Automatic detection of manipulative Consent Management Platforms and the journey into the patterns of darkness

Marius Pedersen[1], Frode Guribye[1] and Marija Slavkovik[1]

[1]University of Bergen, Norway

Abstract

We study how to automatically classify different types of manipulative interface design pattern for Content Management Platforms (CMPs), also known as a cookie consents. Our approach uses a scraper to extract different features of CMPs. We then classify the CMP, based on these features, into one of five patterns defined specifically for CMPs. We evaluate our automatic "detector" using four different statistical measures. We also consider factors that cause misclassifications and discuss how to potentially avoid them.

Keywords

consent management platforms, dark pattern, automated detection, manipulative design

1. Introduction

We are here concerned with the problem of automatically classifying *dark patterns* from *consent management platforms*, also called *cookie* banners/notices.

A cookie is a short text file that a Web server stores on a user's hard drive [1]. A cookie is used for websites to remember the user and their preferences [2]. Cookies can be used by advertisers for providing personalized advertising that is used to estimate how successful an advertisement is [3].

Cookies are considered to process personal information. The regulation of personal data depends on where the user is situated (jurisdiction). Within the European Union (EU) and the European Economic Area (EEA) the regulation from the EU General Data Protection Regulation (GDPR)[4] applies. To acquire the user's consent, a website will have to use a *Consent management platform* (CMP) to interact with the user and present them with the information on which data is collected and how it is used. Some aspects of the CMP design are regulated by the GDPR. For example, the CMP must contain the option for the user to decline the website's data usage and still be able to avail themselves of a minimum website service[1].

NAIS 2023: The 2023 symposium of the Norwegian AI Society, June 14-15, 2023, Bergen, Norway

Mariuspedersen51@gmail.com (M. Pedersen); frode.guribye@uib.no (F. Guribye); marija.slavkovik@uib.no (M. Slavkovik)

http://conceptbase.sourceforge.net/mjf/ (M. Slavkovik)

[1] https://europa.eu/youreurope/citizens/consumers/internet-telecoms/data-protection-online-privacy/index_en.htm

Having personal data from its users can be directly profitable to website owners [5, 3]. Consequently, there exists an incentive to prefer that users consent to cookie use. Bauer et al., 2021 [6] show that small changes in the design of a CMP significantly increase or decrease the rate of consent. Thus the website owners have an incentive to seek out and use CMP structures that make it more likely that the user consents to its data being processed. Utz et al, 2019 [7] showed that when the CMP's default design is to offer an opt-in consent choice, as required by the GDPR, only 0.1% of users would choose to actively opt-in for allowing third-party cookies.

In an attempt to guide users into giving consent, websites can use various design elements. Design elements and patterns that are created to serve the intent of nudging a user towards a choice that is not in their best interest are called *dark patterns* [8].

A dark pattern exists in a legally grey area. While the use of dark patterns may be regarded unethical, it may not violate existing laws and regulations. An example of a dark pattern that is prohibited by the GDPR regulation is *pre-checked check-boxes*.

Today there are millions of websites[2] and only a few organizations helping to uphold the regulation by reporting those that break it. Some of these enforcers are NOYB[3], and CNIL[4]. However, with today's tools available they will not manage to check each website for dark patterns, or submit lawsuits against each offender, when necessary. For a regulation to be effective there must be a possibility of its efficient enforcement.

A possible solution for enforcing consent regulations is being able to check websites for dark patterns automatically. *Automation* would allow for fast, continuous detection of dark patterns compared to humans. The automated checker would only need URLs of websites to visit them and identify if there are any dark patterns in their CMP. More specifically, as dark patterns differ in type, an automated checker would identify if a dark pattern type is present.

The automation of dark pattern detection has been considered in the literature [9, 10, 11, 12, 13]. A dark pattern is essentially a manipulation aimed to affect a human user. An automated classifier can only process data, not the emotion or impression a human has when encountering a CMP. Thus to identify a dark pattern type, the automated checker needs to be informed which data of the CMP is related to which dark pattern type.

One way to automatically detect dark pattern types is to describe them manually through a set of human identified relevant features. Soe et al [9] used a machine learning based approach on such human labelled features with a moderate success.

An alternative to human labelled feature detection is to use features of the software code of the CMPs. This is the approach we take.

We first analyzed the available definitions of dark pattern types in the literature and selected the ones were most precise and most pertinent to the CMP context. The dark pattern type definitions used in this thesis are those proposed by Soe et, al 2020 [14]. Second, we identified how to extract data from the CMP and website code. Third, we identified which are the relevant code data features that relate to the Soe et, al 2022 [9] dark pattern type definitions. We constructed a prototype automated dark pattern checker and evaluated it on 2000 websites from the Open Page Rank Initiative[5] list of most visited webpages in the world.

[2] https://siteefy.com/how-many-websites-are-there/

[3] https://noyb.eu/en

[4] https://www.cnil.fr/

[5] https://www.domcop.com/openpagerank/frequently-asked-questions

2. Dark Patterns

Dark patterns are a neologism introduced by Harry Brignull[6]. Brignull defined it as “A user interface that has been carefully crafted to trick users into doing things (...) they do not have the user’s interest in mind." Brignull classified 12 different types of dark patterns. We give their definitions in the Table 2 in the Appendix. To make these twelve patterns more “tractable” for use and interrogation by practitioners [8] created five categories which these twelve split into. The five categories have become a staple in dark pattern research and are explained in Table 3 in the Appendix.

The five categories introduced by [8] are not meant to be used for building automated dark pattern classifiers: they are content independent and general. Soe et al 2020 [14] went further in refining and specifying the concept of dark patterns. They focused specifically on dark patterns that occur in CMPs. The eight different dark pattern types that are introduced by Soe et al [14] have a closer connection to how each pattern is represented by features on a CMP. We use these for building our automated detection approach and give their definitions here:

Does not Count is a not a visual design pattern and it is not detectable by humans. It occurs where data is collected although the user has denied consent. This practice was shown in [12, 15, 16]. The legal aspects of this practice are discussed in [12]. Papadogiannakis et al, 2021 [15] found, using advanced ID techniques of ID and information leaking, that more than 75% of all websites have shared the user’s information after the user has rejected all cookies. Bollinger et, al 2022 [16] detect cookies active after rejecting a CMP. They found that 21% of websites still share with third parties after the user clicks “reject all”.

No choice is the CMP design to not give an option to the user to actually deny consent. A common method is that the CMP will only inform the user that the website will use the user’s personal data. An example of a CMP with a “no choice” pattern is given in Figure 7 in the Appendix.

Multiple choice panels is the design pattern to offer more than one CMP choice panels. The user needs to decide which one to click on and it is not clear whether all panels have the same consent accept or reject effect. An example of “multiple choice pannels” is given in Figure 8 in the Appendix.

Choice cascade is the design pattern that involves a cascade of panels, which the user has to navigate through. The user needs to click through at least one instance of “read more” or “learn more” or “settings” or similar to reach the panel where consent can be denied. An example of “choice cascade” is given in Figure 9 in the Appendix.

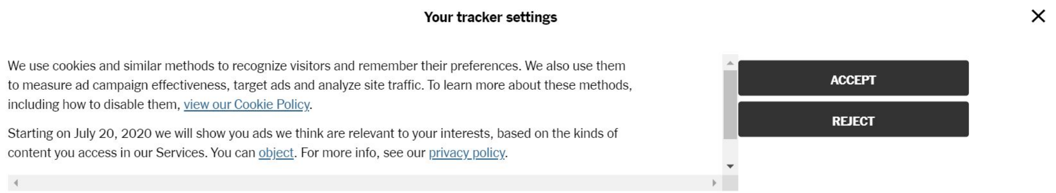
Widget inequality is a pattern of disparity between the consent and refuse options that favors the acceptance choice. This may be accomplished through a variety of design elements, but it is most evident visually when the accept button is intended to be prominent with appealing

[6] https://www.darkpatterns.org/about-us

coloring while the reject button, if there, virtually blends in with its surroundings. These do not have to be color differences, they can also be other design aspects. An example of widget inequality can be found in Figure 10 in the Appendix.

Unlabeled sliders is a dark pattern that occurs when the CMP uses sliders to indicate consent for different cookies or cookie types, but it is not clear what each position on the slider is, which is on and which is off. This pattern may also include check-boxes that appear as if they cannot be changed. Figure 11 in the Appendix shows an example of an unlabeled slider pattern.

Unmarked X is a type of dark pattern that occurs when there is an “X" to close the CMP however, it is not explained whether closing the panel in this way will consent to the CMP or not. Thus, the close function can be misleading. In the example given in Figure 1 we can see the closing X button in the right corner. How this button is interpreted in this example is not specified in the CMP.



Your tracker settings

We use cookies and similar methods to recognize visitors and remember their preferences. We also use them to measure ad campaign effectiveness, target ads and analyze site traffic. To learn more about these methods, including how to disable them, view our Cookie Policy.

Starting on July 20, 2020 we will show you ads we think are relevant to your interests, based on the kinds of content you access in our Services. You can object. For more info, see our privacy policy.

Figure 1: An example of an unmarked x from The New York Times. It is unclear what will happen or how it will interpret the ’X’.

No antonyms is a dark pattern that relies on expressing consent and rejection without using antonyms. The language to accept the cookies is typically clear and direct, while the language to reject them is convoluted. An example, also given in Figre 12 in the Appendix, is to use “Agree and proceed” for one button and “Proceed with required cookies only” for the reject option.

An automatic detection program needs a choice of CMP features to use. For example, if one uses only screenshots of the CMP including or not the surrounding website, one would manage to extract and identify a high percentage of CMPs and would not be reliant on how the website have coded their notice. However, one loses direct link to multiple useful features that are important for some of the dark pattern types. For example, the text feature. You can still extract some text from a screenshot, but you must account for some errors and it will not be fully accurate for all CMPs. Text can be processed with natural language processing to attempt to find some interesting patterns, which we may not want to omit. The automatic detection and analysis approach we use attempts to use both extracted textual features from the CMP code and features extracted from screenshots taken of the CMPs.

We managed to automatically detect no choice, choice cascade, widget inequality, unlabeled sliders, and no antonyms.

The “does not count” dark pattern is already automatically detected in earlier work [12, 15, 16].

The multiple-choice panels pattern raises problems which are harder to solve than the other patterns. The most apparent is that when there are multiple CMPs with buttons for allow or deny, it is hard to identify which CMP is the one that provides the valid reject or accept. What happens if you click accept on one and decline on the other? To solve this one would have to monitor which cookies are in use before and after as with the pattern “Does not count”. We have only seen multiple-choice panel on the https://www.manilatimes.net/ and https://www.bbc.com/ webpages. For both websites there seems to be one notice that is unique, and one created by a CMP provider that is registered at IAB Europe Transparency and Consent Framework[7] which may be a contributing factor in these two websites having two CMPs. To classify this pattern, one will have to create a scraper that locates a CMP, then when it has located one it must store the location of that CMP such that it does not locate that CMP again. Thus, after extracting information from the first located CMP, it needs to exclude the elements under the area of the stored positions as for these elements not interfering from the scrapers search for a second CMP. It then would search for a CMP, however now a decision of which CMP should the other dark pattern types be classified from. If for example the first CMP does use antonyms but the second does not, is it a case of “no antonyms”? Given the low apparant frequency of multiple-choice panels and considerable need for additional effort. we omit this pattern for now.

The dark pattern type unmarked X we excluded also due to its apparent rarity, as well as issues extracting the button, on the websites it was present. Some of these issues are due to it being often stored as a “close" button, where the “X" can be an image. We attempted to use object character recognition (OCR) to extract the x button, but we did not have success.

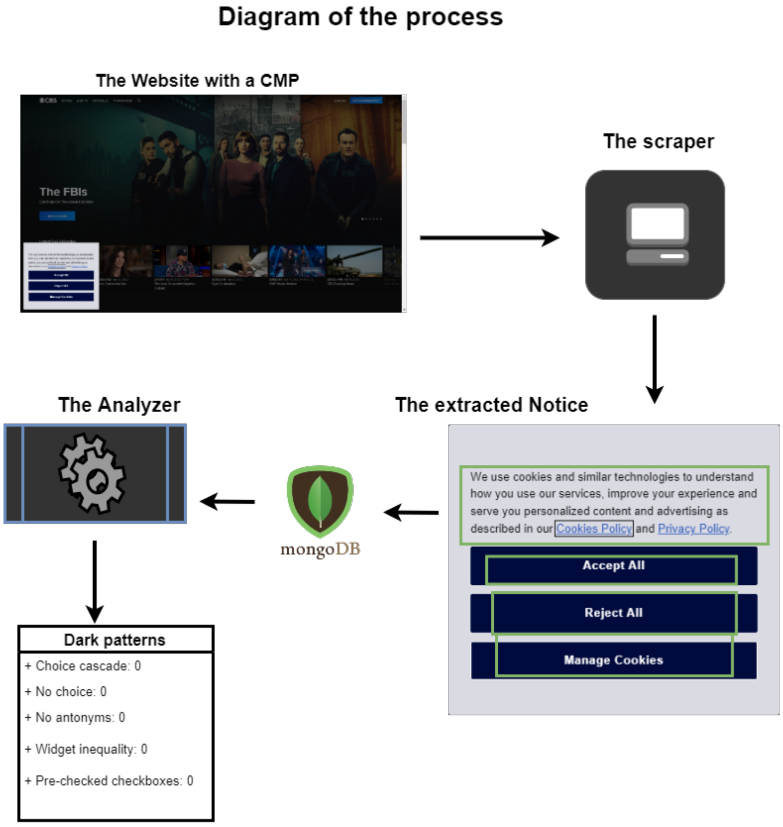
Another issue is how to find out whether the information of closing the notice with the close button is interpreted is in the CMP text.

3. Methodology

To test the automatic detection of the selected dark patterns, we collected CMP features from 2000 websites. The data was collected with the web scraper of [17] which we modified for our use. The [17] scraper was developed to extract certain parameters: the size of the CMP, size of the buttons, the color of the buttons, the number of pre-checked checkboxes, the readability level of the CMP, how long the website saves cookies and if the website redirects the user when the user rejects the cookies. The scraper takes a screenshot of each website it visits and a screenshot of both the initial page of the CMP and the settings page for the CMP. A modified version of the scraper was implemented that altered how the scraper finds the location of the buttons. The modifications increased the number of searchable terms for each type of button; in addition to searching for refuse and accept on both the initial and settings pages, additional searchable links were added when searching for the cookie policy link.

To explain the data extraction from the CMP and website code we require some introduction to HyperText Markup Language (HTML) and how a CMP is coded. A CMP is not an own element on a website but rather a generic tag called “div”. Each element of a CMP is often just given an incomprehensible name of numbers and letters consisting of the generic “div” tags.

[7] https://iabeurope.eu/transparency-consent-framework/



We use cookies and similar technologies to understand how you use our services, improve your experience and serve you personalized content and advertising as described in our Cookies Policy and Privacy Policy.

Figure 2: Diagram of the automated dark pattern checker architecture.

This can make it difficult to detect a CMP and may be one of the reasons why earlier research with automatic detection of CMPs has had such a low recall [10].

The features of the cookie notices that are of interest in the current paper are: the text of the CMP, the optional “settings" / “learn more" page, a consent button the CMP and the “settings" page (if available), a decline/reject button on the CMP and the “settings" page (if available), checkboxes and their checked status (if available). We will refer to these as simply “the features" from now on.

The diagram in Figure 2 depicts the architecture of our automated dark pattern checker. The data from the scraper is stored on a MongoDB[8] cloud database. We constructed a python program to extract the information from a CMP and give a classification to one of the five dark patterns present on the CMP if any.

When the program has obtained its chosen data, mostly textual features and screenshots, it utilizes an object character recognizer (OCR) to retrieve the text from the screenshot[9]. From the OCR, a lengthy string of text is retrieved. This string is analysed to classify different dark patterns. Most notably we used arrays of synonyms and antonyms to identify many of the dark

[8] mongodb.com

[9] https://github.com/tesseract-ocr/tesseract/blob/main/ChangeLog

patterns.

To evaluate the classifications made by the program, taking into consideration both false positives and false negatives, we used the following method for calculating a precision score and a recall score.

Precision is the number of correctly classified instances of dark patters (true positives, TP) divided by TP plus the number of false positives (FP). Precision = TP/TP+FP.

Recall is the number of true positives (TP) divided by TP plus the number of false negatives (FN). $Recall = \frac{TP}{TP+FN}$.

When combined, precision and recall can be used to calculate a measure (F1) of the quality of the classification:

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}.$$

One weakness of this method is that it does not take into account the number of true negatives (TN)[18]. A measure of the accuracy of the classification can thus be calculated by summarizing what is classified correctly (TP + TN) divided by the total of classification options (TP + TN + FP + FN):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

Sampling. The scraper crawled 2000 websites from the top 10 000 websites from "Domcop 10 million most visited websites"[10]. This list is created by "The Open PageRank initiative"[11]. The scraper returned 629 unique CMPs from the 2000 websites. The 629 CMPs returned from the scraper were manually inspected to identify whether the websites actually did contain a CMP or if it was a false positive. 69 false positives were identified in the manual inspection which gives a precision measure of 0.89.

The github with the datasets from the investigation and the code for the programs can be found here: https://github.com/MAP-12/Automatic_detection_of_Dark_patterns. The program that classifies each of the dark patterns is called Analyzer.ipynb. The investigation dataset of CMPs that was returned by the scraper is called cookienotices.csv and can be found in the folder called Data investigation. In this folder you will also find a file called no notices which is the investigation of 101 random sample from the websites the scraper did not find any CMPs. The code of the can be found in the related files in the folder called Scraper, along with a license of free to use from the bachelor thesis [17] that first created the scraper.

4. Results

From the 629 CMPs returned from the scraper, we manually investigated whether the websites actually did contain the CMPs. Each website was accessed in an incognito google Chrome browser window which is what the scraper also uses. We also checked the screenshots from each

[10] https://www.domcop.com/top-10-million-websites

[11] https://www.domcop.com/openpagerank/what-is-openpagerank

website that the scraper took, because the website may have had updated their CMP from the time the scraper ran to the time the investigation took place. The investigation stored whether it was an actual CMP there, what kind of dark pattern is prevalent in the CMP, and if it was from Google. Checking if it was from Google may seem like an odd feature, but it was early in the investigation noted to be very useful as Google had many different websites present on the URL list, which all shared CMPs which where misclassified. It became more important as the checker would manage to correctly classify the new CMP Google has been on track to roll out after the completion of this work in response to the CNIL case against them[12]. The new CMP can be seen in https://blog.google/around-the-globe/google-europe/new-cookie-choices-in-europe/.

From the investigation of the 629 CMPs returned from the scraper, the actual number of CMPs is 560. That means there are 69 false positives. Using the statistic measures from the preliminaries chapter we would ideally calculate recall, precision and F1 score to verify how accurate the scraper classifies that there is a CMP present. The precision is $Precision = \frac{560}{560+69} = 0.890$. However, due to time constraints it was not possible to manually verify all the 2000 websites visited meaning we could not calculate the recall and F1-score of the scraper as we did not have the number of false negatives (the CMPs the scraper did not discover).

We checked the websites for which the scraper indicated that no CMPs were found. As the scraper also takes a screenshot regardless of there is a CMP there or not, it was possible to verify how accurate the scraper was in extracting CMPs. We took a random sample of 101 websites from the non CMP 1 372 websites. We observed that the obstacle for finding CMPs was that websites blocked the scraper or they took too long to load the CMP. From the 101 randomly sampled websites 14 CMPs were found when visited. When we checked the screenshot the scraper took, we observed that in 12 cases the scraper was either blocked, or there was no CMP present on the scraper's screenshot. For the remaining 2 that had a CMP present in both the screenshot and direct observation, one of them was in fully German and the other the scraper missed for unknown reasons, given in Figure 3.

We discovered multiple causes for errors among the 69 false positive cases. One prevalent cause was pages with a link to a privacy and cookie policy on a fixed spot either on the bottom or top of the page. This setup was appearing on all the American government pages. Unfortunately, there were also discovered CMPs which had not been detect had it not been for the method that locates privacy or cookie policy links with fixed positions (See Figure 4 for an example of CMP that had not been detected, and Figure 5 for an example of a misclassification.

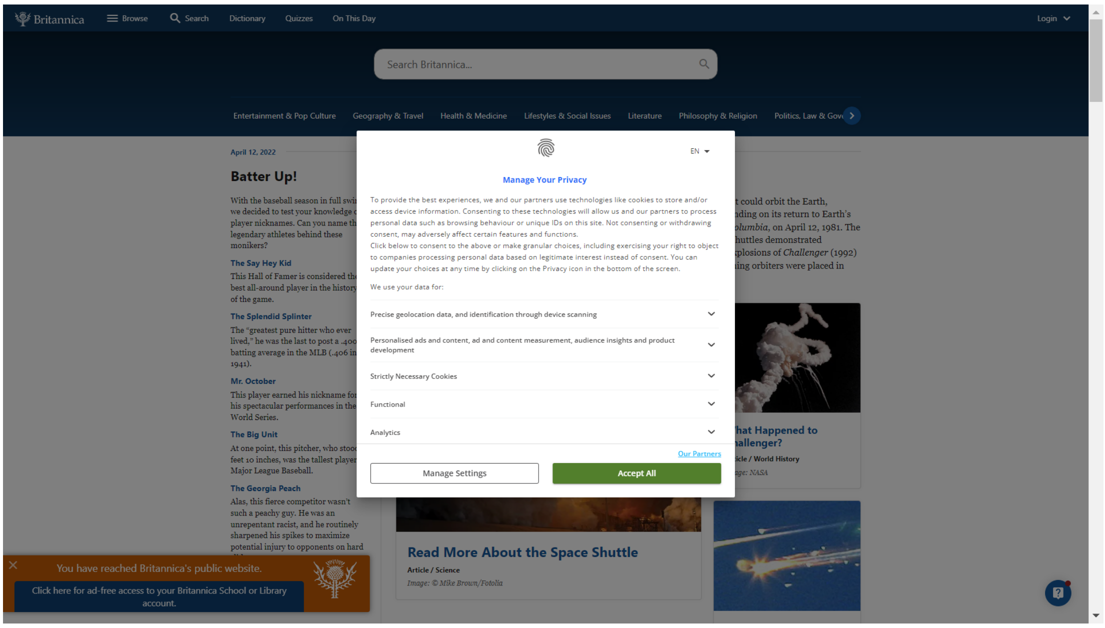
There are also a few instances where there were no CMPs, but the website wanted to use a notification feature of the browser which requires the user to manually allow a pop up. An example of this misclassification can be seen on Figure 6.

Before the removal of the confirmed false positives from the data, the checker detected 238 types of choice cascade, 116 types of no choice, 428 types of no antonyms, 21 types of there being pre-checked checkboxes and 17 types of widget inequality.

After false positives are removed there are left 208 cases of the dark pattern type Choice cascade, 162 types of no choice, 362 types of no antonyms used, 62 cases of pre-checked checkboxes and 15 cases of widget inequality from the 560 CMPs.

Table 1 summarises the results separately for the Google related and non-Google related

[12] https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply



April 12, 2022

Batter Up!

Manage Your Privacy

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. Not consenting or withdrawing consent, may adversely affect certain features and functions.
Click below to consent to the above or make granular choices, including exercising your right to object to companies processing personal data based on legitimate interest instead of consent. You can update your choices at any time by clicking on the Privacy icon in the bottom of the screen.

We use your data for:

Precise geolocation data, and identification through device scanning

Personalised ads and content, ad and content measurement, audience insights and product development

Strictly Necessary Cookies

Functional

Analytics

Our Partners

Manage Settings | Accept All

Read More About the Space Shuttle

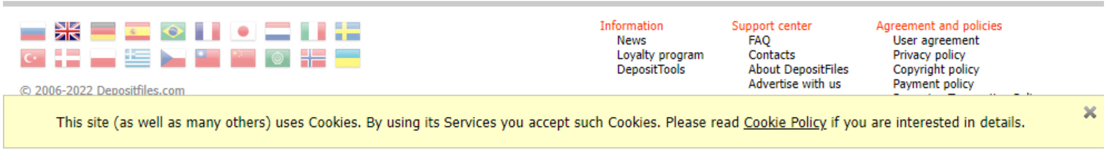
Article / Science

Image: © Mike Brown/Fotolia

You have reached Britannica's public website.

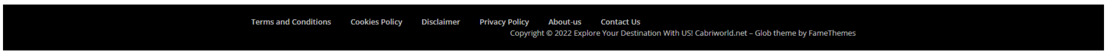
Click here for ad-free access to your Britannica School or Library account.

Figure 3: Example from https://www.britannica.com/ unknown why the scraper miss this



This site (as well as many others) uses Cookies. By using its Services you accept such Cookies. Please read Cookie Policy if you are interested in details.

Figure 4: Example from https://depositfiles.com/



Copyright © 2022 Explore Your Destination With US! Cabriworld.net – Glob theme by FameThemes

Figure 5: Example from https://cabriworld.net/

scrapped examples. Google have 90 related websites in the extracted CMPs.

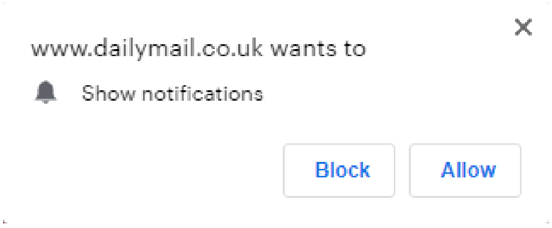
Table 1

Results on Google related CMPs on left-hand side and non Google related CMPs on right-hand side

Type	Accuracy	Precision	Recall	F1-score	Type	Accuracy	Precision	Recall	F1-score
Choice Cascade	0.63	0.55	0.57	0.56	Choice Cascade	0.76	0.65	0.76	0.70
No Choice	0.74	0.65	0.42	0.51	No Choice	0.76	0.65	0.53	0.58
No antonyms	0.83	0.95	0.82	0.88	No antonyms	0.93	0.95	0.95	0.95
Pre-checked checkboxes	0.92	0.53	0.66	0.59	Pre-checked checkboxes	0.90	0.53	0.68	0.60
Widget inequalities	0.99	0.87	1.00	0.93	Widget inequalities	0.99	0.87	1.00	0.93

The world of CMPs moves fast. From running the scraper to coming around to verify the results from the checker, only two weeks later, multiple CMPs where different [13]. These are

[13] The tests were completed 12th of April 2022.



www.dailymail.co.uk wants to

Show notifications

Figure 6: Example from https://dailymail.co.uk/

among others Youtube and Linkedin. Youtube is owned by ALPHABET, Youtube and Google's parent company which got fined for their CMP[14]. Linkedin did not get fined but changed their CMP so that it too follows the regulation that CNIL enforced on Facebook and Alphabet.

5. Discussion

From the data it is apparent that the accuracy score is highest. This comes from the checker being much better at identifying CMPs that do not contain a dark pattern since the TN value is only used in the calculation of the accuracy. From the data it is apparent that the no antonyms and widget inequalities has the highest detection accuracy. These patterns have a precise definition which can be seen to contribute to the successful detection. To check if a CMP used an antonym, it was clearly defined how accept and decline could look. It also helped that most website that had an option to decline had it on the settings page.

Widget inequality was a pattern harder to manually investigate as the threshold of size difference bettween widegets was set to 10% . It was hard to spot if the accept button was 10% or bigger than the decline button. This posed a dilemma as it was such barely noticeable, if it would be better to have a higher threshold. The other problem was that often the character length of the word used for accepting is longer than that of decline, e.g., “approve” and “deny”. This was however expected as the size threshold for it to be a dark pattern or not will require more research.

Features that caused misclassification, where often text based. A good example is instagram.com which the checker classified as having ’no choice’ and ’no antonyms’. It was classified as such since Instagram gave two options on their CMP, see Figure 13 in the Appendix. Instagram uses “Only Allow Essential Cookies” and “Allow Essential and Optional Cookies”. So, the trigger for a decline option is not engaged. This way of posing the options is deceiving in a variety of ways. From having the accept button start the same way as the decline option, letting the accept option be in a bright blue with a bold font while the decline option is the same color and style of the text explaining the cookies above. This can make the option be easily mistaken for just some text of the explanations of cookies and not an option of itself. This is however not a standalone CMP and there are multiple of these CMPs with two options of “only allow”. Other terms we encunterd being used were “use essentials", “proceed with required" or “use necessary". These are bi-terms and with more effort in developing a natural language processing

[14] https://www.cnil.fr/en/use-google-analytics-and-data-transfers-united-states-cnil-orders-website-manageroperator-comply

solutions, some of these could be classified as a decline option. This gives room for improving the false positives for the pattern “no choice” and the false negatives for “choice cascade”.

The decision to use the Open Page Rank Initiative over other more popular website rankings such as the Alexa top websites[15] was due to Alexa having more non English websites on their top listed websites. We also found out that Amazon will close the service 1st of May 2022[16] which may have caused problems for reproducing the work in this project.

A problem with the Open Page Rank Initiative is in the way it ranks websites – not entirely dependent on popularity but how much is the website linked. Therefore, there were some websites that looked quite old and that probably have less than 2000 visitors monthly. The specific number of visitors, is of course, not relevant. We use the number of visitors as a proxy for popularity.

A limitation of our system is that it only uses three antonyms for each of the synonyms for accept and decline. This was a decision based on practicality, and to have a stricter threshold for what is an antonym or not. The stricter threshold helps differentiate between the CMPs that uses clear antonyms for accept and decline, and those that do not but have a decline and accept option. With a too soft threshold on the number of antonyms one can end up classifying every CMP that has an accept and decline option of using antonyms which would make the type obsolete.

Another limitation of our system is handling websites that write about rejecting the cookies in their CMP while not actually offering a proper reject option on the CMP. If the text available to the checker contains a synonym for “reject” the checker will conclude that there is reject an option. This causes a problem since if it were to require the reject option to be on the decline button it would require a scraper that is much more accurate in extracting the decline button for those websites that have it available. If the classification system instead required there to be more than one reject synonym, the analyzer would miss a large portion of those that have been classified of having a decline option and verified as many CMPs only write a synonym for reject once, and that is on the reject button. Unfortunately, some of the CMP type providers that only write about the reject option is Google. This is unfortunate as Google have 90 related websites in the extracted CMPs. These websites follow three different, but similar CMP set ups. The CMPs have a misclassification of having a reject option in the first CMP page which is an antonym. This classifies then the CMP for the most common Google CMP type as not containing any dark patterns.

The other CMP setups Google uses do not have an option to reject and should be classified as a no choice. On these CMPs Google follows a trend with CMPs of only informing the user that they use cookies, and when the user clicks on their cookie policy they are sent to a long page with ’legal speak’ with no option to decline.

There are websites that do not offer any save or reject options on their “settings" page but pre-checked checkboxes or sliders. In the current setup for the checker, it classifies these as having “no choice" and “no antonyms" as it can’t locate a decline button. In one way that is correct as there is not a reject button but instead you have to see to that the checkboxes are un-checked. However, it can be argued that in a way the user has the option to reject by clicking

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

[16] https://support.alexa.com/hc/en-us/articles/4410503838999

on to this kind of settings page then entering the website again if it then stores the user as having rejected.

An issue that we discovered during direct investigation was that even if the scraper's browser language was set to English there where multiple websites that gave the CMP in the language of the website or the language from where the IP-address of the user was (Norway). Our solution was to extend the vocabulary used by the scraper to contain Norwegian synonyms and antonyms, but we did not do the same for other languages. This affected the accuracy scores as well, as we detected at least 33 websites with a different language from Norwegian and English in the 639 CMPs the scraper found. We did not track the exact effect on accuracy that each language had.

An interesting aspect that caused misclassification for the scraper was websites that blocked the scraper. Facebook, for example, had a legal warning in both the robots.txt file and a pop-up warning in the scraper program[17]. However, these warnings have been deemed not accurate by US supreme court in a court ruling where Linkedin attempted to stop a rivaling company scraping information on users from their website[18]. It was ruled that web scraping is allowed and legal if the information is publicly available.

6. Summary and future work

In this paper we explored the possibility to automatically detect dark patterns in CMPs. We focused on detecting 5 of the 8 dark patterns specifically defined for CMPs by [14]. As pointed out in [9], automated detection of dark patterns is hard.

As it can be observed from the evaluations summarised on Table 1, the five dark patterns we focus on are not detected with very high accuracy that can be considered already practical. The scraper managed to extract enough CMPs that we could get an impression of how the analyzer handled different CMPs from different websites. We can conclude that, while it is possible to automatically detect dark patterns, this is not a simple task. The biggest challenge is in working with a "moving target". There are infinite ways to design a CMP, while automated detection requires some framework of predictability. Better CMP regulation can offer such predictability.

There is much future work to be explored. One aspect of the work is software engineering, namely improving the scraper. Another direction is to explore using visual machine learning algorithms in addition to the work we do here. in addition we can also explore using non-supervised learning methods. Lastly, we can refine further the dark pattern definitions into elements that are markup specification and elements that are human detectable.

References

[1] A. M. Hormozi, Cookies and privacy, Information Systems Security 13 (2005) 51–59. doi:10.1201/1086/44954.13.6.20050101/86221.8.

[2] J. Park, R. Sandhu, Secure cookies on the web, IEEE Internet Computing 4 (2000) 36–44. doi:10.1109/4236.865085.

[17] https://www.facebook.com/robots.txt

[18] https://techcrunch.com/2022/04/18/web-scraping-legal-court/

[3] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, Z. Chen, How much can behavioral targeting help online advertising?, in: Proceedings of the 18th International Conference on World Wide Web, WWW '09, Association for Computing Machinery, New York, NY, USA, 2009, p. 261–270. URL: https://doi.org/10.1145/1526709.1526745. doi:10.1145/1526709.1526745.

[4] E. Commission, General data protection regulation, 2018. URL: https://gdpr-info.eu/.

[5] I. N. Cofone, The way the cookie crumbles: online tracking meets behavioural economics, International Journal of Law and Information Technology 25 (2016) 38–62. URL: https://doi.org/10.1093/ijlit/eaw013. doi:10.1093/ijlit/eaw013.

[6] J. M. Bauer, R. Bergstrøm, R. Foss-Madsen, Are you sure, you want a cookie? – the effects of choice architecture on users' decisions about sharing private online data, Computers in Human Behavior 120 (2021) 106729. URL: https://www.sciencedirect.com/science/article/pii/S0747563221000510. doi:https://doi.org/10.1016/j.chb.2021.106729.

[7] C. Utz, M. Degeling, S. Fahl, F. Schaub, T. Holz, (un)informed consent: Studying gdpr consent notices in the field, in: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, CCS '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 973–990. URL: https://doi.org/10.1145/3319535.3354212. doi:10.1145/3319535.3354212.

[8] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, A. L. Toombs, The dark (patterns) side of ux design, in: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA, 2018, p. 1–14. URL: https://doi.org/10.1145/3173574.3174108.

[9] T. H. Soe, C. T. Santos, M. 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, 2022. URL: https://arxiv.org/abs/2204.11836. doi:10.48550/ARXIV.2204.11836.

[10] M. Nouwens, I. Liccardi, M. Veale, D. Karger, L. Kagal, Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence, Conference on Human Factors in Computing Systems - Proceedings (2020) 1–13. doi:10.1145/3313831.3376321. arXiv:2001.02479.

[11] P. Hausner, M. Gertz, Dark Patterns in the Interaction with Cookie Banners, Position Paper at the Workshop "What Can CHI Do About Dark Patterns?" at the CHI Conference on Human Factors in Computing Systems, May 8-13, 2021, Yokohama, Japan 1 (2021) 1–5. URL: http://arxiv.org/abs/2103.14956. arXiv:2103.14956.

[12] C. Matte, N. Bielova, C. Santos, Do cookie banners respect my choice?: Measuring legal compliance of banners from IAB europe's transparency and consent framework, Proceedings - IEEE Symposium on Security and Privacy 2020-May (2020) 791–809. URL: https://arxiv.org/pdf/1911.09964.pdf. doi:10.1109/SP40000.2020.00076. arXiv:1911.09964.

[13] A. Mathur, G. Acar, M. J. Friedman, E. Lucherini, J. Mayer, M. Chetty, A. Narayanan, Dark patterns at scale: Findings from a crawl of 11K shopping websites, Proceedings of the ACM on Human-Computer Interaction 3 (2019). doi:10.1145/3359183. arXiv:1907.07032.

[14] T. H. Soe, O. E. Nordberg, F. Guribye, M. Slavkovik, Circumvention by design - dark patterns in cookie consent for online news outlets (2020). URL: https://doi.org/10.1145/3419249.3420132. doi:10.1145/3419249.3420132.

[15] E. Papadogiannakis, P. Papadopoulos, N. Kourtellis, E. P. Markatos, User tracking in the

post-cookie era: How websites bypass GDPR consent to track users, in: Proceedings of the Web Conference 2021, ACM, 2021. URL: https://doi.org/10.1145%2F3442381.3450056. doi:10.1145/3442381.3450056.

[16] D. Bollinger, K. Kubicek, C. Cotrini, D. Basin, Automating Cookie Consent and GDPR Violation Detection, in: Proceedings of the 31st USENIX Security Symposium, 2022. URL: https://doi.org/10.3929/ethz-b-000525815.

[17] T. Liljedahl Hildebrand, F. Nyquist, Cookies, GDPR and Dark patterns: Effect on consumer privacy (Dissertation), Master's thesis, Blekinge Institute of Technology, Faculty of Computing, Department of Computer Science, 2021. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:bth-21726.

[18] D. Powers, Evaluation: From precision, recall and f-factor to roc, informedness, markedness & correlation, Mach. Learn. Technol. 2 (2008).

Appendix

Definitions of dark patterns

Table 2 gives the definition of the original 12 dark patterns specified by Brignull[19].

Table 2
The initial twelve different dark pattern types defined

Category name	Explanation
Trick question	Is a question framed with the intention to confuse the reader and nudge them into answer in a particular way. Without giving it much thought, you think to answer your opinion you shall click option 'A', but the question is framed in such a manner that you should answer 'B'. An example of this can be "would you *not not like to consent* to our cookie settings?" In this example there is a double negative which is a tactic used to create a confusing question[8].
Sneak into basket	Upon attempting to purchase something on a website, but in your shopping cart you find the website have added a product you have not clicked on.
Roach model	When a service is designed for it being very easily for you to join/buy their services/product but difficult to get out of/cancel. Amazon mentioned earlier as an example on how it is very difficult to unsubscribe.
Privacy Zuckering	When you share more private information than you would want to share. Inspired by the namesake, Mark Zuckerberg with Facebook.
Price Comparison Prevention	Design that have intentionally made it difficult to compare the price of the product you are looking at with another similar item.
Misdirection	Website design that attracts your attention away from anything the website doesn't want you to notice or spend a lot of time thinking about.
Hidden Costs	When you are at the last step for an online purchase but there have been added costs that until now had been hidden from you. Therefore, making your purchase more expensive than what you intended to, but you are now so far in the process you go along with it.
Bait and Switch	The user set out to do one thing but through the design of the website the user end up doing something else often more undesirable, but possibly more beneficiary for the website.
Confirmshaming	The user is guilted into consenting to what the website asks of them due to the wording of the question and/or the answer options.
Disguised Ads	Advertisement which hides that they are advertisement for a product or service. For example the "article ads" which are ads on newspaper websites that looks like other articles, but is in fact just an article for the product or service that they are selling.
Forced Continuity	When the period from free trial goes to paying member, changes without any notification.
Friend Spam	When a service asks for your email or social media for the website/product/service to be in some way beneficial to you, but what ends up happening is that they use it to send 'spam' mails to all your contacts claiming to be from you.

[19] https://www.darkpatterns.org/about-us

All of the dark pattern types mentioned in Table 2 can be read in more depth on the website linked in the footnote.[20]

Table 3 gives the five dark pattern categories from [8].

Table 3
The five dark pattern categories from [8]

Category name	Explanation
Nagging	A design element that is present only to be a nuisance until the user interacts with it in a desired way.
Obstruction	Design that makes a process more difficult than it would inherently be due to either blocking design elements or functionality.
Sneaking	Design whose purpose is to hide or delay information that is relevant to the user. A common place this happens is on online stores, where the store may add an item to the user's shopping cart without the user knowing about it.
Interface Interference	Design that highlights the parts of the interface the website wants you to use and hides other information that is not equally beneficial to the website. Can be seen on CMPs where it will often be a visible button to accept, and the button to decline will not even be on the first page of the CMP, but on another "settings" page.
Forced Action	Design that forces the user to perform a specific action to use or continuing to use certain functionality of the website. Examples of this can for example be websites which requires the user to make an account to use the website.

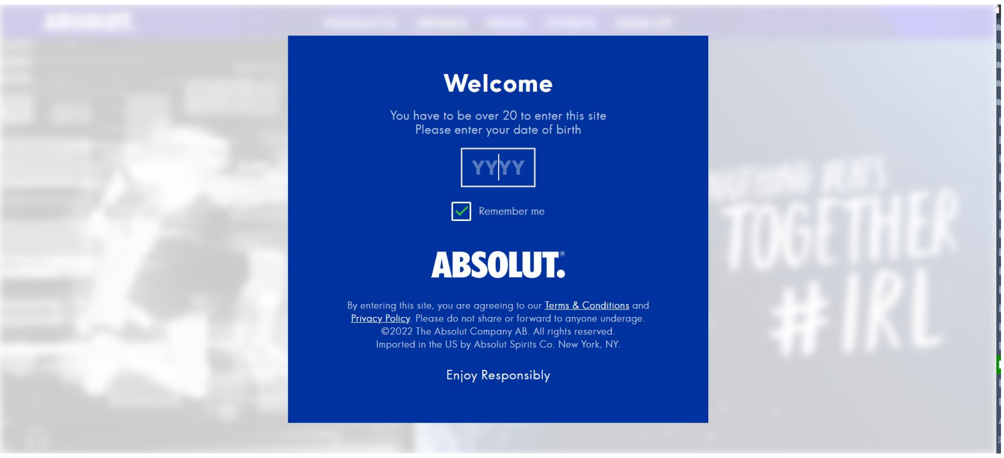
6.1. Visual examples

This section gives visual examples of the 7 CMP dark patterns defined in [14].

An example of "no choice" In this example there is a section that requires the user to write in their date of their birth as it is a site with age restricted content, and then there is text that if you proceed you agree to their privacy and terms & conditions policy. On figure 7 there are no option to reject their privacy policy, and therefore this example contains the dark pattern "no choice".

An example of "no choice" pattern example is given in Figure 8. We see there is a CMP in the middle of the screen giving the user the option to "consent" or "do not consent". At the lower right corner, we see there is another CMP that is also about cookies giving you the option to "agree" or to "read more" about them. With these two CMPs it is unclear if the user have to answer both of them or if only one is enough.

[20] https://www.darkpatterns.org/types-of-dark-pattern



Welcome

You have to be over 20 to enter this site
Please enter your date of birth

YYYY

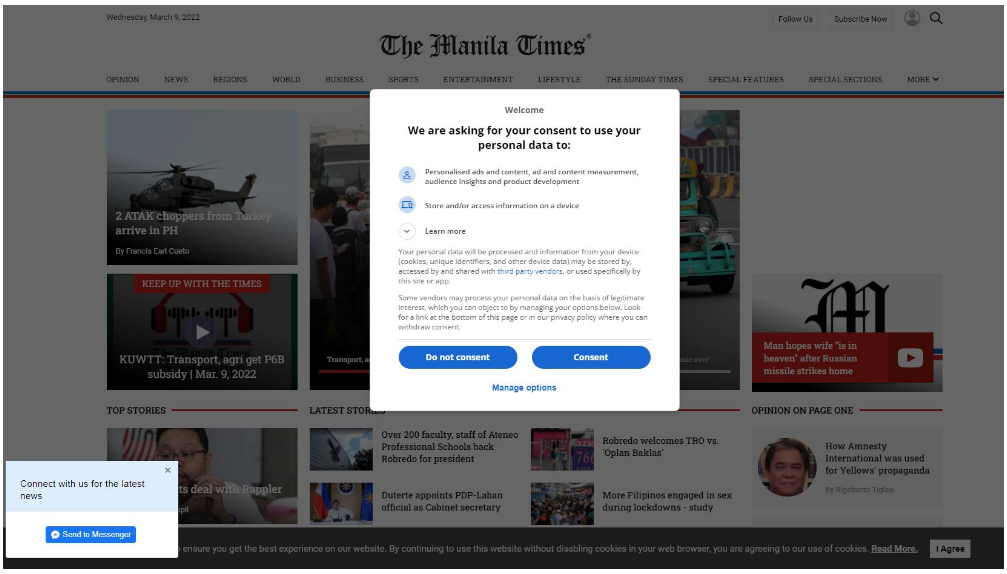
☑ Remember me

ABSOLUT.

By entering this site, you are agreeing to our Terms & Conditions and Privacy Policy. Please do not share or forward to anyone underage. ©2022 The Absolut Company AB. All rights reserved. Imported in the US by Absolut Spirits Co. New York, NY.

Enjoy Responsibly

Figure 7: An example from https://www.absolut.com/en/ where the user have to accept their cookie policy to enable usage of the website



2 ATAK choppers from Turkey arrive in PH
By Francis Earl Cueto

KEEP UP WITH THE TIMES
KUWTT: Transport, agri get P6B subsidy | Mar. 9, 2022

Man hopes wife "is in heaven" after Russian missile strikes home

Over 200 faculty, staff of Ateneo Professional Schools back Robredo for president

Robredo welcomes TRO vs. 'Oplan Baklas'

Duterte appoints PDP-Laban official as Cabinet secretary

More Filipinos engaged in sex during lockdowns - study

How Amnesty International was used for Yellows' propaganda
By Rigoberto Tiglao

Welcome

We are asking for your consent to use your personal data to:

Personalised ads and content, ad and content measurement, audience insights and product development

Store and/or access information on a device

Learn more

Your personal data will be processed and information from your device (cookies, unique identifiers, and other device data) may be stored by, accessed by and shared with third party vendors, or used specifically by this site or app.

Some vendors may process your personal data on the basis of legitimate interest, which you can object to by managing your options below. Look for a link at the bottom of this page or in our privacy policy where you can withdraw consent.

Do not consent | Consent

Manage options

Connect with us for the latest news

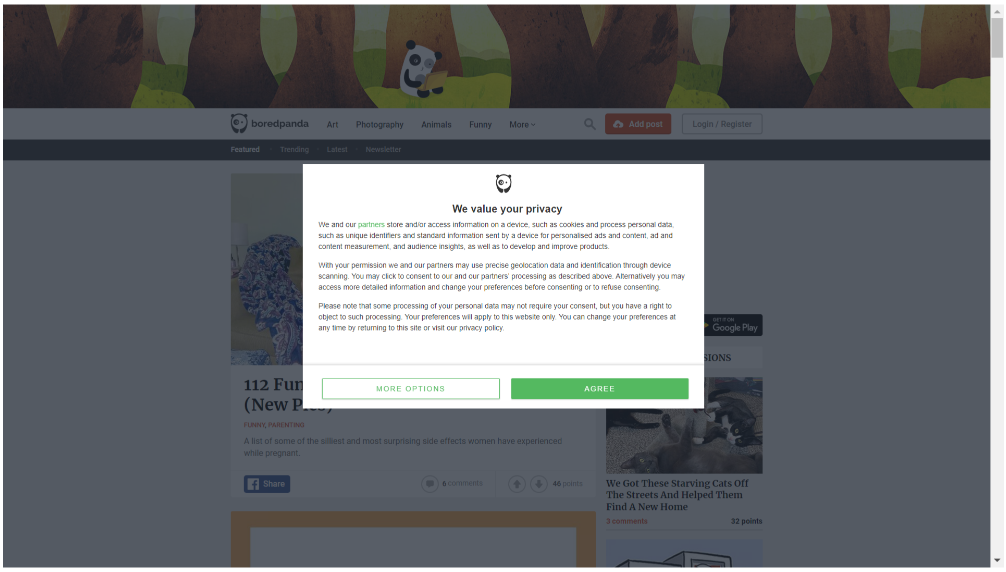
Send to Messenger

ensure you get the best experience on our website. By continuing to use this website without disabling cookies in your web browser, you are agreeing to our use of cookies. Read More. | I Agree

Figure 8: Here we see multiple popup cookie notices https://www.manilatimes.net/

An example of "choice cascade". The user is first presented the options to "Agree" or go to a "more options" page. After the user click the "more options" button there is information on which cookies that is used and "reject all" button next to a "accept all" button.

An example of "Widget inequality" On Figure 10 we have a case where the accept button is bright blue and larger making it more visible than the deny button. These color variations may seem trivial but has shown to have an effect. For example when Google tested 41 shades of blue to use on their links. They reported to have earned 200$ million on going with a more



We value your privacy

We and our partners store and/or access information on a device, such as cookies and process personal data, such as unique identifiers and standard information sent by a device for personalised ads and content, ad and content measurement, and audience insights, as well as to develop and improve products.

With your permission we and our partners may use precise geolocation data and identification through device scanning. You may click to consent to our and our partners' processing as described above. Alternatively you may access more detailed information and change your preferences before consenting or to refuse consenting.

Please note that some processing of your personal data may not require your consent, but you have a right to object to such processing. Your preferences will apply to this website only. You can change your preferences at any time by returning to this site or visit our privacy policy.

MORE OPTIONS

AGREE

112 Fun
(New P

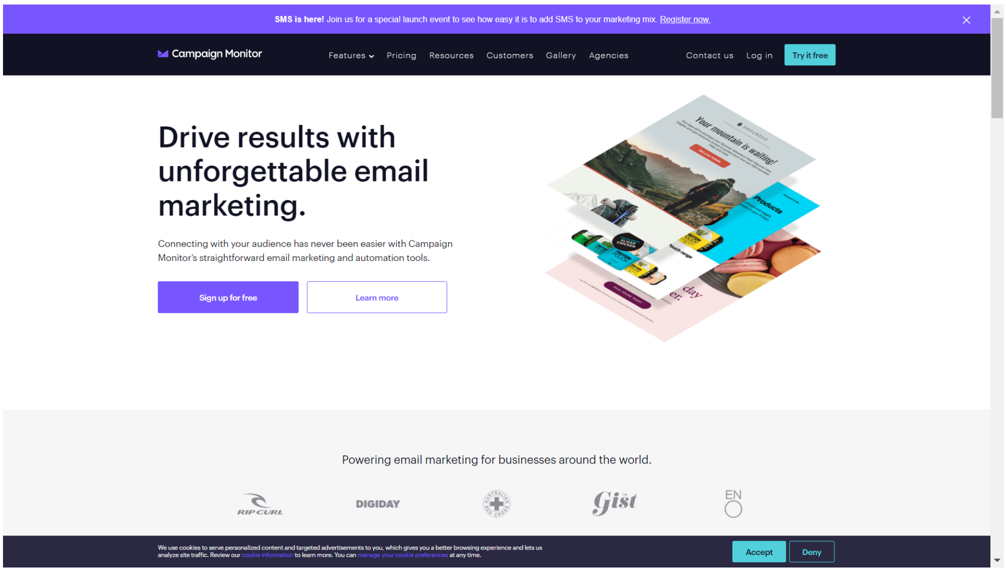
FUNNY, PARENTING

A list of some of the silliest and most surprising side effects women have experienced while pregnant.

We Got These Starving Cats Off The Streets And Helped Them Find A New Home

Figure 9: An example of choice cascade where the user have to go to 'more options' to reject their cookies https://boredpanda.com

purple blue coloring[21].



SMS is here! Join us for a special launch event to see how easy it is to add SMS to your marketing mix. Register now.

Drive results with unforgettable email marketing.

Connecting with your audience has never been easier with Campaign Monitor's straightforward email marketing and automation tools.

Sign up for free

Learn more

Powering email marketing for businesses around the world.

We use cookies to serve personalized content and targeted advertisements to you, which gives you a better browsing experience and lets us analyze site traffic. Review our cookie information to learn more. You can manage your cookie preferences at any time.

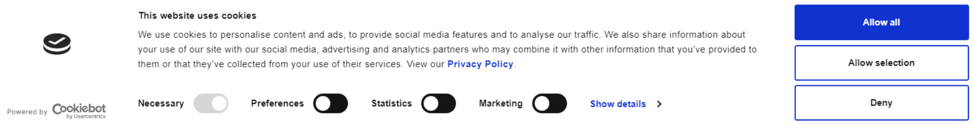
Accept

Deny

Figure 10: An example of widget inequality from https://campaignmonitor.com

[21] https://www.theguardian.com/technology/2014/feb/05/why-google-engineers-designers

An example of "Unlabeled slider" Figure 11 shows an example of an unlabeled slider pattern. slider that can be confusing as it is not clear what represents "on" and what represents "off".

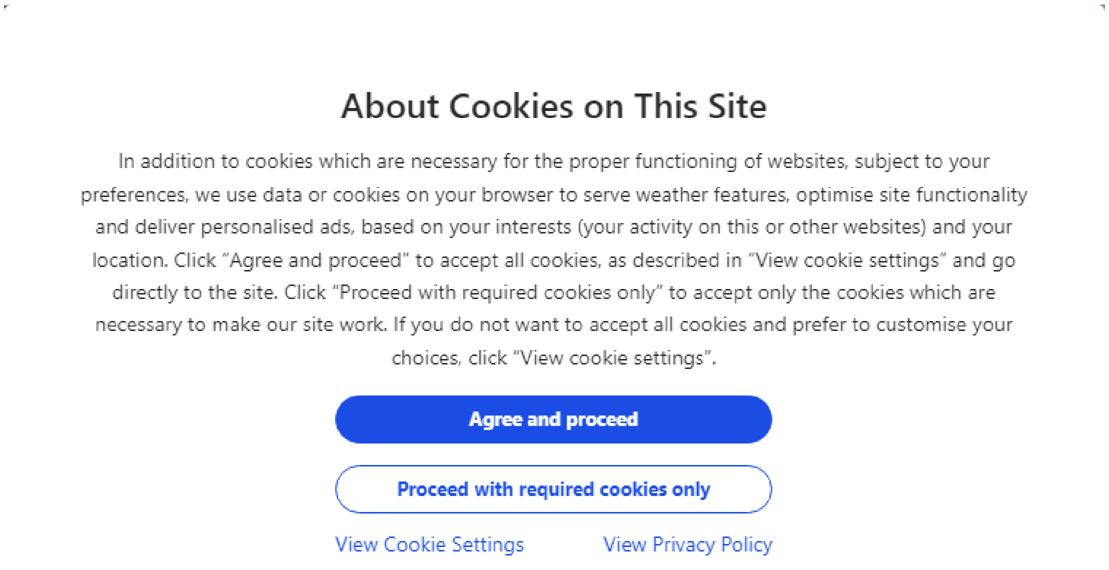


This website uses cookies

We use cookies to personalise content and ads, to provide social media features and to analyse our traffic. We also share information about your use of our site with our social media, advertising and analytics partners who may combine it with other information that you've provided to them or that they've collected from your use of their services. View our Privacy Policy.

Figure 11: An example of a unlabeled slider. From https://www.findhorn.org/

An example of "No antonyms" In Figure 12 we can see that the option for declining the CMP is represented in the option "Proceed with required cookies only". It is not clear whether it will decline the cookie policy, without reading the small text above the options. To avoid this dark pattern, it would be preferred they used for example "decline" and "accept" for the two options.

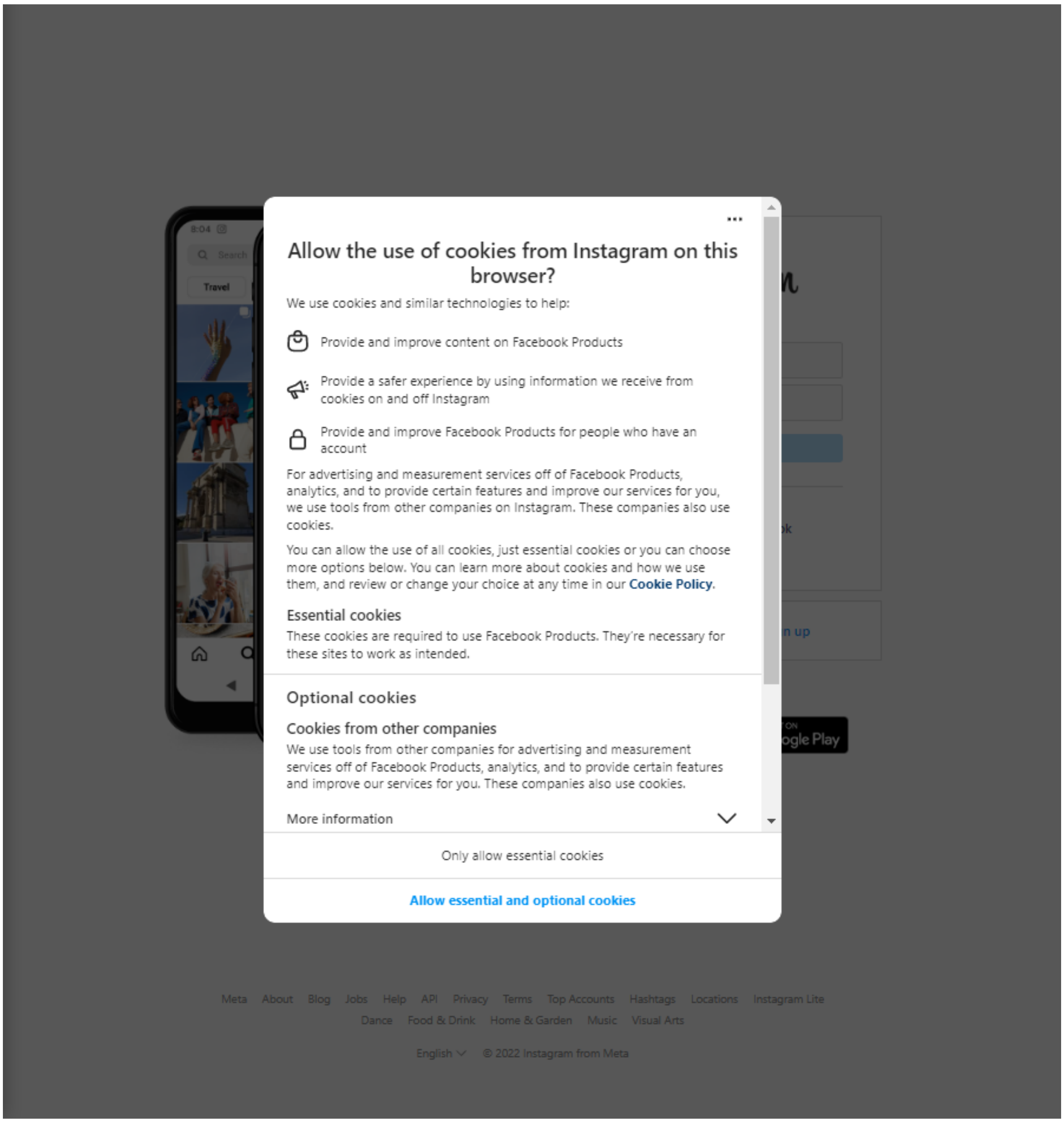


About Cookies on This Site

In addition to cookies which are necessary for the proper functioning of websites, subject to your preferences, we use data or cookies on your browser to serve weather features, optimise site functionality and deliver personalised ads, based on your interests (your activity on this or other websites) and your location. Click "Agree and proceed" to accept all cookies, as described in "View cookie settings" and go directly to the site. Click "Proceed with required cookies only" to accept only the cookies which are necessary to make our site work. If you do not want to accept all cookies and prefer to customise your choices, click "View cookie settings".

Figure 12: An example of no antonyms here from https://www.weather.com

The Instagram CMP example



Allow the use of cookies from Instagram on this browser?

We use cookies and similar technologies to help:

Provide and improve content on Facebook Products

Provide a safer experience by using information we receive from cookies on and off Instagram

Provide and improve Facebook Products for people who have an account

For advertising and measurement services off of Facebook Products, analytics, and to provide certain features and improve our services for you, we use tools from other companies on Instagram. These companies also use cookies.

You can allow the use of all cookies, just essential cookies or you can choose more options below. You can learn more about cookies and how we use them, and review or change your choice at any time in our Cookie Policy.

Essential cookies

These cookies are required to use Facebook Products. They're necessary for these sites to work as intended.

Optional cookies

Cookies from other companies

We use tools from other companies for advertising and measurement services off of Facebook Products, analytics, and to provide certain features and improve our services for you. These companies also use cookies.

8:04
Search
Travel
More information
Only allow essential cookies
Allow essential and optional cookies
n up
ogle Play
Meta About Blog Jobs Help API Privacy Terms Top Accounts Hashtags Locations Instagram Lite
Dance Food & Drink Home & Garden Music Visual Arts
English
© 2022 Instagram from Meta

Figure 13: Example from https://instagram.com