2023 IEEE 8th European Symposium on Security and Privacy (EuroS&P) | 978-1-6654-6512-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/EuroSP57164.2023.00055

DarkDialogs: Automated detection of 10 dark patterns on cookie dialogs

Daniel Kirkman
University of Edinburgh
Edinburgh, UK
D.J.Kirkman@ed.ac.uk

Kami Vaniea
University of Edinburgh
Edinburgh, UK
Kami.Vaniea@ed.ac.uk

Daniel W. Woods
University of Edinburgh
Edinburgh, UK
Daniel.Woods@ed.ac.uk

***Abstract*—In theory, consent dialogs allow users to express privacy preferences regarding how a website and its partners process the user's personal data. In reality, dialogs often employ subtle design techniques known as dark patterns that nudge users towards accepting more data processing than the user would otherwise accept. Dark patterns undermine user autonomy and can violate privacy laws. We build a system, DarkDialogs, that automatically extracts arbitrary consent dialogs from a website and detects the presence of 10 dark patterns. Evaluating DarkDialogs against a hand-labelled dataset reveals it extracts dialogs with an accuracy of 98.7% and correctly classifies 99% of the studied dark patterns. We deployed DarkDialogs on a sample of 10,992 websites, where it successfully collected 2,417 consent dialogs and found 3,744 different dark patterns automatically present on the consent dialogs. We then test whether dark pattern prevalence is associated with each of: the website's popularity, the presence of a third-party consent management provider, and the number of ID-like cookies.**

***Index Terms*—web privacy, consent dialogs, GDPR, user interface, internet measurement**

1. Introduction

Consent interfaces (also known as cookie dialogs) are often presented to users immediately upon visiting a website [55], particularly in the EU [19]. In theory, consent interfaces protect user privacy. Websites ask for a user's permission before processing personal data, which allows privacy conscious users to deny consent and thereby protect their privacy. Collecting consent generates value for the website and its advertising partners [74]. The benefits of collecting consent from users include evidence of legal compliance [62], [63], satisfying advertising partners [56], and even improving sales [25].

In reality, the potential benefits lead websites to design interfaces in a way that makes opting-out more difficult [15]. These design choices are known as *dark patterns*. Obviously manipulative examples of dark patterns include having no opt-out button or ignoring opt-out decisions. More subtle dark patterns include using overly complex language or colour-coding the opt-in button so that it is more prominent. All four patterns are studied in this paper, in addition to a further six patterns. Dark patterns undermine user autonomy [28] and can violate data protection laws [62].

This motivates building a system that can accurately detect dark patterns on webpage consent dialogs. Such a system could be used by well-meaning developers to determine whether consent dialogs extend autonomy to users or undermine it via dark patterns. It could also be used to identify websites that intentionally include dark patterns in order to send a notification about non-compliance or even report the website to authorities. Regulators have issued a small number of fines related to non-compliant dialogs. For example, the French Data Protection Authority (CNIL) recently fined Google ($170 million) and Facebook ($68 million) [14], both in relation to consent dialogs.

While popular websites can be checked with manual analysis, these two fines are the tip of the iceberg given researchers identified thousands of non-compliant dialogs soon after the GDPR [47], [53] and e-Privacy Directive [42] came into effect. Inconsistent regulatory enforcement motivated a non-profit, 'None of Your Busi-

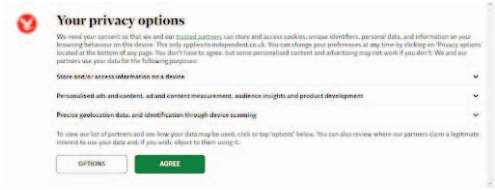

Figure 1: Example of the HighlightedOptIn dark pattern (Source: independent.co.uk)

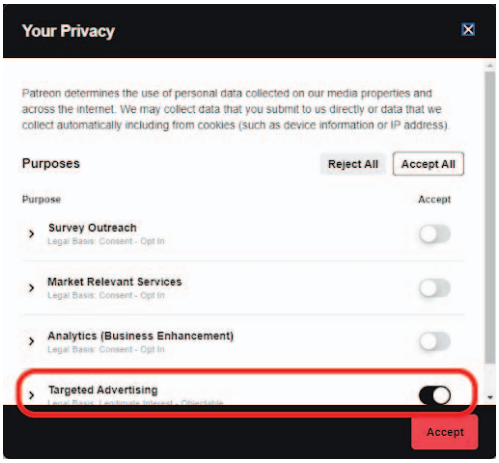

Figure 2: Example of the PreferenceSlider dark pattern (Source patreon.com).

DOI 10.1109/EuroSP57164.2023.00055

ness' (NOYB) [21], to send a written warning about non-compliant dialogs to over 500 companies in May 2021 [21]. 82% of companies failed to comply after a warning, which led NOYB to file 422 complaints with data protection authorities across Europe.

The problem of detecting dark patterns remains unsolved for two reasons. First, prior work has largely focused on extracting a sub-sample of dialogs, namely those designed by third-parties known as consent management providers (CMPs) that are embedded in websites or those that implement a specific legal framework [30]. This allows websites who self-implement dialogs to evade automatic crawls. Second, malicious designers can create new dark patterns that can evade detectors. This creates a similar situation to vulnerability detection in which defenders must continually scan for and fix novel vulnerabilities used by adversaries. This motivates detecting a wide range of dark patterns.

Contributions We build a system, *DarkDialogs*, that makes the following contributions:

- *DarkDialogs* advances the state-of-the-art in both dialog extraction *and* dark pattern detection by implementing a ranking-based approach using CSS selectors to locate consent dialogs and using narrowly-defined patterns to detect dark patterns on consent dialogs.
- We automatically extract arbitrary dialogs and achieve an accuracy of 98.7% in this task, which improves on the only other accuracy rate reported in the literature (91%) [19].
- *DarkDialogs* can automatically detect 10 dark patterns in arbitrary dialogs. The system correctly classified over 99% of the 1,375 pattern instances in the two human-labelled datasets. For comparison, another study achieved accuracy of between 54% and 72% depending on the dark pattern [65].
- We introduce some new dark patterns (e.g. MultipleDialogs) and create novel heuristics for others (e.g. HighlightedOptIn), as well as studying known dark patterns. Detecting this range of dark patterns across arbitrary dialogs provides first evidence that well-resourced actors deploy more subtle dark patterns, which exist in a legal grey area.

The resulting system can be used by website owners to proactively avoid legal issues, privacy advocates to detect and report non-compliant websites, and regulators to collect evidence and issue warnings.

Section 2 identifies historical and contemporary research on consent dialogs. Section 3 introduces the design properties of our system. Section 4 evaluates the system on a hand-labelled dataset. Section 5 presents results from a random crawl of a subset of the Tranco Top 1 million websites. Section 6 discusses the design of our system, as well as the measurement results. Section 7 highlights the conclusions.

2. Related Work

Informed consent has been collected from users since the early days of the Web. More recently, regulations in the EU including the 2009 ePrivacy Directive and the General Data Protection Regulation (GDPR) have issued guidance around when and how to collect consent in the context of cookies and data processing.

Early Consent Interfaces In the 1990s, many browsers voluntary—not due to a legal obligation—requested permission from users before installing cookies. Millet et al. [50] describe how the Netscape Navigator and Internet Explorer browsers collected consent for installing cookies between 1995 and 1999. The authors recommend that browsers should include an "option to decline all" cookies. The impact of such design choices was studied using a $2 \times 2 \times 3$ design [3], which showed that the button text and default are the most important factors. Although the term did not exist, both studies raised concerns about *dark patterns* that nudge users towards consent decisions that reduce user privacy.

2009 ePrivacy Directive The reformed ePrivacy Directive created an obligation for "companies to obtain Internet users" consent before storing or accessing cookies on a device" [5]. This generated legal scholarship [5], [16], [42], [44]. One such legal team conducted a manual analysis of the top 100 Dutch websites [42], which revealed that 56% of websites had dialogs in which rejecting consent was not an option, and just 6% of websites implemented cookie banners that offer a meaningful choice to users—again this is an example of researchers studying dark patterns before the language existed.

GDPR The General Data Protection Regulation created a new reason for websites to collect user consent, namely it serves as a legal basis for processing personal data. Establishing a legal basis is necessary to avoid non-compliance with Article 6 of the GDPR, which could lead to hefty fines, for both first-party websites and also the wider ecosystem of firms who process personal data. The GDPR further specifies requirements and principles for *how* user consent should be collected.

Many studies have investigated consent dialogs since the GDPR came into effect. *HCI studies* test how design choices—now called dark patterns—impact the consent rate [3], [24], [45], [53], [70]. Other studies try to conceptualise and build alternative systems to manage user consent [36], [54], [64], [73], such as privacy preference signals [31]–[33], [75], [76]. Another approach is to measure how real world consent dialogs are designed, which is also the goal of our study. *Scraping studies* detect which dark patterns are deployed by websites.

We conducted a brief literature review of scraping studies. Kretschmer et al. provide a comprehensive literature survey [40]. Table 1 provides a best-effort mapping of empirical studies of dark patterns in consent dialogs to the dark patterns studied in our paper. Such studies follow a common research design but vary in terms of: (1) how and which dark patterns are classified; (2) the sample of websites; and (3) how and which dialogs are extracted.

Although the topic of dark patterns has animated the research community, a precise definition has been elusive [46]. This results in a variety of conceptual and technical definitions of dark patterns in the context of consent dialogs. Reliable technical indicators exist for narrowly defined patterns, such as whether the dialog has a reject button [53] or whether the website sets cookies before a decision is made [42]. Simple technical indicators are not available for the high-level taxonomy of dark patterns introduced by Gray et al. [28]. For example, a classifier

Authors	Study	Year	# Domains	# Dialogs	Automated detection	All dialog types	OnlyOptIn	HighlightedOptIn	ObstructsWindow	ComplexText	MoreOptions	AmbiguousClose	MultipleDialogs	PreferenceSlider	CloseMoreCookies	OptOutMoreCookies
Leenes & Kosta	[42]	2015	100	50	○	●	◐		◐							
Sanchez-Rola et al.	[61]	2019	2000	< 320	○	○	●		●		●					●
Degeling et al.	[17]	2019	6.6k	4.2k	●	○	●							◐		
Eijk et al.	[19]	2019	1.5k	648	◐	●	○		○							
Nouwens et al.	[53]	2020	10k	680	●	○	◐		●		◐			●	●	
Matte et al.	[47]	2020	28.2k	1.4k	◐	○	●							●		●
Hils et al.	[30]	2020	4.2m	414*	◐	○				◐	●					
Krisham et al.	[41]	2021	500	255	○	●	●	●	◐			◐				
Kampanos & Shahandashti	[37]	2021	17.7k	7.5k	●	●	●			◐	●	◐				
Bollinger et al.	[4]	2022	6m	29.4k	●	○										●
Us	-	2022	11k	2k	●	●	●	●	●	●	●	●	●	●	●	●

TABLE 1: A best effort mapping of whether the dark patterns studied in our paper were also studied in the literature. In automated detection column: ● = automated dialog extraction and Dark Pattern detection, ◐ = automated either dialog extraction or Dark Pattern detection, and ○ = manual. In detects all dialogs column: ○ = extracts subset of all dialogs (e.g. those provided by CMPs) and ● = extracts all dialogs. For the dark pattern columns: ● = measured, ◐ = a similar pattern was measured, and ○ = pattern was mentioned but not measured. Prior work also measured dark patterns not described in this table.

built by Soe et al. to detect high-level patterns ("nagging, obstruction, sneaking, interface interference, and forced action" [28]) performed poorly with an accuracy rate of $0.535 - 0.72$ against a baseline of 0.33 [66]. The authors recommend detecting patterns tailored to cookie dialogs. Adopting this approach in our work led to a system with a detection accuracy of 98%+ (see Section 3.5).

Another problem centres on how the sample of websites are collected. Most studies sample the top X websites in a most visited list (e.g. Tranco [59]) with sample sizes ranging from hundreds to millions (Table 1). Alternative approaches include sampling from URLs shared over social-media [30] and directly studying consent dialog service providers [69]. Studies vary in terms of whether they consider all dialogs or a subset. For example, many studies [4], [30], [47], [53] focus on dialogs built by so-called Consent Management Providers (CMPs). This approach allows for easier automation because websites re-use the same CMP code, but means that such studies ignore websites who build their own dialogs.

Manual analysis enables studies of arbitrary dialogs but cannot scale beyond relatively small datasets (e.g. 255 dialogs [41] or 105 popular services [29]). Two prior works automated extraction of arbitrary dialogs, both using a similar approach to ours (see Section 3). However, Kampanos et al. [37] do not evaluate the performance of the automated detector that they built. This means a high rate of false negatives could explain why they find a "lower prevalence [of dialogs] than that of the earlier study" [37, p. 220]. Eijk et al. [19] automate detection and achieve a total accuracy rate of 91% (for comparison our system achieves 98%). Notably, the authors also discuss two of the dark patterns that our study does, but do not report on the fraction of websites that display these patterns. Although initial steps have been made towards automatically detecting dark patterns in arbitrary dialogs, we are not aware of any studies that design and fully evaluate such a system, which is the main contribution of our study.

3. DarkDialogs Design

To detect dark patterns on cookie consent dialogs we built a system to automatically find dialogs on webpages. The system also collects and classifies the clickable elements on the dialog and the cookies set by the page before and after clicking on each element. Finally, it also searches the dialog for dark patterns. The system has been designed as 4 modules and a control module. Figure 3 shows the main steps in each module and the workflow of the system.

The system uses Selenium's[1] web scraping Python library with the ChromeDriver to load and interact with websites using the Chrome browser. We selected Chrome because it is the most widely used browser [72] and has minimal default privacy settings compared to other browsers such as Firefox [52]. We also used a UK-based VPN to ensure constant location as some webpages change their presentation based on the geolocation of the user.

3.1. Dialog detection

Cookie dialogs exhibit a wide range of designs, wording, and options. These variations makes them challenging to locate automatically on potentially very complex

1. https://www.selenium.dev/

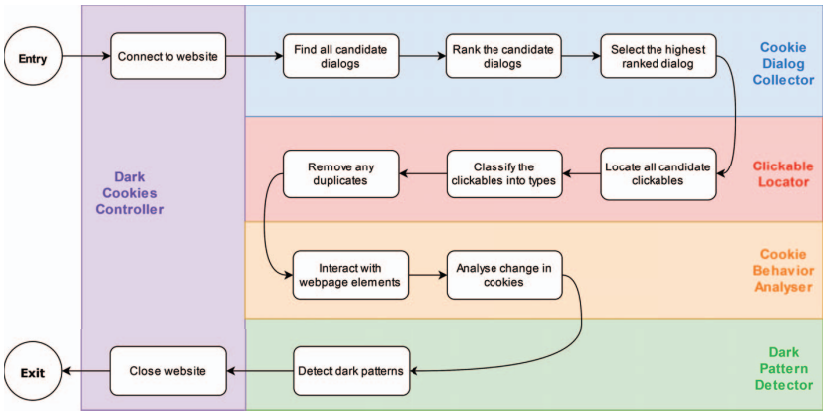

Figure 3: The main actions taken by the designed system to analyse a single website.

websites. To further complicate the situation, some ad-blocking software also block cookie dialogs which websites do not like. As a result, many websites regularly change the technical design and setup of their cookie dialogs to make it hard for them to be found and blocked. For example, the *google.hk* website changes the cookie dialog HTML attributes on every load and loads the cookie content via JavaScript to make it extra hard to locate the dialog. By considering dialog features other than the CSS selectors, such as text and screenshots, our system was able to successfully detect this cookie dialog. In short, finding cookie dialogs is a non-trivial problem with an active adversary (websites) that regularly changes tactics.

The process we have developed to find a cookie dialog can be broken down into the following steps:

1) Find all candidate dialogs which are potential elements on the web page that could be cookie dialogs.
2) Rank candidate dialogs from most to least likely to be a cookie dialog.
3) Select the highest-ranked candidate dialog as the valid cookie dialog. By "valid" we mean that this candidate is the best representation of a cookie dialog present on the webpage.

3.1.1. Finding all candidate dialogs. To locate dialog candidates we make heavy use of CSS Selector Lists. CSS allows developers to *select* HTML elements using a range of pattern types known as *CSS Selectors*. Selectors are made up of a wide variety of CSS Properties such as the element ID, class, or position in the DOM structure. Several browser plugins allow users to block cookie dialogs from appearing, to locate the dialogs plugins typically use crowdsourced CSS Selector lists of known cookie dialog elements and common cookie dialog patterns. They also typically allow users to update the shared lists, thereby making the lists a crowdsourced shared resource that is regularly updated. In this work, we make use of two frequently updated CSS Selector lists: *EasyList* [57] and *I don't care about cookies* [39].

To locate elements on a web page that could be cookie dialogs we use the following selectors:

1) **Domain-specific CSS selectors:** CSS selectors from a CSS selector list which are only valid for identifying a cookie dialog on a particular domain.
2) **General CSS selectors:** CSS selectors which are commonly associated with cookie dialogs.
3) **Custom CSS selectors:** CSS selectors developed as part of this project to search div tags containing common terms. The full list of Custom CSS Selectors is included in Appendix A.2.
4) **iframes:** HTML element tags that are used to embed an external element in the current web page. Cookie dialogs are commonly contained within iframes so the system initially marks all such included content as a candidate dialog.

3.1.2. Ranking all the candidate dialogs. Both prior work [51], [58] and our own observations find that CSS selectors are a good way to find dialogs, but their overall accuracy is not high. Also the intuition that a match against a domain-specific list will be more accurate than a match against a general list does not hold reliably. The cause of these issues is likely the changes websites regularly make to prevent ad-blockers from blocking content. These observations lead us to use the method of first collecting a set of all possible candidates and then using ranking to decide which, if any, is most likely to be a cookie dialog.

After identifying a candidate dialog the system automatically collects the following features about it: how it was located, text content, element's HTML, and a screenshot of the element. Screenshots were collected via Selenium which can screenshot specific elements using CSS selectors. If the element is hidden or otherwise not visible to the user, Selenium will return an error.

All candidates are then assigned a score starting at 0 and then adjusted using the Ranking Factors summarised in Table 2. We will briefly describe the criteria behind each ranking factor in the remainder of this section. The full criteria for each Ranking Factor is complex and not essential to understand the rest of the paper but it is provided in Appendix A.1. Candidates are completely removed if a screenshot cannot be taken (likely user-invisible), or if they are identical to another candidate. They are also seriously penalised for containing no text, as most real dialogs contain text. Prior work [51], [58] successfully used manually curated N-grams of common words and phrases found on cookie dialogs to identify real

cookie dialogs, we adapt their lists and approach to adjust scores upward for dialogs containing such common words and phrases (Appendix A.3). Scores of dialogs located using Domain-specific and General CSS selectors are also adjusted upward, as these were at some point human validated, though that may have changed. Finally, it is common for the candidate list to contain HTML children and parent elements. So we give a mild penalty to the child in such cases as non-text elements such as a dialog close icon may reside in the parent element making the parent a better candidate choice if all other features are equal.

3.1.3. Selecting the highest-ranked candidate dialog. The valid dialog is selected to be the highest scoring dialog with a positive score. Overall the scoring approach is designed to only give positive scores to candidate dialogs that match a CSS selector list or contain common cookie dialog words or phrases and give negative scores to candidates that are unlikely to be a cookie dialog. Therefore, if there are no positively scored candidates, the website is marked as not having a cookie dialog present.

3.2. Consent Management Providers

A **Consent Management Provider (CMP)** is a company that provides cookie dialogs and cookie technology for a website [34]. In a previous study, Hils et al. [30] developed fingerprints for 6 major CMPs. We updated and expanded upon the fingerprints created by Hils et al. to cover cookie dialogs from 13 major CMPs. Fingerprints were based on CSS Selectors, text content, and hostname of cookie dialogs. The full list of fingerprints is provided in Appendix B. For each valid cookie dialog, we compared it against the fingerprint list and if there was a match, labelled it as from that CMP. We then manually validated all CMP labels. If a fingerprint yielded a False Positive or False Negative, we adapted or removed the fingerprint.

3.3. Clickable Location & Classification

For each valid cookie dialog, the system attempts to locate and classify all the clickable elements.

A set of custom CSS selectors (Appendix C.2) are first used to locate all possible clickable elements in the dialog. Then each clickable is classified as one of the clickable types listed below.

1) **Opt-in option:** button which allows the user to consent to all Cookies. Usually worded affirmatively using words such as ‘Accept’, ‘Yes’ or ‘OK’. However, depending on how the consent dialog has phrased the wording of this button may be inverted with the Opt-out button.
2) **Opt-out option:** button which allows the user to reject the collection of Cookies depending on their preference. Usually worded negatively using words such as ‘Reject’, ‘No’ or ‘Opt-Out’.
3) **More options:** button which redirects the user to another dialog where there are more preference options available. Does not include links to cookie policies which don’t have any options.
4) **Preference Slider:** a single element that allows users to consent or reject certain types of cookies before confirming this choice using the Confirm Preferences button. Includes check boxes or toggle switches. We also distinguish between preference sliders that are enabled or disabled. Many websites have a checkbox for ”essential cookies” which will be enabled by default, this usually cannot be switched off by the user since they are essential, this does not count as an enabled preference slider.
5) **Confirm Preferences:** a button used to confirm the cookies preference made using sliders. In some cases, the Opt-out button may take its place.
6) **Close option:** button that allows the user to close the cookie dialog without selecting an option. Commonly worded as ’X’ or ’Close’. Does not include cases where the text of the dialog says this button could count as in opt-in button. For example, if the dialog said ”By clicking the close button you agree to all cookies” then this would count as an opt-in button not a close button.
7) **Policy Link:** link to the website’s Privacy or Cookie Policy page which does not contain any options to select consent preference.

Figure 4 provides an annotated example of clickables on a consent dialog.

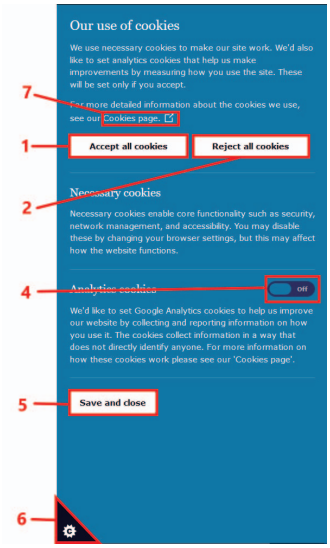

Figure 4: Examples of Clickables (highlighted in red) on a cookie dialog (Source: https://ico.org.uk/) Note: numbers correspond to the list of clickables in Section 3.3.

Classification is based on a custom list of keywords (see Appendix C.1) derived from the study by Petronyte [58] who improved the set of keywords identified by Molnar [51]. We similarly expanded the set during the implementation of this system. Any clickables that did not contain one of the keywords were disregarded. Since all of our keywords were generated in English we translated any non-English text using a Google Translate API. Using any online translation service will not always give us the grammatically correct translation. However, for this project, we are generally only interested in checking for the presence of keywords in any translated text. So this accuracy of translation will suffice. To improve the clas-

Candidate Feature	Ranking Factor	Score Adjustment
Collection Method	Found using Domain-Specific CSS selector	+10
	Found using General CSS selector	+5
Text Content	Contains common cookie dialog N-grams	Unigram = +1 Bigram = +2 ...
	Word count less than 5	-20
	Word count greater than the average plus 100	-20
	Contains no text	-100
Element HTML	Substring of another candidate	-1
	Same as another candidate	Removed
Screenshot	Candidate could not been screenshotted	Removed

TABLE 2: Table showing the factors and score adjustments used to rank candidate dialogs.

sification accuracy we apply case folding and stemming to any text before classification. During implementation, it was observed that keywords alone were not enough to locate all of the "close option" and "preference slider" elements so we developed a custom set of CSS selectors to detect these clickable types (full lists in Appendix C.2). To remove duplicate clickables, we compare the HTML content of candidate clickables. Duplicates can occur when multiple CSS selectors locate the same element on a web page.

3.4. Measuring Cookie Setting Behaviour

To measure the change in cookies set by the browser after interaction with a clickable, we used a combination of Selenium and the Chrome DevTools Protocol command `Network.getAllCookies` [27] which returns all cookies currently stored by the Chrome browser. Cookies are collected under the following scenarios: after the initial page load, after clicking the opt-in option, after clicking the opt-out option, and after clicking the close option.

During development we observed that it took up to 30 seconds for all cookies to be set. So to ensure all cookies have been set, the system waits 30 seconds. After 30 seconds, the system compares the cookies every 10 seconds until there are no changes in the cookies set and only then do we collect cookies.

Between each collection the browser cache was completely cleared and the website reloaded to ensure that prior interaction would not impact results.

Collected cookies were also labelled as first- or third-party based on if the domain that set them was the same as the one the system visited (first) or a different domain (third). The system also looked at the cookie values to determine if they are likely to be a unique identification number (ID-like).

Sanchez-Rola et al. [61] distinguish ID-like cookies using the `zxcvbn` password strength algorithm which roughly measures entropy.

To allow tracking across websites, ID-like cookies must also be persistent, that is they must not deleted immediately after the browsing session is closed. In their study Englehardt et al. [20] consider cookies with an expiry date of over 90 days to be identifier cookies. Typically, tracking ID-like cookies have a long expiration date set to allow a more complete user profile to be constructed. To determine if a cookie is ID-like it must meet the following criteria:

1) The `zxcvbn` [20] score of any component of the value attribute of the cookie is greater than 3 (this is equivalent to 100 million guesses and is the suggested value chosen by Sanchez-Rola et al. [61]).
2) The cookie is set to expire more than 90 days after it was set.

3.5. Dark Pattern Detection

The concept behind dark patterns is somewhat subjective, and most of the categorisations proposed in prior work are similarly framed more in terms of human judgement than at the operational level needed for automation. Our challenge therefore was to take existing identified dark patterns found in literature (Table 1) and convert them into a set of patterns which can be automatically detected with high reliability. Below we describe the set of dark patterns the system is able to identify along with the impact the pattern has on the user, its category (Gray et al. [28]), its legal implications, and the criteria used to automatically identify it. As part of the legal implications we consider PECR, GDPR, and any guidance from European data protection authorities. Screenshot examples of dark patterns are included in Appendix D.3.

OnlyOptIn. *Only the opt-in option is present on the initial cookie dialog.*

Category: Forced Action.

Impact on the user: The user has no choice but to accept all cookies.

Legal implications: Contradicts EDPB guidance that says "users should have an opportunity to freely choose between the option to accept some or all cookies or to decline all or some cookies" making it non-compliant with the GDPR [18]. In the UK the Information Commissioner's Office (ICO) uses this dark pattern as an example of a cookie dialog design that is non-compliant with the PECR [35].

Criteria: Satisfies all of the following:

1) There is an opt-in button present.
2) There is not a opt-out or more options button present.

HighlightedOptIn. *The background colour of the opt-in button leads to it being highlighted more compared to the opt-out button.*

Category: Interface Interference.

Impact on the user: The opt-in button stands out more to the user so they are more likely to click the opt-in button without considering other options.

Legal implications: The ICO singles out this dark pattern as an example of a design non-compliant with the PECR: "[a] consent mechanism that emphasises 'agree' or 'allow' over 'reject' or 'block' represents a non-compliant approach, as the online service is influencing users towards the 'accept' option" [35].

Criteria: If the clickable is light then it will stand out on a darkly coloured dialog. Conversely, if a clickable is dark then it will stand out on a lightly coloured dialog.

To determine the relative brightness, we convert dialog and clickable screenshots to grey-scale. We will get a grey-scale value of between 0 (Dark) and 255 (Light) which is the approximate brightness. During testing, it was determined that considering any average value over 170 be "light" and otherwise "dark" was a good threshold. However, further work is needed to analyse the effectiveness of this threshold. To match this pattern the dialog must match one of the following:

1) The background of the cookie dialog is "light" AND the opt-in button is "dark" AND the opt-out button/more options is "light".
2) The background of the cookie dialog is "dark" AND the opt-in button is "light" AND the opt-out button/more options is "dark".

ObstructsWindow. *Dialog obstructs most of the window.*

Category: Obstruction.

Impact on the user: User may be hindered or entirely prevented from interacting with the rest of the web page without first interacting with the cookie dialog.

Legal implications: Constitutes a 'consent wall' which means it blocks access to the website until the user expresses their choice regarding consent. Santos et al. [62] argue that consent walls are an "unnecessary disruption to the use of a website/app" and are non-compliant with GDPR.

Criteria: The area (length $\times$ width) of the dialog is greater than 60% of the area of the visible webpage.

ComplexText. *Dialog text is difficult to read.*

Category: Interface Interference.

Impact on the user: Text content of the dialog is difficult to understand to an extent that the average user may struggle to comprehend it.

Legal implications: Using overly complicated language on consent mechanisms is regarded as bad practice and being non-compliant with the GDPR. The EDPB says that "The requirement for clear and plain language means that information should be provided in as simple a manner as possible, avoiding complex sentence and language structures" [18].

Criteria: The **Flesch-Kincaid (FK)** reading ease test is a metric used to evaluate the readability of a passage of text. The FK test was originally developed for the US Navy [38] to assess the complexity of their technical manuals. It has been widely used as a metric for automatically calculating readability, including in programs such as Microsoft word [49]. FK scores are numeric values between 0 and 100, where a higher score indicates simpler text and a lower score indicates more complex text. The FK score for a passage of text can be calculated using the following equation [38] :

$$FK = 206.835 - 1.015(\frac{\text{Total\_words}}{\text{Total\_sentences}}) - 84.6(\frac{\text{Total\_syllables}}{\text{Total\_words}})$$

The process of accurately determining the number of syllables in a word automatically is a challenging task so we use the Python `textstat` [1] module to calculate FK scores in the system. The FK score defines any passage of text with a score of 50 or less as being difficult to read and requiring a college-level education to be understood. Thus, we define the text of any cookie dialog which has an FK score of 50 or less as being difficult to read.

MoreOptions. *Dialog hides some options behind a more options button.*

Category: Interface Interference.

Impact on the user: The user is required to navigate to additional cookie dialog interfaces to select their preferences, possibly pushing users to accept all cookies via the initial interface as it requires substantially less effort.

Legal implications: The ICO singles out this dark pattern as an example of a design non-compliant with the PECR and says "A consent mechanism that doesn't allow a user to make a choice would also be non-compliant, even where controls are located in a 'more information' section" [35]. The French CNIL recently fined Google and Facebook for having this dark pattern on their cookie dialogs. CNIL ruled that Google and Facebook "offer a button allowing the user to immediately accept cookies. However, they do not provide an equivalent solution (button or other) enabling the Internet user to easily refuse the deposit of these cookies. Several clicks are required to refuse all cookies, against a single one to accept them" [14].

Criteria: Dialog contains a "more options" button AND the dialog does not contain a "opt-out" button.

AmbiguousClose. *Ambiguous close button is present on the dialog.*

Category: Interface Interference.

Impact on the user: The impact of the close button may be ambiguous. Users may not know if clicking the button will opt-in, opt-out, or make another cookie setting choice on their behalf.

Legal implications: The GDPR requires consent to be given by a user by "clear affirmative action". The EDPB suggests that "Controllers must avoid ambiguity and must ensure that the action by which consent is given can be distinguished from other actions" [18]. The ambiguous nature of the close button makes it unclear if clicking results in consenting.

Criteria: a close option button is present on the dialog.

MultipleDialogs. *Multiple distinct cookie dialogs present on a page.*

Category: Interface Interference.

Impact on the user: The user may be uncertain about which cookie dialog they should interact with and different dialogs may have different options or instructions available.

Legal implications: The GDPR requires consent to be informed and unambiguous [23]. In cases where a website presents multiple cookie dialogs with a different range of consent choices, it would be hard for users to be informed what their true choices were. If a user consented to one dialog but declined another, then it would be ambiguous if they had provided consent or not.

Criteria: There is more than one candidate dialog with a score greater than 0 and the HTML content of at least two such dialogs is different.

PreferenceSlider. *At least one preference slider is enabled by default.*

Category: Sneaking.

Impact on the user: Many sliders may be present leading users to miss those that are enabled, checking and adjusting the sliders also adds additional effort for the user.

Legal implications: The GDPR says that "pre-ticked boxes or inactivity should not therefore constitute consent" [23] meaning that any preference sliders which are enabled by default cannot be used to gain consent for non-essential cookies. In 2019 the Court of Justice of the EU ruled against `planet49.com` who used a pre-ticked box to obtain consent for cookies saying "consent given in the form of a preselected tick in a checkbox does not imply active behaviour" [7].

Criteria: Many dialogs have a preference slider for "necessary cookies" which cannot be interacted with and is enabled by default, we do not count this type of preference slider as a dark pattern. Sliders that do count satisfy both of the following:

1) Can be interacted with by the user.
2) Are enabled before any interaction from the user.

CloseMoreCookies. *Clicking the close button leads to more ID-like cookies being set.*

Category: Sneaking.

Impact on the user: Closing the dialog, which is ambiguous, resulted in more cookies that are likely to be associated with tracking.

Legal implications: To comply with the lawfulness principle of the GDPR the user must provide their consent for any non-essential cookies. As discussed previously, the ambiguous close button does not provide the "clear affirmative action" required for consent under the GDPR [23]. By setting non-essential cookies without gaining consent to the proper standard websites may be in violation of GDPR.

Criteria: Cookies are used for a wide range of purposes, so for this dark pattern we only look at ID-like cookies which are the most likely to be used for unique identification of the user. The criteria is that the number of ID-like cookies increases after the close button is clicked.

OptOutMoreCookies. *More ID-like cookies are set regardless of the opt-out button being clicked.*

Category: Sneaking.

Impact on the user: Despite the user opting out to reject cookies the website has disregarded their choice and set more cookies that are likely associated with tracking.

Legal implications: To comply with the lawfulness principle of the GDPR the user must provide their consent for any non-essential cookies [22]. The EDPB says that "if the individual decided against consenting, any data processing that had already taken place would be unlawful" meaning this dark pattern would be non-compliant with the GDPR [18].

Criteria: Similar to CloseMoreCookies we focus on ID-like cookies, resulting in a criteria that the number of ID-like cookies increases after the opt-out option is clicked.

4. Dialog Collection and System Validation

4.1. Datasets

We created three sets of websites which were drawn from the 2021 Tranco list [59] of most popular websites and then run through the system. The two smaller sets of 500 websites were human-labelled and used to validate system accuracy. A set of 10K was then collected to provide a wide view of cookie dialog behaviour. The sets are detailed below:

Top500: the top 500 most popular websites collected between 22/01/22 and 29/01/22.

Rand500: random subset of 500 websites chosen from the top 1 million websites, websites collected on 14/02/22. The Top500 websites were excluded to avoid overlap.

Rand10k: random subset of 10,000 websites chosen from top 1 million websites and collected between 15/02/22 and 24/02/22. This dataset does overlap with the Top500 and Rand500 websites.

Data11k: union of the prior three sets with duplicates removed, resulting in 10,992 websites. For brevity we use 11k throughout the paper. Where duplicates were found we kept the human-labelled domains (Top500 and Rand500 sets).

4.2. Manual Human Labelling

Consent dialogs are quite easy for a human to identify given their difference from other content. To create a set of ground-truth labels the lead researcher manually went through all web pages in the Top500 and Rand500 sets.

For efficiency, they used the User Interface of the DarkDialogs tool that showed the website and the list of potential consent dialogs, clickables, and dark patterns previously identified by the system as well as a way to manually add any element not automatically identified.

The researcher verified that the system had ranked the correct candidate consent dialog first, otherwise they entered the CSS selector for the correct dialog. Or if there was no consent dialog, they indicated that. Secondly, the researcher validated or adjusted the identified clickable

elements, ensuring that the system had correctly identified them on the dialog and correctly classified their type (i.e. opt-in option). Finally, the researcher reviewed the identified dark patterns and similarly validated or adjusted the system's assessment of them.

In the case of more than one consent dialog, the researcher performed the above for all dialogs. Though in later sections we only consider the top ranked dialog the system identified and consider the system accurate if it listed any valid dialog first.

To assess the accuracy of the lead researcher's manual dark pattern labelling, a second researcher manually went through a random subset of 10% of the dialogs from both the Top500 and Rand500 datasets ($n = 45$) using screenshots collected by the system. To ensure similar interpretation of the patterns, the two researchers first discussed the criteria for identifying patterns. The second researcher then looked at screenshots of the cookie dialogs from the random subset and recorded observed patterns. The two researchers then met to discuss the differences in their results. Out of the 45 dialogs, researchers disagreed on 7. After discussion, three were agreed in the lead researcher's favour. The remaining 4 disagreements highlight the complex nature of some dark patterns (screenshots of the 4 cases are provided in appendix D.1). In one case a button in the upper right of the dialog was worded "continue without accepting" which is functionally similar but not quite the same as an ambiguous close dialog (AmbiguousClose). There was also a case where the opt-in and opt-out buttons had different colours, but it was unclear if one was really highlighted more than the other (HighlightedOptIn). In the following sections we use the agreed labels.

4.3. Dialog Collection Accuracy

We use the manual annotations of the Top500 and Rand500 datasets to compute the accuracy of the automated system.

As shown in Table 3, both Top500 and Rand500 had 6 instances each of disagreement between the system and the human labels. In 8 of the cases, the disagreement was about if a dialog existed at all, and in the remaining 4 cases the system had identified the cookie dialog, but had ranked a non-cookie element as more probably the cookie dialog for the page. Overall the system correctly located cookie dialogs in 98.7% of the pages which loaded correctly.

For comparison, Petronyte [58] reported an accuracy of 93.0% using a combination of just CSS selectors and keyword filtering. This result suggests that our ranking based system (described in Section 3.1.2) was successful in improving cookie dialog collection accuracy.

4.3.1. Clickable Location & Classification. The system was 100% accurate at locating all available clickables on cookie dialogs. A single cookie dialog can have more than one clickable, for example "Accept" and "Reject" buttons which we count as 2 clickables. Multiple occurrences of the same clickable on a dialog are not counted repeatedly, so there can either be 0 or 1 instances of each clickable defined in Section 3.3. In the Top500 dataset, 1211 of the 1273 total clickables were correctly classified, meaning the system had a 95.1% clickable classification accuracy.

	Loaded	Dialog	Disagreement	Cookies
Top500	469	231	6	220
Rand500	467	98	6	95
Rand10k	9199	2092	-	1990
Data11k	10127	2417	-	2301

TABLE 3: Data collection results for all four datasets for how many websites loaded at all (Loaded), how many had a cookie dialog (Dialog), how many had a disagreement between the system and the human in regards to the dialog (Disagreement), and finally of those with a dialog how many had at least one cookie on initial load (Cookies).

In the Rand500 dataset, 639 of the 656 total clickables were correctly classified, meaning the system had a 97.4% clickable classification accuracy. We did not observe any instances where the system unable to detect any clickables.

4.3.2. Dark Pattern Detection. There can be multiple dark patterns present on each dialog so we count the number of distinct dark patterns present on each dialog. For example, a dialog could have both the HighlightedOptIn and AmbigiousClose dark patterns present which we count as 2 distinct dark patterns. Multiple occurrences of the same dark pattern on a dialog are not counted repeatedly, so there can either be 0 or 1 instances of each dark pattern defined in Section 3.5.

In the Top500 dataset, 810 of 814 total dark patterns were correctly detected and there were 3 false positives. Meaning the system had a 99.0% dark pattern detection accuracy.

In the Rand500 dataset, 558 of 561 total dark patterns were correctly detected and there was 1 false positive. Meaning the system had a 99.2% dark pattern detection accuracy.

Interestingly, in both datasets, all the false positives appeared for MultipleDialogs (Multiple Distinct Cookie Dialogs present on a page), which may suggest that the criteria for detecting duplicate dialogs was too relaxed. In future work, more stringent criteria could be applied for this dark pattern. For example, the size and position of a candidate dialogs could also be compared to help reduce the false positive rate.

4.4. Analysis

For the main analysis we use the Data11k dataset. The decision to include the Top500 dataset in analysis was made due to the large impact these websites have on users which makes them particularly important to study. In recognition of oversampling of highly popular websites, our analysis below explicitly controls for Tranco rank, or pulls out the Top500 dataset separately for comparison.

Using the labels automatically generated by the system and ignoring the human generated labels. Because human labels are only available for 1k of the websites, and because the labels are skewed towards the more popular websites we did not want to use them as they add an extra possible confound linked to website popularity. However, as we show below, the system is able to find the correct cookie dialog in 98.7% of cases and is therefore quite accurate.

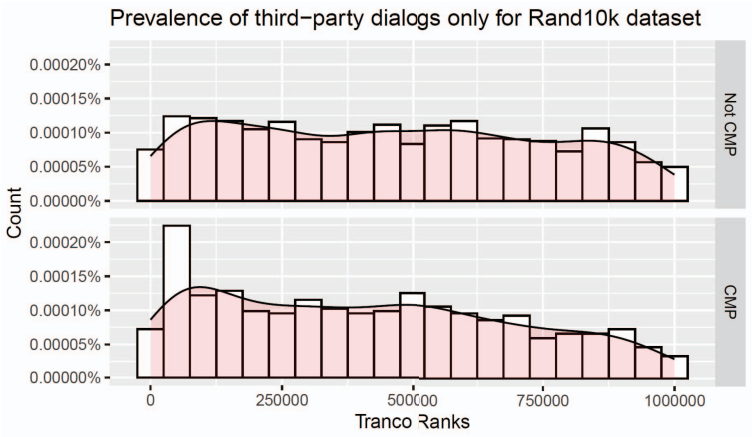

Figure 5: The prevalence of consent dialogs on each fraction of the Tranco Top 1 million websites using the Rand10k dataset. These are divided into those websites whose dialog is provided by a TopCMP (*CMP*) and those who use a niche CMP or self-implement the dialog (*Non-CMP*).

5. Results

5.1. Detecting Dialogs

As summarised in Table 3 our system attempted to collect cookie dialogs from a total of 10,992 websites. As is common with Internet measurement studies, a minority (865) of websites failed to load any content during collection of our datasets. Of the websites that loaded successfully, we detected a cookie dialog on 2,417 websites. A cookie dialog was not detected on the remaining 7,706 websites. In the majority of these cases, this will be due to the website not requiring cookie dialog because they do not process user personal data or do so with a legitimate interest. Alternatively, the website may choose not to display a cookie dialog or not be aware of their obligation to display a cookie dialog. A small fraction (2 websites in Top500 & Random500 datasets) displayed a cookie dialog that DarkDialogs was unable to detect.

Figure 5 displays the prevalence of dialogs across the Tranco Top 1 million websites, divided into two plots based on whether the dialog was created by a popular third-party provider (a CMP). Moderately popular websites (ranked from 50k-150k) are most likely to embed a consent dialog. The long tail of unpopular websites either do not need to collect consent or are unaware of their legal obligations.

Figure 5 shows a difference between which parties embed consent interfaces from popular CMPs. The spike in dialogs in the bottom graph shows that moderately popular websites (ranked from 50k-100k) are especially likely to embed a CMP's consent interface. This result runs counter to the intuition that websites with less resources rely on third-parties to implement dialogs.

5.2. Detecting Clickables

Figure 6 shows a breakdown of the percentage of websites with at least one occurrence of each clickable type. We can see that the opt-in button appeared on a higher percentage of websites compared to the opt-out button. This result suggests that many websites choose to make it easier for a user to opt-in to all cookies and harder for them to opt-out from cookies. A major difference between the datasets can be observed for the 'more options' clickable which is much more prevalent in the Top500 websites. The existence of an opt-out button is similarly uncommon for all three datasets, meaning that a significant share of cookie dialogs provide no way for users to opt-out.

5.3. Detecting Dark Patterns

This subsection asks which dark patterns are present in the $2,417$ dialogs that were identified and extracted. In total our system identified 3,744 dark patterns across in the combined datasets (Data11k).

Prevalence Table 4 shows the prevalence of the 10 dark patterns in our sample of $2,417$ dialogs. Some dark patterns have no value if OnlyOptIn is true. For example, it is impossible to test whether opt-out leads to more cookies if there is no opt-out button. When calculating the prevalence of such patterns, we throw away all data for which OnlyOptIn is true leaving just $1,600$ observations. To calculate the prevalence across all accessible websites in our sample, the figures should be multiplied by $\frac{n}{10123}$.

Using overly complex language (ComplexText) is by far the most common dark pattern, which occurs in almost half of all dialogs. Although MoreOptions has a higher prevalence, this pattern can only occur on dialogs that already have an opt-out button, which is around two thirds. 18% of dialogs have an ambiguous close button, and 12% set more cookies if the user closes the dialog. Dialogs that obstruct the window are comparably rare (3%).

A third of dialogs have no opt-out (OnlyOptIn). When websites implement an opt-out button, the majority of those dialogs place the opt-out behind multiple layers (MoreOptions) and a quarter highlight the opt-in button with colour (HighlightedOptIn). Just 3% have a pre-ticked box or preference slider for non-necessary cookies (PreferenceSlider). 12% set more cookies after the user has opted-out (OptOutMoreCookies).

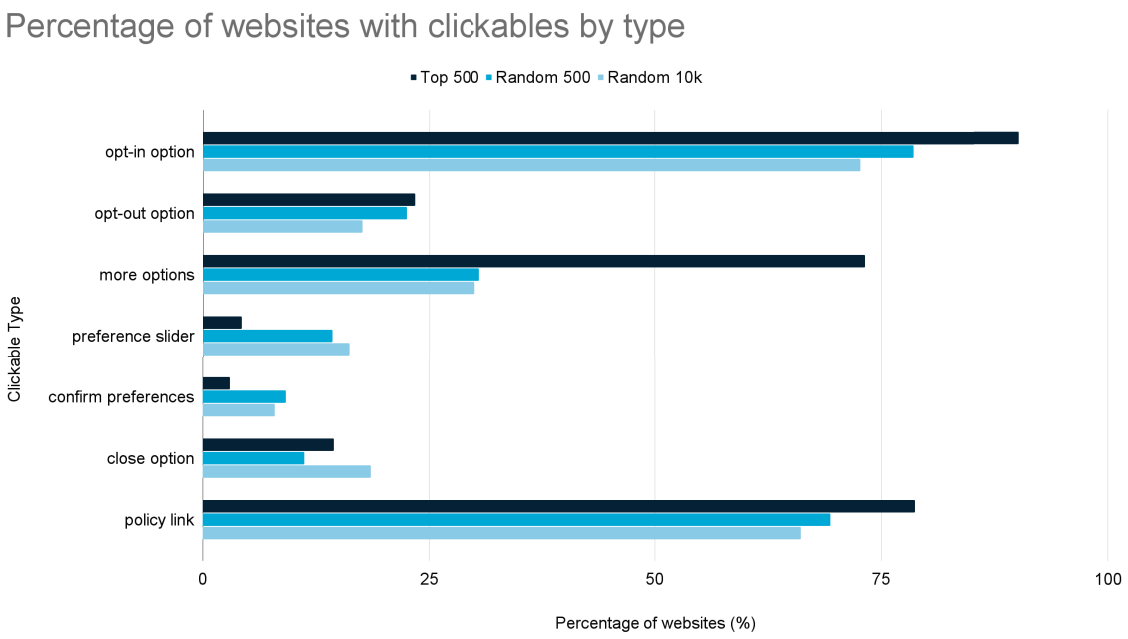

Figure 6: Graph showing percentage of websites with at least one occurrence of each clickable type.

Dark Pattern	n	Prevalence
OnlyOptIn	2,417	33.8%
HighlightedOptIn	1,600	26%
ObstructsWindow	2,417	3.2%
ComplexText	2,417	47.9%
MoreOptions	1,600	51.2%
AmbiguousClose	2,417	17.5%
MultipleDialogs	2,417	7.9%
PreferenceSlider	1,600	2.9%
CloseMoreCookies	2,417	2.7%
OptOutMoreCookies	1,600	11.8%

Explanatory Variable	n	Mean	St. Dev.	Min	Max
tranco_rank	2,417	407,119	304,814	1	999,802
id_like_cookies	2,417	5.4975	10.87	0	122
Top_CMP	2,417	0.298	0.46	0	1

TABLE 4: The prevalence of each dark pattern on dialogs and summary statistics for the independent variables in the logistic regressions.

Which websites adopt which dark patterns It is interesting to understand whether certain kinds of websites adopt specific dark patterns. For example, one might hypothesize that popular websites display less dark patterns because they are more likely to be investigated and punished due to their visibility. We run a series of logistic regressions to explore whether three variables explain whether specific dark patterns are adopted.

We fit the following model with the prevalence of the i-th dark pattern as the dependent variable:

$$\text{logit}(P_i) = \beta_{i,0} + \beta_{i,1}X_1 + \beta_{i,2}X_2 + \beta_{i,3}X_3$$

The independent variables are as follows:

X_1 is log(Tranco Rank) of the website.

X_2 number of ID-like cookies that were initially set.

X_3 is 1 if the dialog was designed by one of 13 major CMPs (see Section 3.2) and 0 otherwise.

Table 4 contains summary statistics for each variable. Note, a mean of 0.298 for the variable *Top_CMP* means that 29.8% of dialogs are designed by a Top CMP.

Each equation allows us to understand the relationship between the independent variables and the prevalence of dark pattern i. The coefficient of the first variable, $\beta_{i,1}$ would be positive if less popular websites are more likely to implement dark pattern i. The coefficient $\beta_{i,2}$ would be positive if websites with more ID-like cookies are more likely to adopt dark pattern i. The coefficient $\beta_{i,3}$ would be positive if popular CMPs are more likely to implement dark pattern i. These results could help regulators prioritise investigations into specific types of websites.

Throughout we will report on the effect size ($\beta_{i,j}$) and the associated statistical significance. Given that we have run ten regressions for each variable, we advise against reading too much into individual effects even if they are statistically significant at a $p < 0.05$ level. Instead readers should focus on effects that are consistent across multiple dark patterns.

It was clear that the independent variable *id_like_cookies* had little explanatory power. One might expect that websites with more tracking-like cookies were more likely to implement dark patterns because this provided the websites with a legal basis for their tracking. However, there was no statistically significant relationship in most cases, with the exception that websites with higher *id_like_cookies* were less likely to include a dialog with a MoreOptions button and less

TABLE 5: Exploring which types of websites are more likely to embed specific dark patterns. Website popularity and the presence of a CMP are the strongest predictors, but the effect depends on the specific dark pattern.

	Log odds ratio for the prevalence of each dark pattern									
	OnlyOptIn	**HighlightedOptIn**	**ObstructsWindow**	**ComplexText**	**MoreOptions**	**AmbiguousClose**	**MultipleDialogs**	**PreferenceSlider**	**CloseMoreCookies**	**OptOutMoreCookies**
log(tranco_rank)	0.139^{***}	-0.070^{***}	-0.422^{***}	0.001	-0.230^{***}	0.016	0.082^{*}	0.089	−0.002	-0.247^{***}
	(0.022)	(0.021)	(0.034)	(0.017)	(0.024)	(0.023)	(0.037)	(0.068)	(0.052)	(0.023)
id_like_cookies	0.006	-0.014^{*}	−0.011	−0.003	-0.022^{***}	0.005	0.002	0.003	0.011	−0.008
	(0.004)	(0.006)	(0.014)	(0.004)	(0.005)	(0.005)	(0.007)	(0.012)	(0.009)	(0.008)
TopCMP	-1.556^{***}	1.385^{***}	−0.150	0.774^{***}	0.935^{***}	0.113	0.263	1.357^{***}	−0.130	−0.012
	(0.121)	(0.121)	(0.267)	(0.091)	(0.110)	(0.116)	(0.160)	(0.324)	(0.283)	(0.166)
Observations	2,417	1,600	2,417	2,417	1,600	2,417	2,417	1,600	2,417	1,600
Log Likelihood	−1,416.411	−839.950	−261.238	−1,636.776	−1,006.368	−1,121.349	−664.218	−201.523	−298.397	−526.820
Akaike Inf. Crit.	2,840.822	1,687.899	530.476	3,281.553	2,020.736	2,250.699	1,336.436	411.046	604.794	1,061.640

*p<0.05; **p<0.01; ***p<0.001

likely to include a highlighted opt-out.

The variables that could reliably explain variance in the prevalence of dark patterns were, respectively, the popularity of websites and the presence of a top CMP. However, there was no simplistic conclusion like "more popular websites are more likely to deploy dark patterns", which motivates a fine-grained analysis.

The first dark pattern (OnlyOptIn) concerns whether dialogs have only one button, an opt-in. This dark pattern is less frequently implemented by more popular websites and also less frequently by the top CMPs (both $p < 0.001$). This result suggests that websites and intermediaries with more resources are less likely to implement this specific dark pattern, possibly because it is so unambiguously non-compliant with data protection law (see Section 3.5). The directions of these relationships are reversed when it comes to the more subtle dark pattern (HighlightedOptIn) for which less regulatory guidance and past enforcement options are not available—popular websites and TopCMPs are more likely to implement HighlightedOptIn.

The biggest effect of the popularity of websites concerns whether the dialog blocks the entire screen (ObstructsWindow). Popular websites are especially likely implement such banners. They are also more likely to set more cookies after a user opts out (OptOutMoreCookies) and to implement a dialog with a more options button (MoreOptions) rather than a 1-click opt-out. When considering the effect sizes for *TopCMP*, it seems the most significant impact of top CMPs is in preventing websites implementing OnlyOptIn and helping websites implement HighlightedOptIn and a PreferenceSlider.

Summary Just 23.9% of websites in our sample embedded a cookie dialog. More popular websites are more likely to embed a dialog (see Figure 5), and moderately popular websites are even more likely to rely on a third-party (CMP) to design the dialog. These two variables, the popularity of the website and the presence of a CMP, had the most predictive power regarding whether specific dark patterns were deployed by a website. However, the direction of this relationship varied by dark pattern. Popular websites and CMPs were less likely to embed a dialog with no opt-out option (OnlyOptIn), which was the most egregious dark pattern. However, they were more likely to implement subtle dark patterns, such as visually highlighting the opt-in button to make it more attractive (HighlightedOptIn). These results can be used by regulators to direct and prioritise investigations. For example, less popular websites should be investigated to find examples of OnlyOptIn, but more popular should be investigated for ObstructsWindow. The tool can also be used to notify websites about the presence of dark-patterns in dialogs.

6. Discussion

We discuss building the system, our measurement results, potential future applications, and limitations.

6.1. System

Designing and implementing *DarkDialogs* achieved two broad goals: (1) automatically extracting arbitrary (e.g. CMP agnostic) dialogs with a high accuracy rate; (2) automatically detecting dark-patterns by opting for narrowly defined definitions, which can be manually mapped to high-level concepts. Our approach to identifying cookie dialogs diverged from influential early studies [30], [47], [53] that detected dialogs via fingerprints linked to specific CMPs, in which a custom fingerprint must be developed for the 100+ CMPs that exist. Instead our detector was designed based on a random sample of dialogs created by a mixture of popular CMPs, niche CMPs and individual websites. The resulting dialog detector is CMP-agnostic with an accuracy rate of 98.7%, which means our automated crawl extracts the vast majority of dialogs on the Web. We estimate that a cookie dialog exists on 20.9% of websites in the Tranco Top 1 million websites. For comparison, Hils et al. [30] estimate the number of CMP-designed dialogs to be 9.25% of the Tranco Top 10k. The discrepancy is because many websites self-implement/use a niche CMP when embedding a consent dialog, neither of which could be detected by Hils et al. [30]. Future work should adopt our approach to avoid the sampling bias from only extracting dialogs from popular CMPs.

Having extracted the dialog, the next design choice was what conceptual level to define the patterns we seek to detect. Prior work tried to detect high-level categories like

"nagging" or "sneaking" [65], which achieved accuracy rates of $0.535 - 0.72$. Our system achieves much higher classification accuracy by focusing on detecting narrow dark patterns. Doing so raises the question of what is lost with this approach. Most fundamentally, there are many more technical ways of implementing dark patterns than there are high-level categories (e.g. in the well-cited taxonomy [28]). This situation creates a game of cat and mouse in which nefarious websites and CMPs rotate through the infinite number of technical ways of implementing a dark pattern. For example, we chose an arbitrary threshold for light/dark when detecting HighlightedOptIn and an adversary could design a dialog that highlights a specific button, but does so just below this threshold. This is reminiscent of cybersecurity in which adversaries continually find new vulnerabilities to exploit systems, with defenders generally one step behind.

We did not manage to automate the detection of all dark patterns that we identified within the time frame of this project. This was mainly due to the remaining dark patterns being difficult to detect automatically with reasonable accuracy and in some cases "human judgement" was required (see Appendix D.2).

6.2. Measurement

Like most of the prior work, we sampled websites using the Tranco Top list [59]. However, we randomly sampled 10K of the top 1 million websites rather than the two alternatives: (1) sampling only the top 10K and (2) sampling all 1 million. The first alternative would have provided a limited perspective on the long-tail of websites. For example, Table 5 shows that the top 10K websites are unusually likely to implement a top CMP's dialog relative to the long tail. The second alternative would have been computationally and time expensive, which is not available to under-resourced research labs. A sample of 10K websites provides enough statistical power (see Table 5). Collecting a larger sample incurs a financial/climate cost that is arguably unnecessary.

Beyond just detecting dark patterns, we showed that each dark pattern could be associated with website characteristics, but that no clear pattern existed. Sometimes this was because we devised somewhat crude metrics, such as counting ID-like cookies to detect how likely a website is to engage in tracking. An alternative approach would be to detect other tracking indicators [2]. Future work could also pair our measurement system with more sophisticated statistical modelling, such as collecting longitudinal data.

The most interesting regression results concern the relationship between dark pattern prevalence and website popularity/CMP presence. Intuitively, more popular websites and the top CMPs have more resources to spend on compliance with privacy regulations. Indeed, both the popularity of websites and the presence of a CMP are associated with lower prevalence of having dialogs with only an opt-in button, which is among the most egregious patterns. This result suggests those resources are used to comply with privacy regulations. However, this relationship is reversed for more subtle and legally ambiguous dark patterns, such as highlighting one option. One interpretation is that the compliance resources are used to build dialogs that undermine user autonomy without attracting too much regulatory attention.

6.3. Applications

While designed for research, our system has other potential uses. Website owners could voluntarily use it to audit their cookie dialogs, particularly those provided by CMPs who refuse to accept liability for legal requirements. Taking the most well-resourced CMP as an example, Google allows websites using AdSense to choose from several cookie dialog layouts [26]. However, Google explicitly says to developers "It is your responsibility to make sure your messages meet legal requirements". This follows the "Developers Are Responsible" trend which was observed by Tahei et al. [68] as being used by Google, Amazon, Facebook, and Twitter in mobile Ad Network consent banners. Similarly, Mhaidli et al. [48] found that developers picked the default options when creating an Ad Network consent banner and developers believed it was the responsibility of Ad Networks to address privacy concerns. Our system would allow developers to audit their websites, at least in theory.

Realistically, website owners will not be aware of our system. Some may even maliciously include dark patterns. In such cases, *DarkDialogs* could also be used by privacy advocates for responsible disclosure [71]. This approach was taken by the pro-privacy NGO noyb in a campaign launched in 2022 [21]. However, their campaign focused only on dialogs implemented by one CMP. Our system would allow them to notify all websites regardless of which CMP they use.

We considered directly contacting websites about the dark patterns that were detected by *DarkDialogs*. Doing so would be comparable to researchers notifying system owners about security vulnerabilities [6], [43], [60], [67]. One major difference is that system owners (typically) want to fix security vulnerabilities and would therefore be grateful for the opportunity to remediate the issue, in contrast websites may have intentionally introduced dark patterns to nudge users towards opt-in consent [15], [74]. Consequently, we did not want to risk the website threatening us with legal action, which occurs even with security notifications that (in theory) help the website. The organisation who did issue notifications about dark patterns is comfortable with court cases given its founder, Max Schrems, won notable cases under EU law about transferring personal data to servers in the United States [21].

The system and its approach could be deployed for social science research. For example, the variables and statistical model we explored in the regression analysis were simplistic. Future work could use our system to collect a panel data-set (e.g. measuring presence of dark patterns over time) and use an event window-study design to detect how specific legal judgements impacted the prevalence of specific dark patterns. For example, intuitively one might expect the French DPA fining Google/Facebook millions of euros for the design of their dialogs to have sent a message leading other websites to fix non-compliant dialogs.

6.4. Limitations

A limitation of our system is that it cannot automatically opt-out from cookies if doing so requires having to navigate through several layers of options. This situation is due to complexity and variation in the layout of cookie dialogs which means there is no trivial way to opt-out from all cookies. Ideally, we would want to find a way to automate this process as it would allow us to fully assess the cookie setting behaviour of cookie dialogs automatically.

The main limitation of our system is that the dialog and dark pattern techniques we have developed are heuristic in nature, by which we mean they are not 100% accurate. Consent dialogs vary massively in layout and the methods used to avoid detection. In addition, websites are constantly changing their consent dialog designs in response to regulatory and advertising demands. This makes detecting dark patterns on consent dialogs a constant battle between detection systems and website designers. DarkDialog's use of crowdsourced CSS selector lists will help to keep the system up to date with the latest avoidance methods used by websites. However, CSS selector lists are reliant on the goodwill of users to keep them updated and this may change in the future. We encourage researchers to build upon the techniques in our paper to detect novel dark patterns. To aid this goal, we have open-sourced our code repository.

7. Conclusion

Our goal was to build a system that could automatically extract cookie dialogs and detect the presence of ten dark patterns. The system successfully extracted 98.7% of the dialogs in a hand labelled sample. Notably, our system detects any dialog regardless of who designed it, whereas prior work [30], [32], [47], [53] relied on detecting third-parties (CMP) or dialogs designed under specific frameworks. We show that prior work underestimates the number of cookie dialogs on the Web. For example, Hils et al. [30] estimate the number of CMP-designed dialogs to be 9.25% of the Tranco Top 10k with this number falling for samples of less popular websites. Our CMP-agnostic approach detects a dialog on 20.9% of websites in a random sample of the top 1 million websites, with the figure even higher among more popular websites (see Figure 5).

Turning to our goal of detecting dark patterns in dialogs, the system correctly classified over 99% of the 1,375 patterns in the two human-labelled datasets (Top500 and Rand500). We opted to detect narrowly-defined patterns with clear heuristics. For comparison, a prior study detecting high-level pattern categories like "forced action" and "nagging" achieved achieved accuracy rates of $0.535 - 0.72$ [66]. We believe these results show that building designer-agnostic dialog detectors and detecting narrowly-defined dark patterns represents the best approach to developing data protection auditing tools.

Our final contribution deployed our system on a random sample of the Tranco Top 1 million websites [59]. The most frequently detected patterns were OnlyOptIn, HighlightedOptIn, ComplexText, and MoreOptions. Interestingly, embedding a dialog provided by a CMP was positively associated with four dark patterns and negatively associated with just one. Similarly, more popular websites were positively associated with four dark patterns (at the $p < 0.001$ level) and negatively associated with just one. This preliminary evidence suggests that dialog designers with more resources are more likely to include dark patterns. Future work could combine our measurement system with more sophisticated statistical modelling to further probe this idea.

Data Availability

We released the source code of the DarkDialogs system in a public repository, along with system installation and usage instructions[2]. The datasets collected as part of this project are also available in a public repository[3], which includes a description of the collection method, the date range in which the data was collected, and a thorough description of the schema of the datasets.

Acknowledgements

We thank the anonymous reviewers and the members of the Technology Usability Lab in Privacy and Security (TULiPs) for their insightful comments. The first author wrote this paper as part of a summer internship funded by the Laboratory for Foundations of Computer Science (LFCS) institute at the University of Edinburgh.

References

[1] Shivam Bansal and Chaitanya Aggarwal. pypi textstat. https://pypi.org/project/textstat/, 2022. Last accessed 27 March 2022.

[2] Reuben Binns. Tracking on the web, mobile and the internet of things. *Foundations and Trends in Web Science*, 8(1–2):1–113, 2022.

[3] Rainer Böhme and Stefan Köpsell. Trained to accept? a field experiment on consent dialogs. In *In Proceedings of the 2010 CHI Conference on Human Factors in Computing Systems (pp. 2403-2406).*, pages 2403–2406, 2010.

[4] Dino Bollinger, Karel Kubicek, Carlos Cotrini, and David Basin. Automating cookie consent and GDPR violation detection. In *31st USENIX Security Symposium (USENIX Security 22)*. USENIX Association, 2022.

[5] Frederik Zuiderveen Borgesius. Behavioral targeting: A European legal perspective. *IEEE Security & Privacy*, 11(1):82–85, 2013.

[6] Orçun Çetin, Carlos Gañán, Lisette Altena, Samaneh Tajalizadehkhoob, and Michel Van Eeten. Tell me you fixed it: Evaluating vulnerability notifications via quarantine networks. In *2019 IEEE European Symposium on Security and Privacy (EuroS&P)*, pages 326–339. IEEE, 2019.

[7] CJEU. JUDGMENT OF THE COURT (grand chamber) 1 october 2019. https://curia.europa.eu/juris/document/document.jsf?text=&docid=218462&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=1447493, 2019. Last accessed 08 July 2022.

[8] CMS-Legal. GDPR enforcement tracker. https://www.enforcementtracker.com/ETid-1134, 2022. Last accessed 21 July 2022.

[9] CMS-Legal. GDPR enforcement tracker. https://www.enforcementtracker.com/ETid-1135, 2022. Last accessed 21 July 2022.

2. https://github.com/DarkDialogs/OpenScience
3. https://doi.org/10.7488/ds/3475

[10] CMS-Legal. GDPR enforcement tracker. https://www.enforcementtracker.com/ETid-1140, 2022. Last accessed 21 July 2022.

[11] CMS-Legal. GDPR enforcement tracker. https://www.enforcementtracker.com/ETid-220, 2022. Last accessed 21 July 2022.

[12] CMS-Legal. GDPR enforcement tracker. https://www.enforcementtracker.com/ETid-369, 2022. Last accessed 21 July 2022.

[13] CMS-Legal. GDPR enforcement tracker. https://www.enforcementtracker.com/ETid-426, 2022. Last accessed 21 July 2022.

[14] CNIL. Cookies: the CNIL fines Google a total of 150 million euros and Facebook 60 million euros for non-compliance with French legislation. https://www.cnil.fr/en/cookies-cnil-fines-google-total-150-million-euros-and-facebook-60-million-euros-non-compliance, 2022. Last accessed 07 July 2022.

[15] Lorrie Faith Cranor. Cookie monster. *Communications of the ACM*, 65(7):30–32, 2022.

[16] Frederic Debusseré. The EU e-privacy directive: a monstrous attempt to starve the cookie monster? *International Journal of Law and Information Technology*, 13(1):70–97, 2005.

[17] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. *Network and Distributed System Security Symp.*, 2019.

[18] EDPB. Guidelines 05/2020 on consent under regulation 2016/679. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf, 2020. Last accessed 07 July 2022.

[19] Rob van Eijk, Hadi Asghari, Philipp Winter, and Arvind Narayanan. The impact of user location on cookie notices (inside and outside of the European union). In *Workshop on Technology and Consumer Protection (ConPro'19)*, 2019.

[20] Steven Englehardt and Arvind Narayanan. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security.*, page 1388–1401, 2016.

[21] NOYB European Center for Digital Rights. noyb files 422 formal GDPR complaints on nerve-wrecking "cookie banners". https://noyb.eu/en/noyb-files-422-formal-gdpr-complaints-nerve-wrecking-cookie-banners, 2021. Last accessed 21 August 2022.

[22] GDPR. Art. 6 GDPR lawfulness of processing. https://gdpr-info.eu/art-6-gdpr/, 2020. Last accessed 07 July 2022.

[23] GDPR. Recital 32 conditions for consent. https://gdpr-info.eu/recitals/no-32/, 2020. Last accessed 07 July 2022.

[24] Julia Giese and Martin Stabauer. Factors that influence cookie acceptance. In *International Conference on Human-Computer Interaction*, pages 272–285. Springer, 2022.

[25] Miguel Godinho de Matos and Idris Adjerid. Consumer consent and firm targeting after GDPR: The case of a large telecom provider. *Management Science*, 68(5):3330–3378, 2022.

[26] Google. About GDPR consent messages. https://support.google.com/adsense/answer/10961068?hl=en&visit_id=637937479705879588-2959457591&ref_topic=10924670&rd=1, 2022. Last accessed 08 July 2022.

[27] Google. Chrome devtools protocol. https://chromedevtools.github.io/devtools-protocol/tot/Network/#method-getAllCookies, 2022. Last accessed 27 March 2022.

[28] Colin M Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L Toombs. The dark (patterns) side of UX design. In *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-14).*, pages 1–14, 2018.

[29] Johanna Gunawan, Amogh Pradeep, David Choffnes, Woodrow Hartzog, and Christo Wilson. A comparative study of dark patterns across web and mobile modalities. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–29, 2021.

[30] Maximilian Hils, Daniel W Woods, and Rainer Böhme. Measuring the emergence of consent management on the web. In *Proceedings of the ACM Internet Measurement Conference*, pages 317–332, 2020.

[31] Maximilian Hils, Daniel W Woods, and Rainer Böhme. Conflicting privacy preference signals in the wild. *arXiv preprint arXiv:2109.14286*, 2021.

[32] Maximilian Hils, Daniel W Woods, and Rainer Böhme. Privacy preference signals: Past, present and future. *Proceedings on Privacy Enhancing Technologies*, 2021(4):249–269, 2021.

[33] Soheil Human, Harshvardhan J Pandit, Victor Morel, Cristiana Santos, Martin Degeling, Arianna Rossi, Wilhelmina Botes, Vitor Jesus, and Irene Kamara. Data protection and consenting communication mechanisms: Current open proposals and challenges. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 231–239. IEEE, 2022.

[34] IAB. TCF for CMPs. https://iabeurope.eu/tcf-for-cmps/, 2022. Last accessed 21 July 2022.

[35] ICO. How do we comply with the cookie rules? https://ico.org.uk/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies/how-do-we-comply-with-the-cookie-rules/, 2022. Last accessed 13 June 2022.

[36] Vitor Jesus and Harshvardhan J Pandit. Consent receipts for a usable and auditable web of personal data. *IEEE Access*, 10:28545–28563, 2022.

[37] Georgios Kampanos and Siamak F Shahandashti. Accept all: The landscape of cookie banners in Greece and the UK. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pages 213–227. Springer, 2021.

[38] J.P. Kincaid, R.P. Fishburne Jr., R.L. Rogers, and B.S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. In *Naval Technical Training Command Millington TN Research Branch*, 1975.

[39] Daniel Kladnik. I don't care about cookies 3.3.8 get rid of cookie warnings from almost all websites! https://www.i-dont-care-about-cookies.eu/, 2021. Last accessed 27 March 2022.

[40] Michael Kretschmer, Jan Pennekamp, and Klaus Wehrle. Cookie banners and privacy policies: Measuring the impact of the GDPR on the web. *ACM Transactions on the Web (TWEB)*, 15(4):1–42, 2021.

[41] Chiara Krisam, Heike Dietmann, Melanie Volkamer, and Oksana Kulyk. Dark patterns in the wild: Review of cookie disclaimer designs on top 500 German websites. In *European Symposium on Usable Security 2021 (EuroUSEC '21),*, page 8, https://doi.org/10.1145/3481357.3481516, 2021.

[42] Ronald Leenes and Eleni Kosta. Taming the cookie monster with dutch law–a tale of regulatory failure. *Computer Law & Security Review*, 31(3):317–335, 2015.

[43] Frank Li, Zakir Durumeric, Jakub Czyz, Mohammad Karami, Michael Bailey, Damon McCoy, Stefan Savage, and Vern Paxson. You've got vulnerability: Exploring effective vulnerability notifications. In *USENIX Security Symposium*, volume 16, 2016.

[44] Joasia Luzak. Much ado about cookies: The European debate on the new provisions of the eprivacy directive regarding cookies. *European Review of Private Law*, 21(1), 2013.

[45] Eryn Ma and Eleanor Birrell. Prospective consent: The effect of framing on cookie consent decisions. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–6, 2022.

[46] Arunesh Mathur, Mihir Kshirsagar, and Jonathan Mayer. What makes a dark pattern... dark? design attributes, normative considerations, and measurement methods. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–18, 2021.

[47] Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do cookie banners respect my choice?: Measuring legal compliance of banners from iab europe's transparency and consent framework. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 791–809. IEEE, 2020.

[48] Abraham H Mhaidli, Yixin Zou, and Florian Schaub. "we can't live without them!" app developers' adoption of ad networks and their considerations of consumer risks. In *USENIX Symposium on Usable Privacy and Security (SOUPS)*, https://www.usenix.org/conference/soups2019/presentation/mhaidli, 2019.

[49] Microsoft. Get your document's readability and level statistics. https://support.microsoft.com/en-us/office/get-your-document-s-readability-and-level-statistics-85b4969e-e80a-4777-8dd3-f7fc3c8b3fd2, 2021. Last accessed 8 Mat 2022.

[50] Lynette I Millett, Batya Friedman, and Edward Felten. Cookies and web browser design: Toward realizing informed consent online. In *In Proceedings of the 2001 CHI Conference on Human Factors in Computing Systems (pp. 46-52.*, pages 46–52, 2001.

[51] B. Molnar. Measuring the cookie-setting behaviour of web pages showing privacy warnings). In *University of Edinburgh. Undergraduate Thesis.*, 2020.

[52] Mozilla. Comparing firefox browser with google chrome. https://www.mozilla.org/en-US/firefox/browsers/compare/chrome/, 2022. Last accessed 27 March 2022.

[53] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In *In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (pp. 1-14).*, pages 1–13, 2020.

[54] Harshvardhan J Pandit, Christophe Debruyne, Declan O'Sullivan, and Dave Lewis. Gconsent-a consent ontology based on the GDPR. In *European Semantic Web Conference*, pages 270–282. Springer, 2019.

[55] Paulina J Pesch, Harshvardhan J Pandit, Vitor Jesus, and Cristiana Santos. Consent 2022: 2nd international workshop on consent management in online services, networks and things. In *Companion Proceedings of the Web Conference 2022*, pages 509–513, 2022.

[56] Paulina Jo Pesch. Drivers and obstacles for the adoption of consent management solutions by ad-tech providers. In *2021 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 269–277. IEEE, 2021.

[57] R. Petnel. Easylist. https://easylist.to/, 2022. Last accessed 27 March 2022.

[58] S. Petronyte. Measure the cookie setting behavior of web pages showing cookie privacy warnings. In *University of Edinburgh. Undergraduate Thesis.*, 2021.

[59] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. Tranco a research-oriented top sites ranking hardened against manipulation. https://tranco-list.eu/, 2021. Last accessed 19 October 2021.

[60] Elsa Rodríguez, Susanne Verstegen, Arman Noroozian, Daisuke Inoue, Takahiro Kasama, Michel Van Eeten, and Carlos H Gañán. User compliance and remediation success after iot malware notifications. *Journal of Cybersecurity*, 7(1):tyab015, 2021.

[61] Iskander Sanchez-Rola, Matteo Dell'Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos. Can i opt out yet? GDPR and the global illusion of cookie control. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, pages 340–351, 2019.

[62] Cristiana Santos, Nataliia Bielova, and Célestin Matte. Are cookie banners indeed compliant with the law? https://arxiv.org/pdf/1912.07144.pdf, 2020.

[63] Cristiana Santos, Midas Nouwens, Michael Toth, Nataliia Bielova, and Vincent Roca. Consent management platforms under the GDPR: processors and/or controllers? In *Annual Privacy Forum*, pages 47–69. Springer, 2021.

[64] Arianna Schuler Scott, Michael Goldsmith, Harriet Teare, Helena Webb, and Sadie Creese. Why we trust dynamic consent to deliver on privacy. In *IFIP International Conference on Trust Management*, pages 28–38. Springer, 2019.

[65] Than Htut Soe, Oda Elise Nordberg, Frode Guribye, and Marija Slavkovik. Circumvention by design-dark patterns in cookie consent for online news outlets. In *Proceedings of the 11th nordic conference on human-computer interaction: Shaping experiences, shaping society*, pages 1–12, 2020.

[66] Than Htut Soe, Cristiana Teixeira Santos, and Marija Slavkovik. Automated detection of dark patterns in cookie banners: how to do it poorly and why it is hard to do it any other way. *arXiv preprint arXiv:2204.11836*, 2022.

[67] Ben Stock, Giancarlo Pellegrino, Christian Rossow, Martin Johns, and Michael Backes. Hey, you have a problem: On the feasibility of large-scale web vulnerability notification. In *USENIX Security Symposium*, volume 16, 2016.

[68] Mohammed Tahaei and Kami Vaniea. "developers are responsible": What ad networks tell developers about privacy. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI '21 Extended Abstracts),*, https://groups.inf.ed.ac.uk/tulips/papers/tahaei2021AdNetworkLBW.pdf, 2021.

[69] Michael Toth, Nataliia Bielova, and Vincent Roca. On dark patterns and manipulation of website publishers by CMPs. *Proceedings on Privacy Enhancing Technologies*, (3):478–497, 2022.

[70] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (Un)informed consent: Studying GDPR consent notices in the field. In *CM SIGSAC Conference on Computer and Communications Security (CCS '19)*, 2019.

[71] Christine Utz, Matthias Michels, Martin Degeling, Ninja Marnau, and Ben Stock. Comparing large-scale privacy and security notifications. *Proceedings on Privacy Enhancing Technologies*, 2023.

[72] Steven Vaughan-Nichols. What's the most popular web browser in 2021? https://www.zdnet.com/article/most-popular-web-browser-in-2021/, 2021. Last accessed 27 March 2022.

[73] Joachim Vogt, Alina Stöver, Nina Gerber, Christin Cornel, Mona Henz, Karola Marky, and Verena Zimmermann. Website operators are not the enemy either-analyzing options for creating cookie consent notices without dark patterns. *Mensch und Computer 2022-Workshopband*, 2022.

[74] Daniel W Woods and Rainer Böhme. The commodification of consent. *Computers & Security*, 115:102605, 2022.

[75] Sebastian Zimmeck and Kuba Alicki. Standardizing and implementing do not sell. In *Proceedings of the 19th Workshop on Privacy in the Electronic Society*, pages 15–20, 2020.

[76] Sebastian Zimmeck, Oliver Wang, Kuba Alicki, Jocelyn Wang, and Sophie Eng. Usability and enforceability of global privacy control. *Proceedings on Privacy Enhancing Technologies*, 2:1–17, 2023.

A. Dialog Detection details

A.1. Dialog Detection Factor Criteria

In section 3.1.2 we briefly described the factors that we use to rank to candidate dialogs. In this section we will describe each factor in detail, it's importance in identifying a cookie dialog, the criteria used to detect the factor, and justify the weight assigned to this factor.

Ranking Factor: Found using a General CSS selector.
Importance: General CSS selectors identify CSS tags that are commonly used in cookie dialogs across the web. As they come from publicly validated lists they are more likely to locate a cookie dialog.
Criteria: The CSS selector used to locate the candidate dialog was a General CSS Selector.
Weight: +5 they are from publicly validated lists which makes them more important than dialogs identified from iframes or Custom CSS-selectors. iframes were also assessed to be less important as there is typically a wide range of content that can be contained within an iframe not just cookie dialogs.

Ranking Factor: Found using a Domain-Specific CSS Selector.
Importance: domain specific selectors are from publicly validated lists and as they specify the exact website, they were deemed to be more important than those found using General CSS selectors and are assigned a higher weight.
Criteria: The CSS selector used to locate the candidate dialog was a Domain-Specific CSS Selector.
Weight: +10 as they specify the exact website, they were deemed to be more important than those found using General CSS selectors and were assigned a higher weight.

Ranking Factor: Contains common cookie dialog N-grams
Importance: Looking for phrases that commonly appear on cookie dialogs helps distinguish genuine cookie dialogs from random elements that were selected on the web page.
Criteria: From the text of 100 manually collected cookie dialogs, a set of N-grams was generated up to the 5-gram level. These were manually curated to ensure that they only contained N-grams which help identify a cookie dialog from other elements. The full list is included in appendix A.2. Since all our N-grams were generated in the English language we must first translate any non-English text. For each N-gram found in the candidate's text, a score is added based on the length of the N-gram. The performance of this factor could be increased by generating N-grams from a larger set of cookie dialogs.
Weight: unigram = +1, bigram = +2 ... A weight is added for each N-gram that is present on the cookie dialog and the weight added depends on the length of the N-gram. Longer N-grams are considered to be more important at identifying a cookie dialog. For example, the uni-gram "cookie" is less important than the tri-gram "we use cookies" at identifying the dialog as the uni-gram is a sub-string of the tri-gram.

Ranking Factor: Word count less than 5.
Importance: Elements with less than 5 words were generally found to be clickable elements rather than cookie dialogs.
Criteria: Checked if the total number of words in the candidate dialog is less than 5.
Weight: -20 generally found to be clickable elements rather than cookie dialogs so are given a negative weighting to lower their rank.

Ranking Factor: Contains no text.
Importance: Elements that have no text are very unlikely to be a cookie dialogs. Cookie dialogs generally have some text explaining what the user should do.
Criteria: Checked if the total number of words in the candidate dialog is 0.
Weight: -100 Candidates that contain no text are very unlikely to be a cookie dialog so were afforded a very large negative weighting. Note: It probably would have been better to fully remove these candidates rather than give them a large negative weight.

Ranking Factor: Word count greater than average plus 100.
Importance: Looking at the average number of words across all candidates allows us to assess whether a dialog is exceptional long compared to other candidates on the same web page. We initially tried lowering the score of candidates with more than the average number of words but in many cases, the correct dialog was just above the average number of words. In many cases, candidates that had an exceptionally large number of words did contain the cookie dialog but also included other non-relevant content.
Criteria: We calculated the average number of words across all candidates that were detected on the web page. Checked if the total number of words in the candidate dialog is more than the average plus 100.
Weight: -20 we assessed this factor to have roughly the same importance as the "Length less than 5" factor, as it helps remove non-relevant content.

Ranking Factor: Sub-string of another candidate.
Importance: The Outer HTML is a string made up of the HTML element itself, including its attributes, start tag, and end tag. By checking this string we can assess the similarity of candidates to each other in a more precise way than comparing the text content. If a candidate's Outer HTML is a sub-string of another candidate's Outer HTML then the first candidate is a child of the second candidate. Child candidates are more likely to be clickable elements or incomplete parts of a cookie dialog.
Criteria: We compare the Outer HTML of the candidate with all other candidate dialogs. We add the negative weight each time the candidate is a sub-string of another candidate. This score should be applied after duplicate candidates have been removed as this will increase the effectiveness of this factor.
Weight: -1 we apply a small negative to reduce the chances of locating incomplete parts of a cookie dialog,

this will ensure the parent candidate is ranked higher than any child candidates. We don't want a large weight as this candidate could still contain the cookie dialog and should not be completely disregarded.

Ranking Factor: Same as another candidate.
Importance: We Remove any candidates that are duplicates as there is no point assessing the same dialog more than once. Duplicates can occur when multiple CSS selectors locate the same element on a web page. An example of this would be the HTML element `<div class = "cookie-dialog">` which will match both the CSS selectors `div[class*="cookie"]` and `div[class*="dialog"]` resulting in finding the same clickable twice.
Criteria: We compare the Outer HTML of the candidate with all other candidate dialogs. If the Outer HTML is the same then we have selected the same dialog twice. We remove any duplicate candidates, giving preference to keeping candidates selected by Domain-Specific and General CSS selectors.
Weight: One of the candidate dialogs is removed.

Ranking Factor: Candidate could not be screenshotted.
Importance: Selenium only allows you to screenshot elements that are visible to the user. If a candidate is not visible to the user then this is not likely a cookie dialog.
Criteria: If Selenium could not take a screenshot of the dialog or the returned screenshot was blank.
Weight: Candidate dialog is removed.

A.2. Custom Dialog CSS Selectors

CSS Selectors
div[class*="gdpr"]
div[class*="Cookie"]
div[class*="cookie"]
div[class*="Privacy"]
div[class*="privacy"]
div[class*="Policy"]
div[class*="policy"]
div[class*="Consent"]
div[class*="consent"]
div[class*="Notice"]
div[class*="notice"]
div[class*="Dialog"]
div[class*="dialog"]
div[id*="gdpr"]
div[id*="Cookie"]
div[id*="cookie"]
div[id*="Privacy"]
div[id*="privacy"]
div[id*="Policy"]
div[id*="policy"]
div[id*="Consent"]
div[id*="consent"]
div[id*="Notice"]
div[id*="notice"]
div[id*="Dialog"]
div[id*="dialog"]
div[data-project*="cmp"]
div[id*="privacy"]
div[id*="Privacy"]
div[id*="cmp"]

TABLE 6: Custom CSS Selectors used to detect dialogs in the DarkDialogs system.

A.3. Dialog N-grams

N Level	N-grams
Uni-grams	cookies cookie track tracking
Bi-grams	use cookies cookies and cookies to we use accept all any time at any you agree learn more manage preferences
Tri-grams	we use cookies at any time use cookies and use cookies to cookies and similar use of cookies learn more about and our partners and similar technologies our cookie policy
4-grams	we use cookies to use cookies and similar cookies and similar technologies you can change your access information on a and or access information at any time by information on a device or access information on store and or access
5-grams	access information on a device and or access information on store and or access information use cookies and similar technologies ad and content measurement audience and content measurement audience insights audience insights and product development content measurement audience insights and improve your experience on our measurement audience insights and product

TABLE 7: N-grams of common words and phrases found on cookie dialogs used to by the DarkDialogs system to identify real dialogs.

B. Content Management Provider Fingerprints

Table 8 shows the fingerprints used to identify dialogs from Content Management Providers. Content Management Provider Fingerprints can be divided into 3 types:

1) **Text:** check the text content of the cookie dialog to see if the fingerprint value is a sub-string.
2) **CSS Selector:** check if the CSS selector is present within the HTML content of the cookie dialog. We used the Python *bs4* library to parse the HTML of the cookie dialog and to check if the CSS selector was present.
3) **Hostname:** check if a hostname/domain used by the CMP to save cookies is present. We check the HTML content to see if the hostname is a sub-string.

CMP	Fingerprint Type	Fingerprint Value
LiveRamp	Text	To provide the best experiences, we and our partners use technologies like cookies to store and/or access device information. Consenting to these technologies will allow us and our partners to process personal data such as browsing behaviour or unique IDs on this site.
OneTrust	CSS selector	id=”onetrust
Quantcast	CSS selector	qc-cmp2-consent-info
TrustArc	CSS selector	truste-
Cookiebot	CSS selector	CybotCookiebotDialog
Cookiebot	CSS selector	uc-banner-content
Crownpeak	CSS selector	_evidon_banner
CookieYes	CSS selector	cookie-law-info-bar
Didomi	CSS selector	didomi-notice
Osano	CSS selector	osano-cm-window
CookieYes	CSS selector	cky-consent
Termly	CSS selector	termly-styles
DataPrivacy Manager	CSS selector	hs-eu-cookie-confirmation
Ezoic	CSS selector	ez-cookie-dialog
Google	CSS selector	fc-dialog-container
Quantcast	CSS selector	cmpbox
Quantcast	Hostname	quantcast.mgr.consensu.org
OneTrust	Hostname	cdn.cookielaw.org
TrustArc	Hostname	consent.trustarc.com
Cookiebot	Hostname	consentcdn.cookiebot.com
LiveRamp	Hostname	gdpr.privacymanager.io
Crownpeak	Hostname	c.evidon.com

TABLE 8: Table showing the CMP Fingerprints.

C. Clickable Location & Classification details

C.1. Clickable Keyword Combinations

Opt-in		Opt-out		more options
¬ (no ∨ no ∨ don’t) ∧	accept	Opt-in keyword ∧	essenti	adjust
	activ		necessari	advanc
	agre		requir	choic
	allow	essenti	∨ cooki onli	configur
	assent	necessari		custom
	confirm	requir		customis
	consent	(not ∨ no ∨ don’t) ∧	accept	manag
	continu		consent	option
	enabl		track	personali
	enter	disagre		prefer
	fine	reject		purpos
	ok	declin		set
	okay	refus		tool
	opt-in	deactiv		let me choos
	proceed	continu without accept		select cooki
	understand			
	understood			
	yes			
got it				
i am happi with all cooki				

TABLE 9: keyword Combinations for Opt-in, Opt-out and more options clickables.

Preference Slider	Confirm Preferences		Close Option	Policy Link
on	save		close	privaci
off	submit		dismiss	polici
	Opt-in keyword ∧	select	exit	notic
		choic	x	here
		custom	×	cooki
		prefer	hide	vendor
				partner
				use of cooki
				data protect
				terms servic
				read more
				learn more
				tell me more
				more inform
				see detail
				more detail

TABLE 10: keyword Combinations for Preference Slider, Confirm Preferences, Close Option and Policy Link clickables.

C.2. Custom Clickable CSS Selectors

Clickable Type	CSS Selectors
All Clickables	button a [role=’button’] input[type=’submit’] input[type=”checkbox”] span[class*=”button”] span[class*=”Button”] span[class*=”btn”] span[class*=”btn”] [class*=”close”] [class*=”Close”] [class*=”button”] [class*=”Button”] [data-tracking-opt-in-learn-more] [data-tracking-opt-in-accept] [class*=”settings”] [id*=”custom”] [id*=”accept”] [class*=”settings”] [class*=”custom”] [class*=”accept”]
Close Buttons	[class*=”close”] [class*=”Close”] [aria-label*=”close”] [aria-label*=”Close”] svg
Preference Sliders	input[type=”checkbox”]

TABLE 11: Custom CSS Selectors used to detect clickables in the DarkDialogs system.

D. Dark Pattern Detection details

D.1. Manual validation disagreements

Figures 7 & 8 show the two cases where a button word “continue without accepting” was found. Figures 9 & 10 show the cases where opt-in and opt-out buttons had different colours, but it was unclear if one was really highlighted more than the other.

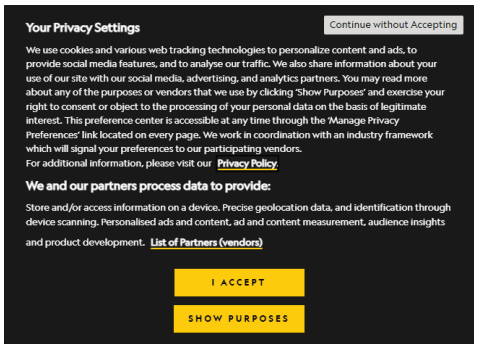

Figure 7: Example of a button worded "continue without accepting" (Source nationalgeographic.com

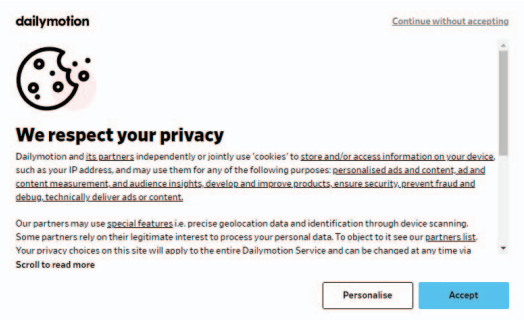

Figure 8: Example of a button worded "continue without accepting" (Source dailymotion.com

Figure 9: Example where opt-in and opt-out buttons had different colours, but it was unclear if one was really highlighted more than the other (Source golf-alcanada.com

Figure 10: Example where opt-in and opt-out buttons had different colours, but it was unclear if one was really highlighted more than the other (Source mailchimp.com

D.2. Criteria for additional dark patterns

This appendix provides additional details about other dark patterns that we identified but were unable to find a way to detect automatically (within the scope of this project).

Dark Pattern 10: Takes more clicks to Opt-out than Opt-in or Opt-out option is not visible.
Category: Obstruction
Impact on the user: Additional effort is required by the user to opt-out than to opt-in which may make sway some users to simply opt-in rather than undergo the additional effort.
Criteria: The number of clicks it took the user to opt-out is more than the number it took them to opt-in during the manual interaction with the cookie dialog.

Dark Pattern 12: Poorly Labelled preference sliders
Category: Sneaking
Impact on the user: The user may be unsure about what one or more preference sliders do when enabled which may lead to them unintentionally opting in for more cookies that they wish to.
Criteria: This is decided by the user during manual input of Dark Patterns. General guidance is that the user believes that labels for one or more preference sliders are unclear to the extent that their purpose is ambiguous.

Dark Pattern 13: In the context of the Cookie Dialog text the standard meaning of the Opt-in and Opt-out buttons is inverted.
Category: Interface Interference
Impact on the user: The typical layout of a cookie dialog has the opt-in button worded affirmatively and the opt-out button worded negatively. Switching this standard ordering may lead to the user unintentionally selecting more cookies than they wished to.
Criteria: This is decided by the user during manual input of Dark Patterns. General guidance is that the user believes that the context of the opt-in button is worded negatively and the opt-out button worded affirmatively to the extent that it may confuse other users.

Dark Pattern 14: Opt-out button is named to guilt the user for selecting it
Category: Interface Interference
Impact on the user: The user may feel guilty or that they are missing out on a feature by opting out of non-essential cookies.
Criteria: This is decided by the user during manual input of Dark Patterns. General guidance is that the user believes that they or other users would feel guilty for selecting the opt-out button.

Dark Pattern 15: When the user clicks the Opt-out button they are asked to confirm their choice
Category: Nagging
Impact on the user: The user has to undergo additional effort when opting out of non-essential cookies.
Criteria: The opt-out button should be followed by an additional action to confirm the user's choice and the opt-in button should not have this additional action.

One noteworthy dark pattern that we missed from our taxonomy is the absence of a cookie dialog on websites that set non-essential cookies. Krisham et al. [41] argues for this dark pattern, they say it is impossible for websites to inform users about cookies without a cookie dialog and that not having a cookie dialog violates existing law. The Spanish Data Protection Authority has fined 6 different websites that set non-essential cookies but do not have cookie dialogs [8]–[13].

D.3. Examples of dark patterns

Figure 11: Example of OnlyOptIn dark pattern (Source: twitter.com)

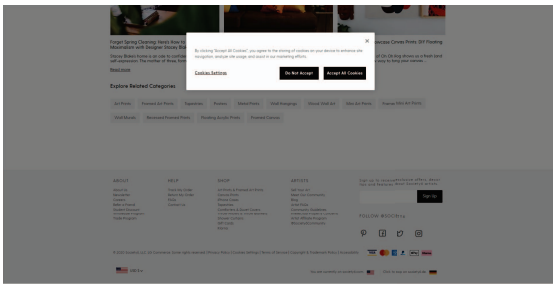

Figure 12: Example of ObstructsWindow. dark pattern (Source: webstaqram.com)

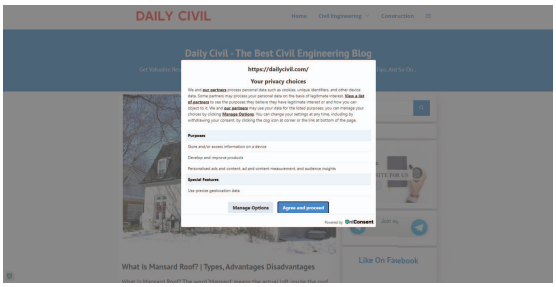

Figure 13: Example of ComplexText dark pattern, this dialog had an FK score of 47.62 (Source: dailycivil.com)

Figure 14: Example of MoreOptions dark pattern (Source: ipsos.com)

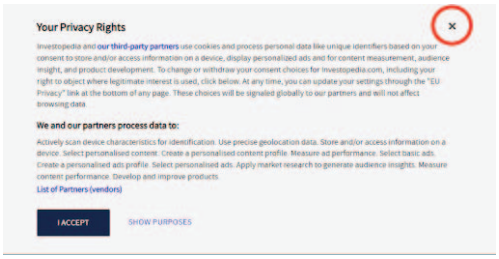

Figure 15: Example of Ambiguous Close dark pattern (Source: investopedia.com)

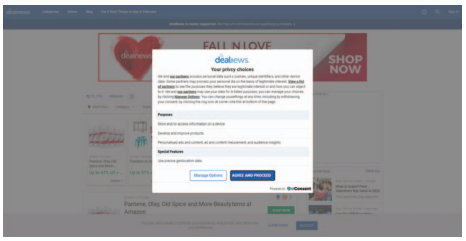

Figure 16: Example of MultipleDialogs dark pattern (Source: dealnews.com)