and also sheds light on how these tools are used in practice to give users actionable knowledge they can use to pro-actively protect their privacy.

We observe that PrivacyCheck users have a tendency to check the same privacy policies, presumably the most commonly used services or the most important sources of privacy concerns. There is, generally, a good number of URLs investigated by the PrivacyCheck user base more than once: among 534 calls to PrivacyCheck in a given period of time, only 366 (68%) were on unique URLs.

We find it fascinating that the user base of PrivacyCheck was disproportionately interested in running it on a variety of software service privacy policies (online social networks, large software companies, and predominantly smaller software companies) versus any other market sector/category. We think that the reason is the sheer amount of information such services can collect from their users.

Finally, we found that a typical PrivacyCheck user investigates 1.8% of all the new websites he/she visits in a year. Previous literature estimates a typical (non PrivacyCheck) web user investigates 1% of the privacy policies of new websites visited. We concluded that PrivacyCheck has the potential to increase the number of times a user consults a privacy policy and, consequently, increase their knowledge about an organization's privacy commitments.

[11] `https://curlie.org`

Copyright 2021. The Center for Identity. All rights reserved.

4.4 Other Related PET Tools

In this section, we review tools, services, and privacy enhancing technologies that help users protect their privacy, but do not focus on summarizing *privacy policy text*, in the following categories:

1. Privacy seals require web page operators to enroll in order to evaluate their privacy policies.
2. New formats encourage web page operators to adopt machine-readable notation of privacy policies to be automatically interpreted.
3. Crowd sourced services have an online community that reads and rates privacy policies.
4. Tracking monitors observe web pages in action, instead of investigating their privacy policies.

4.4.1 Privacy Seals

Privacy seals are logos of organizations or agencies that evaluate and rate privacy policies. For example, TRUSTe [73] (now TrustArc) is a data privacy management company that examines privacy policies and helps businesses align their privacy policies with legal requirements. While the information provided by TRUSTe is manually extracted, TRUSTe and other similar services suffer from two major drawbacks: (1) they significantly lack comprehensive web coverage; even though TRUSTe owns the majority of the market share among all similar services it covers only roughly 55,000 in one million web pages[12], and (2) web page operators have to sign up and potentially pay for such services which hinders their universal adoption.

Another privacy seal is provided by Better Business Bureau (BBB) [9], a non-profit organization that provides free business reviews. However, BBB accredited businesses pay a fee for accreditation review.

Overall, researchers have expressed concerns with privacy seals in general: insufficient scrutiny of privacy seal organizations, negative self-selection of websites that participate in a seal, and users' ignorance regarding privacy seals [38].

4.4.2 New Formats

The Platform for Privacy Preferences Project (P3P) [77] is a standard for websites to express their privacy policies in a both human and machine-readable format. Following such standards enables automatic interpretation of privacy policies. P3P and similar standards require web page operators to adopt new formats. As a result, P3P has always suffered from lack of industry participation. Consequently, the P3P working group was closed in 2006 and P3P 1.1 was never finalized [17].

[12] `https://www.datanyze.com/market-share/security/truste-market-share`

Copyright 2021. The Center for Identity. All rights reserved.

Currently, P3P is used in only 70,000 in one million web pages[13]. The failure of P3P in attracting industry participation, however, is not limited to the number of websites that do not support it. A large fraction of the websites that support P3P chose to include a minimal and mis-representative version of their privacy policy just to prevent Internet Explorer, the major web browser supporting P3P, from blocking their cookies. In fact, thousands of websites were found to use an identical erroneous policy recommended by a Microsoft support website to avoid cookie blocking by Internet Explorer [17].

Researchers have attempted visualizing privacy policies represented in P3P (e.g., Privacy Bird extension for Internet Explorer [7, 18]), some with only little success in improving the comprehension [60]. Internet Explorer 6.0 included a feature to textually present privacy policies formatted as P3P.

Nutrition Label [37, 36, 17] introduced another new format that asks website operators to consolidate their privacy policies in a one page standardized format inspired by the nutrition facts panel found on food and drug packages. Complying with this new format also places an additional, unwanted burden on website operators. In 2020, Apple, Inc. moved to adopt the Nutrition Label for privacy policies of its mobile apps[14].

4.4.3 Crowd Sourced Services

Terms of Service; Didn't Read (ToS;DR) [72] is a free software project that started in 2012 to address the problem that very few users actually read the terms of service for websites they use. In this project, an online community of volunteers read, discuss, and rate privacy policies. The ratings and discussions are available online and as free software in the form of browser extensions for Mozilla Firefox, Google Chrome, Apple Safari, and Opera. Even though privacy policies addressed in this project are read and rated by humans and discussed thoroughly, the in-depth and occasionally long discussions pose a new challenge to the usefulness of the ratings: one might as well read, selectively, the original privacy policy itself. Furthermore, the coverage of ToS;DR is even more limited than privacy seals and new formats [72].

4.4.4 Tracking Monitors

Ghostery [29] is a software available as free browser extensions for Mozilla Firefox, Google Chrome, Internet Explorer, Opera, and Apple Safari. Ghostery tracks cookies, tags, web bugs, pixels, and beacons and then notifies the user of their presence as well as the companies that operate them, giving the user the choice to make informed decisions about blocking them. Similarly, Adblock Plus [1] is a free extension that blocks adds and disables tracking. Adblock Plus is available for Android, Google Chrome, Mozilla Firefox, Internet Explorer, Opera, and Apple Safari among other browsers. Ghostery, Adblock Plus, and

[13] `http://trends.builtwith.com/docinfo/P3P-Policy`
[14] `https://developer.apple.com/app-store/app-privacy-details`

Copyright 2021. The Center for Identity. All rights reserved.

other similar services are fundamentally different from PrivacyCheck and ML-based PET tools, in that they focus on the actions a website takes, instead of the legal privacy policy it posts. Hopefully, such actions are aligned with the privacy policy, but that is not guaranteed. In addition, tracking monitors do not indicate the usage of the information gathered from users.

5 Privacy Regulations

As we mentioned earlier, regulatory bodies around the world have long enforced requirements on online privacy policies. Over the past few years, newer laws have gone into effect to protect user privacy. The most important examples of such laws are the General Data Protection Regulation (GDPR)[15] in the European Union and the California Consumer Privacy Act (CCPA)[16] in the United States.

5.1 Regulations Governing Privacy Policies

In this section, we list some recent and old privacy policy regulations, with an emphasis on the United States (US) and the European Union (EU).

- The GDPR is the newest regulation in the EU law on data protection and privacy. The key principles of the GDPR are (1) Lawfulness, fairness, and transparency, (2) Purpose limitation, (3) Data minimization, (4) Accuracy, (5) Storage limitation, (6) Security, and (7) Accountability.
- The California Consumer Privacy Act (CCPA) is a state statute to enhance privacy rights and consumer protection for residents of California.
- The International Safe Harbor Privacy Principles were principles (overturned in 2015) developed to prevent private organizations within the European Union or United States which store customer data from accidentally disclosing or losing personal information.
- The Federal Trade Commission (FTC) Fair Information Practice Principles (FIPP) are recommendations, though not legally enforced, for maintaining privacy-friendly, consumer-oriented data collection practices and include Notice, Choice, Access, and Integrity.
- The Children's Online Privacy Protection Act (COPPA) protects the personal information of children under 13.
- The Gramm-Leach-Bliley Act (GLBA) requires financial institutions (i.e. companies) to explain their information-sharing practices to their customers and to safeguard sensitive data.

[15] `https://gdpr-info.eu`
[16] `https://oag.ca.gov/privacy/ccpa`

Copyright 2021. The Center for Identity. All rights reserved.

- The Health Insurance Portability and Accountability Act (HIPPA) applies to health care providers, suppliers and vendors (business associates).
- The Telephone Consumer Protection Act (TCPA) regulates the collection of information by telephone service providers.
- The Privacy Act of 1974 is the primary law in the US that governs government collection, maintenance, use, and dissemination of PII by federal agencies.
- The Freedom of Information Act (FOIA) governs the collection, maintenance, use, and dissemination of PII that is maintained in systems of records by federal agencies.
- Federal Information Security Management Act (FISMA) mandates that each federal agency implements an information security program for the information and information systems that support the operations and assets of the agency.
- The Electronic Communications Privacy Act (ECPA) restricts the government's access and disclosure of electronic communication.

5.2 The Effect of Privacy Regulations such as the GDPR on Online Privacy Policies

The research community has looked into quantifying the effect of privacy policy regulations on the landscape of online privacy. Interestingly, the PET tools we discussed earlier have found a new use in addition to assisting final consumers in understanding privacy policies: researchers have leveraged these tools to automatically analyze huge corpora of privacy policies and study their statistics. For instance, both Polisis and PrivacyCheck have been utilized to study the effect of the GDPR on the privacy landscape [44, 84, 85].

The GDPR is considered by some to be the most important change in data privacy regulation in 20 years. The European Union GDPR privacy law applies to any organization that collects and processes the personal information of EU citizens within or outside the EU. While the GDPR is a European Union law, it covers any organization that collects or processes EU citizen data independent of the organization's location. Due to the global nature of commerce and people's movements, the GDPR drove businesses around the world to make important decisions and changes regarding how they collect and process their employee's and customers' PII.

The GDPR went into effect on May 25, 2018. As a result, many companies that do business completely or partially in the EU or handle EU citizens' data updated their online privacy policies around the same time in order to comply with the GDPR [21]. It is also important to note that the GDPR has already motivated other advancements in privacy regulation and continues to do so around the globe as consumers demand their data rights[17]. Evaluating the

[17] California Consumer Privacy Act (CCPA), `https://www.caprivacy.org`

Copyright 2021. The Center for Identity. All rights reserved.

actual effect of the GDPR on online privacy policies of companies is a significant research question, which has not received the attention it deserves.

Degeling and colleagues are among the few who studied a sizable set of privacy policies (from more than 6,500 websites) across the EU [21] with the help of automation, and compared policies before and after the GDPR. That work, however, focused exclusively on the consent to use cookies in privacy policies. Libert et al. [43] studied privacy policies of seven countries of the EU before and after the GDPR went into effect, but also focused on cookies only.

A closely related work comes from Linden et al. [45] which examines policies from both inside and outside the EU. While they use over 6,000 privacy policies, with the exception of the visual representation evaluation that is done by Amazon MTurk users, the rest of the compliance assessment is done automatically. The automation is in turn based on Polisis [31], a deep learning tool developed by the authors. They concluded that, even though the GDPR has prompted a general overhaul in privacy policies, many policies still do not meet several GDPR requirements. The authors of Claudette [15] compare pre- and post-GDPR privacy policies with a data mining tool, but their work is limited to only few policies (from 14 [15] to 50 policies [46]).

In a recent work [84], we quantified the progress the GDPR has made in improving privacy policies around the globe. We leveraged our data mining tool, PrivacyCheck, to automatically compare three corpora (totaling 550) of privacy policies, pre- and post-GDPR. In addition, to evaluate the current level of compliance with the GDPR around the globe, we manually studied the policies within two corpora (450 policies).

We applied PrivacyCheck to compare the privacy policies before and after the GDPR through PrivacyCheck's ten questions to assess the impact of the GDPR. We manually study these privacy policies after the GDPR went into effect to measure how close the policies are to full compliance with the GDPR. We distill another ten questions, previously not supported by PrivacyCheck, directly from the GDPR.

In this work, we examined the landscape of online privacy policies, to evaluate the effect of the GDPR as well as to paint a clearer picture of data privacy and best practices regarding privacy policies and their level of compliance with the GDPR. The verdict is that, with modest changes, most of the privacy policies were able to satisfy many but not all GDPR requirements. The most notable non-compliance of the investigated privacy policies was found when policies fail to indicate compliance with a GDPR requirement, either affirmatively or negatively. Consequently, the most notable non-compliance with the GDPR results when an organization lacks transparency and explicit disclosure of their processing and protection of consumer personal information.

Based on the manual examination of policies, when non-compliance does appear, it is often in failing to explicitly indicate compliance. We identify the following areas for improvement in privacy policies:

- Many policies fail to mention whether they encrypt data while at rest.
- Many do not mention if and when they notify the supervisory authority

Copyright 2021. The Center for Identity. All rights reserved.

in case of a data breach.

- Some US-based policies protect children under 13 as per US regulation, and hence do not necessarily protect children between 13 and 16 years of age.
- Privacy policies lack a 100% consensus to require informed consent.

We reported that the GDPR has made progress in protecting user data, but more progress is necessary—particularly in the area of giving users the right to edit and delete their information—to entirely fulfill the GDPR's promise. We also observed that the GDPR encourages sharing user data with law enforcement, and, as a result, many policies have facilitated such sharing after the GDPR. Finally, we saw that when there is non-compliance with the GDPR, it is often in the form of failing to explicitly indicate compliance, which in turn speaks to an organization's lack of transparency and disclosure regarding their processing and protection of personal information.

We found that websites have modestly changed their privacy policies after the GDPR. Changes have often been geared toward compliance. This is not surprising since the penalties for noncompliance far exceed prior regulations. Overall, the landscape of privacy policies after the GDPR is promising when considering all the privacy factors measured by PrivacyCheck. For many factors, over 90% of the privacy policies we considered were in compliance with the GDPR.

6 Conclusions and Future Work

Personal data or PII–the "little i"–is directly tied to an individual human–the "big I"– giving individuals benefits or posing threats. Thus, we must protect personal data like we protect humans.

In today's world, where a huge amount of information is collected about individuals, achieving privacy through anonymity is extremely difficult if not impossible and achieving such anonymity will involves losing access to many goods, systems, and services. As anonymity is near impossible and interferes with full participation in a society, privacy rights over one's personal data are essential. The rights of agency and control are central to these protections.

To protect *individual users'* privacy rights, in particular when *companies* collect user data, privacy laws put in place by *regulators* are critical. In fact, users, companies, and regulators are the three main stakeholders required to protect privacy rights.

Privacy enhancing technologies, such as PrivacyCheck, are critical to ensuring individuals can implement privacy controls and defend their privacy rights. These privacy enhancing technologies can also serve both companies and regulators, to gauge the adherence to privacy laws.

Laws are changing to offer more privacy protections. The two most recent examples are the GDPR in the European Union and the CCPA in the United

Copyright 2021. The Center for Identity. All rights reserved.

States. As these privacy laws change, it is paramount to study their tangible effects to guide further regulation, ensure compliance, and inform and empower users.

All in all, the privacy enhancing technologies and privacy regulations have made progress in educating users and enforcing data protection regulations. More work is necessary to enable and empower users to make informed decisions with respect to privacy policies. For future work, we envision answering multiple sets of questions with PrivacyCheck to better inform users to understand their privacy risks and provide actionable knowledge to use their usage and buying power to select products and services offering by privacy-respecting organizations. Privacy enhancing tools like PrivacyCheck aim to give users the agency and control necessary to protect and defend their privacy rights. Finally, pertaining to the privacy regulations like the GDPR, more work is necessary, particularly in the area of granting users the right to edit, update, and delete their data, to entirely fulfill the GDPR's promise.

References

[1] AdblockPlus. Adblock plus surf the web without annoying ads!, 2021.

[2] Ryan Amos, Gunes Acar, Elena Lucherini, Mihir Kshirsagar, Arvind Narayanan, and Jonathan Mayer. Privacy policies over time: Curation andanalysis of a million-document dataset. *arXiv preprint arXiv:2008.09159*, 2020.

[3] Benjamin Andow, Samin Yaseer Mahmud, Wenyu Wang, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Tao Xie. Policylint: investigating internal privacy policy contradictions on Google play. In *28th USENIX Security Symposium*, pages 585–602, 2019.

[4] Benjamin Andow, Samin Yaseer Mahmud, Justin Whitaker, William Enck, Bradley Reaves, Kapil Singh, and Serge Egelman. Actions speak louder than words: Entity-sensitive privacy policy and data flow analysis with policheck. In *29th {USENIX} Security Symposium ({USENIX} Security 20)*, pages 985–1002, 2020.

[5] Annie I Anton, Julia Brande Earp, Qingfeng He, William Stufflebeam, Davide Bolchini, and Carlos Jensen. Financial privacy policies and the need for standardization. *IEEE Security & privacy*, 2(2):36–45, 2004.

[6] Steven Arzt, Siegfried Rasthofer, Christian Fritz, Eric Bodden, Alexandre Bartel, Jacques Klein, Yves Le Traon, Damien Octeau, and Patrick McDaniel. Flowdroid: Precise context, flow, field, object-sensitive and lifecycle-aware taint analysis for android apps. *Acm Sigplan Notices*, 49(6):259–269, 2014.

[7] AT&T. Privacy bird, 2002.

Copyright 2021. The Center for Identity. All rights reserved.

[8] Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, et al. Finding a choice in a haystack: Automatic extraction of opt-out statements from privacy policy text. In *Proceedings of The Web Conference 2020*, pages 1943–1954, 2020.

[9] BBBOnline. Better business bureau, 2021.

[10] Jasmine Bowers, Bradley Reaves, Imani N Sherman, Patrick Traynor, and Kevin Butler. Regulators, mount up! analysis of privacy policies for mobile money services. In *Thirteenth Symposium on Usable Privacy and Security ({SOUPS} 2017)*, pages 97–114, 2017.

[11] Vanessa Bracamonte, Seira Hidano, Welderufael B Tesfay, and Shinsaku Kiyomoto. Evaluating the effect of justification and confidence information on user perception of a privacy policy summarization tool. In *ICISSP*, pages 142–151, 2020.

[12] Duncan H Brown and Jeffrey Layne Blevins. The safe-harbor agreement between the united states and europe: A missed opportunity to balance the interests of e-commerce and privacy online? *Journal of Broadcasting & Electronic Media*, 46(4):549–564, 2002.

[13] Duc Bui, Kang G Shin, Jong-Min Choi, and Junbum Shin. Automated extraction and presentation of data practices in privacy policies. *Proceedings on Privacy Enhancing Technologies*, 2021(2):88–110, 2021.

[14] Jiyoung Cha. Information privacy: a comprehensive analysis of information request and privacy policies of most-visited web sites. *Asian Journal of Communication*, 21(6):613–631, 2011.

[15] Giuseppe Contissa, Koen Docter, Francesca Lagioia, Marco Lippi, Hans-W Micklitz, Przemysław Pałka, Giovanni Sartor, and Paolo Torroni. Claudette meets GDPR: Automating the evaluation of privacy policies using artificial intelligence. 2018.

[16] Elisa Costante, Yuanhao Sun, Milan Petković, and Jerry Den Hartog. A machine learning solution to assess privacy policy completeness: (short paper). In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, pages 91–96, 2012.

[17] Lorrie Faith Cranor. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *J. on Telecomm. & High Tech. L.*, 10:273, 2012.

[18] Lorrie Faith Cranor, Praveen Guduru, and Manjula Arjula. User interfaces for privacy agents. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(2):135–178, 2006.

Copyright 2021. The Center for Identity. All rights reserved.

[19] Mary J Culnan. Georgetown Internet privacy policy survey: report to the Federal Trade Commission. *Washington, DC: Georgetown University, The McDonough School of Business*, 1999.

[20] Gitanjali Das, Cynthia Cheung, Camille Nebeker, Matthew Bietz, and Cinnamon Bloss. Privacy policies for apps targeted toward youth: descriptive analysis of readability. *JMIR mHealth and uHealth*, 6(1):e3, 2018.

[21] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy... now take some cookies-measuring the gdpr's impact on web privacy. *Informatik Spektrum: Vol. 42, No. 5*, 2019.

[22] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. *ACM Transactions on Computer Systems (TOCS)*, 32(2):1–29, 2014.

[23] Tatiana Ermakova, Annika Baumann, Benjamin Fabian, and Hanna Krasnova. Privacy policies and users' trust: does readability matter? In *Proceedings of the Twentieth Americas Conference on Information Systems*, page 12, 2014.

[24] Tatiana Ermakova, Benjamin Fabian, and Eleonora Babina. Readability of privacy policies of healthcare websites. *Wirtschaftsinformatik*, 15, 2015.

[25] Benjamin Fabian, Tatiana Ermakova, and Tino Lentz. Large-scale readability analysis of privacy policies. In *Proceedings of the International Conference on Web Intelligence*, pages 18–25, 2017.

[26] Kassem Fawaz, Thomas Linden, and Hamza Harkous. The applications of machine learning in privacy notice and choice. In *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, pages 118–124. IEEE, 2019.

[27] FTC. Privacy online: A report to congress, 1998.

[28] FTC. Privacy online: Fair information practices in the electronic marketplace: A federal trade commission report to congress, 2000.

[29] Ghostery. Join over 40 million ghostery users and download the web's most popular privacy tool., 2021.

[30] Mark A Graber, Donna M D Alessandro, and Jill Johnson-West. Reading level of privacy policies on internet health web sites. *Journal of Family Practice*, 51(7):642–642, 2002.

Copyright 2021. The Center for Identity. All rights reserved.

[31] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G Shin, and Karl Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX Security Symposium*, pages 531–548, 2018.

[32] Hamza Harkous, Kassem Fawaz, Kang G Shin, and Karl Aberer. Pribots: Conversational privacy with chatbots. In *Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016*, 2016.

[33] Carlos Jensen and Colin Potts. Privacy policies as decision-making tools: an evaluation of online privacy notices. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 471–478, 2004.

[34] Daniel Kales, Christian Rechberger, Thomas Schneider, Matthias Senker, and Christian Weinert. Mobile private contact discovery at scale. In *28th USENIX Security Symposium (USENIX Security 19)*, pages 1447–1464, 2019.

[35] Jerry Kang. Information privacy in cyberspace transactions. *Stan. L. Rev.*, 50:1193, 1997.

[36] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W Reeder. A nutrition label for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, page 4. ACM, 2009.

[37] Patrick Gage Kelley, Lucian Cesca, Joanna Bresee, and Lorrie Faith Cranor. Standardizing privacy notices: an online study of the nutrition label approach. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 1573–1582. ACM, 2010.

[38] Alfred Kobsa. Privacy-enhanced web personalization. In *The adaptive web*, pages 628–670. Springer, 2007.

[39] Ron Kohavi. Mining e-commerce data: the good, the bad, and the ugly. In *International conference on Knowledge discovery and data mining*, pages 8–13. ACM, 2001.

[40] Vinayshekhar Bannihatti Kumar, Abhilasha Ravichander, Peter Story, and Norman Sadeh. Quantifying the effect of in-domain distributed word representations: A study of privacy policies. In *AAAI Spring Symposium on Privacy-Enhancing Artificial Intelligence and Language Technologies*, 2019.

[41] Yuanxiang Li, Walter Stweart, Jake Zhu, and Anna Ni. Online privacy policy of the thirty Dow Jones corporations: Compliance with FTC fair information practice principles and readability assessment. *Communications of the IIMA*, 12(3):5, 2012.

[42] Yuanxiang Li, Walter Stweart, Jake Zhu, and Anna Ni. Online privacy policy of the thirty dow jones corporations: Compliance with ftc fair information practice principles and readability assessment. *Communications of the IIMA*, 12(3):5, 2012.

Copyright 2021. The Center for Identity. All rights reserved.

[43] Timothy Libert, Lucas Graves, and Rasmus Kleis Nielsen. Changes in third-party content on european news websites after gdpr. 2018.

[44] Thomas Linden, Hamza Harkous, and Kassem Fawaz. The privacy policy landscape after the gdpr. *arXiv preprint arXiv:1809.08396*, 2018.

[45] Thomas Linden, Rishabh Khandelwal, Hamza Harkous, and Kassem Fawaz. The privacy policy landscape after the gdpr. *Proceedings on Privacy Enhancing Technologies*, 2020(1):47–64, 2020.

[46] Marco Lippi, Przemysław Pałka, Giuseppe Contissa, Francesca Lagioia, Hans-Wolfgang Micklitz, Giovanni Sartor, and Paolo Torroni. Claudette: an automated detector of potentially unfair clauses in online terms of service. *Artificial Intelligence and Law*, 27(2):117–139, 2019.

[47] Chang Liu and Kirk P Arnett. Raising a red flag on global www privacy policies. *Journal of Computer Information Systems*, 43(1):117–127, 2002.

[48] Aleecia M McDonald and Lorrie Faith Cranor. the cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4:543, 2008.

[49] David B Meinert, Dane K Peterson, John R Criswell, and Martin D Crossland. Privacy policy statements and consumer willingness to provide personal information. *Journal of Electronic Commerce in Organizations*, 4(1):1, 2006.

[50] George R Milne and Mary J Culnan. Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing*, 18(3):15–29, 2004.

[51] George R Milne, Mary J Culnan, and Henry Greene. A longitudinal assessment of online privacy notice readability. *Journal of Public Policy & Marketing*, 25(2):238–249, 2006.

[52] James H Moor. Towards a theory of privacy in the information age. *ACM Sigcas Computers and Society*, 27(3):27–32, 1997.

[53] Razieh Nokhbeh Zaeem, Safa Anya, Alex Issa, Jake Nimergood, Isabelle Rogers, Vinay Shah, Ayush Srivastava, and K Suzanne Barber. Privacycheck v2: A tool that recaps privacy policies for you. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 3441–3444, 2020.

[54] Jonathan A Obar and Anne Oeldorf-Hirsch. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1):128–147, 2020.

[55] Yue Pan and George M Zinkhan. Exploring the impact of online privacy disclosures on consumer trust. *Journal of retailing*, 82(4):331–338, 2006.

Copyright 2021. The Center for Identity. All rights reserved.

[56] Robert C Post. Three concepts of privacy. *Geo. LJ*, 89:2087, 2000.

[57] Stephen A Rains and Leslie A Bosch. Privacy and health in the information age: A content analysis of health web site privacy policy statements. *Health communication*, 24(5):435–446, 2009.

[58] Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah A Smith. Unsupervised alignment of privacy policies using hidden markov models. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 605–610, 2014.

[59] Rima Rana, Razieh Nokhbeh Zaeem, and K Suzanne Barber. An assessment of blockchain identity solutions: Minimizing risk and liability of authentication. In *2019 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 26–33, 2019.

[60] Robert W Reeder, Patrick Gage Kelley, Aleecia M McDonald, and Lorrie Faith Cranor. A user study of the expandable grid applied to p3p privacy policy visualization. In *Proceedings of the 7th ACM workshop on Privacy in the electronic society*, pages 45–54. ACM, 2008.

[61] Having Regard. Recommendation of the council concerning guidelines governing the protection of privacy and transborder flows of personal data. 1980.

[62] Sasha Romanosky, Rahul Telang, and Alessandro Acquisti. Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2):256–286, 2011.

[63] Randy Ryker, Elizabeth Lafleur, Bruce McManis, and K Chris Cox. Online privacy policies: An assessment of the fortune e-50. *Journal of Computer Information Systems*, 42(4):15–20, 2002.

[64] Norman Sadeh, Alessandro Acquisti, Travis D Breaux, Lorrie Faith Cranor, Aleecia M McDonalda, Joel R Reidenbergb, Noah A Smith, Fei Liu, N Cameron Russellb, Florian Schaub, et al. The usable privacy policy project. Technical report, Technical Report, CMU-ISR-13-119, Carnegie Mellon University, 2013.

[65] Zeinab Karake Shalhoub. Content analysis of web privacy policies in the gcc countries. *Information Systems Security*, 15(3):36–45, 2006.

[66] Daniel J Solove. *Understanding privacy*. Harvard University Press, May, 2008.

[67] Mukund Srinath, Shomir Wilson, and C Lee Giles. Privacy at scale: Introducing the privaseer corpus of web privacy policies. *arXiv preprint arXiv:2004.11131*, 2020.

Copyright 2021. The Center for Identity. All rights reserved.

[68] Nili Steinfeld. "I agree to the terms and conditions":(how) do users read privacy policies online? an eye-tracking experiment. *Computers in human behavior*, 55:992–1000, 2016.

[69] Ali Sunyaev, Tobias Dehling, Patrick L Taylor, and Kenneth D Mandl. Availability and quality of mobile health app privacy policies. *Journal of the American Medical Informatics Association*, 22(e1):e28–e33, 2015.

[70] Welderufael B Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, and Jetzabel Serna. I read but don't agree: Privacy policy benchmarking using machine learning and the eu gdpr. In *Companion Proceedings of the The Web Conference 2018*, pages 163–166. International World Wide Web Conferences Steering Committee, 2018.

[71] Welderufael B Tesfay, Peter Hofmann, Toru Nakamura, Shinsaku Kiyomoto, and Jetzabel Serna. Privacyguide: towards an implementation of the EU GDPR on internet privacy policy evaluation. In *Proceedings of the Fourth ACM International Workshop on Security and Privacy Analytics*, pages 15–21. ACM, 2018.

[72] ToS;DR. Terms of service; didn't read, 2012.

[73] TRUSTe. Truste, 2021.

[74] Janice Y Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study. *Information systems research*, 22(2):254–268, 2011.

[75] European Union. European union law.

[76] Pelayo Vallina, Álvaro Feal, Julien Gamba, Narseo Vallina-Rodriguez, and Antonio Fernández Anta. Tales from the porn: A comprehensive privacy analysis of the web porn ecosystem. In *Proceedings of the Internet Measurement Conference*, pages 245–258, 2019.

[77] W3C. The platform for privacy preferences 1.1 (p3p1.1) specification, 2006.

[78] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N Cameron Russell, et al. The creation and analysis of a website privacy policy corpus. In *Annual Meeting of the Association for Computational Linguistics*, pages 1330–13340, 2016.

[79] Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A Smith, and Frederick Liu. Crowdsourcing annotations for websites' privacy policies: Can it really work? In *Proceedings of the 25th International Conference on World Wide Web*, pages 133–143, 2016.

Copyright 2021. The Center for Identity. All rights reserved.

[80] Razieh Zaeem and K. Barber. Comparing privacy policies of government agencies and companies: A study using machine-learning-based privacy policy analysis tools. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART,*, pages 29–40. INSTICC, SciTePress, 2021.

[81] Razieh Nokhbeh Zaeem, Ahmad Ahbab, Josh Bestor, Hussam H. Djadi, Sunny Kharel, Victor Lai, Nick Wang, and K. Suzanne Barber. Privacycheck v3: Empowering users with higher-level understanding of privacy policies. In *20th Workshop on Privacy in the Electronic Society (WPES 21)*, 2021. Under Submission.

[82] Razieh Nokhbeh Zaeem, Safa Anya, Alex Issa, Jake Nimergood, Isabelle Rogers, Vinay Shah, Ayush Srivastava, and K Suzanne Barber. Privacycheck's machine learning to digest privacy policies: Competitor analysis and usage patterns. In *2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 291–298. IEEE, 2020.

[83] Razieh Nokhbeh Zaeem and K Suzanne Barber. A study of web privacy policies across industries. *Journal of Information Privacy and Security*, 13(4):169–185, 2017.

[84] Razieh Nokhbeh Zaeem and K Suzanne Barber. The effect of the GDPR on privacy policies: Recent progress and future promise. *ACM Transactions on Management of Information Systems*, 2020.

[85] Razieh Nokhbeh Zaeem and K Suzanne Barber. Comparing privacy policies of government agencies and companies: A study using machine-learning-based privacy policy analysis tools. In *ICAART (2)*, pages 29–40, 2021.

[86] Razieh Nokhbeh Zaeem and K Suzanne Barber. A large publicly available corpus of website privacy policies based on dmoz. In *the 11th ACM Conference on Data and Application Security and Privacy*, 2021. To Appear, retrieved from `https://tinyurl.com/ybsq9qc8`.

[87] Razieh Nokhbeh Zaeem, Suratna Budalakoti, K Suzanne Barber, Muhibur Rasheed, and Chandrajit Bajaj. Predicting and explaining identity risk, exposure and cost using the ecosystem of identity attributes. In *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, pages 1–8. IEEE, 2016.

[88] Razieh Nokhbeh Zaeem, Rachel L German, and K Suzanne Barber. Privacycheck: Automatic summarization of privacy policies using data mining. *ACM Transactions on Internet Technology (TOIT)*, 18(4):53, 2018.

[89] Razieh Nokhbeh Zaeem, Monisha Manoharan, and K Suzanne Barber. Risk kit: Highlighting vulnerable identity assets for specific age groups. In *2016 European Intelligence and Security Informatics Conference (EISIC)*, pages 32–38. IEEE, 2016.

Copyright 2021. The Center for Identity. All rights reserved.

[90] Razieh Nokhbeh Zaeem, Monisha Manoharan, Yongpeng Yang, and K Suzanne Barber. Modeling and analysis of identity threat behaviors through text mining of identity theft stories. *Computers & Security*, 65:50–63, 2017.

[91] Jim Zaiss, Razieh Nokhbeh Zaeem, and K Suzanne Barber. Identity threat assessment and prediction. *Journal of Consumer Affairs*, 53(1):58–70, 2019.

[92] Xiaoni Zhang, Sakaguchi Toru, and Max Kennedy. A cross-cultural analysis of privacy notices of the global 2000. *Journal of Information Privacy and Security*, 3(2):18–36, 2007.

[93] Sebastian Zimmeck and Steven M. Bellovin. Privee: An architecture for automatically analyzing web privacy policies. In *23rd USENIX Security Symposium*, pages 1–16, San Diego, CA, Aug 2014. USENIX Association.

[94] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N Cameron Russell, and Norman Sadeh. Maps: Scaling privacy compliance analysis to a million apps. *Proceedings on Privacy Enhancing Technologies*, 2019(3):66–86, 2019.

[95] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman Sadeh, Steven Bellovin, and Joel Reidenberg. Automated analysis of privacy requirements for mobile apps. In *2016 AAAI Fall Symposium Series*, 2016.

[96] Sebastian Zimmeck, Ziqi Wang, Lieyong Zou, Roger Iyengar, Bin Liu, Florian Schaub, Shomir Wilson, Norman M Sadeh, Steven M Bellovin, and Joel R Reidenberg. Automated analysis of privacy requirements for mobile apps. In *NDSS*, 2017.

Copyright 2021. The Center for Identity. All rights reserved.

WWW.IDENTITY.UTEXAS.EDU