IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 11, 2021, accepted May 17, 2021, date of publication May 26, 2021, date of current version June 7, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3083897

Will EU's GDPR Act as an Effective Enforcer to Gain Consent?

JUNHYOUNG OH[1], JINHYOUNG HONG[2], CHANGSOO LEE[3], JEMIN JUSTIN LEE[4], (Member, IEEE), SIMON S. WOO[5,6], AND KYUNGHO LEE[1]

[1]School of Cybersecurity, Korea University, Seoul 02841, Republic of Korea
[2]School of Law, Korea University, Seoul 02841, Republic of Korea
[3]Korea Investment, Seoul 07220, Republic of Korea
[4]Center for Information Security Technology (CIST), Korea University, Seoul 02841, Republic of Korea
[5]Department of Applied Data Science, Sungkyunkwan University, Suwon 16419, Republic of Korea
[6]College of Computing and Informatics, Sungkyunkwan University, Suwon 16419, Republic of Korea

Corresponding author: Kyungho Lee (kevinlee@korea.ac.kr)

This work was supported in part by the Grant of Korean Health Technology Research and Development Project, Ministry of Health and Welfare, Republic of Korea, under Grant HI19C0866, and in part by the Institute for Information Communications Technology Promotion (IITP) Grant funded by the Ministry of Science and ICT (MSIT) of Korea government under Grant 2018-0-00261.

ABSTRACT Since the GDPR was implemented in 2018, organizations that collect data from the EU residents are required to receive the user's consent. Organizational measures to ensure that the organizations are compliant to the recently enacted GDPR are still abstract and ambiguous. Moreover, data subjects and controllers have demanded the practice of obtaining consent from organizations. By observing the case law and guidelines related to the GDPR provisions, we deduced four consent conditions. Then, we examined how online service provider's websites are making efforts to implement the GDPR framework. For this, we identified key characteristics of these websites, such as the existence of consent buttons. In order to help the data subjects obtain consent, we proposed an automatic tool that can check the consent conditions by checking the websites. Our study examined 10,000 websites for 26 days using the Python libraries with the tool automatically crawling the website information and analyzes the HTML structure according to the specified conditions. In addition, this tool crawls the privacy policy of each website. Moreover, it automatically determines whether it meets the four consent conditions by calculating it according to the formula defined in the consent condition. To evaluate the tool's accuracy, the researchers manually analyzed 500 websites and compared the manual analysis with the results of the tool's automatic analysis. We found that this tool differentiates itself through qualitative comparisons with other GDPR meters.

INDEX TERMS GDPR, privacy policy, consent, privacy.

I. INTRODUCTION

Can we say that online users have their rights over their personal information? If we consider policies, standards and laws, we can see that users have their own privacy rights. But do users actually enjoy their privacy rights?

The proliferation and the popularity of web services, social media platforms, big data, and search optimization websites have allowed relevant entities to leverage an enormous amount of user data for the recommendation, advertisement, and price adjustment services they provide [1]. Several scandals surrounding tech giants, such as Facebook and Cambridge Analytica, have clearly demonstrated the lack of appropriate data governance and enforcement of the sanctions on data privacy and user consent [2]. Google holds 85.86% of the market share of search engines worldwide, while Facebook holds 60.68% of the market share of social media platforms [3]. This shows the possibility for these companies to utilize their large amounts of data to produce personalized products or services. With the recent introduction of the European Union's General Data Protection Regulation (GDPR) [4] and the Cambridge Analytica incident, data protection agencies around the world are making efforts to enforce strong data protection mechanisms. Under the GDPR, the uniform data protection regulation of the EU, if a personal data breach occurs, data controllers can be fined

The associate editor coordinating the review of this manuscript and approving it for publication was Donghyun Kim.

up to €10 million, or 2% of the worldwide annual turnover, for lower-level penalties, and €20 million, or 4% of the worldwide annual turnover, for upper-level penalties [4].

According to the GDPR, data subjects hold various rights, including the *right to be informed*, *right to erasure*, and *right to restriction of processing*, which are highly relevant to data subjects themselves [5]. Therefore, data controllers should guarantee these rights, and data subjects should be able to exercise their rights at any time (for example, according to *right to erasure*, a controller needs to delete personal information without unreasonable delay when the data subject so desires. There are many derogations from exercising this right under certain conditions described in Article 17). Together, these rights clearly demonstrate the importance of data subjects' intention or willingness in terms of data processing. Naturally, the consent of data subjects deserves careful review as it enables the commencement of data processing based on the '*will*' of the data subjects.

According to the GDPR, consent is the basic legal basis for processing personal data (Rec. 40, Art. 6). GDPR provides the data subject with the right to change and withdraw consent at any time (Art. 7-3). Moreover, in order for the data controllers to obtain consent, they must provide information such as the purpose of use, retention period, and processing process of personal data [6]. They are also responsible for proving that the data controller itself has complied with various conditions related to consent [7]. Consent in the GDPR requires high quality and emphasizes freely provided features. However, many webservice providers are not sure if their business is GDPR compliant. For example, IAB Europe presents the Transparency & Consent Framework, which is widely used in the online advertising industry [8]. However, this framework was also pointed out by the French data protection authority CNIL to lack consent verification.

Unfortunately, GDPR provisions regarding consent include some ambiguous words and their recitals lack detailed explanations of conditions of consent. This may cause divergent interpretations of the GDPR provisions by data controllers and raise the possibility of non-compliance. For example, Article 7 of the GDPR requires the conditions of consent to be ''clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language'' and ''freely given'' [4]. However, these terms of conditions are somewhat equivocal and lack precise criteria. Consequently, many service providers may not guarantee whether their consent and privacy policies meet the requirements of the GDPR provisions. To avoid this confusion, the ambiguity and uncertainty within the GDPR text leads to the need for more detailed investigation and analysis.

In this study, we analyze GDPR provisions and recitals as well as relevant EU guidelines to propose quantifiable consent conditions to check whether website providers are compliant with the GDPR. We then evaluate the extent to which various popular web service providers meet these conditions. Two researchers systematically analyzed 500 websites to find out whether they met the consent condition and classified the type. In addition, we developed the tool that automatically determines whether each website meets the consent conditions. The automated tool was developed in the form of organizing the html structure into a tree structure, and then searching for terms and symbols suitable for each consent condition. This tool achieved a high accuracy of approximately 96%. Using this tool, 10,000 websites were analyzed to determine how many met our consent conditions. The tool was compared with other existing GDPR meters to analyze the pros and cons of the tool.

Our contributions are summarized as follows:

- Four quantifiable consent conditions were deduced by analyzing the GDPR and related guidelines.
- An automated tool was developed to process compliance determination, and the accuracy was verified by comparing the results of manual website analysis. Ten thousand websites were subjected to the processing.
- This tool differentiates itself from other GDPR meters [9]–[11] and achieved 96% accuracy.

II. BACKGROUND AND RELATED WORK

European Union Member States have taken the lead in international regulation for data protection, reshaping and influencing policies in other parts of the world. What follows are the main EU initiatives for data protection. The Organization for Economic Cooperation and Development (OECD) provided the Privacy Guideline in 1980 [12], which outlines privacy in eight different principles [13] and has been widely adopted by a number of states in their domestic legislation. The guideline states that data controllers must acquire the consent of data subjects when collecting their personal data [14]. Similarly, the Council of Europe (CoE) data protection Convention 108 was adopted in 1981, and later served as a reference for the EU Data Protection Directive. Prior to the EU's GDPR, the European Directive 95/46/EC was enacted in 1995 and adopted by European Union Member States. According to the directive, explicit consent is required and such consent means ''freely given specific and informed indication'' [15]. The Privacy and Electronic Communications Directive (2002/58/EC) introduced privacy and data protection issues in a new way by addressing the privacy elements in electronic communication aspects [16]. The GDPR, which builds on the previous European Directive 95/46/EC, was enacted in 2016 and came into force in 2018 across all EU member states. States outside the EU are also subject to the regulation as they adopt extraterritorial application and adequacy tests. The GDPR provides six core privacy principles: fairness and lawfulness, purpose limitation, data minimization, accuracy, storage limitation, and integrity and confidentiality [17].

Among these principles, lawfulness can be achieved by six legal bases of personal data processing. Consent, one of those six lawful bases, is worth examining as it is necessary to achieve lawfulness related to data subjects' diverse rights, as mentioned above. Article 4 of the GDPR provides

a definition of consent, while Article 7 specifies the conditions for valid consent. According to Article 4 paragraph 11, ''consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement'' [4]. In addition, Article 7 provides that consent should be ''clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language'', and confirms that ''consent is freely given'' [4].

Diverse research methods have attempted to examine whether privacy policies are adequately provided to data subjects before consent. Prior studies have argued that consent should be explicitly stated to help the data subject make a conscious decision about the processing of their personal data [18]. Obar and Oeldorf-Hirsch [19] discovered that users usually do not spend much time reading privacy policies, and user studies have shown that users are not aware of the contents of privacy policies. Mcdonald *et al.* [20] measured psychological acceptability in various formats of privacy policy. Fabian *et al.* [21] analyzed privacy policy and measured the readability of users through various readability metrics, including the Flesch reading ease score, which was also adopted in our study. Daniel [22] studied the dilemma of consent by dividing cognitive and structural perspectives and stressed the importance of consent for privacy norms to be organized. Reeder *et al.* [23] developed a tool to visualize privacy policies, claiming that it is more accurate than natural language. Pointing out that traditional methods have limitations in making users read privacy policies, Tabassum *et al.* [24] used comics to induce users to read the privacy policy, and used eye trackers to measure the extent to which they did so.

Since the enforcement of the GDPR in 2018, several studies have been conducted to evaluate data flow and check if cookies information was acquired in accordance with the GDPR, which requires user consent. Sanchez-Rola *et al.* [25] examined 2,000 websites and analyzed their cookies information; they found that 92% of the websites set and track identifiable cookies. Iordanou *et al.* [26] analyzed how much information was transferred internationally through a web tracking service. They argued that as most of the cookies information is exchanged between European countries, it falls under the jurisdiction of the GDPR.

Degeling *et al.* [27] surveyed 6,579 websites to compare the situation before and after the GDPR and to refine the types of cookie consent. Most studies related to the GDPR tend to focus on cookies or the overall framework of the GDPR. Trevisan *et al.* [28] argued that many websites did not comply with the GDPR by using cookies information without data subjects' consent. According to the research, an average of 49% of websites use cookies information without prior user consent, for profiling and other purposes. This is a meaningful study that reveals a violation of the GDPR, but it only focuses on whether consent was obtained rather than the legality of the consent. In this context, our study seeks to study the ''forms'' and ''legality'' of consent from the outset.

Utz *et al.* [29] systemized various forms of cookie consent and analyzed whether each form of consent actually affected users. We studied the agreement among members and therefore, the details are slightly different. Utz's study will provide insights when we create consent conditions in the next chapter, and it also provides a good basis for some consent conditions.

Nouwens *et al.* [30] set three conditions: explicit consent, accepting all is as easy as rejecting all, and no pre-ticked boxes. They analyzed whether the UK's top 10,000 websites met these conditions. While their conditions are similar to ours, they seem to lack sufficient policy and legal bases. Our paper presents four consent conditions and provides sufficient evidence through GDPR recitals, guidelines, and judicial precedents.

Matte *et al.* [31] examined the consent regarding cookie banners. They analyzed 22,949 websites and found those that saved information even without their users' consent on cookie banners. They also found cases where websites used pre-ticked boxes in connection with cookie banners, and where some websites collected information even when users refused to give their consent. Their research has made a huge contribution in that they conducted large-scale research on cookie banner consent through automated tools. But the study's limitation is that it only targeted cookie banners.

III. CONSENT CONDITIONS IN THE GDPR

Unfortunately, online privacy policies often consist of layered rules and jargon-laden legal phrases. As a result, data subjects often provide their consent to data controllers without knowing what exactly that consent entails. This undesirable phenomenon is also caused by giant companies. For instance, ''Google'' was fined by the France National Data Protection Commission in January 2019 for not properly obtaining consent from data subjects [32]. The decision was once again upheld at the French supreme administrative court following the appeal by Google LLC [33]. The focus of the decision was the matter of how, rather than if, the user consent was acquired. This case clearly demonstrates that consent requirements within the GDPR regime should be examined with due care.

Since the GDPR binds EU members and affects non-EU states' privacy policies, our study attempts to resolve a specific aspect of the privacy issue based on the GDPR. Our work aims to analyze GDPR provisions regarding consent and to derive four conditions to assess the websites of main web service providers. However, prior studies have demonstrated that the GDPR is deficient in accurately specifying organizations' right to process personal data and individuals' right to prevent certain processing [27], [34]. GDPR recitals and relevant guidelines fill this gap by providing more detailed explanations and sometimes specific examples. Although recitals and guidelines are not legally binding, they are meaningful documents in that they are prepared by the same organization as the GDPR text and are worth data controllers' (web service providers) reference. In this context,

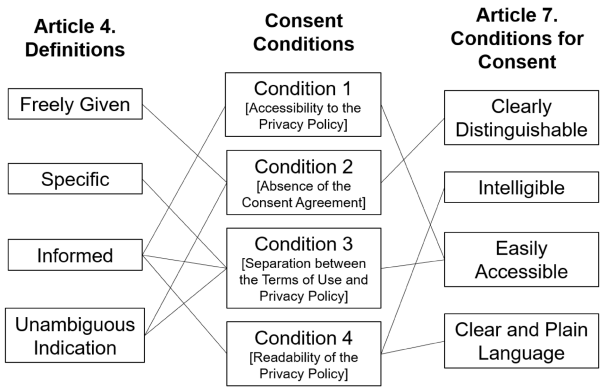


FIGURE 1. Relationship between conditions and articles of the GDPR.

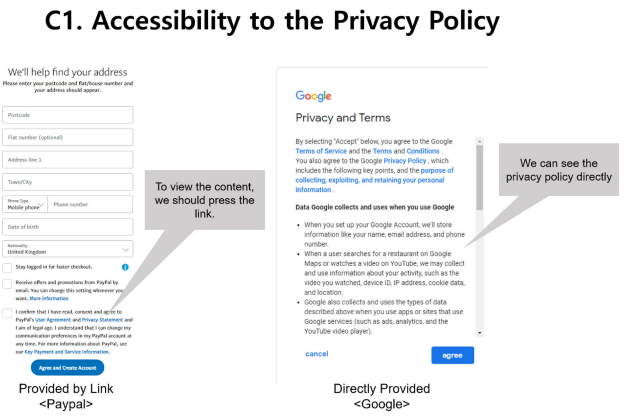


FIGURE 2. Explanation of consent Condition 1.

this study refers to GDPR recitals as well as the "Guidelines 05/2020 on consent under Regulation 2016/679 [35]" and the "Guidelines on transparency under Regulation 2016/679 [36]." Sometimes, we also refer to case law.

The final goal of our study is to examine and observe if the world's main websites are compliant with the consent form required by the GDPR legal regime. To achieve this research purpose, we propose the following four key quantifiable conditions to determine whether the valid consent forms are provided on the websites of data controllers and are properly presented to data subjects (users). These conditions are all designed to reflect the requirements within the GDPR regime:

A. CONDITIONS

- **Condition 1. Accessibility to the Privacy Policy**. Data subjects must have easy access to the privacy policy, as presented directly or by link.
- **Condition 2. Absence of the Consent Agreement**. A consent button exists for the privacy policy.
- **Condition 3. Separation between the Terms of Use and Privacy Policy**. The data subject can agree to the privacy policy and the terms of use separately **(Condition 3-1. Separately Consent)**. The privacy policy and terms of use exist separately **(Condition 3-2. Separately Exist)**.
- **Condition 4. Readability of the Privacy Policy** . The privacy policy is easily readable, which means that the Flesch reading ease score of the Privacy policy is at least 50.

Next, we present detailed rationales for constructing each consent condition. They are primarily based on the GDPR provisions, GDPR recitals, and the two relevant guidelines.

1) RATIONALE FOR C1

Article 4 of the GDPR articulates that consent needs to be "informed." In addition, Article 7 paragraph 2 of the GDPR requires that the request for consent should be "in an easily accessible form." This "informed" consent includes clear "visibility" of information regarding data processing. Recital 32 of the regulation makes it clear that "informed" entails the obligation to make consent requests clear and unnecessarily disruptive to users. According to the UK Information Commissioner's Office, this includes developing user-friendly layered information, and just-in-time consents [37]. "Guidelines 05/2020 on consent under Regulation 2016/679" state that "providing information to data subjects prior to obtaining their consent is essential in order to understand what they are agreeing to," and adds that information should be "accessible" for the consent to be valid. Further, the "Guidelines on transparency under Regulation 2016/679" provide that "The 'easily accessible' element means that data subjects should not have to seek out the information; it should be immediately apparent to them where and how this information can be accessed, for example by providing it directly to them, by linking them to it." We consider that this accessibility criterion requires that privacy policies be provided directly on the sign up page or other areas of the website, or by a link. *Therefore, a controller may not meet this provision if the privacy policy is not provided in an appropriate manner such as directly or as a link.* There has been some judicial decisions relevant to the condition 1. In December 2019, the Dutch Data Protection Agency issued a fine of €525,000 to the Dutch Tennis Association for selling its members' personal data without their prior "informed" consent. Predicating its decision on the GDPR provisions, including Article 5 and 6, the Agency adopted a strict approach on the issue of data subjects' prior consent [38]–[40].

2) RATIONALE FOR C2

Article 4 of the GDPR provides that consent should be "freely given" and "unambiguous." Recital 32 explains that "silence, pre-ticked boxes or inactivity" fail to provide valid consent. Article 7 paragraph 2 of the GDPR also states that "consent shall be presented in a manner which is clearly distinguishable from the other matters" [4]. "Guidelines 05/2020 on consent under Regulation 2016/679" state that "proceeding with a service cannot be regarded as an active

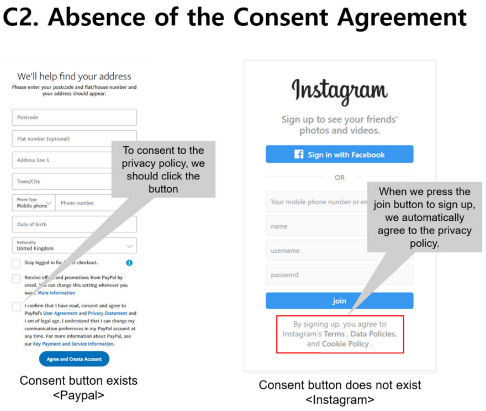


FIGURE 3. Explanation of consent Condition 2.

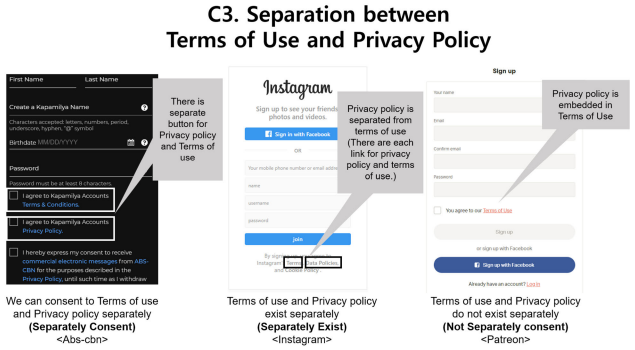


FIGURE 4. Explanation of consent Condition 3.

indication of choice'' in Section 3.4. of it [35]. Furthermore, directly mentioning Recital 32, the document confirms that acquiring user consent using ''scrolling or swiping through a webpage or similar user'' falls under ambiguous consent. Therefore, the existence of a consent button specifically for the privacy policy can be a criterion for valid consent. If a data subject agrees directly with the privacy policy, there is no problem. However, simply pressing a ''Join'' or ''Sign up'' button to use the service does not comprise direct consent to the privacy policy. It can be argued that pressing ''Join'' or ''Sign up'' includes or has the same meaning as agreeing to the privacy policy. However, this can be interpreted as intentionally making the process indistinguishable, making data subjects use the service without being conscious of giving away his or her personal information. In October 2019, the highest European court made its first decision on cookies. In the Planet 49 case, the CJEU found that consent obtained by using a pre-ticked box is invalid, because ''only active behaviour'' meets the unambiguity requirement [41]. *Therefore, a controller may not meet this provision if a consent button does not exist.*

3) RATIONALE FOR C3

The GDPR requires ''specific (Article 4)'' and ''clearly distinguishable (Article 7)'' consent. It is explained in detail in the ''Guidelines 05/2020 on consent under Regulation 2016/679,'' which state that ''the consent of the data subject must be given in relation to specific purposes and that a data subject has a choice in relation to each of them'' in Section 3.2. of it [35]. Therefore, ''choice'' should be clearly given to users. Similarly, the guideline also states that user consent should not be a condition for accessing website services. Consider a website (data controller) that only provides its users (data subjects) with opportunities to consent to the terms of use, not separately to its privacy policies. In this situation, even in cases where a data subject agrees to the terms of use, it is difficult to ensure that the data subject also intends to agree to the privacy policy. He/She may want to agree only to the terms of use and not to the privacy policy. Therefore, to identify the real purpose of users' consent, separate consent buttons should exist for the privacy policy and terms of use. This will ensure service providers' compliance with the GDPR. Support can be found in the ''Guidelines 05/2020 on consent under Regulation 2016/679,'' which provide that ''distinguishable'' means that the consent issue should ''stand out'' or be ''separate.'' In addition, the guideline also suggests that ''consent cannot be obtained through the same motion as accepting the general terms and conditions of a service'' [35]. These are clear expressions. Therefore, accepting only the terms of use does not comply with the guidelines. In August 2018, the Supreme Court of Austria found the absolute nature of the ''specificity'' requirement. The Court held that bundling consent is absolutely prohibited by GDPR and incorporating a consent clause in its general terms and conditions is illegal [42], [43].

4) RATIONALE FOR C4

The GDPR articulates that requests for consent shall use ''clear and plain language (Article 4)'' and should be in ''intelligible form (Article 7).'' The ''Guidelines 05/2020 on consent under Regulation 2016/679'' also elaborate that ''this means a message should be easily understandable for the average person and not only for lawyers.'' The requirement of ''plain language'' is also in line with the definition of consent in the GDPR Article 4, which stipulates that consent is an ''informed'' indication of data subjects. To obtain data subjects' consent to a privacy policy, it should be easily readable to the general public. According to the UK Information Commissioner's Office, clear and plain language means ''easy to understand'' [37]. The Flesch Reading Ease score can be a useful tool to determine whether a privacy policy is easily readable, as the score provides a subjective standard for the readability of reading materials using the number of words, sentences, and syllables. A privacy policy with a Flesch reading ease score of 50–60 points corresponds to 10th to 12th grade level, while a privacy policy with a score below 50 corresponds to college level [44]. Flesch reading ease score of 50 points works as a reference point to ensure a privacy policy is expected to be understandable to the

TABLE 1. Summary of Flesch reading ease score [44].

Score	Grade	Notes
60-70	8th-9th grade	easily understood by 13- to 15-year-old students
50-60	10th-12th grade	Fairly difficult to read
30-50	College	Difficult to read
0-30	College graduate	Very difficult to read

general public. The Flesch reading ease score is calculated by applying the entire privacy policy to the provided equation.

IV. MANUAL ANALYSIS

With our four consent conditions, we analyzed the consent forms of Alexa's top 500 popular websites [45] to determine whether each of the websites meets these quantifiable conditions. We limited the scope of our analysis as follows: this study focuses on the use of personal information through membership registration. In other words, websites that are used without membership registration are excluded. This is because most valuable and sensitive personal information is collected through membership registration, even though it is possible to collect some personal information without active registration (i.e., by using surfing records on websites). We also excluded websites that do not support web services in English, to focus on universal service providers. We excluded websites that use accounts from other popular websites to avoid overlapping calculations (i.e., Blogspot.com was excluded, because its login is possible from Google accounts).

To do so, we approached the target sites with an IP from France, one of the major EU countries. This is because European countries are the most directly affected by the GDPR. January 6, 2019, was set as the standard date for the ranking and forms of the website so that our research did not become subject to the time variable.

Depending on the conditions, two different methods were adopted to analyze the target websites. To determine whether a website met consent conditions 1, 2, and 3, we manually checked if the contents corresponding to each condition appeared while entering the website or registering as a member of that website. This work was conducted by two researchers and cross-validated. When checking consent condition 4, we scrapped the website's privacy policy if that privacy policy came out directly or if the link came out. After saving the scrapped privacy policy, the Flesch reading ease score was calculated using Python's library (textatistic).

For the statistical analysis, we tested the correlation across different cases for each consent condition. We used the Kruskal-Wallis test for categorical variables with a small sample size (expected values less than 5) and the Chi-square (χ^2) test for categorical variables with a large sample size. For all statistical tests, we used a significance level of $p = 0.05$. We further conducted pairwise tests and compared a subset of possible pairs of conditions.

Among the 500 websites, 27 could not be accessed at the time of the experiment. Among these 473 websites,

TABLE 2. Summary of Alexa's Top 500 websites used in our study.

Type	Num. of websites
Not accessible	27
No English	154
Owned by same providers	95
No sign up	46
No service available	29
Valid (can be analyzed)	**149**
Total	500

TABLE 3. Websites analysis integration matrix (Top 50 websites). ✓:supports GDPR, ✗:not supports GDPR.

Webpage	Ranking	C1	C2	C3-1	C3-2	C4
Google.com	1	✓	✓	✓	✓	43.8
Facebook.com	3	✓	✗	✗	✓	42.1
Amazon.com	8	✓	✗	✗	✓	40.4
Yahoo.com	9	✓	✗	✗	✓	40.1
Twitter.com	11	✓	✗	✗	✓	43.4
Reddit.com	13	✓	✗	✗	✓	42
Live.com	14	✓	✓	✗	✓	37.2
Vk.com	15	✓	✓	✓	✓	30.8
Instagram.com	17	✓	✗	✗	✓	42.4
Netflix.com	23	✓	✗	✗	✓	32
Pornhub.com	27	✓	✗	✗	✓	31
Twitch.tv	28	✓	✗	✗	✓	32.9
Linkedin.com	30	✓	✗	✗	✓	45.7
Aliexpress.com	35	✓	✗	✗	✓	34
Ebay.com	38	✓	✗	✗	✓	29.8
Porn555.com	42	✓	✗	✗	✓	38
Livejasmin.com	44	✓	✗	✗	✓	42.2
Xvideos.com	45	✓	✗	✗	✓	33.1
Imdb.com	46	✓	✗	✗	✓	43

9 websites could not be accessed with the EU IP address. In addition, there were 154 websites that did not provide website services in English. For certain websites, we often only log in through our Google account. Since we have already analyzed Google and reflected it in the statistics once, those websites should be excluded from the statistics. There were 95 websites that sign up for membership through other popular websites (mostly Google). Further, 46 websites did not require any membership sign up process but allowed the use of the service without signing up (e.g., onlinevideoconverter.com). Furthermore, 29 websites, including financial websites, could not be serviced in general. Since online banking services are usually related to offline services, we considered this group of websites to be unsuitable for our analysis. Finally, and surprisingly, six of Alexa's top 500 websites did not have a privacy policy, but we analyzed these websites' consent conditions for **C1**–**C4**. Having excluded those inappropriate websites, we investigated the remaining 149.

For **C1 (Accessibility to the Privacy Policy)**, five out of 149 websites provided privacy policy directly, while 138 websites provided privacy policy through links. Only six websites failed to provide their privacy policies. This shows a 96.0% conformance rate for 149 companies in Alexa.com's top 500 sites, as shown in Fig. 5. It is significantly different across other consent conditions (KW test $p = 3.1 \times 10^{-29}$, $\chi^2 = 126.0$).

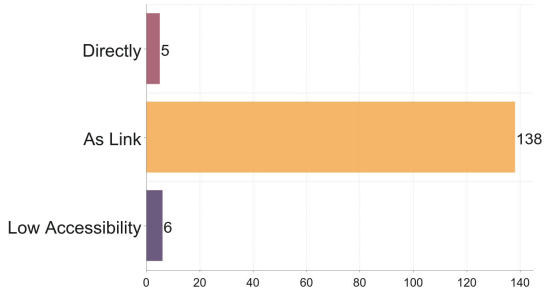


FIGURE 5. Results for Condition 1, where the X-axis presents the number of websites and the Y-axis shows the different consent conditions of websites.

For **C2 (Absence of the Consent Agreement)**, only 42 websites had consent buttons for privacy policy. One hundred and seven websites did not have a consent button, and users could show their inclination to the website's privacy policy only by signing up (e.g., "By signing up, I agree to the Privacy Policy and Terms of Service."). This shows a 28.2% conformance rate for 149 companies in Alexa.com's top 500 sites, as illustrated in Fig. 6. It is significantly different across other conditions (KW test $p = 1.0\times10^{-7}$, $\chi^2 = 28.4$).

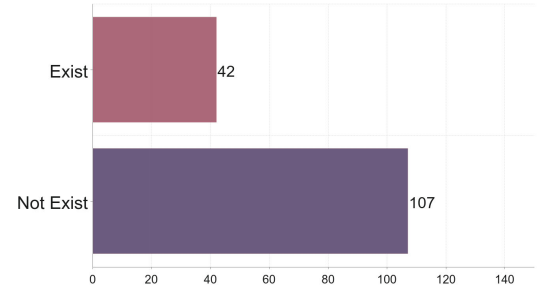


FIGURE 6. Results for Condition 2, where the X-axis is the number of websites and the Y-axis shows the different consent conditions of websites.

For **C3-1 (Separately Consent)**, 26 out of 149 websites allowed data subjects to separately agree to the privacy policy and terms of use, while 123 websites did not. The latter websites provided only a single button, so users had to simultaneously provide their consent to the privacy policy and terms of use. This represents a 17.4% conformance rate for 149 companies in Alexa.com's top 500 sites, as shown in Fig. 7. It is significantly different across other conditions (KW test $p = 1.9\times10^{-15}$, $\chi^2 = 63.1$). For **C3-2 (Separately Exist)**, 139 out of 149 websites have a privacy policy that is separated from terms of use. Only 10 websites have a privacy policy embedded in their terms of use. This shows a 93.3% conformance rate for 149 companies in Alexa.com's top 500 sites, as shown in Fig. 7. It is significantly different across other conditions (KW test $p = 4.2\times10^{-26}$, $\chi^2 = 111.7$).

For **C4 (Readability of the Privacy Policy)**, only five out of 149 websites' privacy policies scored over 50 on the

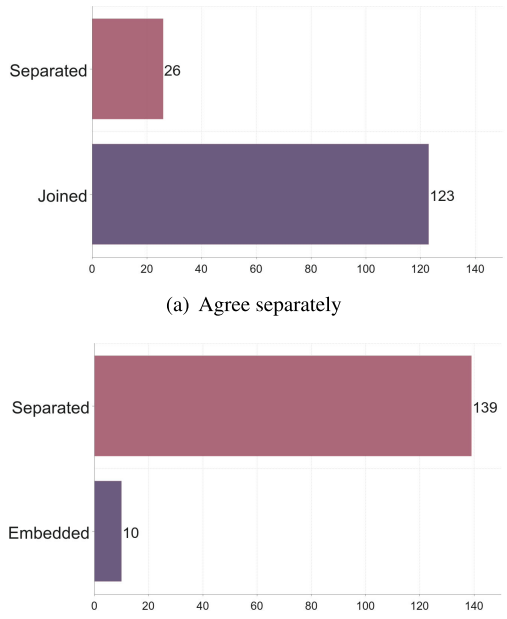


FIGURE 7. Results for Condition 3, where the X-axis comprises the number of websites and the Y-axis shows the different consent conditions of websites.

Flesch reading ease score. This shows a 3.4% conformance rate for 149 companies in Alexa.com's top 500 sites, as shown in Fig. 8. There were 120 websites with Flesch reading ease scores between 30 and 50, while 18 websites scored lower than 30 points. Their mean Flesch reading ease score was 36.3, and the standard deviation was 7.3. We can infer from the figures that the privacy policies of the main websites are generally understandable to college students and beyond. Consent condition 4 shows an overwhelmingly low conformance rate compared to other consent conditions, suggesting that many websites do not care much about the difficulty of the text content.

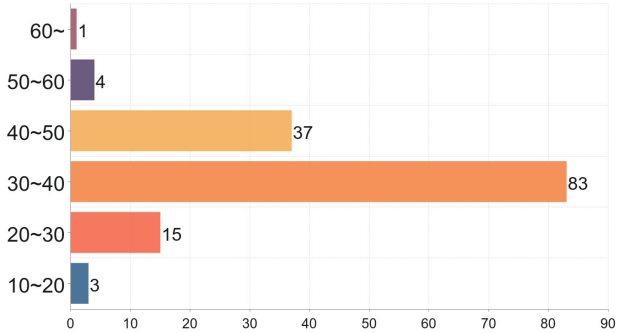


FIGURE 8. Results for Condition 4, where the X-axis is the number of websites and the Y-axis shows the Flesch reading ease score.

A. ADDITIONAL ANALYSIS

Three main differences were found when accessing the websites with an EU and non-EU based IP address. The first

difference is the presence or absence of a consent button. A summary of the results is shown in Table 4. There were consent buttons on three websites when accessing with an EU IP address, but there was no button when accessing the URL with a non-EU IP address. In seven websites, there was no button when accessing it with EU IP address, while when accessing with the non-EU IP address, there was a button.

TABLE 4. Presence of consent button for only one side.

IP address from EU	IP address from non-EU
- Discordapp.com - Yelp.com - Espncricinfo.com	- Netflix.com - Apple.com - Adobe.com - Playstation.com - Samsung.com - Nike.com - Oracle.com - Bhance.net (Consent is divided into "collect and use," "provide to third party," and "transfer to other countries")

The second difference is the forms in which privacy policies are delivered. A summary of the results is shown in Table 5. There are a variety of ways for websites to provide their privacy policies, including texts, drawings, tables, and comics. Most of the websites adopted the same form, regardless of the IP we used. However, some websites provided their privacy policy in different forms depending on the IP used. A typical case is turning texts into a table. In some cases, privacy policy was provided as a table when a website was accessed by EU IP address, while in some other cases, a table was provided when accessing with a non-EU IP address. To be more specific, there were two websites where the privacy policy was presented as a table when accessed from an EU IP address, and just one opposite case. To sum up, there were some, but not many, cases in which privacy policies were presented in different forms depending on the geographical location of the IPs.

TABLE 5. Form of privacy policy.

IP address from EU	IP address from non-EU
- VK.com (personal information is listed in the extra table) - Quora.com (a summary of the privacy policy is listed in the extra table)	- Nike.com (a summary of privacy policy is listed in the extra table)

The third difference is where content is added to the privacy policy. A summary of the results is shown in Table 6. There are seven websites for this type. The additions usually include the legality of data processing, the transfer of data outside the country, and the rights of the data subject.

V. AUTOMATED ANALYSIS

In the previous chapter, we manually analyzed the top 500 websites. Based on this, we developed a tool to automatically check whether a website satisfies the consent conditions. For consent conditions 1, 2, and 3, we entered the sign up windows of each website and checked whether or not they matched the three conditions. The tool used various tags and an HTML structure in a tree format. Since consent condition 4 is about the Flesch reading ease score, privacy policies of each website were crawled and scored. The tool is based on Python TensorFlow library (such as fasttext, selenium, textatistic). To verify that our tool worked, we compared the results obtained with the tool and the results of manual analysis. We then analyzed the top 10,000 websites at *Alexa.com*. We gathered data from these websites by crawling each website's ranking and URL. We used this tool to check that 10,000 websites meet these quantifiable conditions.

A. AUTOMATION METHODOLOGY

Consent conditions 1, 2, and 3 are conditions that can be identified on the sign up screen, while consent condition 4 can be identified using the contents of the privacy policy. Therefore, when analyzing websites, consent conditions 1, 2, and 3 were identified using the HTML structure on the sign up screen, while consent condition 4 was calculated by crawling the privacy policy.

1) CONSENT CONDITION 1. ACCESSIBILITY TO THE PRIVACY POLICY

When analyzing 500 websites in the previous chapter, few cases show privacy policy directly. Most websites provided privacy policies through links or properly. To reflect this tendency, in our tool, we automated cases where the privacy policy was provided through links. In the HTML structure of a website, each component is tagged. Among these tags, the *href* tag is attached to the component connected through the link. The tags that link privacy policies usually contain specific words, such as data policy, privacy policy, privacy, terms, and so on. Based on the websites analyzed in the previous chapter, we determined that consent condition 1 is satisfied when the *href* tag is attached to the word.

2) CONSENT CONDITION 2. ABSENCE OF THE CONSENT AGREEMENT

One of the programming interfaces of HTML documents is the Document Object Model. The Document Object Model (DOM) provides a structured representation and a way for the programming language to access the DOM, making it easy to change [46]. The HTML document based on the DOM has a hierarchical structure with each element functioning as a root, in a so-called tree structure.

Normally, even if there is a text ''privacy'' and a check box in the code, it is not known whether the check box relates to privacy policy. However, in the tree structure, if the element with the text ''privacy'' and the check box have the same

TABLE 6. Privacy policy conditions applicable to the EU IP address.

	Facebook	Yahoo	AliExpress	PayPal	Quora	Yelp	IKEA
Lawful Processing	○	○	○				○
Transfer to Other Countries		○		○		○	
Rights of Data Subject	○		○	○	○	○	
Retention of Personal Data					○		
Use of Personal Data		○	○			○	
Collect Personal Data						○	○
Children's Privacy							○

parent node, we can know that the check box is for the privacy policy. Therefore, consent condition 2 is divided into three types: cases that contain words related to privacy *x*, cases that contain both words related to "privacy" and "terms of use" related to *xy*, and cases that contain terms related to terms *y*. In the case of *x* and *xy*, if it has a parent node such as a check box, it is judged to satisfy consent condition 2. In the case of *y*, even if it has a parent node such as a check box, it is judged that the privacy policy must exist in terms of use to meet condition 2.

3) CONSENT CONDITION 3. SEPARATION BETWEEN THE TERMS OF USE AND PRIVACY POLICY

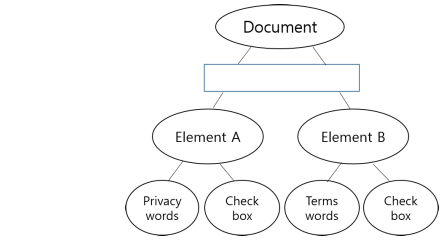
Consent condition 3, as in Consent condition 2, is determined using the parent nodes. As shown in Fig. 9, if *x* and *y* have check boxes and parent nodes, respectively, the website meets consent condition 3-1. However, there are exceptions. As shown in Fig. 9, *x* and *y* may have check boxes and parent nodes, respectively, even if they are not separately consented. Those exceptions include cases where the nodes are in the footers. Therefore, it is checked whether it is in a footer or a different place, and in that case, it is excluded. In addition, we determined that consent condition 3–2 is met if *href* tags are included in *x* and *y*, respectively.

4) CONSENT CONDITION 4. READABILITY OF THE PRIVACY POLICY

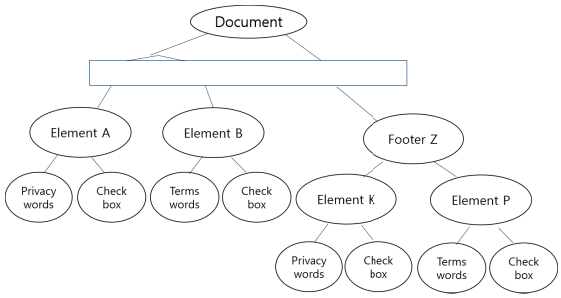
We crawled the content of privacy policy and calculated the Flesch reading ease score using Python's library to check whether consent condition 4 is met. Polisis [47] was used to crawl privacy policies. Polisis retrieves the URL of each website as an input value and analyzes the privacy policy to show various results. Polisis is suitable for this study, because it can retrieve a link to the privacy policy as an output upon entering a website.

B. EFFECTIVENESS VERIFICATION

Although we created a tool to determine if each website meets the consent conditions in an appropriate way, there was no guarantee that it would work properly. Therefore, we tested the accuracy of this tool. In the previous chapter, researchers who fully understood the consent condition manually analyzed 149 websites (selected from the top 500 websites).



(a) Example for C3-1 (Separately Consent)



(b) Example for C3-1 (Separately Consent) using footer

FIGURE 9. Examples for Consent Condition 3.

TABLE 7. Accuracy of the tool.

	C1	C2	C3	C4
Accuracy	95.3%	97.3%	92.6%, 95.9%	94.5%

The results were regarded as the ground truth, and the results from this tool were compared to calculate the accuracy. Since consent conditions 1, 2, and 3 (3–1, 3–2) comprise a binary result, accuracy was calculated by the number of websites with the same result among 149 websites. Since the result of consent condition 4 is a continuous number, the error rate of the Flesch reading ease score was calculated based on the ground truth, and the error results of 149 websites were averaged and calculated. Consent condition 4 was calculated manually using Python for both manual analysis and tool analysis. However, the error in terms of consent condition 4 appears to be due to the fact that the collection of the privacy policy contained some words and sentences that did not match the privacy policy.

Overall, the tool is highly accurate. Consent condition 1 is considered slightly less accurate, because it excludes the direct privacy policy. Consent condition 2 had a fairly high accuracy of 97%. Consent condition 3–1 has a rather low accuracy of 92.6%, because there is a possibility of erroneously determining the parent node, including footers. Consent condition 4 shows an accuracy of 94.5%, because it only needs to calculate privacy policies.

C. RESULT

In the same way that we limited our scope of analysis in the previous chapters, we counted the number of websites excluded from our scope of research. Among the 10,000 websites, 253 could not be accessed at the time of the experiment. There were 2,142 websites that did not provide services in English, while 1,158 provided sign up and log-in services through other sites such as "Google." There were 860 websites that did not require any membership sign up process but allowed the use of the service without signing up. In addition, 457 websites could not be serviced in general, including financial websites. There were two types of error in this automated tool, which were different from those in the manual analysis.

The first error type was in fetching and analyzing the content in the sign up window, which is related to consent conditions 1, 2, and 3. The second error type related to consent condition 4. Analysis for consent condition 4 is to obtain and analyze the privacy policy through Polisis. There was an error related to Polisis. For the first type of error, "WSAETIMEDOUT" was displayed on 85 websites. The error is a case where a technical issue arises in our tool due to a firewall and it is not possible to properly check whether the consent condition is met. There seems to be an error caused by the automatic connection and analysis of the tool. Consent conditions 1, 2, and 3 were analyzed for 5,045 websites. For the second type of error, automated analysis of 976 websites was not conducted. These include websites where the connection was not properly made through polisis, or they are connected to a window that was not a privacy policy. Therefore, 4,154 websites were analyzed for consent condition 4. The number of websites is summarized in Table. 8.

For **C1 (Accessibility to the Privacy Policy)**, 4,566 out of 5,045 websites provide privacy policies through links. Only 479 websites did not provide privacy policies. This shows a 90.6% conformance rate for 1,949 websites, as shown in Fig. 10. This conformance rate is 5.4% lower than that of Alexa.com's top 500 sites. This gap seems to be based on the fact that our tool cannot find cases that show privacy policies with low accessible. Only a very small number of websites show their privacy policies with low accessible.

For **C2 (Absence of the Consent Agreement)**, only 621 websites had consent buttons for privacy policy, while 4,424 websites did not have a consent button, and agreement to the privacy policy was conducted solely by signing up (e.g., "By signing up, I agree to the Privacy Policy and Terms of Service."). This shows a 12.3% conformance rate for 5,045 websites, as shown in Fig. 11. This conformance rate is 15.8% lower than that of Alexa.com's top 500 sites.

TABLE 8. Summary of Alexa's Top 10,000 websites used in our study.

Type	Num. of websites for C1,C2,C3	Num. of websites for C4
Not accessible	253	253
No English	2,142	2,142
Owned by same providers	1,158	1,158
No sign up	860	860
No service available	457	457
First error type(C1,C2,C3)	85	-
Second error type(C4)	-	976
Valid (can be analyzed)	**5,045**	**4,154**
Total	10,000	10,000

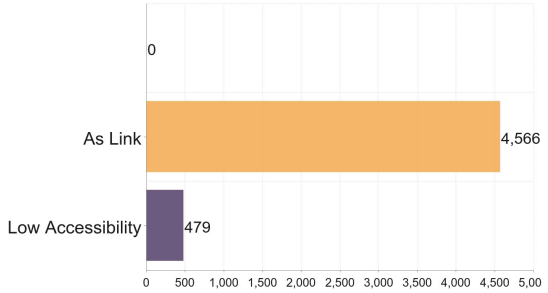


FIGURE 10. Results for Condition 1 by the automated tool, where the X-axis is the number of websites and the Y-axis shows the different consent conditions.

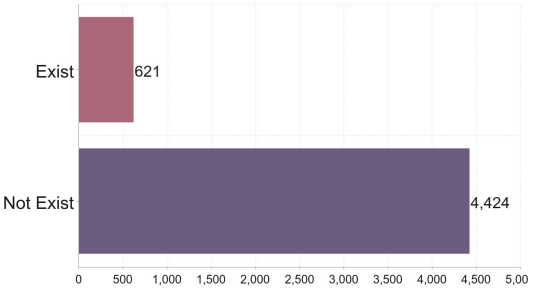


FIGURE 11. Results for Condition 2 by the automated tool, where the X-axis is the number of websites and the Y-axis shows the different consent conditions.

For **C3-1 (Separately Consent)**, 530 out of 5,045 websites allowed data subjects to agree to the privacy policy and terms of use separately, while 1,755 websites did not allow this. This shows a 10.05% conformance rate for 5,045 websites, as shown in Fig. 12. This conformance rate is 7.3% lower than that of Alexa.com's top 500 sites.

For **C3-2 (Separately Exist)**, 4,480 out of 5,045 websites had a privacy policy that was separated from its terms of use. Only 565 websites had privacy policies embedded in their terms of use. This shows an 88.8% conformance rate for 5,045 websites, as shown in Fig. 12. This conformance rate is 4.5% lower than that of Alexa.com's top 500 sites.

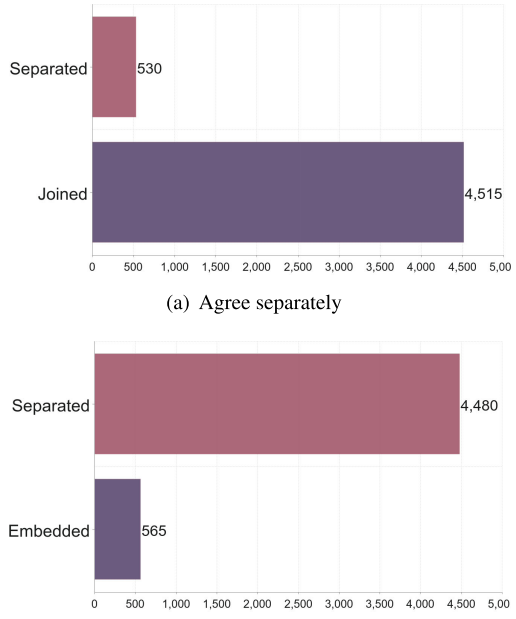


FIGURE 12. Results for Condition 3 by the automated tool, where the X-axis is the number of websites and the Y-axis shows the different consent conditions.

For **C4 (Readability of the Privacy Policy)**, only 59 websites scored over 50 on the Flesch reading ease score. This shows a 1.4% conformance rate, as shown in Fig. 13. This conformance rate is 2.0% lower than that of Alexa.com's top 500 sites. The mean Flesch reading ease score was 37.5, while the standard deviation was 12.8. The 4,154 websites from the top 10,000 websites had a lower conformance rate than the 149 websites from the top 500 websites, but their average Flesch reading ease score was higher, indicating that the overall privacy policy was slightly more readable.

VI. PERFORMANCE EVALUATION OF THE GDPR METERS

There are many tools to ensure that websites comply with the GPDR. The main examples are Normshiled, GDPR-pass, 2GDPR, and EZIGDPR.

Normshiled asks 10 questions and the website operator answers yes or no to indicate if their website meets the GDPR requirements. However, the questions are quite conceptual and comprehensive; therefore, this tool cannot identify if the website really fits each item of the GDPR. For example, one of the questions is as follows:

"Are you sure that your organisation has a privacy policy that covers GDPR requirements? If you are not familiar with the GDPR requirements, the answer itself is meaningless."

Our consent conditions are a specific version of one of Normshield's questions *"7. Do your data subjects from whom you hold and process data give explicit consent on your processes that use their data?"*. The difference is that our consent conditions are clearer, so anyone can find out whether a certain website complies with the GDPR. This can be done either by our automated tool or through manual checking.

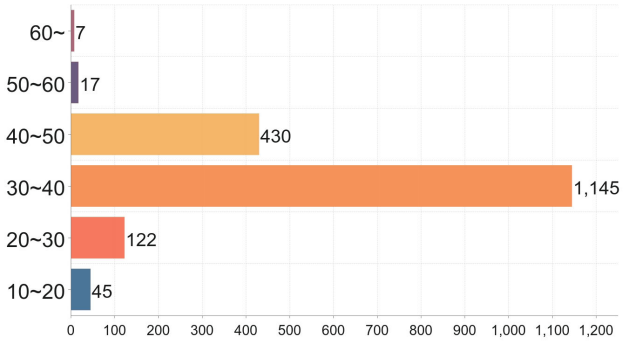


FIGURE 13. Results for Condition 4 by the automated tool, where the X-axis is the number of websites and the Y-axis shows the Flesch reading ease score.

GDPR-pass consists of 11 modules and 91 questions. It asks questions in simple language by taking the details of the GDPR as they are.

Does your organization inform the data subjects that you collect, use, view, or otherwise process their personal data?

This meter is useful for the Data Protection Officer (DPO), who knows the flow of personal information in their services, but is not familiar with the GDPR. DPOs can answer each question of the GDPR-pass, thereby identifying how well the service meets the GDPR. Our tool takes a step forward as we analyze judicial precedents and GDPR guidelines rather than simply explaining and reiterating GDPR items.

2GDPR focuses on cookies. If cookies are collected while loading a page, they are of course collected without consent. Therefore, we believe that the site does not comply with the GDPR. In addition, cookies can be divided into necessary cookies for proper online services and extra personal data, depending on the importance of the data. Furthermore, the tool can determine whether the cookies are automatically transferred overseas. 2GDPR is meaningful in that it performs an automated analysis, but its limitation is that it only involves cookies.

EZIGDPR checks which cookies are collected at each site and uses tags to find possible user tracking resources. In addition, similar to 2GDPR, it can be determined whether each cookie is transferred overseas so that we can determine the website's GDPR compliance status.

Other GDPR meters include SmartSurvey, Cookiebot, and Autoprivacy. Similar to the representative examples above, some of these reveal limitations, because the questions are too inclusive or difficult to answer for ordinary users who are not GDPR experts. Some other GDPR meters are also available for automated analysis of cookies. Compared to the above-mentioned GDPR meters, our meter has several strengths. First, we use specific and precise conditions to determine websites' GDPR conformity, which entails reliability and practical value. Second, our conditions focus on consent in particular. Third, our tool is automated that has more specific contents than other GDPR meters. Finally, our consent conditions and tools are supported by a solid legal basis.

VII. DISCUSSION

A. COMPLIANCE WITH GDPR

The purpose of this study was to examine whether the world's major websites are in compliance with the GDPR in terms of data privacy protection and user consent. Based on our four consent conditions, the data protection aspect was tested by examining the consent conditions of the 5,045 websites (for consent conditions 1, 2, 3) and 4,154 websites (for consent condition 4) under the GDPR jurisdiction. Three of our conditions proved that the majority of controllers did not completely comply with GDPR provisions. For consent conditions 1 and 3–2, most websites were GDPR compliant. This is in contrast with consent conditions 2, 3–1, and 4, where most websites did not observe the GDPR.

The 10,000 websites generally showed lower conformance rates than the top 500 websites. The conformance rates of the two groups for C1 were similar, but there was a large difference between top global websites(from top 500 websites) and general websites(from top 10,000 websites). The conformance rates for C2 and C3-1 were approximately 10% lower for the latter group. For C4, the conformance rate was lower for general web services, but the average readability score was slightly higher for the general web services. The popular web services and general web services showed similar readability scores. Since the conformance rate was too low, 3.4% and 1.4%, respectively, for each group, the gap does not seem to be a meaningful deviation.

Conformance to the consent condition did not vary significantly among website categories. For example, when we consider a website that includes porno video, manga, and chatting service for sex as an adult website, the Flesch reading ease score in the privacy policy of an adult website is 35.0, which is roughly equivalent to the average of the entire website (36.3). Surprisingly, one of the four websites with scores above 40 that met consent conditions 1, 2, and 3 is also an adult website. Even an adult website may admit the fact that it does not neglect privacy issues compared to a general website.

B. WEB ACCESSIBILITY ISSUE

With the enforcement of the GDPR, not only web service providers in Europe but also those that provide services to European citizens make sure their businesses are GDPR Compliant. Combined with the universal nature of online services, websites based outside of Europe should take additional steps to comply with the GDPR. However, some websites use their own methods to avoid being restricted by the GDPR. There are three main ways. First, some websites restrict access from Europe (''Indiatimes.com,'' ''Crunchyroll.com,'' ''Hclips.com,'' ''Hotstar.com''). For example, ''Indiatimes.com'' prevents access to services when accessed from a European IP. When we enter the website, the following banner is displayed:

Hello, we are currently not providing access or use of our website/mobile application to our users in Europe.

Second, some websites (''Foxnews.com,'' ''op.gg,'' ''Slickdeals.net'') restrict sign up. For example, when we access ''Foxnews.com'' with a non-European IP, we can log in and create an account. However, when we access it with a European IP, there is no button to log in or create an account. Similarly, when we access ''Slickdeal.net'' with a European IP, we see a banner confirming the EEA user status without any problem. However, if we click the EEA user button and click the ''Log In'' or ''Sign Up'' button, we cannot log in or create an account. Instead, it displays the following sentence:

You confirmed you are an EEA user, and our site does not support user accounts from the EEA.

Third, some websites (''Target.com,'' ''Bestbuy.com'') ask users to select the country they are currently in when accessing the website, but they are asked to select a limited number of countries, not all countries. For example, when we access ''Bestbuy.com,'' we have to select our country from among Canada, the United States, and Mexico.

C. LIMITATIONS

In this study, we analyzed *Alexa.com*'s top websites, but there are still some limitations.

While analyzing *Alexa.com*'s top 500 websites manually, we found that we were unable to sign up for several financial sites if we did not have a bank account or credit card (e.g., Bankofamerica.com) from certain financial institutions. As a result, nine financial sites were excluded and the financial category was excluded from the analysis.

According to statistics provided by Google, as of January 2021, there are 1,826,089,359 websites worldwide. Therefore, while we can claim that our four consent conditions apply to the majority of websites, it is difficult to argue that they apply to every single website.

We have created an automated tool, but there are still some errors. While this solution is valuable for research purposes, it has low marketability. Bringing it out on the market is another story, as compliance with the GDPR is so critical that even a slight error in this regard can be a serious flaw.

VIII. CONCLUSION

Consent is a means to guarantee privacy-related rights to data subject, and at the same time, it is a procedure from companies to ensure the fair use of personal information. However, if excessively detailed consent is obtained, consent paradox [48] occurs, which is rather ineffective. Therefore, an appropriate level of consent procedure is required, and in this paper, four consent conditions were created based on the GDPR. In addition, in order to confirm whether this consent condition is practically valid from the point of view of the GDPR, relevant guidelines were analyzed and precedents were investigated. In this study, four quantifiable consent conditions were established and tested to determine websites' GDPR compliance. We empirically measured whether actual web services correspond to these user consent conditions. We then created a tool to ensure that each website was

compliant with the GDPR items of consent. Surprisingly, we found that the majority of data controllers, including most popular websites, failed to comply with the regulation by not adopting simple steps, such as presenting a consent button or providing a directly visible privacy policy. There is a limitation in that the tool is not 100% accurate and the consent conditions are not legally enforced criteria. However, there is still a strong necessity for websites to pay more attention to consent. In addition, we sought to overcome some limitations of the existing GDPR meters, such as simply relying on the subjective judgments of respondents, and automated GDPR meters that only analyze cookies. While our tool is not flawless either, it remains under development. Our tool, which checks for the GDPR compliance, can be a reference for future studies. Based on this study, there may be many consent conditions related to consent for personal information in the future. Increasing the number of consent conditions regarding the personal information may help improve the privacy rights of the data subject, while also reducing the company's fear of violating the GDPR.

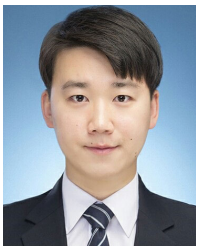
REFERENCES

[1] W. Oh and H. C. Lucas, Jr., "Information technology and pricing decisions: Price adjustments in online computer markets," *MIS Quart.*, vol. 30, pp. 755–775, Sep. 2006.
[2] A. Datoo, "Data in the post-GDPR world," *Comput. Fraud Secur.*, vol. 2018, no. 9, pp. 17–18, Sep. 2018.
[3] (2019). *Social Engine Market Share Worldwide: January 2021*. [Online]. Available: https://perma.cc/TV4D-86QS
[4] *Regulation EU 2016/679 of the European Parliament and of the Council (GDPR)*, Eur. Commission, Europe, 2016.
[5] H. Perera, W. Hussain, D. Mougouei, R. A. Shams, A. Nurwidyantoro, and J. Whittle, "Towards integrating human values into software: Mapping principles and rights of GDPR to values," in *Proc. IEEE 27th Int. Requirements Eng. Conf. (RE)*, Sep. 2019, pp. 404–409.
[6] S. Mittal and P. Sharma, "The role of consent in legitimising the processing of personal data under the current EU data protection framework," *Asian J. Comput. Sci. Inf. Technol.*, vol. 7, no. 4, pp. 76–78, 2017.
[7] C. Tikkinen-Piri, A. Rohunen, and J. Markkula, "EU general data protection regulation: Changes and implications for personal data collecting companies," *Comput. Law Secur. Rev.*, vol. 34, no. 1, pp. 134–153, Feb. 2018.
[8] C. Matte, C. Santos, and N. Bielova, "Purposes in IAB Europe's TCF: Which legal basis and how are they used by advertisers?" in *Proc. Annu. Privacy Forum*. Cham, Switzerland: Springer, 2020, pp. 163–185.
[9] (2019). *GDPR-Pass*. [Online]. Available: https://gdpr-pass.com/
[10] (2019). *2GDPR*. [Online]. Available: https://2gdpr.com/
[11] (2019). *EZIGDPR*. [Online]. Available: https://www.ezigdpr.com/
[12] D. E. O'Leary, S. Bonorris, W. Klosgen, Y.-T. Khaw, H.-Y. Lee, and W. Ziarko, "Some privacy issues in knowledge discovery: The OECD personal privacy guidelines," *IEEE Expert*, vol. 10, no. 2, pp. 48–59, Apr. 1995.
[13] *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*, OECD, Paris, France, 1980.
[14] P. Ashley, C. Powers, and M. Schunter, "From privacy promises to privacy management: A new approach for enforcing privacy throughout an enterprise," in *Proc. Workshop New Secur. Paradigms (NSPW)*, 2002, pp. 43–50.
[15] R. F. Filho and M. Jeffery, "Information technology and workers' privacy: Notice and consent," *Comp. Lab. Law Pol'y J.*, vol. 23, pp. 551–552, 2002.
[16] E. B. Cleff, "Privacy issues in mobile advertising," *Int. Rev. Law, Comput. Technol.*, vol. 21, no. 3, pp. 225–236, Nov. 2007.
[17] M. Goddard, "The EU general data protection regulation (GDPR): European regulation that has a global impact," *Int. J. Market Res.*, vol. 59, no. 6, pp. 703–705, Nov. 2017.
[18] B. W. Schermer, B. Custers, and S. van der Hof, "The crisis of consent: How stronger legal protection may lead to weaker consent in data protection," *Ethics Inf. Technol.*, vol. 16, no. 2, pp. 171–182, 2014.
[19] J. A. Obar and A. Oeldorf-Hirsch, "The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services," *Inf., Commun. Soc.*, vol. 23, no. 1, pp. 1–20, 2018.
[20] A. M. Mcdonald, R. W. Reeder, P. G. Kelley, and L. F. Cranor, "A comparative study of online privacy policies and formats," in *Proc. Int. Symp. Privacy Enhancing Technol. Symp.* Berlin, Germany: Springer, 2009, pp. 37–55.
[21] B. Fabian, T. Ermakova, and T. Lentz, "Large-scale readability analysis of privacy policies," in *Proc. Int. Conf. Web Intell.*, Aug. 2017, pp. 18–25.
[22] J. S. Daniel, "Privacy self-management and the consent dilemma," *Harvard Law Rev.*, vol. 126, pp. 1880–1883, 2013.
[23] R. W. Reeder, P. G. Kelley, A. M. McDonald, and L. F. Cranor, "A user study of the expandable grid applied to P3P privacy policy visualization," in *Proc. 7th ACM Workshop Privacy Electron. Soc.*, 2008, pp. 45–54.
[24] M. Tabassum, A. Alqhatani, M. Aldossari, and H. Richter Lipford, "Increasing user attention with a comic-based policy," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, p. 200.
[25] I. Sanchez-Rola, M. Dell'Amico, P. Kotzias, D. Balzarotti, L. Bilge, P.-A. Vervier, and I. Santos, "Can I opt out yet?: GDPR and the global illusion of cookie control," in *Proc. ACM Asia Conf. Comput. Commun. Secur.*, Jul. 2019, pp. 340–351.
[26] C. Iordanou, G. Smaragdakis, I. Poese, and N. Laoutaris, "Tracing cross border Web tracking," in *Proc. Internet Meas. Conf.*, Oct. 2018, pp. 329–342.
[27] M. Degeling, C. Utz, C. Lentzsch, H. Hosseini, F. Schaub, and T. Holz, "We value your privacy. . . Now take some cookies: Measuring the GDPR's impact on Web privacy," 2018, *arXiv:1808.05096*. [Online]. Available: http://arxiv.org/abs/1808.05096
[28] M. Trevisan, S. Traverso, E. Bassi, and M. Mellia, "4 years of EU cookie law: Results and lessons learned," *Proc. Privacy Enhancing Technol.*, vol. 2019, no. 2, pp. 126–145, Apr. 2019.
[29] C. Utz, M. Degeling, S. Fahl, F. Schaub, and T. Holz, "(Un)informed consent: Studying GDPR consent notices in the field," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, Nov. 2019, pp. 973–990.
[30] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal, "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2020, pp. 1–13.
[31] C. Matte, N. Bielova, and C. Santos, "Do cookie banners respect my choice? Measuring legal compliance of banners from IAB Europe's transparency and consent framework," 2019, *arXiv:1911.09964*. [Online]. Available: http://arxiv.org/abs/1911.09964
[32] *Deliberation of the Restricted Committee San-2019-001 of 21 January 2019 Pronouncing a Financial Sanction Against Google LLC*, Commission Nationale de l'Informatique et des Libertés, French Data Protection Authority, Paris, France, 2019.
[33] *Conseil d'Etat Ruling as the Supreme Administrative Court, no. 430810, Google LLC*, Council State, Ruling Supreme Administ. Court, Paris, France, 2020.
[34] M. Butterworth, "The ICO and artificial intelligence: The role of fairness in the GDPR framework," *Comput. Law Secur. Rev.*, vol. 34, no. 2, pp. 257–268, Apr. 2018.
[35] *Guidelines 05/2020 on Consent Under Regulation 2016/679*, Eur. Data Protection Board, Brussels, Belgium, 2020.
[36] *Guidelines on Transparency Under Regulation 2016/679*, Article 29 Data Protection Working Party, Brussels, Belgium, 2018.
[37] United Kingdom Information Commissioner's Office. (2018). *Guidance on Consent-What is Valid Consent?* [Online]. Available: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/consent/what-is-valid-consent/
[38] Dutch Data Protection Authority (Autoriteit Persoonsgegevens, Dutch DPA). (2019). *Dutch DPA Fines Tennis Association*. [Online]. Available: https://autoriteitpersoonsgegevens.nl/en/news/dutch-dpa-fines-tennis-association
[39] Privacy and Information Security Law Blog. (2020). *Dutch DPA Fines Royal Dutch Tennis Association 525,000 Euros for Illegal Data Sale*. [Online]. Available: https://www.huntonprivacyblog.com/2020/03/06/dutch-dpa-fines-royal-dutch-tennis-association-525000-euros-for-illegal-data-sale/
[40] *Besluit Tot Het Opleggen Van Een Bestuurlijke Boete Decision Toimpose an Administrative Fine*, Dutch Data Protection Authority, Autoriteit Persoonsgegevens, Dutch DPA, Hague, The Netherlands, 2020.
[41] *C-673/17—Planet49, Judgement of the Court (Grand Chamber) of 1 October 2019*, Court Justice Eur. Union, Luxembourg City, Luxembourg, 2019.

[42] *6 Ob 140/18h*, Austrian Supreme Court, Vienna, Austria, 2018.
[43] G. Fritz. (2018). *First Supreme Court Decision on GDPR: Austrian Supreme Court Rules on 'Prohibition of Consent Bundling' in General Terms and Conditions*. [Online]. Available: https://digital.freshfields.com/post/102f7ew/first-supreme-courtdecision-on-gdpr-austrian-supreme-court-rules-on-prohibitio
[44] R. Flesch, "How to write plain english: A book for lawyers and consumers," 2014.
[45] (2019). *Alexa.com's Top Websites*. [Online]. Available: https://https://www.alexa.com/topsites
[46] L. Wood, "Programming the Web: The W3C DOM specification," *IEEE Internet Comput.*, vol. 3, no. 1, pp. 48–54, 1999.
[47] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin, and K. Aberer, "Polisis: Automated analysis and presentation of privacy policies using deep learning," in *Proc. 27th USENIX Secur. Symp. (USENIX Security)*, 2018, pp. 531–548.
[48] B. Bergemann, "The consent paradox: Accounting for the prominent role of consent in data protection," in *Proc. IFIP Int. Summer School Privacy Identity Manage.* Cham, Switzerland: Springer, 2017, pp. 111–131.

JUNHYOUNG OH received the B.S. degree from Korea University, in 2017, where he is currently pursuing the combined master's and Ph.D. degrees with the Graduate School of Information Security. He is also a member of the Risk Management Laboratory, Korea University. His research interests include privacy, de-identification, anonymization, and data analysis.

JINHYOUNG HONG received the M.A. degree in law and the B.A. degree in economics from Korea University, where she is currently pursuing the Ph.D. degree in law. She is also a Researcher with the Cyber Law Centre, Korea University. Her research interests include cybersecurity and international law, the law of international responsibility, and data protection. She was a recipient of the Global Ph.D. Fellowship granted by Korean Government.



CHANGSOO LEE received the B.S. degree from Korea University. His major is computer science, and intensified course completion of major is also computer science. He participated in AI-RUSH Contest handled by Naver Corporation and got the First Prize for the image-classification part as a Mel 1100 Spectrogram Image Classifier.

JEMIN JUSTIN LEE (Member, IEEE) received the B.A. degree from the Glion Institute of Higher Education, Switzerland, in 2013, and the M.S. and Ph.D. degrees in electrical and computer engineering from Yonsei University, Republic of Korea. He is currently a Postdoctoral Researcher with Korea University, Republic of Korea. His research interests include cyber security, cyber threat intelligence, cloud computing, network optimization, and cyberwarfare.

SIMON S. WOO received the M.S. and Ph.D. degrees in computer science from the University of Southern California, Los Angeles, the M.S. degree in electrical and computer engineering from the University of California at San Diego, San Diego, and the B.S. degree in electrical engineering from the University of Washington, Seattle. He was a Technical Staff Member (Technologist) for nine years at the NASA's Jet Propulsion Laboratory (JPL), Pasadena, where he was conducting research in the satellite communications, networking, and cybersecurity areas. He also worked at Intel Corporation and Verisign Research Lab. Since 2017, he has been a Tenure-Track Assistant Professor with SUNY Korea, South Korea, and a Research Assistant Professor with Stony Brook University. He is currently a Tenure-Track Assistant Professor with the SKKU Institute for Convergence and the Department of Applied Data Science and Software, Sungkyunkwan University.

KYUNGHO LEE received the Ph.D. degree from Korea University. He was a Former CISO at Naver Corporation. He was also a CIO, a CISO, and a CPO with Korea University. He is currently a Professor with the Graduate School of Information Security, Korea University, where he has been leading the Risk Management Laboratory, since 2012.