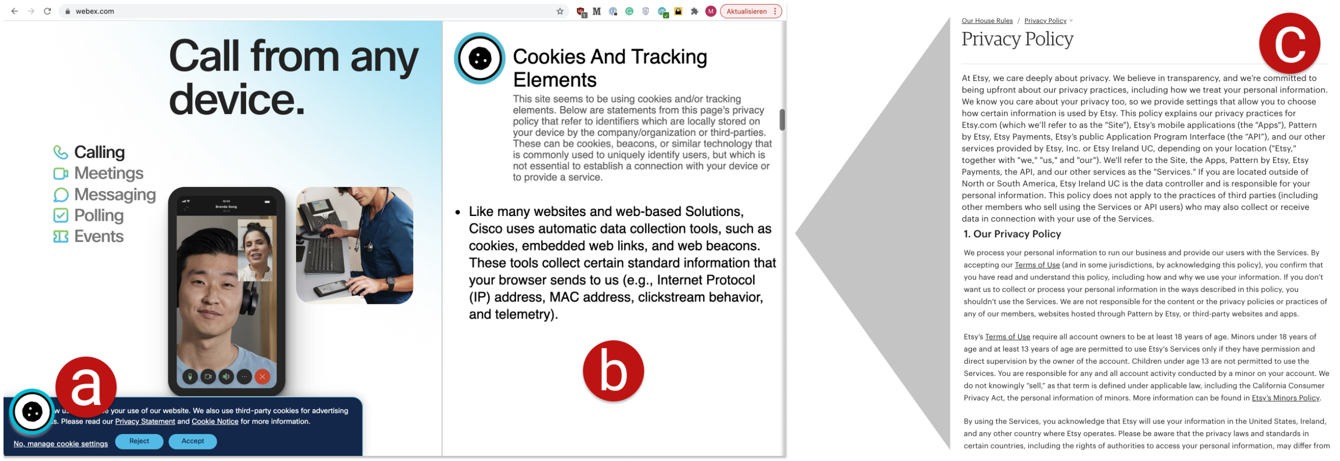
Automating Contextual Privacy Policies: Design and Evaluation of a Production Tool for Digital Consumer Privacy Awareness

Maximiliane Windl
LMU Munich
Munich, Germany
maximiliane.windl@ifi.lmu.de

Niels Henze
University of Regensburg
Regensburg, Germany
niels.henze@ur.de

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Sebastian S. Feger
LMU Munich
Munich, Germany
sebastian.feger@ifi.lmu.de



w u… your use of our website. We also use third-party cookies for advertising …s. Please read our Privacy Statement and Cookie Notice for more information.

Cookies And Tracking Elements

This site seems to be using cookies and/or tracking elements. Below are statements from this page's privacy policy that refer to identifiers which are locally stored on your device by the company/organization or third-parties. These can be cookies, beacons, or similar technology that is commonly used to uniquely identify users, but which is not essential to establish a connection with your device or to provide a service.

- Like many websites and web-based Solutions, Cisco uses automatic data collection tools, such as cookies, embedded web links, and web beacons. These tools collect certain standard information that your browser sends to us (e.g., Internet Protocol (IP) address, MAC address, clickstream behavior, and telemetry).

Privacy Policy

At Etsy, we care deeply about privacy. We believe in transparency, and we're committed to being upfront about our privacy practices, including how we treat your personal information. We know you care about your privacy too, so we provide settings that allow you to choose how certain information is used by Etsy. This policy explains our privacy practices for Etsy.com (which we'll refer to as the "Site"), Etsy's mobile applications (the "Apps"), Pattern by Etsy, Etsy Payments, Etsy's public Application Program Interface (the "API"), and our other services provided by Etsy, Inc. or Etsy Ireland UC, depending on your location ("Etsy," together with "we," "us," and "our"). We'll refer to the Site, the Apps, Pattern by Etsy, Etsy Payments, the API, and our other services as the "Services." If you are located outside of North or South America, Etsy Ireland UC is the data controller and is responsible for your personal information. This policy does not apply to the practices of third parties (including other members who sell using the Services or API users) who may also collect or receive data in connection with your use of the Services.

1. Our Privacy Policy

We process your personal information to run our business and provide our users with the Services. By accepting our Terms of Use (and in some jurisdictions, by acknowledging this policy), you confirm that you have read and understand this policy, including how and why we use your information. If you don't want us to collect or process your personal information in the ways described in this policy, you shouldn't use the Services. We are not responsible for the content or the privacy policies or practices of any of our members, websites hosted through Pattern by Etsy, or third-party websites and apps.

Etsy's Terms of Use require all account owners to be at least 18 years of age. Minors under 18 years of age and at least 13 years of age are permitted to use Etsy's Services only if they have permission and direct supervision by the owner of the account. Children under age 13 are not permitted to use the Services. You are responsible for any and all account activity conducted by a minor on your account. We do not knowingly "sell," as that term is defined under applicable law, including the California Consumer Privacy Act, the personal information of minors. More information can be found in Etsy's Minors Policy.

By using the Services, you acknowledge that Etsy will use your information in the United States, Ireland, and any other country where Etsy operates. Please be aware that the privacy laws and standards in certain countries, including the rights of authorities to access your personal information, may differ from

Figure 1: Overview of *PrivacyInjector*, a contextual privacy policy tool that automatically detects and displays concrete privacy policy information relevant in the context of use. In the above example, a user navigated to the homepage of webex. *PrivacyInjector* identified those segments in the lengthy policy document (c) that are relevant to cookies and tracking elements. An information icon (a) is displayed on the cookie banner. When a user selects the icon, a sidebar (b) with the extracted information snippets is displayed.

ABSTRACT

Users avoid engaging with privacy policies because they are lengthy and complex, making it challenging to retrieve relevant information. In response, research proposed contextual privacy policies (CPPs) that embed relevant privacy information directly into their affiliated contexts. To date, CPPs are limited to concept showcases. This work evolves CPPs into a production tool that automatically extracts and displays concise policy information. We first evaluated the technical functionality on the US's 500 most visited websites with 59 participants. Based on our results, we further revised the tool to deploy it in the wild with 11 participants over ten days. We found that our tool is effective at embedding CPP information on websites. Moreover, we found that the tool's usage led to more reflective privacy behavior, making CPPs powerful in helping users understand the consequences of their online activities. We contribute design implications around CPP presentation to inform future systems design.

CCS CONCEPTS

• **Security and privacy** → **Human and societal aspects of security and privacy**; • **Human-centered computing** → *Web-based interaction*; *Human computer interaction (HCI)*.

KEYWORDS

privacy, privacy policies, online services, contextual privacy

ACM Reference Format:

Maximiliane Windl, Niels Henze, Albrecht Schmidt, and Sebastian S. Feger. 2022. Automating Contextual Privacy Policies: Design and Evaluation of a Production Tool for Digital Consumer Privacy Awareness. In *CHI Conference*

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3517688

on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3491102.3517688

1 INTRODUCTION

Most of us use a multitude of online services daily, leading to numerous digital traces. Data practices of online providers increasingly raise concerns as the collected data is often used for personalization and targeted advertising or misused for malicious actions, such as price, search, or gender discrimination [12, 27]. Moreover, the Facebook-Cambridge Analytica data scandal showed how user data could even become the target of stealth and criminal activity [18]. Incidents like this often result in users losing their trust in companies that collect data, which leads to financial losses, as users tend to avoid them in the future [49].

Privacy policies are one of the most common ways to inform users about data practices. They describe what data is accessed, how they are processed, and which options are available for users to control these practices. This is referred to as the principle of notice and choice [30, 50] or notice and consent [1]. However, privacy policies are often ineffective, as users do not read and fail to understand them [32]. While most users, about 74%, do not read privacy policies at all, if they read them, the average reading time for the entire policy is 73 seconds, which is way too short to read and understand the whole text [32]. The reasons for this include their length, their abstract legal language, and that they are often hidden somewhere at the end of the website [13, 25, 28, 32, 38].

Researchers and developers worked on improving the representation of privacy policies, for example, by standardizing them through segmentation into categories [48] or by using P3P, a machine-readable format for privacy policies [11]. However, there has been no wide-spread adoption of standardized formats and approaches, leading to issues and challenges around *machine accessibility* of privacy policies. Other approaches included using privacy icons to visualize the information [14, 17, 29] or displaying privacy policies similar to nutrition labels [19]. Unfortunately, most online providers refrained from incorporating these ideas into their policies mainly because the financial benefit of collecting user data outweighed their benefits. In response, researchers explored supplementing privacy policies in their current state, for example, by developing tools, such as a privacy-aware shopping engine [45]. One of the most promising concepts from recent years are contextual privacy policies (CPPs). They aim to make privacy policies more accessible by displaying relevant snippets directly in their corresponding contexts [13, 33, 34]. This decreases the amount of text that has to be read at once, helps users understand the abstract wordings of the policies, and prevents service providers from hiding important information behind inconspicuous links. A concept showcase demonstrated that users preferred this contextual representation over the lengthy policies [34].

We implemented PrivacyInjector, a production AI tool for digital consumer privacy awareness. The system, depicted in Figure 1, consists of three main parts: 1) a process to automatically extract the relevant text snippets from the privacy policies, 2) a process to automatically identify suitable contexts, and 3) a representation of the CPPs, which is unintrusive and allows their automatic injection on web pages without interfering with page layouts.

We report on two empirical studies. First, we evaluated the system's technical functionality and effectiveness on the 500 most visited websites in the US with 59 participants. We found that PrivacyInjector performed well at text classification and identifying, placing, and displaying CPP elements. Applying the insights from the first study, we conducted a second *in the wild* study with 11 participants who used the extension for ten days on a regular basis. We found that the tool's usage led to more reflective privacy behavior. Additionally, we describe design implications related to the interpretation and presentation of CPP segments. Our research showed that it is possible to create and display CPPs automatically and that they are effective in helping users understand the consequences of their online activities.

Our work makes three key contributions:

- We present the design and architecture of PrivacyInjector, the first production AI tool for CPPs. Moreover, we provide our text classifiers, the code for the classifiers, and our extension's source code to enable future research.
- We report on a study investigating how effective PrivacyInjector is at detecting, segmenting, and placing CPPs in the heterogeneous web service landscape, based on an evaluation of the US's top 500 websites with 59 participants. Further, we present a detailed error and coverage analysis that is expected to benefit future developments.
- Finally, we present findings from a ten-day longitudinal mixed-method study with 11 participants and show how CPPs promote privacy awareness and more reflective behavior. Further, we present design implications related to the interpretation and presentation of CPP segments and discuss needs and requirements related to machine accessibility of privacy policies.

2 RELATED WORK

Information *privacy* refers to individuals' desire to control their data, in particular related to access, sharing, and use [2]. Privacy should be of great interest to users, as online companies use collected data for tracking and personalization that has shown to lead to various forms of discrimination, including price, search, and gender discrimination [12, 15, 27].

This section reflects on the value of online privacy and the challenges around privacy communication and awareness. We further present work related to CPPs and highlight how our development and evaluation of PrivacyInjector advances research and application in this domain. We conclude with a summary and presentation of research questions that guided our exploration.

2.1 Challenges with Privacy Policies

Privacy policies are still a key resource for communicating data practices. In principle, users should make informed decisions based on policy reviews. This principle of informing about data practices and asking for consent - commonly through a checkbox - is called the principle of notice and choice [30, 50] or notice and consent [1]. Notice and choice are necessary, but privacy policies fail this requirement most of the time [5, 9, 43]. This is mainly caused by users not understanding them the way they are currently presented or not reading them in the first place [32, 40]. Research has described the following four key challenges that represent barriers in users' engagement with privacy policies.

Length of Privacy Policies. Obar [32] found that most users, namely 74%, do not read privacy policies at all, and those that do invest on average 73 seconds. Clearly, this is not enough time. McDonald and Cranor [25] estimated that it would take users 201 hours to read all the privacy policies they encounter in one year.

Difficult Language. Websites often use complex legal language when describing their practices, requiring at least a two-year college education to understand them [24]. This way, privacy policies meet their legal requirements, yet often conflict with users' requirements.

Abstract Wording. It is difficult for most users to link the abstract texts to their concrete actions [13]. This is especially problematic when the privacy policies are shared across multiple platforms, as for Google and YouTube [8].

Inconsistent Placement. Policies are often linked at the bottom of a web page, requiring a certain effort from users to find them [38]. Yet, retail sites often place privacy policies on the top, and financial sites tend to have a link at the top and bottom [38]. This inconsistency decreases the accessibility of privacy policies, as users have to find them first.

In conclusion, significant changes need to be made, especially in terms of length, phrasing, and placement of privacy policies, to achieve accessible policies that enable actual notice and consent. PrivacyInjector addresses the length, placement, and abstract wording challenges by presenting only information relevant in the context of service use.

2.2 Improving Privacy Policies

Previous research tried to incorporate alterations directly into the actual privacy policies, for example, by introducing *privacy by design*, which proposes design principles to increase users' control over their private information and to give companies a competitive advantage [6, 20, 47]. Other examples included design guidelines [37] and design spaces [41] that help create more concise privacy information. Yet, the commercial benefit from the collection of user data still seems to outweigh the benefits that may result from better privacy protection, so that providers mainly refrain from embracing these suggestions.

In response, researchers developed tools to complement privacy policies. Cranor et al. [10] developed *Privacy Bird*, a privacy agent for the Platform of Privacy Preferences (P3P), which is a machine-readable format for privacy policies developed in 1999 by the World Wide Web Consortium [39]. But, it was suspended later on because major browsers did not support the standard, and it was deemed too difficult for the average internet user to understand [7]. Kelley et al. [19] developed a standardized table format for privacy policies, similar to nutrition labels on groceries. Tsai et al. [45] developed *Privacy Finder*, a prototype of a shopping search engine that displayed a summary of the policy next to the search results. Harkous et al. [16] implemented PriBot, a chatbot allowing users to ask privacy-related questions in free form. Additionally, there are privacy agents that make privacy-relevant decisions on the users' behalf [21–23, 46].

Despite previous efforts, most proposed solutions have drawbacks that prevented their adoption in practice. P3P was suspended, making *Privacy Bird* redundant and *Privacy Finder* never left its prototypical state. However, the nutrition labels have meanwhile been applied in practice as Apple made Privacy Nutrition Labels mandatory[1].

2.3 Contextual Privacy Policies (CPPs)

CPPs display privacy information in their corresponding contexts. Patil et al. [36] stated that privacy feedback is the most effective when it is "delivered [...] at the right time at a manageable rate." CPPs support both requirements: the amount of text displayed at once is reduced, and the information is displayed at the right time in their context of use. Bergmann [3] showed that users' privacy-awareness increases significantly when the privacy information is displayed in their corresponding contexts. Additionally, the theory of contextual integrity considers the context as essential in determining whether information collection and distribution is appropriate or not [31].

The concept of CPPs was introduced [13] and validated [33] by previous work. Ortloff et al. [34] evaluated the CPPs using a concept showcase for seven different websites. They manually inspected each website to identify suitable placements for the CPPs. They also manually extracted and processed the privacy policies, which does not scale considering that there are millions of websites. In contrast, we present PrivacyInjector, a scalable production AI-CPP system, and demonstrate its real-world feasibility. We developed a process to automatically generate and display CPPs, including the automatic recognition of privacy policies, identification of contexts, extraction of relevant information, and unobtrusive presentation.

An essential step in automating the CPPs is the creation of appropriate text snippets. Previous research has already developed approaches to process privacy policies automatically. Zimmeck and Bellovin [51] developed *Privee*, an architecture to analyze privacy policies automatically. Mysore Sathyendra et al. [30] developed classification models that can identify opt-out choices in privacy policies. The major drawback of both is that they only allow a limited set of queries on the policies and thus do not allow a complete annotation. Harkous et al. [16] overcame this limitation by developing *Polisis*, an automated framework for the analysis of privacy policies. They used hierarchical neural network classifiers that allow for high-level and fine-grained queries. *Polisis'* approach of annotating policies seems to be the most promising one, compared to the one of Zimmeck and Bellovin [51] and Mysore Sathyendra et al. [30], since it allows to analyze the entire policy and provides fine-grained annotations. Therefore, we build on this approach in defining suitable text snippets for the real-world implementation of CPPs.

2.4 Summary and Research Questions

There are several barriers to privacy policies, including their length [25, 32], difficult legal language and phrasing [13, 24, 38], and often inconsistent and inconspicuous placement [38]. Previous attempts at addressing those issues included: standardizing privacy policies by introducing P3P [39]; presenting design principles and guidelines for better representation of privacy policies [6, 20, 37, 47]; developing a nutrition label for privacy policies [19]; and implementing a privacy-aware shopping search engine [45]. But, most of those

[1] https://www.apple.com/privacy/labels/

efforts have not yet managed to contribute to a positive change on a larger scale.

Among the most promising ideas from recent years are CPPs [13, 33, 34]. They place concise text snippets from privacy policies directly in their affiliated contexts when they become relevant. This reduces the text to be read at once and makes the privacy policies less abstract because users learn how their actions trigger data processing. Feth [13] introduced the concept of CPPs. Later, Ortloff et al. validated their acceptance [33] and evaluated the concept in situ [34]. The concept showcase reported by Ortloff et al. [34] is only implemented for seven different websites by manually inspecting each website to identify suitable placements for the CPPs and manually extracting and processing the privacy policies. Clearly, this manual approach does not scale across millions of websites and frequently updated privacy policies that have on average 2700 words [32]. In contrast to Ortloff et al. [34], we present the scalable CPP system PrivacyInjector and report on its real-world feasibility. We developed a process to automatically generate and display CPPs, including the automatic recognition of privacy policies, identification of contexts, extraction of relevant information, and unobtrusive presentation.

Inspired by the work of Harkous et al. [16] who provide a framework for the hierarchical annotation of privacy policies, we aimed to evolve CPPs through the design of PrivacyInjector from their current prototypical state to a real-world browser extension. Our work was guided by the following two research questions.

RQ1: How effective is PrivacyInjector at identifying and displaying contextual privacy policies?
Our work is motivated by the need to evolve CPPs from their current prototypical state into production tools. Therefore, we need to evaluate PrivacyInjector across a wide variety of diverse websites and users.

RQ2: How do users interact with PrivacyInjector on a regular basis?
PrivacyInjector is designed to alter the presentation of websites by placing relevant policy snippets into their context of use. As this will impact interaction experience, we need to better understand how users experience PrivacyInjector on a regular basis, in order to understand how to prepare such a functional tool for daily interaction.

3 SYSTEM

In this section, we describe the implementation of PrivacyInjector, a functional web browser extension for Google Chrome and Firefox that is depicted in Figure 1. We released the code for the creation of our text classifiers and word-embeddings, as well as our extension's source code through our GitHub repositories to enable future research[2].

As depicted in Figure 2, the system comprises two parts: the front end, which identifies the links to the privacy policies **(1)**, as well as the contexts for the text snippets **(5)**, and the back end which holds the segmenter **(2)** and the classifiers **(3)**. The implementation of the segmenter and the classifier was based on the work by Harkous et al. [16]. The JavaScript front end contains the code of the browser extension. The extension gets triggered every time a user visits a new website. The back end was written in Python and uses the framework Flask to communicate with the front end. Next, we detail the implementation related to the four key steps: (1) identification of privacy policy URL; (2) segmentation and annotation; (3) segment classification; (4) identification of contexts and (5) display.

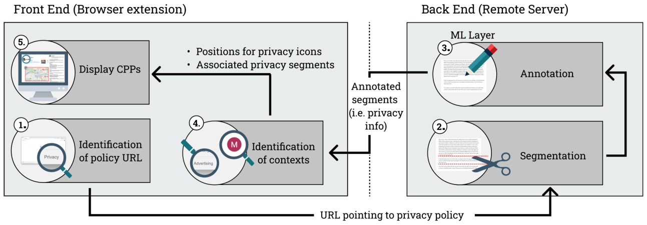

Front End (Browser extension)
Back End (Remote Server)
5.
Display CPPs
Positions for privacy icons
Associated privacy segments
1.
Privacy
Identification of policy URL
4.
M
Advertising
Identification of contexts
Annotated segments (i.e. privacy info)
ML Layer
3.
Annotation
2.
Segmentation
URL pointing to privacy policy

Figure 2: Architecture overview of PrivacyInjector. The extension is triggered once the user navigates to a webpage. In the first step (1), PrivacyInjector identifies the URL of the site's privacy policy. Next (2), the policy is segmented, before (3), the segments get annotated. Then (4), the contexts are identified and finally (5), the annotated segments, i.e. the actual privacy info snippets, are placed, represented by an icon bubble, on the webpage in the location where they are contextually relevant.

3.1 Identification of Privacy Policy URL

To identify the privacy policies, we allocated all links on the website, saved them to a list and applied a set of keywords containing words that are usually part of privacy policy URLs, such as *privacy*, *privacypolicy*, or *legal*. We constructed this list manually by visiting the 50 most visited websites in the United States[3] and recording words of relevant URLs. When more than one URL matched with words from the keywords list, we scraped the HTML content of all matching pages and ranked them according to their likelihood. We did this by applying regular expressions to the potential policies, searching for words or phrases typically contained in privacy policies. We identified those words and phrases similar to the way we constructed the URL keyword list using the same 50 websites as before. The keywords we identified were *privacy policy/statement/notice*, *policy/policies*, *personal information*, *SSL/encrypt/safeguard*, *third party*, *choice opt out/in*, *location*, *advertisements*, *collect/collection*, *share/disclose*, *children*, *cookie*, and *safe harbor*. **This process results in most cases in a single privacy policy URL.**

3.2 Segmentation and Annotation

The classifiers need fine-grained, thematically matching paragraphs to predict the data practices of each segment. Therefore, we processed the policy by obtaining the text content from the policy's website, handling lists, cleaning the HTML from irrelevant elements, and performing the actual segmentation of the text. We used Google Chrome in headless mode to scrape the policy's text. Using headless mode enabled us to benefit from the capabilities of a real browser, such as executing JavaScript, while not having to deal with a user interface.

[2] https://github.com/Maxikilliane/CPP-extension-new.git; https://github.com/Maxikilliane/polisis-classifiers.git

[3] https://www.alexa.com/topsites

In most cases, lists are prefaced with an introductory statement needed to grasp the meaning of the individual list items. An example of that could be a list that is prefaced with *You will be asked to provide information about yourself, which includes:* and has short bullet points, such as *Your name* or *Contact information.* That is why we merged all list items with their introductory statements to receive unambiguous sentences. Similarly, we removed all elements that did not contain the actual privacy policy's text, such as script tags or CSS instructions, which we identified by checking 50 privacy policies. This led to the additional removal of menu items, headers, footers, breadcrumbs, and sidebars.

Most websites could be segmented by breaking up the page according to their *<div>* and *<p>* tags. However, sometimes this failed to return any segments at all or the returned segments turned out to be very long. No segments were, for example, returned when the website was not using the classical HTML structure, like Facebook. Whenever the first step failed to return adequate segments, we used an algorithm, which uses word-embeddings to split documents into coherent parts, where coherence is defined as the accumulated weighted cosine similarity of the segment's words to the mean vector of this segment. By adequate segments, we mean that segments were returned and that these segments had a suitable length of one to four sentences. The code for this algorithm was taken from the *textsplit*[4] library but used the same domain-specific word-embeddings as the classifiers that are described later on. **At the end of this process, we have identified multiple clean policy segments.**

3.3 Segment Classification

The classification was inspired by the work of Harkous et al. [16]. The segment classification was done leveraging two key stages: 1) an *unsupervised* part, where the domain-specific word-embeddings were built from unlabeled data, and 2) a *supervised* part, where text classifiers were trained using Convolutional Neural Networks (CNNs).

Word-Embeddings. Unlike traditional text classifiers, we use word-embeddings instead of relying only on keywords and their frequencies. A significant drawback of the latter is their lower generalization power. An example is the following example taken from Harkous et al. [16]: The two sentences *"Your data gets deleted when you disable your account."* and *"Your data gets erased when you disable your account."* are classified. If our training set only contained the word *delete*, but not *erase*, the classification results turn out to be significantly different for the two sentences. Even though *delete* and *erase* are synonyms, so the classification results should be almost identical. Word-embeddings overcome this issue by extracting generic word vectors so that the word vectors of two segments containing semantically similar words are close in the vector space. That way, the classifiers also consider words not included in the training set, as long as they were part of the large corpus used for the training of the word-embeddings [16]. We used the MAPS Policies Dataset [52] to create a corpus of privacy-related words. This dataset consists of URLs from 441,626 privacy policies collected from apps in the Google Play Store. We checked that dataset and removed all duplicate URLs, which left us with about 150,000 URLs. We used a python script to scrape the text content of all those policies. The resulting text file was then cleaned from redundant spaces and special characters and transformed to lower case. We used fastText[5] to train the actual word-embeddings.

Hierarchical Multi-Label Classifiers. Following the work of Harkous et al. [16], we used hierarchical multi-label classifiers for the prediction of the data practices. The hierarchical structure is defined by the labeled dataset used for the prediction of the data practices because it subdivides these practices into high-level categories, attributes, and attribute values. We used the dataset created by Wilson et al. [48] which consists of 115 privacy policies that were annotated by law school students, resulting in about 23,000 annotated data practices. They comply with the hierarchy of privacy policies by assigning high-level categories (e.g., "First-Party Collection"), as well as attributes (e.g., "Purpose") and attribute-values (e.g., "Advertisement"). Altogether, Wilson et al. [48] identified 10 high-level categories, 22 attributes, and 130 attribute-values. A simplified visualization of that hierarchy is shown in Figure 3. This hierarchy resulted in one classifier for predicting the high-level category and one classifier for each of the 22 attributes. We note that attribute classifier selection depends on the predicted high-level category. That means, for example, that if the high-level category was predicted as *Third Party Collection*, only the attribute classifiers that are leaves of that category qualify. In this case, these would be, among others, *Action*, *Information Type*, and *Purpose*. Multi-label means a prediction for each of the possible values so that one segment has multiple predictions. We assumed that a probability above 0.5 or 50% means that the data practice was present in the segment.

Models Training. We trained one classifier for the high-level category and one classifier for each of the 22 attributes. Table 1 shows the evaluation metrics on the testing set for the category classifier. All values were micro-averaged per label to predict not only the presence but also the absence of a label. **After this step, all policy segments are annotated with their affiliated data practices.**

3.4 Identification of Contexts

Contexts are areas on the website where data practices apply. For example, data practices concerning advertisements become active when users interact with online advertisements. For this work, we focused on six different categories of data practices. These were: Advertisement information, information about users with accounts, financial information, cookie information, information about data collection and sharing with social media platforms, and location information. We chose these categories because they are, on the one hand, widespread practices that have a high likelihood of being covered in the privacy policy. On the other hand, these categories are also often represented as elements on the website.

We developed several techniques to identify those contexts. First, we visited 50 websites to identify contexts suitable to display CPP information and which reoccurred on most websites. This step returned, for example, log-in and sign-up elements for account information, social media icons and social media log-ins for social media information, and shopping carts and input fields for debit

[4] https://github.com/chschock/textsplit

[5] https://fasttext.cc/

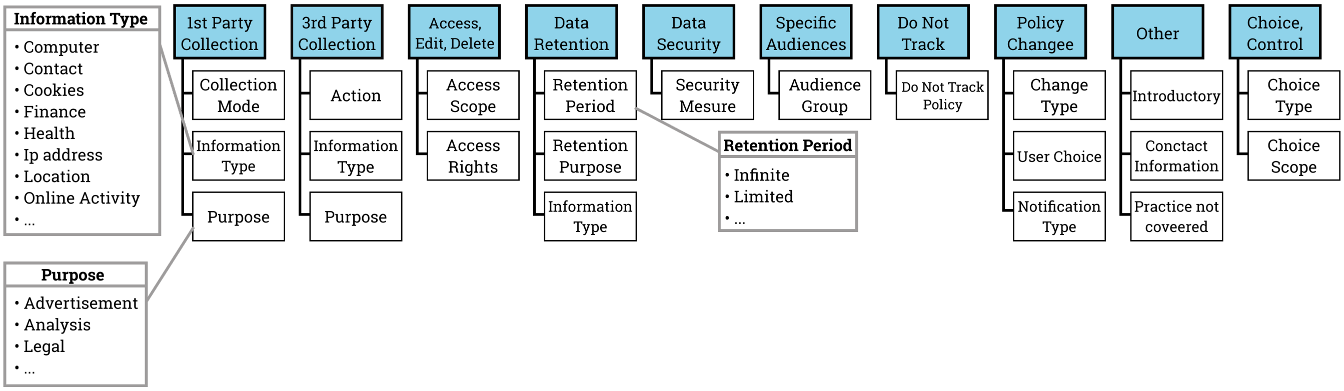


Figure 3: The mandatory values of the privacy policy taxonomy, as proposed by Wilson et al. [48]. The blue boxes are the high-level categories, the boxes below are the attributes, and the boxes with the gray frame show some examples of attribute-values.

Table 1: The classification results (F1, Precision, Recall) for the high level category of the test set.

Category	F1	Prec.	Rec.	Category	F1	Prec.	Rec.
1st Party Collection/Use	0.8	0.87	0.75	Do Not Track	1.0	1.0	1.0
3rd Party Collection/Sharing	0.77	0.86	0.69	Policy Change	0.75	0.86	0.67
User Access, Edit and Deletion	0.69	0.73	0.65	User Choice/Control	0.71	0.7	0.72
Data Retention	0.39	0.54	0.31	Introductory/Generic	0.65	0.72	0.59
Data Security	0.63	0.7	0.58	Practice not covered	0.39	0.59	0.29
International/Specific Audiences	0.9	0.96	0.84	Privacy contact information	0.72	0.84	0.63
Average	0.7	0.78	0.64	Average	0.7	0.79	0.65

card data for financial information. In the next step, we investigated the HTML structure of those elements to identify HTML tags, CSS classes, and reoccurring text contents typically used for these elements. We found that in several cases, contexts can be identified by searching for keywords in URLs. For example, account elements were identified by searching for the keywords *account*, *login*, *profile*, or *me*. Other strategies employed to identify contexts were searching for HTML-ids, which was especially successful for account elements, advertisements, and financial elements. **This step gives us areas on the websites where we can display the CPP information.**

3.5 Display of the CPP-information

We used an iterative design approach involving three design experts to determine how the CPP information should be displayed. The expert group consisted of one professional UI designer and two professors teaching UI/UX design courses. Already the first version of PrivacyInjector used floating and draggable bubbles to display the CPP information. When the user clicked on the bubble, a scrollable textbox attached to the bubble appeared, including a heading, an icon depicting the category, and the information from the privacy policy as bullet points.

All three experts criticized the presentation of the policy snippets in the form of text boxes attached to the bubbles because they were considered too small and sometimes disappeared behind website elements. In response, experts suggested implementing the text boxes as sidebars. They further suggested giving a short explanation about why information is displayed, such as: "The site seems to be using [*your location*]. Below are statements from this page's privacy policy that refer to [*location information*]." Therefore, the final design version used sidebars and short descriptions of the displayed information, as depicted in Figure 1. **After this step, small draggable bubbles appear on the website in the previously defined contexts.**

4 METHOD

We conducted two empirical studies to answer our research questions. Figure 4 provides an overview of key study characteristics and their temporal aspects. Study I focused on evaluating how effective PrivacyInjector is at extracting and displaying concise privacy policy information snippets across a large and diverse sample of web services and users. We recruited 59 participants who interacted with a subset of the US top 500 websites and the privacy information displayed by PrivacyInjector. In total, we reduced the pool of 500 websites to 354 to filter explicit and dubious content. We conducted a pilot study with five participants to exclude errors in the installation process and to verify that users understood the questionnaires. In response, we created a video tutorial. Further details on the study procedure are provided in the corresponding Section 5. Study participants checked on each website whether or

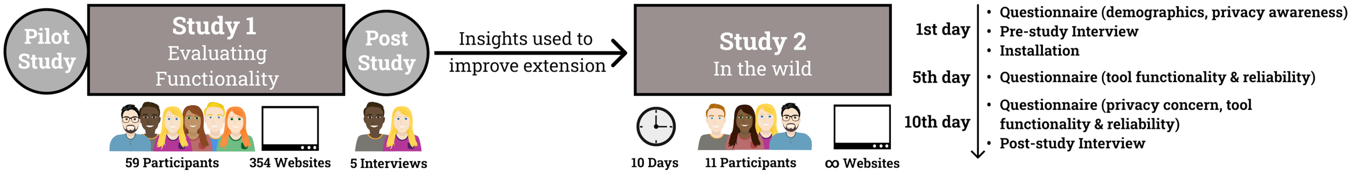


Figure 4: A high-level overview of our method. We conducted two empirical studies to investigate the *technical functionality* of PrivacyInjector (Study I) and the *interaction experience* during regular tool use (Study II).

not, and how accurately, privacy information was detected and displayed, that the context was correctly identified, and that extracted snippets matched the context of use. Based on our findings and brief post-study interviews with selected informants, we further improved the CPP tool for regular use.

We released the latest version of PrivacyInjector as Chrome and Firefox browser extensions and recruited 11 new participants for a ten-day mixed-methods study in the wild (Study II). Initially, we asked the participants about their understanding and concern of online privacy. Afterward, the participants installed the extension and were tasked to use their internet browser as usual. Participants could tailor PrivacyInjector to their needs, i.e., they could pause the extension on specific websites, pause the tool for websites until updates were detected, or pause the tool completely. Detailed information on these modes and how we logged tool interaction are provided in the corresponding study Section 6. We asked participants to rate tool functionality and reliability after five and ten days to investigate if constant exposure to the tool and its information impacted users' perceptions of it. Finally, we conducted a brief post-study interview to gather more information on the interaction experience and users' expectations towards future systems.

5 STUDY I: VALIDATING THE FUNCTIONALITY

To answer RQ1 on the effectiveness of PrivacyInjector at communicating CPPs, we evaluated the tool across a diverse set of websites, including 354 out of the 500 most frequently used websites in the United States. Reasons for excluding some of the websites are detailed below. We recruited 59 participants in this between-group study, where each participant assessed PrivacyInjector's accuracy across six different websites from the 354 websites pool. Given the size of the pool, no website was assessed by more than one participant.

Each participant answered one questionnaire for every CPP element on the website. These questionnaires contained four items assessing the placement, the texts, the visibility of the element, and the sidebar. The participants used their own computers and laptops for the study. That way, we also verified that the extension worked across diverse setups.

5.1 Website Selection

We used the US's 500 most visited websites[6] as a basis for our investigation. Out of this pool, we initially excluded websites that showed illegal, explicit, or violent content (18) and websites that were not in English (19). Further, we pre-processed all remaining websites before we started the study and found that some websites had cryptic links to their policies, making it difficult for PrivacyInjector to identify the document. An example of that is alibaba.com[7]. Another reason for which our extension failed to recognize the privacy policy was when it was presented unconventionally, for example, in a pop window or as an interactive walkthrough[8].

We manually checked all URLs detected by PrivacyInjector in the annotation process and identified four main reasons for failure: (1) the privacy policy could not be identified (61); (2) the website did not have a privacy policy (10); the website could not be reached (not secure, IP address could not be found) (23); and (4) the website was region-restricted (15). We removed websites from our pool that were classified according to these criteria. This left us with a total of 354 (out of 500) websites. A complete table containing the 61 URLs for which PrivacyInjector failed to identify the privacy policy, including the reason for why the privacy policy could not be identified, as well as possible solutions to solve these issues can be found in Table 3.

5.2 Participants

We recruited the participants via mailing lists of our institution and through convenience sampling, resulting in a diverse participant pool. In total, we recruited 59 people for the study (29 female, 30 male). Their ages ranged from 18 to 66 (M = 31.2, SD = 10.36). 24 of the participants were students. Most (14) of them studied computer science or media informatics, three studied lectureships, two business administration, and one each medicine, media production, social science, economics, and geoecology. 34 worked in various fields, including IT (8), entertainment (4), media (3), automotive (3), finance (2), education (2), health care (2), engineering (2), retail (2), administration (2), real estate (1), law (1), fashion (1), service (1), and one person was retired. The students were compensated with course credits[9], and a 30$ Amazon voucher was raffled among all remaining participants.

5.3 Procedure

Once a participant signed up, we sent an email with a consent form, installation instructions with pictures, instructions on how to proceed with the questionnaires, a video tutorial, and a list of six URLs to be visited. We instructed the participants to visit one

[6] https://www.alexa.com/topsites

[7] https://rule.alibaba.com/rule/detail/2034.htm?spm=a2700.8293689.0.0.500267afpGugNR

[8] https://admin.typeform.com/to/dwk6gt/

[9] The students have to earn a certain amount of study credits towards the completion of their degree, where one hour equals one course credit. The participation is anonymous and the students receive the same amount of compensation, no matter their responses.

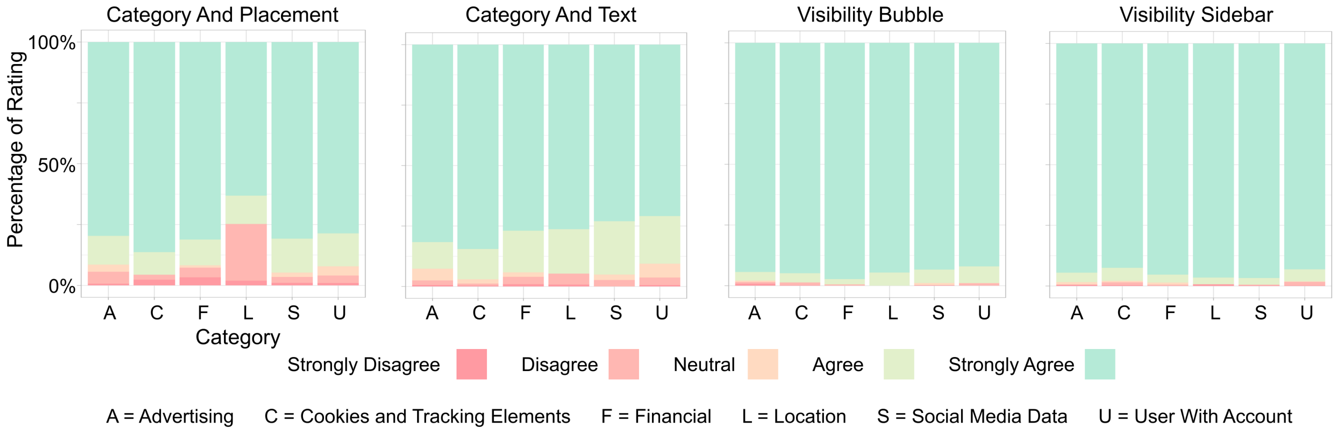


Figure 5: The percentage values of participants' ratings to our four statements on a five-point Likert scale. We framed the statements so that "strongly agree" means that it worked very well, and "strongly disagree" means that it did not work at all.

URL after another and complete the questionnaires after each page visit. The extension automatically focused on the area where the bubble was located and highlighted the corresponding bubble in green. When the user clicked on "Next Bubble," it scrolled to the next bubble.

We asked the participants to rate agreement to four statements for each bubble on a 5-point Likert scale ranging from 1 (=strongly disagree) to 5 (=strongly agree). We asked about the visibility of the bubble, the visibility of the sidebar, whether the bubble's category matched the UI element on which it was placed, and whether the texts shown in the sidebar matched the category. Refer to Appendix A for the exact wording of the statements.

5.4 Results

We analyzed how our participants rated the statements for each bubble grouped by the CPP information type: *User with account*, *Financial*, *Cookies and Tracking Elements*, *Social Media Data*, *Advertising*, and *Location*. Figure 5 shows the percentage values of the participants' ratings and Table 2 shows the mean, median, and SD for the four statements. We analyzed the data using the Kruskal–Wallis test. When significant effects were revealed, we used Wilcoxon's rank-sum test with Bonferroni corrections for pairwise posthoc analyses.

Match of bubble category and UI element. Overall, this statement received high ratings with all categories having over 77% of their responses as either "agree" or even "strongly agree", meaning that the placement of the bubble indeed matched with its information type, see Figure 5. This is also underlined by the high values for the mean (4.33) and median (5), see Table 2. Only the placement of *Location* elements stands out, with 46.67% of its ratings being "disagree" or "strongly disagree". We performed a Kruskal–Wallis test which showed a significant effect between the conditions ($H(5) = 15.598$, $p = .008$). Through post-hoc tests we found significantly lower ratings for the *Location* category compared to almost all other categories, except for the *Financial* category (Advertising: $p < .012$; Cookies and Tracking Elements: $p < .021$; Social Media Data: $p < .006$; User With Account: $p < .006$). We investigated the websites on which the *Location* bubble was rated as misplaced and found that in some cases, the *Location* context was identified by searching for the plain text *location*. However, the word *location* often appeared in the continuous text where it had nothing to do with the actual data collection. Figure 6a shows an example of a misplaced bubble. In comparison, Figure 6b shows an example for a correctly placed bubble.

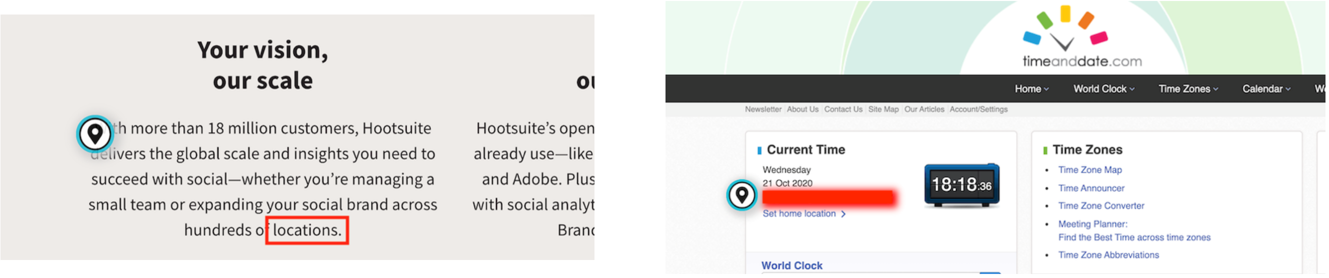
Match of text and category. Overall, the text classification worked well, as more than 82.6% of the ratings for all categories were either "agree" or even "strongly agree." The high values for the mean (4.42) and median (5) underline these results. This is also in line with the evaluation metrics of the text classifiers, which mostly showed high values for precision, recall, and F1 (see Table 1). By performing a Kruskal–Wallis test, we found a significant effect between the conditions ($H(5) = 18.868$, $p = .002$). The post-hoc analysis revealed that the information type *User With Account* was significantly different from *Advertising* and *Cookies and Tracking Elements* (Advertising: $p < .044$; Cookies and Tracking Elements: $p < .005$).

Visibility of bubbles. Altogether, the visibility of the bubbles received high ratings, with 96.65% of all ratings being either "agree" or even "strongly agree", see Figure 5. These positive results are also underlined by the high values for the mean (4.82) and median (5). However, the *Advertising* bubbles were not visible in 4% of all cases. This is because *Advertising* bubbles were almost always placed on online advertisements, which are often dynamic, meaning they automatically change or reload multiple times while the user is on the websites. Since the bubbles get only drawn once when the page first loads, it occasionally happened that the bubble got placed on an advertisement that later reloaded, and thus, the bubble disappeared. A Kruskal-Wallis test revealed no significant differences between the conditions.

Visibility of sidebars. Except for some rare occasions (40 out of 1136 bubbles = 3.5%; mean = 4.82; median = 5), the sidebar was rated as perfectly visible, see Figure 5. By manually investigating the websites on which the sidebars were not visible, we found that these were covered by floating elements, such as cookie banners. Even though we assigned a very high CSS z-value to the sidebar to

Table 2: Mean, Median, and SD for the four statements subdivided by category.

	Category & Placement			Category & Text			Visibility Bubble			Visibility Sidebar		
Category	**Mean**	**Mdn**	**SD**	**Mean**	**Mdn**	**SD**	**Mean**	**Mdn**	**SD**	**Mean**	**Mdn**	**SD**
Advertising	4.35	5	1.15	4.5	5	0.98	4.74	5	0.88	4.82	5	0.70
Cookies	4.36	5	1.32	4.67	5	0.78	4.83	5	0.65	4.76	5	0.77
Financial	4.08	5	1.50	4.39	5	1.10	4.93	5	0.38	4.89	5	0.46
Location	3.43	4	1.55	4.37	5	1.13	4.93	5	0.25	4.83	5	0.75
Social Media	4.44	5	1.09	4.5	5	0.83	4.85	5	0.56	4.89	5	0.55
User With Account	4.37	5	1.12	4.32	5	1.04	4.84	5	0.53	4.79	5	0.70
Average	4.33	5	1.19	4.42	5	0.99	4.82	5	0.62	4.82	5	0.68



Your vision, our scale

ou

h more than 18 million customers, Hootsuite elivers the global scale and insights you need to succeed with social—whether you're managing a small team or expanding your social brand across hundreds o locations.

Hootsuite's open already use—like and Adobe. Plus with social analyt Branc

(a) Misplaced Bubble for *location*

Current Time

Wednesday
21 Oct 2020
[redacted]
Set home location

18:18:36

Time Zones

- Time Zone Map
- Time Announcer
- Time Zone Converter
- Meeting Planner: Find the Best Time across time zones
- Time Zone Abbreviations

World Clock

(b) Correctly Placed Bubble for *location*

Figure 6: (a) An example of a misplaced bubble for the information-type *Location* on https://hootsuite.com/. (b) An example of a correctly placed bubble for the information-type *Location* on https://timeanddate.com/. The location is concealed for blind review.

verify that it is positioned in front of all other elements, in some cases, the banners of the website had an even higher z-value so that they eventually covered our sidebar. Using a Kruskal-Wallis test, we found no significant differences between the conditions.

5.5 Error and Coverage Analysis

We conducted error and coverage analyses to understand the limitations of our keyword-based matching approach, discover to which extent we currently cover the privacy policies' information, and provide directions for future research. The error analysis and the second part of the coverage analysis were conducted by two experts in usable privacy who have extensive experience developing web applications. We note that the experts' backgrounds differed in terms of seniority (i.e., one junior and one senior researcher), gender, and primary expertise (i.e., one researcher focuses on AI in usable privacy while the other studies user communication of privacy information.

5.5.1 Error Analysis. The experts systematically reviewed all bubbles that were rated as misplaced by our participants (i.e., all bubbles rated with less than 3 (=neutral)) by opening each website with a misplaced bubble and noting the reason. Thereby, they identified three main causes.

Semantic mismatches were the most common issue and accounted for 75.23% of all misplaced bubbles. It happened when our keyword-based matching approach failed by mistakingly matching the wrong contexts. PrivacyInjector, for example, identified cookie banners by searching for the keyword "cookie," which led to several mistakes on recipe websites. Similar mismatches, where the meaning of a word differed depending on the context, happened with HTML-ids, classes, and links.

The second most common issue related to *reloading elements* with 18.35%. Several providers use advertisements that reload multiple times while on the website. Whenever such an element disappeared, the bubble we placed on the advertisement disappeared with it.

In 6.42% participants gave *wrong judgments* since sometimes, it was not immediately apparent that a bubble was correctly placed. For example, amazon.com has a link in the footer named "self publish with us" that links to the login for merchant accounts. We assume that participants did not view this as a login element and therefore judged the bubbles as misplaced.

Even though our keyword-based matching approach was successful in most cases, it also led to some misplacements, as described in this section. We envision further improving the keyword-based matching through experience and feedback from a larger user base to remedy the semantic mismatches. The issue with reloading elements can be fixed by introducing an update method. Altogether, this error analysis helped us improve our tool for Study II and longitudinal use as described in Section 5.6.

5.5.2 Coverage Analysis. We further analyzed to which extent PrivacyInjector currently covers the privacy policies' information and studied what part of the remaining policy segments could be

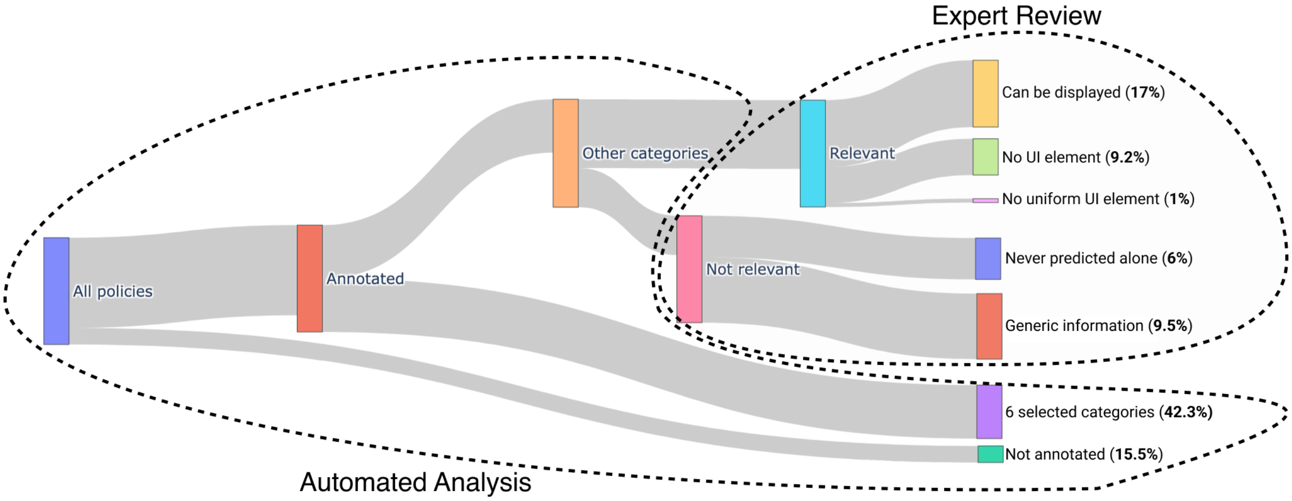


Figure 7: Our coverage analysis of policy segments is separated into an automated analysis and a subsequent expert review. The automated analysis revealed the percentage of segments for which our classifiers predicted at least one data practice (*Annotated*) and of those the segments that did not belong to our 6 selected categories and thus are not shown so far (*Other categories*). Of those segments not shown so far, a subsequent expert coverage review revealed the percentage of segments that contained relevant information that should be displayed (*Relevant*). Of those segments with relevant information, the experts judged if they can or why they can not be displayed (*Can be displayed, No UI element, No uniform UI element*). Finally, of those segments that did not contain relevant information (*Not relevant*), the experts reported why the information is not relevant (*Never predicted alone, generic information*).

displayed using our approach on the US' top 500 websites. As depicted in Figure 7, we used two analysis strategies, an automatized analysis and a subsequent expert review. The expert review was conducted by two researchers, as described in the section intro, and included several rounds of discussion. Here, two researchers judged whether a policy segment was privacy-relevant for the users. We note that this manual relevancy check was highly inclusive, meaning segments were registered as relevant in this step if at least one researcher nominated a segment for further analysis. We considered that including false-positive segments in this inclusive process would cause fewer issues for future analyses and developments than excluding false negatives. Future researchers will make their own decisions based on our analysis and their own experiences that are likely to extend into other languages and contexts.

First, within the automated analysis, we calculated the number of policy segments that were annotated by our classifiers and found that only 15.5% of all policy segments were not annotated with data practices by our classifiers. Next, we calculated the percentage of policy segments currently displayed by PrivacyInjector, i.e., that were annotated with one of our six selected categories. Thereby we found that these account for 42.3% of all policy segments.

Next, within the expert review study, two researchers evaluated whether the remaining segments that were annotated but not displayed so far described privacy-relevant practices that should be displayed. Afterward, we checked if these segments could be displayed using our approach (i.e., matching them with a user-facing UI element). We found that 17% of all policy segments could potentially be displayed using our system. In contrast, 9.2% did not have a suitable UI element on the website to display the information because they, for example, described security measures performed by the providers, such as educating their employees about privacy. Additionally, 1% did not have a uniform UI element that could be automatically detected to display the information. With uniform UI elements, we mean the element was represented across multiple websites similarly, such as, for example, advertisements or cookie banners. Elements for which we could not detect a suitable element were opt-in and opt-out choices for data collections since those are implemented differently depending on the provider; for example, some manage these choices by email, while others have, for example, a link in their privacy policy.

Finally, we describe why the experts consider some of the segments currently not displayed by PrivacyInjector as not relevant. Here, 9.5% refer to generic practices (for example, annotated with unspecified) and thus do not add value. Since we employ a multi-label classifier, we also have 6% of policy segments that are never predicted alone. Those, for example, describe more precisely how data is collected or whether the collected data is anonymized. Since these segments are always predicted together with another data practice, they do not require a separate display option.

We provide a list containing the category, the number of occurrences, the importance of information, and if and how they can be displayed in the supplementary material.

5.6 Post-Study Interviews and Improvements for Longitudinal Use

We conducted semi-structured post-study interviews with five participants to determine what we needed to improve about PrivacyInjector for longitudinal use. We selected five participants of different ages ranging from 22 to 65 and various professions to account for more diverse insights. Two participants were students of medicine and computer science, and two were working, one as a producer, the other as a property manager. Another participant was retired. We found that we had to include the option of disabling the extension for specific websites or disabling the extension until updates were detected. Besides, participants stated that the number of bubbles was excessive on some websites. For example, Amazon placed multiple log-in and purchase elements on each subsite, and on each of these elements, bubbles were placed. So we reduced the number of bubbles to a maximum of three per category and site. Finally, we also improved the placement of the bubbles according to our findings from the error analysis by, e.g., no longer identifying the words "location" and "cookie" in the continuous text and by taking care to match keywords only as whole words (e.g., to prevent errors like the word "skip" matched with the keyword "IP").

5.7 Summary

Although the identification of the privacy policies worked well, we could not identify the privacy policy for 61 websites. In those cases, the privacy policy either had a different base-URL, was unconventionally presented, e.g., in form of a pop-up or interactive walkthrough, opened in a new tab, was not linked on the landing page, or the URL leading to the privacy policy did not contain any of the previously identified keywords, see Table 3. One possible solution is integrating a text field into the extension's pop-up to rely on users to enter the links to missing privacy policies.

We were worried that using low-level heuristics to identify the contexts might raise problems. However, as Figure 5 and Table 2 show, the placement worked well with a mean rating well above 4 out of 5 for all categories except *Location*. To improve the placements in the future, we conducted an error analysis that revealed three main reasons for misplaced elements, with semantic mismatches accounting for most. Since most misplacements could be traced back to our keyword-based matching approach, using more sophisticated heuristics should be explored and can, once done, be easily incorporated into the system because of its previously mentioned modularity. Also, the statement of whether the contexts match the texts received high ratings, with 85.74% being either "agree" or "strongly agree." This is in line with the evaluation metrics of the text classifiers, which also showed high values for precision, recall, and F1. Except for some rare occasions, the bubbles and sidebars appeared and were fully visible, where we planned them to be, leading us to conclude that we successfully evolved the concept of CPPs into a functional, automated browser extension through PrivacyInjector.

Additionally, we conducted a coverage analysis to study the potential of PrivacyInjector . We found that, even though we only display six categories so far, we already cover 42.3% of the relevant information. In preparation for the second study, we performed post-study interviews and improved our extension based on the findings of these interviews and the error analysis.

6 STUDY II: IN-THE-WILD

After we improved the extension based on the findings from Study I, we evaluated the extension's acceptance and influence on participants' privacy awareness by employing a mixed-methods approach. Eleven participants installed the extension for ten days on their computers and kept on using them as usual. They participated in pre-and post-study interviews and filled out questionnaires.

In the pre-study interview, we assessed how participants informed themselves about privacy regulations, how they usually dealt with privacy policies, and asked about their online privacy concerns. In the final interview, we asked about their experience using the extension, especially what surprised them and what they liked and disliked about the tool. Further, we assessed the participants' privacy awareness using the 7-point Likert scale item proposed by Schaub et al. [42], and the tool's functionality and reliability scales described by McKnight et al. [26]. We also tracked whenever participants disabled the extension or only enabled it for updates. The study lasted ten days, during which we instructed the participants to keep using their computers as usual.

6.1 Participants

We recruited the participants via mailing lists of our institution and by word of mouth. Eleven participants took part in the study (8 female, 3 male). Their ages ranged from 18 to 50 ($M = 28.27$, $SD = 8.83$). Eight had a bachelor's degree, two had a master's degree, and one did not have a college degree. Eight of the participants were students, and three worked respectively as project manager, computing services specialist, and producer. A detailed overview of the participants is provided in Table 4. We note that our sample consisted mainly of well-educated, tech-savvy individuals and students. However, Sotirakopoulos et al. [44] showed that a more representative sample did not lead to significantly different results in the field of usable privacy and thus rejected concerns about student samples leading to less generalizable results. The active time required for the interviews and surveys ranged from 60 to 90 minutes. Besides that, participants kept on using their computers as usual. We compensated the participants with 17$.

6.2 Procedure

We first jointly installed the extension on the participants' computers and answered any questions. We then conducted the pre-study interview, and the participants filled out a questionnaire to assess their privacy awareness and to record demographic information. Afterward, the participants used their computers as usual for ten days. After five days, we sent another questionnaire to the participants to assess the tool's reliability and functionality. At the end of the study, the participants answered another questionnaire assessing the tool's reliability and functionality and privacy awareness again. We wanted to see whether the perception about the tool's reliability and functionality changed over time and whether the tool's usage influenced participants' privacy awareness. We also did a post-study interview to gain further insights into the tool's usefulness. We asked if participants experienced any surprises, what they liked and disliked about the extension, and inquired about necessary improvements.

6.3 Qualitative Results

We conducted one pre- and one post-study interview with each of the 11 participants. We fully transcribed those recordings ourselves to familiarize us even further with the data. Additionally, we studied transcribed recordings before starting the coding process. We used Thematic Analysis for the data analysis [4] and Atlas.ti as analysis software. Initially, two authors independently performed open coding of two interview transcriptions. The two authors discussed their codes, merged, and cleaned them, and created corresponding code groups. This code tree guided the analysis of the remaining interviews. In total, we created 72 codes and seven code groups. We further discussed code assignments and code group titles and finally discussed themes based on the set of code groups. The full Atlas.ti code group report is available as supplementary material.

This process resulted in the following three themes: Awareness, Presentation, and Interaction. The theme **Awareness** relates to factors influencing users' perception of privacy relevance. In this context, participants also described actions taken to protect their privacy. This theme also covers indications of how privacy awareness was impacted by PrivacyInjector. **Presentation** covers a wide range of suggestions and comments regarding the presentation of policy snippets and interpretations that impacted users' **Interaction** experience. In this theme, we report on concrete experience around regular tool usage, the tool's perceived usefulness, and the role of serendipitous discovery.

6.3.1 *Awareness.*

Most of our participants stated, before using PrivacyInjector, that they were concerned, to varying degrees, about their online privacy. Informants who reported concrete concerns were mostly relating to social media privacy concerns, to the risk of identity theft, and to heightened awareness in professional contexts. Additional service-related factors impacting their awareness and concern included trust in the service provider, the service frequency of use, and the website type (e.g. financial vs. entertainment). Those who reported on concrete privacy actions referred to rejection of cookie policies, refusal to store sensitive data (e.g. credit card information), and not using social media. Yet, participants also stated that barriers related to engagement with privacy policies and privacy-preserving actions kept them from making informed decisions at the moment. They cited lengthy and complicated policies, lack of time and/or interest, and lack of choice/alternatives as key barriers.

Our post-study interviews showed that privacy awareness changed for several of the participants. For example, P2 and P7 explicitly referred to the benefits of constant recall. P7 provided a concrete example: *"what was cool is that it resurfaced stuff that I already knew, but it was a nice reminder that this is indeed what they are collecting. Yeah, I had forgotten about that. That Google also uses my stuff for Google Play Store and shopping."* Several informants further reported that the tool impacted their awareness regarding hidden ads that were disguised as information. And P5 reported on an experience related to a website that wanted to share location data with third parties, stating that this information presented by PrivacyInjector led to the decision to leave the website immediately. The consensus of our participants was that the extension increased awareness and helped to make informed decisions. Yet, they also described challenges, requirements, and opportunities, as outlined in the following themes.

6.3.2 *Presentation.*

The participants extensively commented on the presentation of policy snippets. While they generally stressed the value of presenting only those parts of a privacy policy relevant to the current context of use, several participants asked for even more concise information. Some examples:

> I would click on it and read through the headlines. [...] And then it was a bit too much, and I was like, ok, that's enough, I want to now actually use the website and not read through all of that. – P10
>
> If you could further filter the texts. Within this already shorter version, if you could make an iconic representation of things that need to be flagged. It is not in my interest to know everything, but if you could filter certain keywords that are trigger points and you could focus on them. – P7

Topics around visualization have come up repeatedly and were connected to the call for more interpretations. For example, P3 stated that a *"summary would also be really nice. Like this is dangerous. Look out for that. And the other thing is the icons could be colored like green, yellow and red... or some other color scheme"*. Interestingly, this idea of a color scheme was echoed by several other informants describing a traffic light interpretation. Further, some participants stressed that concrete updates needed to be better visible. P3 provides an account of this, emphasizing that it would be more useful *"if it highlighted that thing because it's the same when they send you an email that their policy changed and send you the whole 10000-word text. Like I don't know what the difference is and what it means. It would be useful if it showed what exactly changed and what that means."*

Finally, we perceived a call for more customization by some participants. Here, P6 referred to the number of bubbles displayed on some pages as overwhelming. In this context, P3 added that it *"would be interesting if you could hide certain icons. For example, if there are certain ones where I think yeah, I would expect them to be here, while there are other ones where I want to be reminded that when I do this when I click here, they are collecting this data."* This is an interesting call for customization that goes beyond our current website-based configuration mechanisms and impacts interaction experience, as the following theme shows.

6.3.3 *Interaction.*

The informants provided rich accounts of their experience interacting with PrivacyInjector regularly over ten days. Foremost, they described how this interaction impacted their awareness and ability to make informed decisions, as reflected in the previous theme Awareness.

Given this positive impact, P7 expressed disappointment that the tool was not yet running on mobile device browsers. Participants further reported on reasons for pausing the extension on some websites. Those reasons included recall, especially on frequently used pages, and websites on which the tool did not work. However, only one informant (P3) reported experiencing functional issues. However, P3, P5, and P9 reported occasionally moving an icon bubble because it was inconveniently positioned over a button or content.

Study participants highlighted several areas where they find PrivacyInjector most useful. These were: at registration/login to learn about account practices, when information was surprising, and on frequently used sites as users expect to leave most data there. Further, we documented a link between a perceived increase in usefulness as users spent more time with the tool.

This is closely related to the notion of serendipity in the interaction with PrivacyInjector's information. The serendipitous discovery was related to surprising information. For example, P1 described being surprised about YouTube requesting location data. Other participants were surprised to see that certain website elements were actually hidden ads. In particular, P4 and P8 reported clicking on bubbles out of curiosity to see what the service provider tracked. P4 reported on an insight perceived as shocking:

> For Duolingo, they save everything you ever say into the microphone, and I didn't expect it. So it was shocking to read everything about that. – P4

Finally, participants demonstrated how serendipitous discovery ultimately turned into effective interpretation, e.g.:

> Like in the beginning, I would be more curious. I would be like: Oh, what does this little button do? What is this symbol? What is this symbol? And then, as I've gotten used to it, I could actually figure out what I was interested in. So, initially, I was just clicking around randomly, like what is this data, what is this data. And later, it was like, oh, I know what that symbol does, but why is that here on the webpage and so on. So I think it got actually more useful the more I used it because I could actually pinpoint the parts where I was most interested – P3.

6.4 Quantitative Results

We conducted an exploratory quantitative analysis to investigate whether the tool's usage influenced participants' privacy concern (pre-and post-study) and their assessment of the tool's functionality and reliability (middle and post-study). The participants answered all items on a 7-point Likert scale. We acknowledge the limited interpretability of this analysis due to the small sample size. Therefore, the results merely serve to complement our qualitative findings.

Privacy Concern. We used a Wilcoxon signed-rank test to explore changes in privacy concerns since the data was not normally distributed. After the tools' usage, the privacy concern is slightly greater ($M = 5.27, SD = 1.42$) than before ($M = 4.9, SD = 1.3$). However, this difference was not significant. We investigated the privacy concern by participant and found changes in both directions. For three participants, the tool's usage made them more privacy concerned, while for one participant, it became less. Our qualitative findings help to reason about this. While some participants were surprised by the number of information websites collect and store about them, others reported a positive effect as they expected even more personal information to be collected.

Functionality and Reliability. Participants filled out a questionnaire assessing the tool's functionality and reliability once after five days and a second time at the end of the study, after ten days, see Figure 8. Since the tool's functionality data were not normally distributed, we used a Wilcoxon signed-rank test to test for significant changes in the functionality assessments. Even though this test did not reveal significant differences, we still see that participants rated the tool's functionality higher post-study ($M = 5.67, SD = 0.1$) than after five days ($M = 5.18, SD = 0.16$). We used a t-test for tool reliability since the data was normally distributed. Though not significant, we found that the reliability assessments were higher after ($M = 5.36, SD = 0.27$) than during ($M = 4.73, SD = 0.5$) the study. Our qualitative data undermine these findings. Several participants stated that they learned to interact with the tool over time since they already knew which bubbles would appear and what kind of information they would typically show. Thus, they could use this knowledge to interact with the tool effectively and, for example, only interact with those bubbles they were most interested in. This made the tool more reliable and functional over time. Due to the higher functionality ratings after the study, we also conclude that the tool did not disturb the browsing experience and annoyed the users but provided useful information beyond the first curious exploration phase.

7 DISCUSSION

We reported on two empirical studies investigating functionality and user experience of PrivacyInjector, the first production CPP tool. In the following, we discuss our findings through the lenses of our two research questions (RQs) and provide design implications in the context of RQ2.

RQ1: How effective is PrivacyInjector at identifying and displaying contextual privacy policies?

We conducted a technical validation study across the US's most visited websites with 59 participants to answer this research question. Overall, the study yielded promising results. PrivacyInjector successfully identified the privacy policies on most websites, and the participants gave high ratings for the tool's functionality. Even though the choice to use relatively low-level heuristics to identify the contexts might raise skepticism at first, the placement worked well with a mean above 4 out of 5 for all categories except *Location*. Yet, we still wanted to identify the reasons for misplaced CPP elements to improve the extension for the second study and provide directions for future research. Thus, we conducted an error analysis and found three main reasons, with semantic mismatches of keywords being the most common one. Based on these results, we argue that using more sophisticated heuristics should be explored and can, once done, be easily incorporated into the system because of the previously mentioned modularity. Additionally, we note that it is a limitation that each website was only assessed by one participant and hence can be prone to errors if a participant simply got the assessment wrong. We decided to do this to be able to cover a larger pool of websites.

To study the potential of PrivacyInjector, we also determined to which extent our tool currently covers the privacy policies' information and to which extent it could potentially cover the remaining information. Thus, we conducted a coverage analysis which revealed that our six selected categories already cover 42.3% of all policy segments. Furthermore, our inclusive manual expert review showed that an additional 17% of all policy segments could be integrated in PrivacyInjector as they can be matched with UI elements. To this end, we released all analysis details in the supplementary materials to enable future research to independently investigate what additional categories they want to make use of. Finally, we note that around 10.2% of the segments that were classified as relevant could not be assigned to a UI element. A potential solution

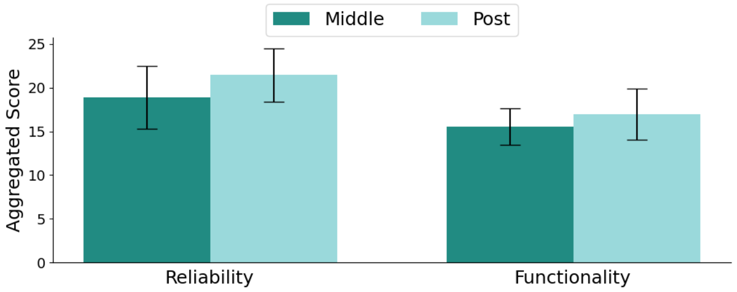


Figure 8: Tool Functionality and Reliability grouped by condition (middle and post). Error bars show mean confidence intervals.

would be integrating links into the sidebars to provide supplementary information. For example, when displaying information related to data collection and storage, there could be a link entitled "Linked secondary policy information."

Since our analysis of non-compatible websites (Table 3) showed that unconventional interaction patterns, cryptic policy URLs, forced user interaction, and missing policy links, represent barriers for any systematic effort to analyze and communicate policy segments contextually, we encourage service providers to foster *machine* accessibility of their privacy policies. This call for immediate actions to improve *machine* accessibility is connected to our vision of future service providers creating pre-annotated policy files that they can send along their website resources and original policies. Such a scenario could be enabled by future privacy supportive tools like PrivacyInjector that offer public service interfaces. This would further contribute to user privacy, as user interaction with a website would not trigger calls to third party privacy tools anymore.

RQ2: How do users interact with PrivacyInjector on a regular basis?
Based on the encouraging results of Study I and the subsequent improvements made to PrivacyInjector, we recruited 11 participants to use the extension on a regular basis for ten days. To provide users with control, we implemented additional mechanisms for deactivating the tool on individual websites, either entirely or until updates were detected. These mechanisms were rarely used: six participants deactivated a total of 13 websites. Nevertheless, our interviews showed that the option of deactivating the CPP tool on specific websites, characterized by high usage frequency, a large number of privacy info bubbles, or low perceived risk of sharing sensitive data, was crucial for individual testers. These findings suggest that **providing control on the level of individual websites is necessary** to ensure a good user experience across a large population and long-term usage. This is part of a bigger **call to design for granularity** which extends into the personalization of CPP tools. In particular, study participants expressed that they wanted to manage notifications based on website type and type of privacy information.

Our pre-study interviews revealed that most participants were generally aware of privacy risks, and many employed personal strategies to cope with these risks. These included leaving social media platforms, blocking services from storing their credit card details, and assessing individual page risks by their frequency of use and the sensitivity of data stored. Given this established awareness around online privacy, PrivacyInjector did not significantly impact privacy concerns. Instead, our qualitative analysis showed that PrivacyInjector empowered participants to review rules and make informed decisions. The participants emphasized that the tool both served as a reminder for already known privacy threats and further provided new insights, for example, by revealing advertisements that were disguised as information. Given those findings, we see an opportunity to provide users with tools needed to translate established privacy concerns into informed behavior change.

The study participants reported several issues, including page elements covered by info bubbles and excessive numbers of bubbles displayed on individual websites. Regarding the excessive number of bubbles on some websites, participants felt that the bubbles were unintrusive enough to not lead to information fatigue since they could selectively decide when they wanted more information. The participants showed agreement towards the functionality and reliability of PrivacyInjector. This assessment was not negatively impacted as tool usage continued. Instead, sampling experiences between days five and ten showed a positive trend towards users' perceptions of functionality and reliability. Part of this can likely be explained by the observed transition from an exploratory phase to more effective and conscious tool usage. Further, we argue that PrivacyInjector's perceived value outweighed encountered issues, as it created awareness for specific actions, acted as a recall mechanism, and helped identify and separate hidden ads from content.

Our informants discussed several suggestions for tool improvement that were connected to the *presentation* of CPP information. One of these related directly to PrivacyInjector's key strength: concisely communicating privacy information. While the participants acknowledged this as a strength, several asked for even more condensed and precise information that can be reviewed with minimal effort. Highlighting keywords and introducing iconic representations could, together with more advanced (natural language) analysis of policy segments, provide strategies to tackle the **need for an even more concise presentation of policy segments**.

In this context, policy analysis might allow to **provide users with further interpretation of displayed policy segments**. This call for interpretation was echoed by several participants, suggesting that the system could automatically detect dangerous data practices and highlight them accordingly. They further wished for the system to **interpret the impact of policy updates**. As our findings showed, this call for interpretations requires tool developers to **further design and explore different types of privacy information visualizations**. Several of our study participants referred to traffic lights as possible metaphors for visualization of interpretations. In this context, we have to note that Oser et al. [35] found that users preferred binary indicators over traditional three-state traffic lights in their study on smart device security and privacy risk communication. Yet, we argue that smart device users might have fewer options to act immediately on security and risk warnings, i.e., keep the device connected or remove it from the network. Instead, online service users might switch to different services, delete particular data, or abandon websites at no financial cost. These considerations should be taken into account in the future design of CPP info representations.

As part of the final interview, participants discussed how they explored privacy bubbles out of curiosity and surprise. They described several accounts of translating learned experiences into their understanding of online privacy. As service developers, we can motivate this exploration by linking additional resources that further explain data practices. This might conflict with users' desire to design even more concise policy representations. Thus, we argue that future CPP tool developments should **investigate design for unintrusive serendipitous discovery**.

7.1 Limitations and Future Work

We see a strong opportunity for future CPP tools to transform how online service users interact with privacy policy information and expect future systems to enable users to make more informed decisions. To support this, we release the code for the creation of our text classifiers and word-embeddings, as well as our extension's source code through our GitHub repositories[10]. In this context, we suggest researchers and developers to study and explore our findings and design implications, particularly regarding the presentation and interpretation of CPPs.

To support future developments and research, we outline the limitations of PrivacyInjector and our study. Through our error analysis, we found cases for which our approach to identifying contexts failed. We hope that this understanding will help future researchers and tool developers to create even more robust applications and that service providers improve *machine* accessibility. Further, we note that our extension currently only identifies and analyzes privacy policies written in English. This does not mean that PrivacyInjector is limited to websites displayed in English. Instead, the extension identifies English privacy policies from multi-language websites and displays the original English policy segments into the context of any language version of that site. This allows the use of many websites and services used worldwide (e.g., Google, Facebook, Gmail, YouTube, Apple, Amazon). There is an opportunity for future systems to integrate additional languages and to study how effective translations of English policy snippets are across a wide set of languages.

Finally, we argue that future CPP tools incorporating our findings and design implications should be systematically tested with users long-term. Our Study II investigated the user experience of participants over ten days. Limiting the duration of this study was necessary to sample and analyze experiences with an entirely new type of tool. Running this study over a more extended period would have risked exposing users to undesired interaction experiences that could have alienated them from reviewing policy information. Thus, we see this research as a necessary intermediate step towards long-term investigations.

8 CONCLUSION

We successfully evolved CPPs from a concept showcase to a production AI tool. We detail the development of PrivacyInjector and provide the source code to enable further development and research. By conducting two studies, we first evaluated PrivacyInjector's technical functionality with 59 participants on 354 of the US's most visited websites, improved the extension based on the findings, and conducted a second, ten-day in the wild study to investigate how users effectively interact with the tool. We find that PrivacyInjector is capable of showing CPP information on most websites. Further, we discovered that PrivacyInjector eases the process of reviewing privacy information and empowers users to make informed decisions. While we did not find a significant influence on privacy concerns, the tool helped users translate their existing privacy concerns into concrete behavior change. The interviews also revealed various aspects that can benefit from further development. Suggestions included reducing the number of bubbles and lengths of texts or interpreting and visualizing the severity of privacy threats. We outline these insights as design implications that we hope will guide future systems design.

ACKNOWLEDGMENTS

This work has been funded by the German Federal Ministry of Education and Research (BMBF) under Grant No. 01IS18036A. The authors of this work take full responsibilities for its content.

REFERENCES

[1] Solon Barocas and Helen Nissenbaum. 2009. On notice: The trouble with Notice and Consent. *Proceedings of the Engaging Data Forum: The First International Forum on the Application and Management of Personal Electronic Information*, 7 pages. https://ssrn.com/abstract=2567409

[2] France Bélanger and Robert E Crossler. 2011. Privacy in the digital age: a review of information privacy research in information systems. *MIS quarterly* (2011), 1017–1041.

[3] Mike Bergmann. 2009. Testing Privacy Awareness. In *The Future of Identity in the Information Society*, Vashek Matyáš, Simone Fischer-Hübner, Daniel Cvrček, and Petr Švenda (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 237–253.

[4] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI Research: Going Behind the Scenes.* Morgan & Claypool Publishers, 51–60. https://doi.org/10.2200/S00706ED1V01Y201602HCI034

[5] F. H. Cate. 2010. The Limits of Notice and Choice. *IEEE Security Privacy* 8, 2 (March 2010), 59–62. https://doi.org/10.1109/MSP.2010.84

[6] Ann Cavoukian. 2012. Privacy by Design. *IEEE Technology and Society Magazine* 31, 4 (2012), 18–19.

[7] Electronic Privacy Information Center. 2000. Pretty Poor Privacy: An Assessment of P3P and Internet Privacy. https://epic.org/reports/prettypoorprivacy.html. Last accessed: 2020-08-21.

[8] Sushain K. Cherivirala. 2018. UsablePrivacy.org Explore Google.com. https://explore.usableprivacy.org/youtube.com/?view=machine. Last accessed: 2020-08-21.

[9] Lorrie Faith Cranor. 2012. Necessary but not sufficient: Standardized mechanisms for privacy notice and choice. *Journal on Telecommunications and High Technology Law* 10 (2012), 273.

[10] Lorrie Faith Cranor, Praveen Guduru, and Manjula Arjula. 2006. User Interfaces for Privacy Agents. *ACM Trans. Comput.-Hum. Interact.* 13, 2 (June 2006), 135–178. https://doi.org/10.1145/1165734.1165735

[11] Lorrie Faith Cranor and Lawrence Lessig. 2002. *Web Privacy with P3P.* O'Reilly & Associates, Inc., Sebastopol, CA, USA.

[12] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2015. Automated Experiments on Ad Privacy Settings. *Proceedings on Privacy Enhancing Technologies* 2015, 1 (2015), 92 – 112. https://doi.org/10.1515/popets-2015-0007

[13] Denis Feth. 2017. Transparency through Contextual Privacy Statements. In *Mensch und Computer 2017-Workshopband*, Manuel Burghardt, Raphael Wimmer,

[10] https://github.com/Maxikilliane/CPP-extension-new.git; https://github.com/Maxikilliane/polisis-classifiers.git

Christian Wolff, and Christa Womser-Hacker (Eds.). Gesellschaft für Informatik e.V., Regensburg. https://doi.org/10.18420/muc2017-ws05-0406

[14] Armin Gerl. 2018. Extending layered privacy language to support privacy icons for a personal privacy policy user interface. In *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. BCS Learning & Development Ltd., Swindon, GBR, 1–5.

[15] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-Commerce Web Sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 305–318. https://doi.org/10.1145/2663716.2663744

[16] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G. Shin, and Karl Aberer. 2018. Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning. In *Proceedings of the 27th USENIX Conference on Security Symposium* (Baltimore, MD, USA) *(SEC'18)*. USENIX Association, USA, 531–548.

[17] Leif-Erik Holtz, Katharina Nocun, and Marit Hansen. 2011. Towards Displaying Privacy Information with Icons. In *Privacy and Identity Management for Life*, Simone Fischer-Hübner, Penny Duquenoy, Marit Hansen, Ronald Leenes, and Ge Zhang (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 338–348.

[18] Jim Isaak and Mina J. Hanna. 2018. User Data Privacy: Facebook, Cambridge Analytica, and Privacy Protection. *Computer* 51, 08 (August 2018), 56–59. https://doi.org/10.1109/MC.2018.3191268

[19] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. 2009. A "Nutrition Label" for Privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (Mountain View, California, USA) *(SOUPS '09)*. ACM, New York, NY, USA, Article 4, 12 pages. https://doi.org/10.1145/1572532.1572538

[20] Marc Langheinrich. 2001. Privacy by Design — Principles of Privacy-Aware Ubiquitous Systems. In *Ubicomp 2001: Ubiquitous Computing*, Gregory D. Abowd, Barry Brumitt, and Steven Shafer (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 273–291.

[21] Hosub Lee and Alfred Kobsa. 2017. Privacy preference modeling and prediction in a simulated campuswide IoT environment. In *2017 IEEE International Conference on Pervasive Computing and Communications (PerCom)*. 276–285. https://doi.org/10.1109/PERCOM.2017.7917874

[22] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun (Aerin) Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. 2016. Follow My Recommendations: A Personalized Privacy Assistant for Mobile App Permissions. In *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*. USENIX Association, Denver, CO, 27–41. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/liu

[23] Bin Liu, Jialiu Lin, and Norman Sadeh. 2014. Reconciling Mobile App Privacy and Usability on Smartphones: Could User Privacy Profiles Help?. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. Association for Computing Machinery, New York, NY, USA, 201–212. https://doi.org/10.1145/2566486.2568035

[24] E. Maris, Timothy Libert, and Jennifer R. Henrichsen. 2019. Tracking sex: The implications of widespread sexual data leakage and tracking on porn websites. *ArXiv* abs/1907.06520 (2019).

[25] Aleecia M McDonald and Lorrie Faith Cranor. 2008. The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society* 4 (2008), 543–568. http://hdl.handle.net/1811/72839

[26] D. Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 2002. Developing and Validating Trust Measures for e-Commerce: An Integrative Typology. *Information Systems Research* 13, 3 (2002), 334–359. https://EconPapers.repec.org/RePEc:inm:orisre:v:13:y:2002:i:3:p:334-359

[27] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. Detecting Price and Search Discrimination on the Internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks* (Redmond, Washington) *(HotNets-XI)*. Association for Computing Machinery, New York, NY, USA, 79–84. https://doi.org/10.1145/2390231.2390245

[28] George R. Milne, Mary J. Culnan, and Henry Greene. 2006. A Longitudinal Assessment of Online Privacy Notice Readability. *Journal of Public Policy & Marketing* 25, 2 (2006), 238–249. https://doi.org/10.1509/jppm.25.2.238

[29] Ben Moskowitz, Mozilla, and Aza Raskin. 2011. Privacy Icons project (beta release). https://wiki.mozilla.org/Privacy_Icons. Last Accessed 4-November-2019.

[30] Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck, and Norman Sadeh. 2017. Identifying the Provision of Choices in Privacy Policy Text. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, Copenhagen, Denmark, 2774–2779. https://doi.org/10.18653/v1/D17-1294

[31] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Washington Law Review* 79 (2004), 119–158.

[32] Jonathan A. Obar. 2016. The Biggest Lie on the Internet: Ignoring the Privacy Policies and Terms of Service Policies of Social Networking Services. *SSRN Electronic Journal* (2016). https://doi.org/10.2139/ssrn.2757465

[33] Anna-Marie Ortloff, Lydia Güntner, Maximiliane Windl, Denis Feth, and Svenja Polst. 2018. Evaluation kontextueller Datenschutzerklärungen. In *Mensch und Computer 2018 - Workshopband*, Raimund Dachselt and Gerhard Weber (Eds.). Gesellschaft für Informatik e.V., Bonn.

[34] Anna-Marie Ortloff, Maximiliane Windl, Valentin Schwind, and Niels Henze. 2020. Implementation and In Situ Assessment of Contextual Privacy Policies. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (Eindhoven, Netherlands) *(DIS '20)*. Association for Computing Machinery, New York, NY, USA, 1765–1778. https://doi.org/10.1145/3357236.3395549

[35] Pascal Oser, Sebastian Feger, Paweł W. Wozniak, Jakob Karolus, Dayana Spagnuelo, Akash Gupta, Stefan Lüders, Albrecht Schmidt, and Frank Kargl. 2020. SAFER: Development and Evaluation of an IoT Device Risk Assessment Framework in a Multinational Organization. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 4, 3, Article 114 (Sept. 2020), 22 pages. https://doi.org/10.1145/3414173

[36] Sameer Patil, Roberto Hoyle, Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2015. Interrupt Now or Inform Later?: Comparing Immediate and Delayed Privacy Feedback. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 1415–1418. https://doi.org/10.1145/2702123.2702165

[37] Andrew S. Patrick and Steve Kenny. 2003. From Privacy Legislation to Interface Design: Implementing Information Privacy in Human-Computer Interactions. In *Privacy Enhancing Technologies*, Roger Dingledine (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 107–124.

[38] Robert W. Proctor, M. Athar Ali, and Kim-Phuong L. Vu. 2008. Examining Usability of Web Privacy Policies. *International Journal of Human-Computer Interaction* 24, 3 (2008), 307–328. https://doi.org/10.1080/10447310801937999

[39] Joseph Reagle and Lorrie Faith Cranor. 1999. The Platform for Privacy Preferences. *Commun. ACM* 42, 2 (Feb. 1999), 48–55. https://doi.org/10.1145/293411.293455

[40] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Carnor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, and Florian Schaub. 2014. Disagreeable Privacy Policies: Mismatches Between Meaning and Users Understanding. *Berkeley Technology Law Journal* 30, 1 (2014). https://doi.org/10.15779/Z384K33

[41] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. 2015. A Design Space for Effective Privacy Notices. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 1–17. https://www.usenix.org/conference/soups2015/proceedings/presentation/schaub

[42] Florian Schaub, Aditya Marella, Pranshu Kalvani, Blase Ur, Chao Pan, Emily Forney, and Lorrie Faith Cranor. 2016. Watching them watching me: Browser extensions impact on user privacy awareness and concern. In *NDSS workshop on usable security*.

[43] Paul M Schwartz and Daniel Solove. 2009. Notice & Choice. In *The Second NPLAN/BMSG Meeting on Digital Media and Marketing to Children*. 7 pages.

[44] Andreas Sotirakopoulos, Kirstie Hawkey, and Konstantin Beznosov. 2011. On the Challenges in Usable Security Lab Studies: Lessons Learned from Replicating a Study on SSL Warnings. In *Proceedings of the Seventh Symposium on Usable Privacy and Security* (Pittsburgh, Pennsylvania) *(SOUPS '11)*. Association for Computing Machinery, New York, NY, USA, Article 3, 18 pages. https://doi.org/10.1145/2078827.2078831

[45] Janice Y. Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. 2011. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research* 22, 2 (2011), 254–268. https://EconPapers.repec.org/RePEc:inm:orisre:v:22:y:2011:i:2:p:254-268

[46] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David Wagner, and Konstantin Beznosov. 2017. The Feasibility of Dynamically Granted Permissions: Aligning Mobile Privacy with User Preferences. In *2017 IEEE Symposium on Security and Privacy (SP)*. 1077–1093. https://doi.org/10.1109/SP.2017.51

[47] Daricia Wilkinson, Saadhika Sivakumar, David Cherry, Bart P Knijnenburg, Elaine M Raybourn, Pamela Wisniewski, and Henry Sloan. 2017. User-tailored privacy by design. In *Proceedings of the Usable Security Mini Conference*. https://www.ndss-symposium.org/wp-content/uploads/2017/09/usec2017_03_4_Wilkinson_paper.pdf

[48] 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, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh. 2016. The Creation and Analysis of a Website Privacy Policy Corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1330–1340. https://doi.org/10.18653/v1/P16-1126

[49] Jochen Wirtz, May O. Lwin, and Jerome D. Williams. 2007. Causes and consequences of consumer online privacy concern. *International Journal of Service Industry Management* 18, 4 (2007), 326–348.

[50] Kuang-Wen Wu, Shaio Yan Huang, David C. Yen, and Irina Popova. 2012. The effect of online privacy policy on consumer privacy concern and trust. *Computers in Human Behavior* 28, 3 (2012), 889–897. https://doi.org/10.1016/j.chb.2011.12.008

[51] Sebastian Zimmeck and Steven M. Bellovin. 2014. Privee: An Architecture for Automatically Analyzing Web Privacy Policies. In *Proceedings of the 23rd USENIX Security Symposium, San Diego, CA, USA, August 20-22, 2014*, Kevin Fu and Jaeyeon Jung (Eds.). USENIX Association, 1–16. https://www.usenix.org/conference/usenixsecurity14/technical-sessions/presentation/zimmeck

[52] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel Reidenberg, N. Cameron Russell, and Norman Sadeh. 2019. MAPS: Scaling Privacy Compliance Analysis to a Million Apps. *Proceedings on Privacy Enhancing Technologies* 2019, 3 (2019), 66 – 86. https://content.sciendo.com/view/journals/popets/2019/3/article-p66.xml

A APPENDIX

The statements we asked about in Study I, rated on a 5-point Likert scale ranging from 1 (=strongly disagree) to 5 (=strongly agree):

- S1: "The bubble is completely visible (you see a bubble highlighted in green and no parts of it are covered by elements of the website)."
- S2: "The text of the sidebar is completely readable (no text is covered by elements of the website)."
- S3: "The sidebar's category (heading/icon) matches with the placement of the bubble (the element highlighted in red), e.g. the bubble containing information about advertisements is placed on advertisements."
- S4: "The bullet points match the category/heading/icon, e.g. the bullet points contain information about advertisements when the heading is Advertising."

Table 3: Reasons for which PrivacyInjector failed to automatically detect the privacy policies of 61 of the US' top 500 websites, the count, and the URLs affected, as well as possible solutions (analysis conducted November '21).

Reason	Num	URLs	Possible Solution
Active interaction required to get to start page/English version	6	bestbuy.com, office365.com, quora.com, norton.com, istockphoto.com, spotify.com, delta.com, fidelity.com	Requires real user interaction
Privacy Policy not in English	1	soundcloud.com	Works for users locatedin English speaking countries
Unconventional layout	7	ancestry.com, feedly.com, disneyplus.com, adobe.com, mozilla.org, typeform.com, studentaid.gov	Offering option to manually enter the policies' URLs
Different base URL of privacy policy (e.g., blogspot's privacy policy leads to google's privacy policy)	27	blogspot.com, match.com, mapquest.com, gotowebinar.com, giphy.com, steamcommunity.com, nextdoor.com, patreon.com, shareasale.com, pinimg.com, ikea.com, doordash.com, wikimedia.org, syf.com, instagram.com, steampowered.com, att.net, service-now.com, trello.com, wsj.com, citi.com, thesaurus.com, citibankonline.com, alibaba.com, ksl.com, surveymonkey.com, access.wa.gov	Offering option to manually enter the policies' URLs
Privacy policy opens in new tab	5	robinhood.com, accuweather.com, onlyfans.com, lowes.com, behance.net, spectrum.net	Offering option to manually enter the policies' URLs
Unusual keyword in privacy policy link	1	worldometers.info	Offering option to manually enter the policies' URLs
Cryptic privacy policy URL	2	npr.org, staples.com	Offering option tomanually enter the policies' URLs
Privacy policy not linked on start page	5	pinterest.com, nih.gov, archive.org, imgur.com, unsplash.com	Users followingadditional navigation paths could trigger automatic identification. Otherwise: Offering option to manually enter the policies' URL
Unknown (works now (Nov '21) but did not at time of scraping)	7	grubhub.com, sba.gov, biblegateway.com, chewy.com, wikihow.com, ballotpedia.org, apartments.com	-

Table 4: The ID, age, gender, education, and occupation of the participants of Study II.

ID	Age	Gender	Education	Occupation
1	50	Male	Apprenticeship	Producer
2	18	Female	Bachelor's Degree	Student
3	23	Female	Bachelor's Degree	Graduate Student
4	24	Female	Bachelor's Degree	Student
5	23	Female	Bachelor's Degree	Student Art Education, Art Director
6	28	Female	Bachelor's Degree	Student
7	32	Male	Master's Degree	Project Manager
8	28	Female	Bachelor's Degree	Student
9	25	Male	Bachelor's Degree	Student Environmental Technology
10	23	Female	Bachelor's Degree	Student and Working Student in Product Management
11	37	Female	Master's Degree	Computing Services Specialist