Cookie Banners, What's the Purpose?
Analyzing Cookie Banner Text Through a Legal Lens

Cristiana Santos*
Utrecht University

Arianna Rossi*
SnT, University of Luxembourg

Lorena Sánchez Chamorro
University of Luxembourg

Kerstin Bongard-Blanchy
University of Luxembourg

Ruba Abu-Salma
King's College London

ABSTRACT

A cookie banner pops up when a user visits a website for the first time, requesting consent to the use of cookies and other trackers for a variety of purposes. Unlike prior work that has focused on evaluating the user interface (UI) design of cookie banners, this paper presents an in-depth analysis of *what cookie banners say* to users to get their consent. We took an interdisciplinary approach to determining *what cookie banners should say*. Following the legal requirements of the ePrivacy Directive (ePD) and the General Data Protection Regulation (GDPR), we manually annotated around 400 cookie banners presented on the most popular English-speaking websites visited by users residing in the EU. We focused on analyzing the purposes of cookie banners and how these purposes were expressed (e.g., any misleading or vague language, any use of jargon). We found that 89% of cookie banners violated applicable laws. In particular, 61% of banners violated the purpose specificity requirement by mentioning vague purposes, including "*user experience enhancement*". Further, 30% of banners used positive framing, breaching the freely given and informed consent requirements. Based on these findings, we provide recommendations that regulators can find useful. We also describe future research directions.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**;

KEYWORDS

Usable security and privacy; cookie banners; ePD; GDPR; transparency; consent; purpose specification; framing

1 INTRODUCTION

When users residing in the EU browse the web, they encounter a large number of banners prompting them to "*click accept to consent to the use of cookies as described in our cookie policy*". These banners appear because, according to the ePrivacy Directive (ePD) [29] and the General Data Protection Regulation (GDPR) [35], website operators, regardless of where they are based, must inform users located in the EU of the collection of their personal data. User consent is needed only when cookies and similar tracking technologies are used for *unnecessary purposes*, such as advertising. Yet, website operators are required to be transparent and clearly explain the purpose of the use of cookies even if these cookies are necessary for the website to operate (i.e., *necessary purposes*), such as authentication and security [54]. Without knowing the specific purpose(s)

of the use of cookies, users cannot decide whether to consent to the collection of their personal data.

While prior studies have evaluated the user interface (UI) design of cookie banners to determine how design influences users' consent decisions [9, 37, 38, 49, 71], little attention has been given to the textual elements of banners, which could likewise unlawfully steer users toward particular choices. There has been no in-depth analysis of what cookie banners say (and should say) and, hence, website owners can request user consent at their own discretion [7, 42]. They may employ technical jargon [67, 71], vague and ambiguous language, and positive or negative framing. The resulting lack of transparency undermines users' ability to understand why their data is collected and what risks are involved (Recital 39 GDPR, [20, 50, 55]), hindering informed decisions and unlawfully nudging users toward giving their consent [12, 13, 19, 23, 34, 46].

As part of ongoing work aimed at promoting transparency, lawfulness, and user-friendliness of cookie banner UIs, this paper addresses the following: *What are the purposes of cookie banners, and how are these purposes expressed?* We combine expertise in data protection law, human-computer interaction (HCI), linguistics, and computer science to evaluate whether cookie banner text complies with the ePD and GDPR legal requirements concerning purposes and consent. To this end, we collected a corpus of about 1,300 cookie banners presented on the most popular English-speaking websites visited by users residing in the EU. We randomly selected and manually coded 407 of these banners, finding that 89% of banners violated at least one legal requirement concerning processing purposes and consent. In particular, 20% of banners violated the purpose availability requirement as they did not mention any processing purpose. More than 50% of banners unlawfully mentioned the widely used but rather vague purpose: "*user experience enhancement*". Further, 30% used positive framing, breaching the freely given and informed consent requirements. Our findings suggest that many cookie banners use unlawful and questionable practices to obtain user consent.

Our contributions are as follows:

(1) We analyzed relevant legal sources and extracted six legal requirements explaining how cookie banner text should describe purposes of data collection and use;
(2) We mapped legal requirements and their violations to observable linguistic features;
(3) We empirically demonstrated that the wording used in cookie banners did not comply with ePD and GDPR;
(4) We provided a set of recommendations that regulators and policymakers can find useful.

---

*Both authors contributed equally to this research.

2 RELATED WORK

Within the security and privacy domain, prior work has sought to assess the *comprehensibility* of alert messages and warnings [4, 25–27, 30, 31, 66, 68], privacy policies [69, 72], contractual terms [10, 48], browser disclosures [1, 2], and security and privacy advice on the web [58]. These studies have generally shown that privacy-related text is long and difficult to understand since it is overly complex and full of legalese [61], casting doubt on how informed users are when they make decisions with regard to the collection of their personal data. Moreover, the typical use of vague quantifiers (e.g., 'certain', 'some') [59] and modality markers (e.g., 'may', 'might') in privacy-related text makes it challenging for users to assess the data collection practices of organizations [57]. Further, scholars and regulators have shown that certain linguistic strategies may influence users' online decisions by toying with users' emotions [36]; e.g., shame [44], guilt [18], blame [19], or fear [15].

With regard to the analysis of cookie banner text, Utz et al. [71] showed that purposes were expressed in generic terms in almost half of the analyzed banners (e.g., "*to improve user experience*") and were unspecified in one out of six banners. In their empirical work, Hausner et al. [39] argued that positive framing (e.g., "*Yes, I am happy!*") could be used to nudge users toward giving their consent, whereas configuration options used to refuse or manage consent were expressed neutrally in banners. Similarly, Kampanos et al. [41] showed that most banners presented "affirmative" options that could nudge users toward consenting to tracking, whereas options like "*Read more*" and "*I do not accept*" were less prevalent. The remaining literature on how users' consent to the use of cookies is requested exclusively focused on the UI design of cookie banners [9, 37, 38, 49, 71].

We build on prior work [37, 45, 62] and employ user-centric transparency criteria [55] to establish a benchmark that can be used to assess whether consent banner text is ePD- and GDPR-compliant. We use an inductive approach to investigating transparency issues (ambiguity, vagueness, technical jargon, misleading statements, and framing) through expert annotation of banner text.

3 METHODOLOGY

Data collection. Using the Tranco list [43], we created a dataset of about 1,300 cookie banners presented on the most visited English-speaking websites by users residing in the EU (in March 2020). We used the Polyglot library [5] to detect the website language and then scrape English-speaking websites. The resulting set included both European and non-European domains. To *scrape websites*, we used OpenWPM [28], a web privacy measurement framework based on Selenium [64]. It allowed full-page rendering before analysis and enabled taking screenshots of specific site elements.

To *detect cookie banners*, we followed three steps: segmentation, scoring, and tree traversal. First, we segmented webpages into small segments and built a segment tree [44] based on the segments' HTML tag and text. Second, we assigned a score to each segment based on its inner text using a vocabulary set that we created by analyzing cookie banner content. We ranked tree leaf segments according to their scores. Third, we used the highest-scoring segments to traverse our segment tree. We performed bottom-up and top-down tree traversals. We captured HTML elements that contained

cookie banners. To reduce false positives (i.e., websites with no banners), we used the segment scores to decide whether a cookie banner existed based on a threshold we set. We then manually filtered out any remaining false positives.

Legal requirements applicable to banner text. We analyzed various legal documents (ePD, GDPR, case law, regulatory decisions, and guidelines of non-binding sources like the European Data Protection Board (EDPB) and Data Protection Authorities (DPAs)) and extracted six legal requirements applicable to cookie banner text [62] (the requirements and their violations are described in Table 1 and further detailed in Table 3). The first two requirements (R1, R2) mandate that the *purposes* of personal data processing should be described in an explicit and specific manner (Article 5(1)(b) GDPR) [6, 33]. The other requirements (R3, R4, R5, R6) relate to the validity of *consent*, which should be intelligible, expressed in clear and plain language, freely given, and informed (Articles 4(11), 7(2)(4) GDPR) [13]. Without explicit and specific purposes and without valid consent, websites may be found to infringe the GDPR's principle of lawfulness (Article 6 (1)(a)), which would render any subsequent data processing unlawful and subject to heavy fines (Article 83 (5)(a) GDPR).

Legal Requirement	Violation
R1 Purpose explicitness	
R.1.1 Availability	Absence of purpose [32, 50, 55]
R.1.2 Unambiguity	Ambiguous intent [6]
R.1.3 Shared common understanding	Inconsistent purposes [6].
R2 Purpose specificity	Vague or general purposes [6, 55]
R3 Intelligible consent	
R.3.1 Non-technical terms	Presence of technical jargon [20, 24, 55]
R.3.2 Conciseness	Prolixity [14, 20, 24]
R4 Consent with clear and plain language	
R.4.1 Straightforward statements	Misleading expressions [20, 22, 24, 55, 56]
R.4.2 Concreteness	Indefinite qualifiers [13, 55]
R5 Freely given consent	Pressure to provide consent [13, 19, 34, 55]
R6 Informed consent	Absence of essential information about data processing [20, 22]

Table 1: A description of six legal requirements applicable to cookie banner text.

Banner text coding. The legal requirements we extracted (and their violations) were mapped to codes that we used to capture observable textual elements. Based on annotating our cookie banner set (described next), we identified five main codes: purpose of banners (see Table 2), framing, misleading language, vagueness, and technical jargon.

We – a multidisciplinary team of five researchers with diverse expertise in data protection law, computer science, linguistics, and HCI – iteratively coded a set of 150 banners (three iterations in total) using MAXQDA[1]. We weekly met to develop our codebook until we reached good interrater agreement calculated for each pair of coders. Agreement ranged from 0.71 to 0.8 (Cohen's kappa coefficient) for all six pairs. We used the final codebook to annotate 407 banners that we randomly selected from the dataset we created. We analyzed the first layer of cookie banners without considering

---

[1]MAXQDA: https://www.maxqda.com/

the second layer where the cookie policy can usually be found. Our choice was motivated by prior studies showing that many users simply disregard the second layer of consent requests [47, 70, 71] and make decisions exclusively based on the first layer. Additionally, transparency requirements mandate that the first layer of banners should give users a clear overview of data collection and processing [55] and that the second layer should be consistent with the first one [55].

4 FINDINGS AND DISCUSSION

We present the findings of analyzing the text of 407 cookie banners. We also discuss the compliance of these banners with the six legal requirements we extracted. We found that 80% of banners explained the purpose of data collection and processing. More than one-half of banners did not use misleading wording to explain purposes of processing, and about two-thirds of banners did not employ framing. Further, around 90% of banners did not use vague language and technical jargon. However, 89% of banners violated at least one out of six legal requirements considered in this study, as detailed below.

Absence of purposes. 20% of banners did not mention the purpose of data processing although several DPAs [8, 40, 51, 52] mandate transparent disclosure of purposes even for *strictly necessary* cookies that do not require user consent. Hiding the reason for data processing violated the purpose availability (R1.1) and informed consent (R6) requirements.

Categories of purpose. We identified eight different categories of purpose described in cookie banner text. We explain these categories in order of occurrence (from the most to the least recurring): user experience enhancement, analytics, advertising, custom content, service offering, essential functionalities, social media features, and profiling. The exact distribution of categories is shown in Table 2. We identified our categories by annotating banner text. Future work can map these categories to the ones created by several DPAs [13, 24, 51].

Wording used to describe purposes. We found a wide range of terms used to describe the purpose of data collection and processing, which often did not clearly match the eight categories we identified. For example, *Advertising* was also referred to as marketing (content), targeted/tailored ads, ad(s) delivery/personalization/measurement, and promotional offers. *Essential functionalities* were referred to as basic functions/functioning, operation of website, optimal website provision, and user preferences. Hence, the use of different terms to describe the same purpose did not comply with the common understanding (R1.3) requirement and raised the question about whether users could map different terms to the same concept.

Most often mentioned purposes. The *user experience enhancement* purpose appeared in 61% of banners that explained the purpose(s) of data processing. However, it was unclear how cookies improved the user experience of website visitors. It was also unclear whether certain cookies were necessary for the website to operate properly (e.g., adapting the presentation of website content to the user screen size [53, 54]). DPAs and the EDPB [6, 24] explain that such wording should not be used due to its vagueness and ambiguity. Thus, almost one-half of banners breached the specificity (R2), unambiguity (R1.2), and plain language (R4) requirements.

Analytics was the second most mentioned purpose, appearing in 33% of banners that described the purpose(s) of data processing.

Since data could be collected and processed by first and/or third parties in aggregate or anonymously, the chosen wording violated the specificity (R2) requirement. We recall that third-party analytics entail the risk of cookie synchronization between different websites and, thus, that of profiling [21].

Profiling was mentioned in 8% of all banners that explained the purpose(s) of data processing, but it was rarely mentioned explicitly. More often, based on our interpretation of banner text, we recognized implicit mentions of the profiling purpose: "*[...] combine it with other information that you've provided to them or that they've collected from your use of their services*", violating the unambiguity (R1.2) and specificity of purposes (R2) requirements.

Multiple purposes. Some banners described a host of different processing purposes using a single sentence. For example, "*[...] to derive insights about the audiences who saw ads and content*" bundles up three purposes – analytics, advertising, and profiling – into one sentence. This violated the requirement of purpose specificity (R2) and possibly that of unambiguity (R1.2).

Misleading statements. Misleading statements used to describe purposes were identified in 42% of banners. They included descriptions that were vague, confusing, and ambiguous; were framed positively or negatively; instilled false beliefs; or concealed important information (i.e., deceiving the user [17]). Examples included the following: "*We use cookies that do not contain personal data about you in order to personalize content and ads*". Some statements were misleading due to framing, vagueness, or use of technical jargon.

Framing of purposes. Positive framing was used in 30% of banners describing purposes through the use of superlatives like 'best' and 'most optimal' (e.g., "*We use cookies to deliver the best possible web experience*"). Positive framing was mainly used to describe the *user experience enhancement* purpose, claiming that cookies optimized website performance or improved user experience. Highlighting the positive aspects of consenting to cookie processing provided a partial view, making users pay less attention to other aspects that could be deemed negative (e.g., targeted advertising) [17] but key to making informed decisions. Thus, positive framing violated the freely given (R5) and informed consent (R6) requirements.

Negative framing was only used in 2% of banners, mainly to warn users of the loss of functionalities if users did not consent to the use of cookies; e.g., "*If you're not happy with this, we won't set these cookies but some nice features of the site may be unavailable*". When choices are framed negatively, they may put pressure on users by exploiting loss aversion [3] and nudge them toward consenting [19], especially when it is unclear which functionalities will be lost. Our previous study [15] showed that people may develop wrong mental models of the consequences of (not) consenting to data collection and processing. Therefore, both positive and negative framing may nudge users toward complying with the service provider's wishes [34], violating the freely given consent (R5) requirement.

Necessary vs. unnecessary cookies. Additionally, we found that cookie banner text did not explain the difference between necessary and unnecessary cookies. Therefore, non-expert users may be misled to believe that all cookies are necessary for websites to operate properly [11, 65, 67]. For example, "*This website or its third-party tools use cookies, which are necessary to its functioning and required to achieve the purposes illustrated in the cookie policy*"

Purpose category	Occur.	Example
User experience enhancement	61%	*[...] uses cookies to ensure you get the best experience on our website.*
Analytics	33%	*We use cookies to analyze our traffic.*
Advertising	27%	*Our site is using cookies for advertising purposes.*
Custom content	22%	*This site uses cookies to help personalize content.*
Service provision	15%	*This website uses cookies to provide its services.*
Essential functionalities	14%	*[...] uses cookies to ensure a comprehensive presentation and functionality of the website.*
Social media features	11%	*We use cookies to provide social media features.*
Profiling	8%	*This site use[s] profiling cookies to send you advertising based on your preferences.*

Table 2: A description of eight data processing purposes we identified based on analyzing cookie banner text.

is a misleading statement because it claims that all cookies are key to proper website functionality, making users' consent uninformed and violating the specificity (R2), unambiguity (R1.2), clear and plain language (R4), and informed consent (R6) requirements.

Vagueness. 11% of banners used vague terms to describe purposes: "*We may share information about your use of the site with third parties; we may use cookies*". Vagueness is *misleading* when it leaves individuals uncertain about the intended meaning of an expression [57, 61], in particular whether cookies are used and whether data is shared with other parties. The *user experience enhancement* and *service provision* purposes – mentioned cumulatively in more than one-half of banners – could also be deemed misleading if proven untrue, especially that vague language does not explain how accepting cookies is beneficial to users [6, 20, 24, 46].

Technical jargon and prolixity. The description of purposes in 9% of banners contained technical jargon, including "*derive insights about the audiences*", "*retargeting cookies*", and "*Google Analytics*", breaching the requirements of intelligibility (R3) and straightforward statements (R4.1). While examples or explanations can clarify technical terms, they can still breach the conciseness (R3.2) requirement given the limited size of cookie banners. Empirical studies should be conducted to assess the comprehensibility of technical terms by different audiences, although some prior studies [16, 63, 67, 71] have shown that the use of technical jargon could leave users in a vulnerable state.

Recommendations. Based on our findings, we provide recommendations that regulators and policymakers can find useful.

Standardization of purposes. Given that people may have different interpretations of the same text, we argue that most consensual purposes (e.g., advertising, statistical analysis, social media features, personalization) and their labels should be standardized, following Privacy by Design [60]. The violations we identified in this work are rooted in the fact that websites can describe purposes at their own discretion. The EDPB, DPAs, and standard committees should standardize purpose categories to minimize legal uncertainty and simplify data processing operations.

Requirements for describing purposes. The legal requirements that currently exist are generic and difficult to operationalize. Therefore, there is a need for a set of requirements that can be used to help define purposes of data processing. We also argue that creating a blacklist of illegal purposes could improve the current situation. A simple nomenclature like '*necessary*' and '*unnecessary*' cookies could improve users' comprehensibility of cookie banner text and help users make informed decisions, provided that details about purposes of data processing can be found in cookie policies. Yet, from a user's point of view, we believe the best solution would

be managing cookie consent options pre-emptively at the browser level or through an automated browser extension.

Language tensions. Best practices and examples of how to clearly refer to data privacy concepts while remaining concise should also be made available to website providers, since attempts to comply with the explicitness, specificity, and plain language legal requirements could also lead to prolixity.

5 CONCLUSION AND FUTURE WORK

In this work, we analyzed the text describing the data processing purposes of 407 cookie banners. We found that 89% of banners violated at least one legal requirement. 20% of banners did not mention any purpose although purpose disclosure is legally mandated. Notably, 67% of banners violated the specificity requirement, and 61% unlawfully provided a vague purpose: "*user experience enhancement*". 31% of banners used framing, breaching the freely given consent requirement. Other identified issues included misleading statements, technical jargon, and vagueness.

We argue that the identified violations do not allow users to be aware of the scope, consequences, and risks (Recital 39 GDPR) of consenting to storing cookies on their devices, especially the privacy-invasive ones and, as a result, breaching the principle of transparency that governs personal data processing (Article 5(1)(a) GDPR). It is not only difficult for lay users to understand cookie banner text, but also experts may find it challenging to parse banner text and map it to relevant legal requirements. This suggests that, besides necessary standardization, a purpose-based consent may neither be user-friendly nor feasible and, hence, we argue that unnecessary cookies should be rejected by default.

We will build on this work and conduct user studies to empirically evaluate the comprehensibility of different textual elements of banners. We will investigate the influence of positive and negative framing on users' choices. We will also seek to create a taxonomy of commonly understandable purposes to facilitate comprehensibility, comparability, and compliance checking. It would also be useful to use natural language processing (NLP) to automate, for example, the detection of misleading text as well as use sentiment analysis to identify positive and negative framing.

REFERENCES

[1] Ruba Abu-Salma. 2020. *Designing User-Centered Privacy-Enhancing Technologies.* Ph.D. Dissertation. UCL (University College London).

[2] Ruba Abu-Salma and Benjamin Livshits. 2020. Evaluating the End-User Experience of Private Browsing Mode. In *Proc. Conference on Human Factors in Computing Systems (CHI)*.

[3] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, et al. 2017. Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online. *ACM Computing Surveys (CSUR)* 50, 3 (Aug 2017), 1–41. https://doi.org/10.1145/3054926

[4] Devdatta Akhawe and Adrienne Porter Felt. 2013. Alice in Warningland: A Large-Scale Field Study of Browser Security Warning Effectiveness. In *Proc. USENIX Security Symposium*.

[5] Rami Al-Rfou. 2015. Polyglot Natural Language Processor. (2015). http://polyglot.readthedocs.org/.

[6] Article 29 Working Party. 2013. *Opinion 03/2013 on Purpose Limitation (WP203).* Technical Report.

[7] Article 29 Working Party. 2018. *Guidelines on Consent under Regulation 2016/679 (WP259 rev.01).* Technical Report. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051.

[8] Autorité de Protection des Données (Data Protection Commission). 2020. *Guidance on Cookies and Other Tracking Technologies.* Technical Report. https://www.autoriteprotectiondonnees.be/cookies.

[9] Jan M. Bauer, Regitze Bergstrøm, and Rune Foss-Madsen. 2021. Are You Sure, You Want A Cookie? – The Effects of Choice Architecture on Users' Decisions About Sharing Private Online Data. (2021). https://doi.org/10.1016/j.chb.2021.106729

[10] Ingolf Becker, Alice Hutchings, Ruba Abu-Salma, Ross Anderson, Nicholas Bohm, Steven J Murdoch, M Angela Sasse, and Gianluca Stringhini. 2017. International Comparison of Bank Fraud Reimbursement: Customer Perceptions and Contractual Terms. *Journal of Cybersecurity* 3, 2 (2017), 109–125.

[11] Omri Ben-Shahar and Carl E Schneider. 2011. The Failure of Mandated Disclosure. *University of Pennsylvania Law Review* (2011), 647–749.

[12] European Data Protection Board. 2018. EDPS Opinion 8/2018 on the Legislative Package "A New Deal for Consumers". (2018). https://edps.europa.eu/sites/edp/files/publication/18-10-05_opinion_consumer_law_en.pdf.

[13] European Data Protection Board. 2020. *Guidelines 05/2020 on Consent under Regulation 2016/679.* Technical Report. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_202005_consent_en.pdf.

[14] European Data Protection Board. 2020. *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default.* Technical Report. https://edpb.europa.eu/sites/edpb/files/files/file1/edpb_guidelines_201904_dataprotection_by_design_and_by_default_v2.0_en.pdf.

[15] Kerstin Bongard-Blanchy, Arianna Rossi, Salvador Rivas, Sophie Doublet, Vincent Koenig, and Gabriele Lenzini. 2021. "I am definitely manipulated, even when I am aware of it. It's ridiculous!" - Dark Patterns from the End-User Perspective. *Proc. ACM Conference on Designing Interactive Systems (DIS)* (2021). https://doi.org/10.1145/3461778.3462086

[16] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. 2016. Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. In *Proc. Privacy Enhancing Technologies (PoPETs).* 237–254.

[17] David M. Boush, Marian Friestad, and Peter Wright. 2009. *Deception in the Marketplace: The Psychology of Deceptive Persuasion and Consumer Self-Protection.* Rouledge.

[18] Harry Brignull. 2010. Dark Patterns. (2010). https://www.darkpatterns.org.

[19] Régis Chatellier, Geoffrey Delcroix, Estelle Hary, and Camille Girard-Chanudet. 2019. Shaping Choices in the Digital World. (2019). https://linc.cnil.fr/sites/default/files/atoms/files/cnil_ip_report_06_shaping_choices_in_the_digital_world.pdf.

[20] Commission Nationale de l'Informatique et des Libertés (French Data Protection Authority). 2019. Deliberation of the Restricted Committee SAN-2019-001 of 21 January 2019 Pronouncing a Financial Sanction against GOOGLE LLC. (2019). https://www.cnil.fr/sites/default/files/atoms/files/san-2019-001.pdf.

[21] Commission Nationale de l'Informatique et des Libertés. 2020. CNIL: Délibération de la Formation Restreinte n° SAN-2020-008 du 18 Novembre 2020 Concernant la Société. (2020). https://www.legifrance.gouv.fr/cnil/id/CNILTEXT000042563756.

[22] Commission Nationale de l'Informatique et des Libertés (French Data Protection Authority). 2020. Deliberation of the Restricted Committee n° SAN-2020-013 of 7 December 2020 Concerning AMAZON EUROPE CORE. (2020). https://www.cnil.fr/sites/default/files/atoms/files/deliberation_of_restricted_committee_san-2020-013_of_7_december_2020_concerning_amazon_europe_core.pdf.

[23] Commission Nationale de l'Informatique et des Libertés (French Data Protection Authority). 2020. On the Practical Procedures for Collecting the Consent Provided for in Article 82 of the French Data Protection Act, Concerning Operations of Storing or Gaining Access to Information in the Terminal Equipment of a User (Recommendation "Cookies and Other trackers"). (2020). https://www.cnil.fr/sites/default/files/atoms/files/draft_

[24] Agencia Española de Protección de Datos. 2020. *Guía Sobre El Uso De Las Cookies.* Technical Report. https://www.aepd.es/sites/default/files/2020-07/guia-cookies.pdf

[25] Rachna Dhamija, J. Doug Tygar, and Marti Hearst. 2006. Why Phishing Works. In *Proc. Conference on Human Factors in Computing Systems (CHI)*.

[26] Serge Egelman, Lorrie Faith Cranor, and Jason Hong. 2008. You've Been Warned: An Empirical Study of the Effectiveness of Web Browser Phishing Warnings. In *Proc. Conference on Human Factors in Computing Systems (CHI)*.

[27] Serge Egelman and Stuart Schechter. 2013. The Importance of Being Earnest [In Security Warnings]. In *Proc. Conference on Financial Cryptography and Data Security*.

[28] Steven Englehardt, Chris Eubank, Peter Zimmerman, Dillon Reisman, and Arvind Narayanan. 2015. OpenWPM: An Automated Platform for Web Privacy Measurement. *Manuscript. March* (2015).

[29] ePD-09 2009. Directive 2009/136/EC of the European Parliament and of the Council. (2009). https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:337:0011:0036:En:PDF.

[30] Adrienne Porter Felt, Alex Ainslie, Robert W. Reeder, Sunny Consolvo, Somas Thyagaraja, Alan Bettes, Helen Harris, and Jeff Grimes. 2015. Improving SSL Warnings: Comprehension and Adherence. In *Proc. Conference on Human Factors in Computing Systems (CHI)*.

[31] Adrienne Porter Felt, Robert W. Reeder, Hazim Almuhimedi, and Sunny Consolvo. 2014. Experimenting At Scale With Google Chrome's SSL Warning. In *Proc. Conference on Human Factors in Computing Systems (CHI)*.

[32] Hamburg Commissioner for Data Protection and Freedom of Information. 2021. *Cross-Border Check: Consents to Websites of Media Companies Are Mostly Ineffective - Improvements Are Required.* Technical Report. https://datenschutz-hamburg.de/pressemitteilungen/2021/06/2021-06-30-medienwebsites.

[33] Imane Fouad, Cristiana Santos, Feras Al Kassar, Nataliia Bielova, and Stefano Calzavara. 2020. On Compliance of Cookie Purposes with the Purpose Specification Principle. In *Proc. International Workshop on Privacy Engineering (IWPE).* https://hal.inria.fr/hal-02567022.

[34] Frobrukerrådet. 2018. Deceived by Design: How Tech Companies Use Dark Patterns to Discourage Us from Exercising Our Rights to Privacy. (2018). https://www.forbrukerradet.no/undersokelse/no-undersokelsekategori/deceived-by-design.

[35] GDPR 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/ec (General Data Protection Regulation). (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:32016R0679.

[36] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The Dark (Patterns) Side of UX Design. In *Proc. Conference on Human Factors in Computing Systems (CHI).* Article 534, 14 pages.

[37] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective. In *Proc. Conference on Human Factors in Computing Systems (CHI).* Article 172, 1-18 pages. https://doi.org/10.1145/3411764.3445779

[38] Paul Graßl, Hanna Schraffenberger, Frederik Zuiderveen Borgesius, and Moniek Buijzen. 2021. Dark and Bright Patterns in Cookie Consent Requests. *Journal of Digital Social Research* 3, 1 (Feb 2021), 1–38. https://doi.org/10.33621/jdsr.v3i1.54

[39] Philip Hausner and Michael Gertz. 2021. Dark Patterns in the Interaction with Cookie Banners. In *Proc. CHI Workshop on "What Can CHI Do About Dark Patterns?"*.

[40] Hellenic Data Protection Authority (HDPA). 2020. *Recommendations on Controllers' Compliance with the Specific Legislation on Electronic Communications.* Technical Report. [English translation by Katerina Tassi] https://iapp.org/media/pdf/resource_center/Greek_DPA_Cookie_Guidance_IAPP.pdf.

[41] Georgios Kampanos and Siamak F. Shahandashti. 2021. Accept All: The Landscape of Cookie Banners in Greece and the UK. (2021). arXiv:cs.CR/2104.05750

[42] Farzaneh Karegar, John Sören Pettersson, and Simone Fischer-Hübner. 2020. The Dilemma of User Engagement in Privacy Notices: Effects of Interaction Modes and Habituation on User Attention. *ACM Trans. Priv. Secur.* 23, Article 5 (2020).

[43] Victor Le Pochat, Tom Van Goethem, Samaneh Tajalizadehkhoob, Maciej Korczyński, and Wouter Joosen. 2019. Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation. In *Proc. Network and Distributed System Security Symposium (NDSS)*.

[44] Arunesh Mathur, Gunes Acar, Michael Friedman, Elena Lucherini, Jonathan R. Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark Patterns at Scale: Findings from a Crawl of 11K Shopping Websites. *Proc. ACM on Human-Computer Interaction*.

[45] Célestin Matte, Cristiana Santos, and Nataliia Bielova. 2020. Do Cookie Banners Respect My Choice? Measuring Legal Compliance of Banners from IAB Europe's Transparency and Consent Framework. In *Proc. IEEE Symposium on Security and Privacy.* 791–809.

[46] Célestin Matte, Cristiana Santos, and Nataliia Bielova. 2020. Purposes in IAB Europe's TCF: Which Legal Basis and How Are They Used by Advertisers?. In *Annual Privacy Forum (APF) (Lecture Notes in Computer Science)*. https://hal.inria.fr/hal-02566891.

[47] Aleecia M. McDonald and Lorrie Faith Cranor. 2009. The Cost of Reading Privacy Policies. *Journal of Law and Policy for the Information Society* (2009).

[48] Steven J. Murdoch, Ingolf Becker, Ruba Abu-Salma, Ross Anderson, Nicholas Bohm, Alice Hutchings, M Angela Sasse, and Gianluca Stringhini. 2016. Are Payment Card Contracts Unfair?. In *Proc. International Conference on Financial Cryptography and Data Security*. Springer, 600–608.

[49] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns After the GDPR: Scraping Consent Pop-ups and Demonstrating Their Influence. In *Proc. Conference on Human Factors in Computing Systems (CHI)*. 1–13. https://doi.org/10.1145/3313831.3376321

[50] Court of Justice of the European Union. 2019. Judgment in Case C-673/17 Bundesverband der Verbraucherzentralen und Verbraucherverbände – Verbraucherzentrale Bundesverband eV v Planet49 GmbH. (2019). http://curia.europa.eu/juris/documents.jsf?num=C-673/17.

[51] Information Commissioner's Office. 2019. *Guidance on the Use of Cookies and Similar Technologies*. Technical Report. https://ico.org.uk/media/for-organisations/guide-to-pecr/guidance-on-the-use-of-cookies-and-similar-technologies-1-0.pdf.

[52] Office of the Data Protection Ombudsman. 2019. *Guidelines on Confidential Communications*. Technical Report. https://www.kyberturvallisuuskeskus.fi/en/our-activities/regulation-and-supervision/confidential-communications.

[53] Article 29 Data Protection Working Party. 2013. *Working Document 02/2013 Providing Guidance on Obtaining Consent for Cookies*. Technical Report. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2013/wp208_en.pdf.

[54] Article 29 Working Party. 2012. *Opinion 04/2012 on Cookie Consent Exemption (WP 194)*. Technical Report. https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2012/wp194_en.pdf.

[55] Article 29 Working Party. 2018. *Guidelines on Transparency Under Regulation 2016/679 (WP260)*. Technical Report.

[56] Garante per la Protezione Dei Dati Personali. 2020. Provvedimento Correttivo e Sanzionatorio nei Confronti di Tim S.p.A. [9256486]. (2020). https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9256486?

[57] Irene Pollach. 2005. A Typology of Communicative Strategies in Online Privacy Policies: Ethics, Power and Informed Consent. *Journal of Business Ethics* 62, 3 (2005), 221.

[58] Elissa M. Redmiles, Noel Warford, Amritha Jayanti, Aravind Koneru, Sean Kross, Miraida Morales, Rock Stevens, and Michelle L. Mazurek. 2020. A Comprehensive Quality Evaluation of Security and Privacy Advice on the Web. In *Proc. USENIX Security Symposium*. 89–108.

[59] Joel R. Reidenberg, Jaspreet Bhatia, Travis D. Breaux, and Thomas B. Norton. 2016. Ambiguity in Privacy Policies and the Impact of Regulation. *Journal of Legal Studies* 45, 2 Supplement (2016), S163–S190.

[60] Nancy Richter, Paul Jackson, and Thomas Schildhauer. 2018. *Entrepreneurial Innovation and Leadership: Preparing for a Digital Future*. Springer.

[61] Arianna Rossi, Rossana Ducato, Helena Haapio, and Stefania Passera. 2019. When Design Met Law: Design Patterns for Information Transparency. *Droit de la Consommation = Consumenterecht: DCCR* 122–123 (2019), 79–121.

[62] Cristiana Santos, Nataliia Bielova, and Célestin Matte. 2020. Are Cookie Banners Indeed Compliant with the Law? Deciphering EU Legal Requirements on Consent and Technical Means to Verify Compliance of Cookie Banners. *Technology and Regulation* (2020), 91–135. https://doi.org/10.26116/techreg.2020.009

[63] Cristiana Santos, Aldo Gangemi, and Mehwish Alam. 2017. Detecting and Editing Privacy Policy Pitfalls on the Web. In *Proc. Workshop on Technologies for Regulatory Compliance*.

[64] Selenium. 2020. Selenium Browser Automation Tool. (2020). https://www.selenium.dev/.

[65] Daniel J. Solove. 2013. Privacy Self-management and the Consent Dilemma. *Harward Law Review* 126, 1880 (2013).

[66] Andreas Sotirakopoulos, Kirstie Hawkey, and Konstantin Beznosov. 2011. On the Challenges in Usable Security Lab Studies: Lessons Learned From Replicating a Study on SSL Warnings. In *Proc. Symposium On Usable Privacy and Security (SOUPS)*.

[67] Joanna Strycharz, Edith Smit, Natali Helberger, and Guda van Noort. 2021. No to Cookies: Empowering Impact of Technical and Legal Knowledge on Rejecting Tracking Cookies. *Computers in Human Behavior* 120 (2021), 106750. https://doi.org/10.1016/j.chb.2021.106750

[68] Joshua Sunshine, Serge Egelman, Hazim Almuhimedi, Neha Atri, and Lorrie Faith Cranor. 2009. Crying Wolf: An Empirical Study of SSL Warning Effectiveness. In *Proc. USENIX Security Symposium*.

[69] Janice Y. Tsai, Serge Egelman, Lorrie Cranor, and Alessandro Acquisti. 2011. The Effect of Online Privacy Information on Purchasing Behavior: An Experimental Study. *Information Systems Research* (2011).

[70] Joseph Turow, Michael Hennessy, and Nora Draper. 2018. Persistent Misperceptions: Americans' Misplaced Confidence in Privacy Policies, 2003–2015. *Journal of Broadcasting & Electronic Media* 62, 3 (2018), 461–478. https://doi.org/10.1080/08838151.2018.1451867

[71] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed Consent: Studying GDPR Consent Notices in the Field. In *Proc. Conference on Computer and Communications Security (CCS)*.

[72] Shomir Wilson, Florian Schaub, Rohan Ramanath, Norman Sadeh, Fei Liu, Noah A. Smith, and Frederick Liu. 2016. Crowdsourcing Annotations for Websites' Privacy Policies: Can It Really Work?. In *Proc. World Wide Web Conference (WWW)*.

6 APPENDIX

The table on the next two pages describes the six legal requirements we extracted from legal sources, what constituted a violation of these requirements, and how we mapped these requirements (and their violations) to the codebook we developed based on our cookie banner text annotation.

Requirement	Requirement definition	Violation	Code (from codebook)	Code definition
R1 Purpose explictiness				
R.1.1 Availability	Purposes should be clearly expressed, revealed, or explained, especially on the first layer of consent banners [32].	Absence of purpose. This requirement stems from the transparency principle (Article 5(1)(a), Recital 39 GDPR); data controllers need to inform users of data processing purposes (Article 13 (1)(c) and Recitals 58, 60 GDPR).	Code: purpose	We applied code when banners did not explain the purpose(s) of data processing.
R.1.2 Unambiguity	Purposes should be unambiguous.	Ambiguous intent. A violation occurs when purposes are defined ambiguously, and there is doubt about their meaning or intent [6].	Codes: vagueness; prolixity; positive framing; negative framing [6]	We applied code when banners used ambiguous wording.
R.1.3 Shared common understanding	Purposes should be comprehensible, regardless of users' cultural or linguistic backgrounds or other special needs involved [6].	Purposes are not comprehensible.	Code: purpose	We applied code when different terms were used to describe the same purpose.
R2 Purpose specificity	Purposes should be precisely identified, clearly defined, and detailed enough to determine what kind of processing is included or excluded within the specified purpose [6, 55].	Violations occur when a purpose is too vague or generic; for instance, "improve users' experience"; "develop new services and products"; "offer personalized services" [6, 55].	Code: purpose; sub-codes: user experience enhancement; analytics; advertising; custom content; service provision; essential functionalities; social media features; profiling	We applied sub-codes based on specific data processing purposes banners described.
R3 Intelligible consent				
R3.1 Non-technical terms	Consent should not contain overly legalistic or technical language.	Use of technical jargon.	Code: technical jargon [20, 24, 55]	We applied code when banners used terms that non-expert users are usually unfamiliar with (e.g., JavaScript, trackers, tracking systems, clients, servers).
R3.2 Conciseness	The first layer of consent requests should be brief but contain sufficient information [14].	A violation occurs when unnecessary details are mentioned, distracting users or causing information overload [20, 24].	Code: prolixity	We applied code when unnecessary details were mentioned in banners.
R4 Consent with clear and plain language				
R4.1 Straightforward statements	Consent requests should describe purposes clearly [13].	A violation occurs when misleading expressions are used: "We use cookies to personalize content and create a better user experience" [20, 22, 24, 55, 56].	Code: vagueness	We applied code when indefinite qualifiers were used in banners: can, may, might, someone, certain information, data, other parties, our vendors, some, any, often, possible, etc. Examples include "We may place cookies …"; "We may use information for …" [13, 55].
R4.2 Concreteness	Consent requests should use accurate and definitive statements.	A violation occurs when deceptive language or indefinite qualifiers are used [13, 55].	Code: vagueness	We applied code when banners did not use clear language and, instead, used indefinite qualifiers, incomplete or ambiguous statements, double negatives, or deceptive practices by hiding information from users (e.g., whether or not certain cookies are necessary for the website to operate properly).

Requirement	Requirement definition	Violation	Code (from codebook)	Code definition
R5 Freely given consent	A request for consent should imply a voluntary choice to accept or decline the processing of personal data (Articles 4 (11), 7(4) GDPR).	Any sort of pressure that nudges users toward consenting [7] (e.g., the use of positive or negative framing [19, 34]).	Code: positive framing; sub-codes: assumed happiness; safety and privacy arguments; compliance and authority arguments; playful arguments; superlatives and better experiences Code: negative framing; sub-codes: worse user experience; loss of functionalities	We applied codes when positive or negative framing was used.
R6 Informed consent	When trackers are used and stored on users' devices, users must be informed and aware of these trackers.	Absence of essential information about data processing.	Code: data type	We applied code when banners mentioned the types of data collected – IP address; geolocation data.

Table 3: Mapping the six legal requirements we extracted from different legal sources (and their violations) to our codebook, which we developed based on annotating the text of 407 cookie banners.