Toward Data Protection by Design: Assessing the Current State of GDPR-Compliance in Web Applications

1st Abdel-Jaouad Aberkane
Ghent University
Ghent, Belgium
abdeljaouad.aberkane@ugent.be

2nd Seppe vanden Broucke
Ghent University
Ghent, Belgium
seppe.vandenbroucke@ugent.be

3rd Geert Poels
Ghent University
Ghent, Belgium
geert.poels@ugent.be

***Abstract*—With the growing popularity of web applications, there is a corresponding need to ensure that they comply with relevant regulations and standards, such as the General Data Protection Regulation (GDPR), which mandates strict guidelines for processing personal data within the European Union (EU). In this paper, we leverage machine learning and natural language processing techniques to gather a dataset of web applications to evaluate their GDPR-compliance by scrutinizing their privacy policies. We present an overview of the current state of GDPR compliance among web applications and identify areas that require attention. The results show that, among other things, web applications have a relatively high level of GDPR-compliance, with most requirements being covered at around 80-90%. Furthermore, web applications in the US and India demonstrate higher compliance with GDPR than European web applications. Also, the findings show that a relatively high amount was spent on IT by organizations that did not meet the considered GDPR requirements. In short, this study reveals that there is still work to achieve GDPR compliance, particularly regarding providing clarity about user rights regarding data processing. By highlighting the areas where compliance falls short, our research offers a starting point for enhancing privacy engineering practices for web applications and establishing a more privacy-centric digital landscape.**

***Index Terms*—GDPR, Privacy, Web Applications, Data Protection, Machine Learning.**

1. Introduction

In recent years, the proliferation of web applications has transformed the digital landscape. The convenience and ubiquity of web applications have made them increasingly popular among users. At the same time, businesses have quickly recognized the potential of these applications to reach a wider audience. The rise of web applications has also been fueled by their flexibility, enabling developers to go beyond the limits of conventional desktop applications. For example, instead of installing a new program, users can simply visit the relevant website [13]. Moreover, web applications can serve multiple users concurrently with greater ease than desktop applications, and they offer simpler management, maintenance, and modification [6]. With the continued growth of web applications, there has been a corresponding need to ensure that these applications comply with relevant regulations and standards, such as the General Data Protection Regulation (GDPR).

The GDPR is a regulation by the European Union that came into effect on May 25, 2018. Its primary objective is strengthening and harmonizing data protection laws across the European Union, replacing the 1995 Data Protection Directive [24]. The GDPR provides a framework for processing personal data, defined as any information that can be used to identify an individual, such as a name or identification number. The regulation sets out rules for collecting, processing, and storing personal data. In addition, it gives individuals more control over their data, including the right to access, correct, and erase their data. Organizations that process the personal data of EU residents must comply with the GDPR, regardless of where the organization is based [24]. Non-compliance with the regulation can result in significant fines and other penalties. In short, the GDPR is designed to enhance personal data protection by giving individuals greater control over their personal information and creating a more harmonized data protection framework across the European Union (EU).

The GDPR has significant implications for web applications that process the personal data of EU residents. Web applications commonly collect and process personal data for various purposes, such as online shopping, social media, and healthcare. To comply with the GDPR, web applications must implement measures such as obtaining user consent, ensuring data security and privacy, and providing transparency in the data processing. However, despite the importance of GDPR-compliance in the web application paradigm, there is a lack of research on the current state of compliance among web applications. This paper aims to fill this gap by assessing the state of GDPR-compliance with web applications based on their privacy policy as privacy policies are the primary means of communicating data processing practices to potential users [22]. We analyze 3 930 privacy policies of organizations in Europe, the United States, and India, considering five key requirements of the GDPR. By identifying areas where web applications fall short in terms of compliance, this research can help inform the development of better privacy engineering practices for web applications. Ultimately, the research presented in this paper contributes to paving the way for data protection by design and the broader goal of creating a more privacy-conscious digital environment.

The paper will be structured as follows. First, Section 2 describes the relevant literature. Next, Section 3 delves into the background of web applications and GDPR re-

quirements. Section 4 describes the research approach of this study. After that, Section 5 presents the results and the limitations and provides pointers to future work. Finally, Section 6 summarizes and concludes this study.

2. Related Work

This section briefly discusses a review of related work pertaining to GDPR-compliance in web applications, identifying three distinct research streams. The first stream centers on developing GDPR-aware web applications. The second stream centers on web applications designed to facilitate GDPR-compliance. The final stream concentrates on GDPR-compliance within the broader context of the Internet, potentially including web applications.

Concerning achieving GDPR-aware web applications, several approaches have been developed. Among these approaches, we find PADRES, a tool for privacy, data regulation, and security developed to analyze web applications and aid in the compliance process [20]. Another example is RuleKeeper, a GDPR-aware personal data policy compliance system that can prevent various GDPR-compliance violations [8].

Another research paradigm related to web applications is literature focusing on web applications as a means to achieve GDPR-compliance. Romansky and Kirilov, for example, work toward a web-based application to clarify the specific requirements of the GDPR [23]. Along the same lines, Nokhbeh Zaeem et al. present *PrivacyCheck v3*, a publicly available browser extension that summarizes privacy policies with machine learning, addressing key questions related to the GDPR [18].

Finally, various studies were discovered when adopting a broader perspective and examining GDPR-compliance on the Internet at large. Degeling et al., for example, analyzed the wave of changes caused by the enactment of the GDPR by analyzing changes in popular websites in the European Union, e.g., 70% of websites updated their existing privacy policies. The authors conclude that the web became more transparent when GDPR came into force [4]. Diving further into GDPR-compliance on the Internet, Muller et al. [17] take a similar approach, i.e., focusing on privacy policies, and conclude that at least 76% of the privacy policies do not comply with at least one of the considered GDPR requirements. Along the same lines, Rahat et al. develop a convolutional neural network model and assess, using this model, the state of GDPR-compliance in privacy policies, concluding that only 3% of companies fully comply with GDPR in their privacy policies [21]. The sentiment of these conclusions is reflected in [16], where Kretschmer et al. conduct a literature survey assessing the impact of the GDPR on users and service providers in the context of the World Wide Web. The authors conclude, among other things, that the GDPR has positively affected privacy on the web by increasing transparency. However, the strict requirements set by the GDPR are only fulfilled by a minority of web services.

In conclusion, to the best of our knowledge, there is a dearth of academic research that comprehensively assesses the state of GDPR-compliance of web applications. Existing literature mainly discusses tools, techniques, and frameworks that can aid in achieving GDPR-compliance in addition to general assessments of the state of GDPR-compliance on the web. While these resources are undoubtedly valuable for developers and businesses seeking to ensure compliance with the GDPR, there is a need for more empirical research on the actual state of GDPR-compliance in the context of web applications to assess the extent to which web applications adhere to GDPR requirements. In this research, we attempt to fill this gap by analyzing over 3 930 privacy policies of web applications on their GDPR-compliance based on the disclosure of GDPR requirements.

3. Background

According to Jazayeri, a web application is an application accessed through a web browser [15]. Expanding upon this definition is the definition of PCMag Encyclopedia, stating that a web application is an application "in which all or some parts of the software are downloaded from the Web every time it runs" and is of three types: browser-based, client-based, and native mobile apps [7]. The latter is the definition followed in this paper.

The first type, i.e., browser-based web applications, follows the definition of Jazayeri. Client-based web applications, on the other hand, run without the browser. Instead, the application is installed on the user's computer or mobile device or is downloaded each session. Lastly, native mobile applications, e.g., Android applications, access the Web for additional information. Since web applications often collect and process large amounts of personal information from users, they must comply with the GDPR, given that they are located in the EU or process user data of EU citizens.

The GDPR sets out a framework for processing personal data by data controllers and processors, including requirements for obtaining valid consent, data minimization, and transparency in data processing activities. The GDPR also introduces new data subject rights, including the right to access, rectify, and erase personal data. To comply with the GDPR, organizations must outline, e.g., through a privacy policy, their data processing practices and procedures, including the lawful basis for processing personal data, the types of personal data processed, the purposes of the processing, and the retention periods for personal data [24]. The GDPR privacy policy must also describe the data subject rights available to individuals and the organization's procedures for fulfilling data subject requests. Additionally, the GDPR requires that organizations implement appropriate technical and organizational measures to ensure the security of personal data and report any data breaches to the relevant supervisory authority and affected individuals. In this study, we assess the state of compliance with web applications by considering the following core requirements derived from the GDPR: the appointment of a Data Protection Officer (**DPO**) or equivalent, disclosing the purpose and legal basis of processing personal data (**Purpose**), disclosing what type of data is acquired (**Acquired Data**), disclosing whether data will be shared with third parties (**Data Sharing**), and, lastly, listing the user's rights, scoped to the right to rectification and the right to erasure in this study (**Rights**) [17].

By assessing the state of compliance among web applications, this study aims to pave the way for further

research focusing on achieving the GDPR principle of data protection by design and by default which comprises implementing appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed [24]. To achieve this, we anticipate that the insights provided by this study can be used as guidelines during the first stage of web application development, namely requirements engineering [10, 14].

4. Research Approach

This section briefly describes the adopted research approach consisting of two phases: data collection and data analysis.

We compiled a dataset of 3 930 organizations with different levels of online presence and their corresponding privacy policies. This data was acquired from the Crunchbase database [3] combined with web scraping techniques to extract the relevant privacy policies. After that, the data set was validated to establish that each organization's privacy policy applied to, among other things, a web application. The irrelevant organizations were not considered for analysis. Next, all collected privacy policies were classified according to the five core requirements of GDPR using five different natural language processing-based classification models. The core requirements included in the classification process were DPO, Purpose, Acquired Data, Data Sharing, and Rights. The resulting models had a document precision of 0.908, 0.908, 0.928, 0.912, and 0.941, respectively.

We employed descriptive statistics—using Python, Tableau, and Power BI—to analyze the compiled dataset of 3 930 organizations' privacy policies and the corresponding compliance with GDPR core requirements. In particular, we focused on GDPR-compliance per geographical region, organization size, and IT expenditure—derived from the Crunchbase data set. Regarding the geographical region, we considered Europe, the United States of America (US), and India. Europe was considered because of the direct effect of the GDPR, the US was considered due to the international presence of its companies on the Internet, and, finally, India was considered due to its software services outsourcing industry [11, 25]. Regarding Europe, a distinction was made between EU member states and those that are not. Following, focusing on the organization's size, we distinguish between small and medium-sized enterprises (SMEs) and large enterprises (LEs) based on the number of employees according to the definition of the European Commission [2]. Since only the number of employees was available, the remaining requirements of the European Commission in defining SMEs, i.e., having either an annual turnover not exceeding 50 million euros *or* an annual balance sheet total not exceeding 43 million euros, were not considered. Finally, the approximate amount a company spends on IT per year was also considered. The results are presented in the next section.

5. Results

This section describes the results of this study by outlining the current state of GDPR-compliance in web applications, followed by a discussion of the results, an outline of some of the limitations of this study, and, finally, pointers toward future work.

TABLE 1. COVERAGE (%) OF GDPR REQUIREMENTS IN PRIVACY POLICIES OF WEB APPLICATIONS

DPO	Purpose	Acquired Data	Data Sharing	Rights
70.20%	89.19%	89.90%	84.83%	68.73%

Table 1 displays the percentage coverage of GDPR requirements in privacy policies of web applications. The results indicate that the Acquired Data requirement has the highest coverage at 89.90%, while the disclosure of the Rights requirement has the lowest coverage at 68.73%. In between, we find that the DPO, Purpose, and Acquired requirements were covered by 70.20%, 89.19%, and 84.83% of the considered organizations, respectively. Overall, the table suggests that web applications have a relatively high level of GDPR-compliance, with most requirements being covered at around 80-90% or more. However, there is still room for improvement, particularly in providing clarity about user rights regarding data processing, which is a fundamental aspect of GDPR-compliance.

TABLE 2. COVERAGE (%) OF GDPR REQUIREMENTS IN PRIVACY POLICIES OF WEB APPLICATIONS PER REGION

Region	DPO	Purpose	Acquired Data	Data Sharing	Rights
Europe	65.75%	88.30%	86.77%	76.83%	74.85%
India	67.96%	91.71%	90.61%	90.61%	50.28%
USA	72.76%	89.47%	91.52%	88.69%	66.82%

Table 2 provides an overview of the GDPR-compliance in privacy policies of web applications across three different geographical regions: Europe, India, and the US. The findings indicate that web applications in the US exhibited the highest level of GDPR-compliance in disclosing the requirement for a Data Protection Officer (DPO). Concerning the Purpose requirement, web applications in India performed the best, followed closely by the US and Europe. Regarding the Acquired Data requirement, web applications in the US performed the best, followed by those in Europe, then India. Regarding the Data Sharing requirement, the results reveal that web applications in India performed the best, followed closely by those in the US and Europe. Finally, European web applications performed the best in terms of meeting the Rights requirement, followed by those in the US and India. Overall, the results of this study indicate that web applications in the US and India exhibit higher levels of compliance with GDPR in four of the five requirements analyzed compared to European web applications. It is worth noting, however, that the differences in compliance rates are relatively small and consistent across regions, except for the Rights requirement, where Indian web applications underperformed their European and American counterparts by a more considerable margin. Further

research is needed to explore the factors that may account for these differences and to evaluate the effectiveness of GDPR implementation in different regions.

TABLE 3. NUMBER OF REQUIREMENTS COVERED IN PRIVACY POLICIES OF WEB APPLICATIONS PER REGION

	Number of requirements covered					
Region	**0/5**	**1/5**	**2/5**	**3/5**	**4/5**	**5/5**
Europe	1.91%	5.35%	7.87%	13.53%	25.84%	45.49%
India	2.21%	3.31%	5.52%	16.57%	34.81%	37.57%
USA	2.17%	4.06%	6.31%	10.90%	22.94%	53.63%

Table 3 presents an analysis of the extent to which web applications in three geographical regions (Europe, India, and the US) comply with GDPR, considering a spectrum ranging from non-compliance (zero of the five GDPR requirements met) to full compliance (all five requirements met). The results indicate that most web applications across all regions fully complied with the considered GDPR requirements, with the highest compliance rate observed in the US, followed by Europe and India. Regarding compliance with four of the five requirements, Indian web applications performed the best, followed by those in Europe and the US. Furthermore, in this case, the Rights requirement was generally the remaining GDPR requirement that was not satisfied. Next, a similar order was found for compliance with three of the five requirements: Indian web application top the list, followed by Europe and the US. Regarding compliance with two of the five requirements, European web applications exhibited the highest compliance rate, followed by those in the US and India. The same was observed for compliance with one of the five requirements, with Europe leading the list; however, now followed by India, then the US. Next, if only one requirement was met, the requirement in question was—in most cases—the requirement of Acquired Data. Finally, a small percentage of web applications failed to meet any of the considered GDPR requirements, with the lowest rate of non-compliance observed in Europe, followed by the US and India.

Fig. 1 provides a visual representation of the average number of GDPR requirements complied with by countries that host at least ten web applications in the European region. The top five best-performing countries, based on this metric, are Bulgaria (4.75), Poland (4.36), Finland (4.34), the Czech Republic (4.29), and the Netherlands (4.14). In contrast, the five countries with the lowest average compliance rates are France (3.71), Switzerland (3.70), Italy (3.67), Ukraine (3.60), and Germany (3.57). Overall, the results show that web applications in Europe meet at least, on average, 3.57 of the considered five GDPR requirements.

Table 4 compares the coverage of GDPR requirements in privacy policies of web applications in EU member states and non-EU member states. The results reveal that web applications in non-EU member states perform better than the EU member states concerning disclosure of the DPO, Acquired Data, Data Sharing, and Rights requirements. Web applications in the EU member states perform

Figure 1. The average number of GDPR requirements that are complied with by European countries that host at least ten web applications.

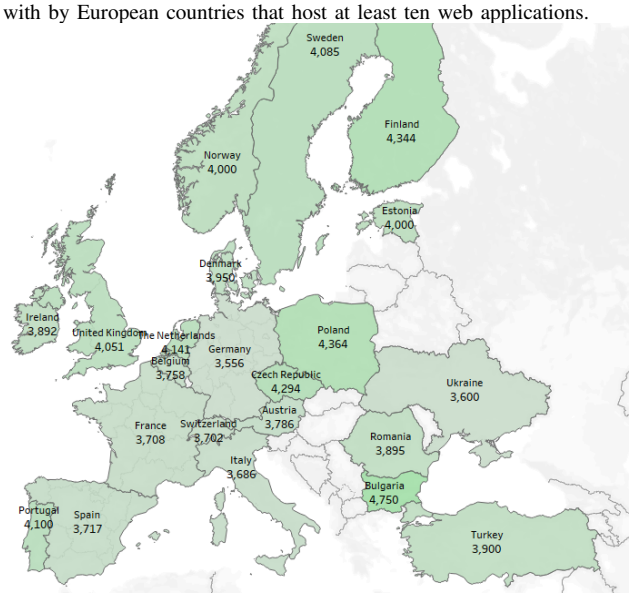

TABLE 4. COVERAGE (%) OF GDPR REQUIREMENTS IN PRIVACY POLICIES OF WEB APPLICATIONS IN EUROPE

Region	**DPO**	**Purpose**	**Acquired Data**	**Data Sharing**	**Rights**
EU member	62.41%	78.72%	88.23%	72.48%	85.67%
Non-EU member	69.65%	70.32%	88.39%	81.92%	88.06%

better in meeting the remaining requirement, i.e., Purpose. Nevertheless, in general, the differences in coverage between the two regions are relatively small, suggesting that GDPR-compliance in privacy policies is somewhat consistent across different regions.

TABLE 5. COMPARISON OF THE COVERAGE (%) OF GDPR REQUIREMENTS IN PRIVACY POLICIES OF WEB APPLICATIONS IN SMES AND LES.

Size	**DPO**	**Purpose**	**Acquired Data**	**Data Sharing**	**Rights**
SME	69.84%	89.01%	89.76%	84.44%	67.50%
LE	74.08%	91.43%	91.43%	88.16%	78.78%

Table 5 summarizes the coverage of GDPR requirements in web applications' privacy policies of SMEs and LEs. The findings indicate that SMEs exhibit a relatively lower level of compliance than their larger counterparts across all GDPR requirements. Specifically, the difference in fulfilling the Rights requirements is noteworthy, while the coverage of the remaining requirements is relatively similar. These outcomes highlight the significance of GDPR compliance for businesses of all sizes, with a special emphasis on SMEs that may require additional assistance to meet the GDPR criteria.

TABLE 6. AVERAGE IT EXPENDITURE PER YEAR PER NUMBER OF GDPR REQUIREMENTS MET.

Number of requirements met	Average IT spent by SMEs	Average IT spent by LEs	Average IT spent
0/5	$66 088 237.06	ND	$62 434 893.89
1/5	$12 774 863.75	$80 616 477.50	$23 109 479.67
2/5	$20 657 822.56	$19 231 948.57	$26 207 828.75
3/5	$32 340 115.15	$38 084 544.29	$32 926 915.30
4/5	$27 295 695.50	$429 968 418	$94 956 495.94
5/5	$24 265 097.35	$187 090 310.19	$69 478 888.17

Table 6 provides a picture of the average IT spending, i.e., the approximate amount a company spends on IT per year, of organizations with the number of GDPR requirements met. When focusing on SMEs, the results show that organizations whose privacy policy did not disclose any of the five considered GDPR requirements spent the most on IT. The least amount was spent in the category of web applications that met two of the five GDPR requirements. Shifting to LEs, the results reveal that the most amount, with a distance, was spent in the category of web applications that met four of the five GDPR requirements, followed by those that met all requirements. The least amount was spent by organizations whose privacy policy met two of the five GDPR requirements. No expenditure data was available on LEs that did not meet any requirements. The last column shows the average IT spent, regardless of the organization's size. Note that this column also includes companies that could not be classified as SME or LE due to the absence of relevant data. Thus, in general, most was spent in the category of privacy policies that met four of the five GDPR requirements. The least amount was spent in the category that met one requirement. Surprisingly, similar to the SMEs, a relatively high amount was spent on IT by organizations that did not meet any of the considered GDPR requirements.

5.1. Discussion

The findings of this study indicate that compliance with the GDPR Rights requirement, specifically the disclosure of the right to delete and rectify user data, is the least met among the considered requirements. This finding is consistent with recent research on facilitating user rights as per the GDPR, even if user data is stored in compliance with the GDPR. The research by Bufalieri et al., for example, demonstrates that a significant proportion of data controllers that handle requests for data access have flaws in identifying users or in their phase of sending the data, thereby exposing users to new threats [1]. Additionally, Di Martino et al. were able to impersonate data subjects and obtained full access to their personal data from data processing organizations, emphasizing the challenges in facilitating data subjects' rights [5].

Moreover, the study results show that organizations based in India perform well with regard to GDPR compliance, despite the GDPR not necessarily applying to its organizations. This finding may be attributed to Indian IT start-ups' investment in GDPR compliance [12], given that the EU has been one of the biggest markets for the Indian outsourcing sector [11]. Similarly, organizations based in the US perform well, potentially due to the robust data protection legal framework in place, such as the California Consumer Privacy Act, which shares similarities with GDPR [19], making it easier for US-based companies to implement GDPR requirements.

The results also reveal that SMEs have challenges complying with the GDPR requirements. These challenges be ascribed to limited resources and expertise [9]. Compliance requires investments in technical and organizational measures, which can be costly for SMEs with limited budgets. Additionally, the GDPR requires SMEs to appoint a Data Protection Officer (DPO), which can be a significant burden for smaller companies with limited human resources. As a result, SMEs may struggle to keep up with the complexity and the costs of GDPR compliance, which may pose a significant risk to their operations and reputation. However, large budgets alone might not be sufficient. The average IT expenditure per number of requirements satisfied, as presented in Table 6, shows that SMEs complying with all considered requirements spent less on IT than SMEs that complied with three or four of the five considered requirements.

5.2. Limitations

This study has several limitations that should be considered when interpreting its findings. First, the assessment of GDPR compliance was based on privacy policies, which may not provide a complete and accurate reflection of the data processing activities of organizations. It is possible that organizations may be complying with the GDPR in practice, even if their privacy policies do not explicitly state this, and vice versa.

Furthermore, this study focused on only five key GDPR privacy policy requirements, which is a limited scope compared to the comprehensive set of requirements established by the GDPR. Therefore, the findings of this study may not provide a complete picture of GDPR compliance across all organizations.

It should also be noted that organizations in the US and India do not necessarily have to comply with the GDPR if they do not process user data from EU citizens. However, organizations from these regions were included to draw comparisons with organizations from Europe, as the latter must comply with the GDPR.

Additionally, the absence of a privacy policy on an organization's website—or technical difficulties encountered during the scraping procedure—resulted in exclusion from the analysis. While this exclusion was necessary to maintain the integrity and consistency of the dataset, it is possible that these organizations may have been compliant with the GDPR in reality. However, this falls out of the scope of this study.

5.3. Future Work

While this study provides valuable insights into the current state of GDPR compliance in web applications regarding the principle of data protection by design and

by default, further research is necessary to build on these findings and address non-compliance issues.

In particular, we intend to develop a recommender-based system that can assist in achieving GDPR compliance by design and by default by identifying non-compliant requirements at the outset of the software development. Such a system could leverage machine learning algorithms to analyze software requirements and identify requirements that potentially contradict the GDPR. Based on this analysis, the system could provide specific recommendations on how to address these issues and ensure GDPR compliance. This type of system has the potential to significantly improve the effectiveness of data protection by design in web applications—or relevant software systems as a whole—as it can help developers and organizations proactively identify and address compliance issues.

In addition, further research could focus on exploring the factors that may account for the differences encountered in GDPR compliance across regions and organizations. This research could help to shed light on the specific challenges that organizations in different regions and sectors face in achieving GDPR compliance. For example, it may be useful to investigate whether differences in cultural attitudes toward data privacy or variations in legal frameworks contribute to differences in GDPR compliance across regions. Furthermore, to mitigate one of the potential biases in our study, future research could concentrate on removing the dependency on privacy policies in assessing the GDPR compliance of organizations. This could involve developing more objective measures of compliance, such as through analyzing data controllers' actual data processing activities rather than relying on self-reported policies.

6. Conclusion

This study has outlined the current state of GDPR-compliance in web applications by assessing related privacy policies. The assessment considers five GDPR core privacy policy requirements, i.e., communicating the contact details of a Data Protection Officer or equivalent, disclosing the purpose and legal basis of processing personal data, disclosing what type of data is acquired, disclosing whether data will be shared with third parties, and listing the user's rights. The results show that web applications have a relatively high level of GDPR-compliance, with most requirements being covered at around 80-90%. However, there is still room for improvement, particularly in providing clarity about data subjects' user rights regarding data processing, which is a fundamental aspect of GDPR-compliance.

Our study also showed that web applications in the US and India exhibit higher levels of compliance with GDPR in four out of the five requirements analyzed compared to European web applications. The differences in compliance rates are relatively small and consistent across regions, except for the Rights requirement, where Indian web applications underperformed their European and American counterparts by a more substantial margin. Furthermore, most web applications across all regions fully complied with the considered GDPR requirements, with the highest compliance rate observed in the US, followed by Europe, then India. Moreover, if four of the five GDPR requirements were met, in most cases, the Rights requirement was the remaining GDPR requirement that was not satisfied. Also, the results show that web applications in Europe meet, considering the five GDPR requirements, on average, 3.57 of the GDPR requirements. Finally, the analysis showed that a relatively high amount was spent on IT by organizations that did not meet any of the considered GDPR requirements.

Overall, our findings suggest that while web applications generally exhibit high levels of compliance with GDPR, there is still scope for improvement, particularly for organizations in the EU, as they must comply with the GDPR. By shedding light on compliance shortcomings, this research can add to the GDPR ideal of data protection by design and by default, as professionals can tailor privacy engineering approaches according to the insights provided by this study.

References

[1] Luca Bufalieri, Massimo La Morgia, Alessandro Mei, and Julinda Stefa. Gdpr: when the right to access personal data becomes a threat. In *2020 IEEE International Conference on Web Services (ICWS)*, pages 75–83. IEEE, 2020.

[2] European Commission. User guide to the SME definition, 2020.

[3] Crunchbase Database, 2022. https://crunchbase.com/ (visited: 2022-11-10).

[4] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy... now take some cookies: Measuring the GDPR's impact on web privacy. *Informatik Spektrum*, 42:345–346, 2019.

[5] Mariano Di Martino, Pieter Robyns, Winnie Weyts, Peter Quax, Wim Lamotte, and Ken Andries. Personal information leakage by abusing the gdpr'right of access'. USENIX, 2019.

[6] Nalaka Ruwan Dissanayake and Kapila Asanga Dias. Web-based applications: extending the general perspective of the service of web. In *10th International Research Conference of KDU (KDU-IRC 2017) on Changing Dynamics in the Global Environment: Challenges and Opportunities. Rathmalana*, 2017.

[7] PCMAG Encyclopedia. Definition of web application, 2023. https://www.pcmag.com/encyclopedia/term/web-application (visited: 2023-03-20).

[8] Mafalda Ferreira, Tiago Brito, José Fragoso Santos, and Nuno Santos. Rulekeeper: Gdpr-aware personal data compliance for web frameworks. In *2023 IEEE Symposium on Security and Privacy (SP)*, pages 1014–1031. IEEE Computer Society, 2022.

[9] M da C Freitas and Miguel Mira da Silva. Gdpr compliance in smes: There is much to be done. *Journal of Information Systems Engineering & Management*, 3(4):30, 2018.

[10] Athula Ginige. Web engineering: managing the complexity of web systems development. In *Proceedings of the 14th international conference on Software engineering and knowledge engineering*, pages 721–729, 2002.

[11] Brijesh Kumar Gupta. General data protection regulation and its impact on indian enterprises. *AKGEC Int J Technol*, 11:28–31, 2020.

[12] Gaurav Gupta and Shaji Joseph. Challenges in corporate governance in the implementation of gdpr for it start-up companies in india. *PalArch's Journal of Archaeology of Egypt/Egyptology*, 17(9):1663–1679, 2020.

[13] Johan Harjono, Gloria Ng, Ding Kong, and Jimmy Lo. Building smarter web applications with html5. In *Proceedings of the 2010 Conference of the Center for Advanced Studies on Collaborative Research*, CASCON '10, page 402–403, USA, 2010. IBM Corp. doi: 10.1145/1923947.1924015. URL https://doi.org/10.1145/1923947.1924015.

[14] Wei Huang, Ru Li, Carsten Maple, Hongji Yang, David Foskett, and Vince Cleaver. Web application development lifecycle for small medium-sized enterprises (smes)(short paper). In *2008 The Eighth International Conference on Quality Software*, pages 247–252. IEEE, 2008.

[15] Mehdi Jazayeri. Some trends in web application development. In *Future of Software Engineering (FOSE'07)*, pages 199–213. IEEE, 2007.

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

[17] Nicolas M. Müller, Daniel Kowatsch, Pascal Debus, Donika Mirdita, and Konstantin Böttinger. On GDPR compliance of companies' privacy policies. In Kamil Ekštein, editor, *Text, Speech, and Dialogue*, pages 151–159, Cham, 2019. Springer International Publishing. ISBN 978-3-030-27947-9.

[18] Razieh Nokhbeh Zaeem, Ahmad Ahbab, Josh Bestor, Hussam H Djadi, Sunny Kharel, Victor Lai, Nick Wang, and K Suzanne Barber. Privacycheck v3: Empowering users with higher-level understanding of privacy policies. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 1593–1596, 2022.

[19] Grace Park. The changing wind of data privacy law: A comparative study of the european union's general data protection regulation and the 2018 california consumer privacy act. *UC Irvine L. Rev.*, 10:1455, 2019.

[20] Fábio Pereira, Paul Crocker, and Valderi R.Q. Leithardt. Padres: Tool for privacy, data regulation and security. *SoftwareX*, 17:100895, 2022. ISSN 2352-7110. doi: https://doi.org/10.1016/j.softx.2021.100895. URL https://www.sciencedirect.com/science/article/pii/S2352711021001515.

[21] Tamjid Al Rahat, Tu Le, and Yuan Tian. Automated detection of GDPR disclosure requirements in privacy policies using deep active learning. *arXiv preprint arXiv:2111.04224*, 2021.

[22] Joel R. Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T. Graves, Fei Liu, Aleecia McDonald, Thomas B. Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, and Florian Schaub. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30 (1):39–88, 2015. ISSN 10863818, 23804742. URL http://www.jstor.org/stable/43917627.

[23] Radi Romansky and Kiril Kirilov. Architectural design and modelling of a web based application for GDPR clarification. In *AIP Conference Proceedings*, volume 2048, page 060006. AIP Publishing LLC, 2018.

[24] The European Parliament and the Council of European Union. REGULATION (EU) 2016/679. *Official Journal of the European Union*, pages 1–2, Apr 2016.

[25] Carol Upadhya. The global indian software labour force: It professionals in europe. *Indo-Dutch Programme on Alternatives in Development*, 2006.