



## A Comparative Study of Privacy Policies in E-Commerce Platforms

Chandrakanth Reddy Borra

Ph.D. Researcher, Department of Information Technology, University of the Cumberlands, USA

**ABSTRACT:** This is the digital era. Every device can generate and process information in the age of Internet of Things (IoT). Security of the new technologies and platforms is debated for various reasons. There are technological platforms that collect data from the users directly while some platforms use the processed data. The focus of this research study is on the privacy of the user data in e-commerce platforms. The privacy policies of e-commerce websites are compared and analyzed to check the different ways user data is collected and processed by careful manipulation of the privacy policies. Privacy policies are not just generic documents with standard terms but there are specific terms that are written using unique style, choice of words, readability scores, etc. so that the data can still be collected but not known to the users. This study aims at analyzing this specific aspect of privacy policies of the e-commerce websites.

**KEYWORDS:** E-commerce websites, privacy policies, user data analysis.

### I. INTRODUCTION

Privacy policies have become an essential part of companies. A user privacy policy essentially defines how the user data is collected and used by the company. E-commerce companies are chosen in this research paper. Three giant e-commerce companies - Amazon, eBay, and Walmart are selected and the privacy policies of these three companies are analyzed. These three companies are essentially e-commerce companies selling various items online to millions of customers throughout the world. They use websites and mobile applications to exhibit and sell their products. They use online platforms to advertise, collect data, process data, and share it if necessary, with other companies. There are many companies that fall into different categories who share and process the data collected by e-commerce companies. The research question is related to how different the privacy is policing of these three companies drafted and how do they ensure privacy to the users' data. The privacy policies of e-commerce companies are different compared to other companies. The policy clearly defines how the data of the users is collected through various platforms, how it is handled, and shared with others. There are different types of interactions a customer engages in with the company through the websites. These interactions generate a large amount of data that could have sensitive information and personal data. There are different levels of sensitivity involved in users' data for which the company should define the policies appropriately (Bennett & Raab, 2017). The privacy policy of an e-commerce company should essentially inform the users about the different ways the private data is collected and how it is used. It should also provide an option for the users to agree or deny the policy before the interaction begins. Users should also be allowed to access the data company has collected and modify it accordingly, if necessary (Chang, Wong, Libaque-Saenz, & Lee, 2018). The policy should ensure that all the rules are strictly followed, and no violation of any kind is permitted in the policy. It is necessary for any company to create and implement the user privacy policy because it is required by the law. The privacy laws are strictly enforced in the United States, EU, Australia and many other countries across the world. The privacy policy not only provides the details on the use of data but also builds the necessary trust with customers on how authentically the company follows the rules. It is also necessary that some applications and software platforms require the privacy policies to be implemented before they can process the user data. Another important requirement of the policy for both users and companies are that they can use the policy for legal purposes. The policy stands as a solid evidence to prove the violation of users or companies. Privacy policy comprises of a list of ways data is collected from the users, how cookies are placed and used by the websites, when the data can be processed or shared with the third-party companies, how the users can edit the information, how the data will be stored in case the company is shut down, and any other similar cases. The companies take time and legal consultation before creating a privacy policy so that there are no clauses left out or misrepresented. Privacy policies of e-commerce companies are more complex because all the business transactions and interactions are online and there are increasing number of software platforms and data processing tools that process the user data using a wide variety of methods. It is a requirement for both users and companies to stay and operate safe using a legally valid document in the form of privacy policy (Chua, Herbland, Wong, & Chang, 2017).



## II. LITERATURE REVIEW

### A. AppAware: A Model for Privacy Policy Visualization for Mobile Applications

This research work is related to reducing the difficulty in understanding the privacy policies. The policies are written in a technical and legal language that is not interpreted appropriately by all the users. There are many applications that take undue advantage of users skipping the analysis of the policy. This study proposes a method to visualize the details of the policy in an easy understandable format. This solution was designed only for android applications. However, the same methodology can be extended to other platforms. The solution is excellent for its simple design. The details are accurately extracted and presented to the users. Reflection - The study is related to how the issue of privacy policy interpretation can be solved. The current research study is related to the privacy policies of the e-commerce industry that also has several points in the policies that are quite difficult to understand. This study can be a solution for the problem.

### B. Privacy Policy Negotiation in Social Media

Social media applications have many multimedia data that is shared and viewed by millions of users. The recent privacy laws and policies restrict the users to access all the information if it is blocked by the users in the privacy settings. There are restricted interactions due to this conflict. A negotiation mechanism is proposed in this study. The study proposes two stages of negotiation method that use three heuristics. The results were satisfactory, and the heuristics could be modelled for higher accuracy. The social media interactions provided faster results with this method. The solution can be extended to other applications. The study is a technical solution that processes the information sharing conflict between the privacy policies. The relationship between the current study and this research is related to how a technical solution can be designed and implemented for a negotiation between the current policies and user interaction. A deeper analysis of this study will reveal how the methodology can be designed for developing a technical solution.

### C. The Creation and Analysis of a Website Privacy Policy Corpus

The research study proposes a method for automatically extracting and analyzing the privacy policies of the websites. This method designed a machine learning model with natural language processing mechanism to understand and interpret the text. The solution is an excellent method of diving deep into the policies of the websites. This method is implemented with good results. The research study analyzed over 115 policies and the annotations were also involved in the analysis. The annotations were processed, and the complexity and structure of the system were input to the NLP and machine learning processes. The solution has good results and found to be effective. The study is a good start for analyzing the privacy policies of the e-commerce websites. The solution is also applicable to the e-commerce websites that have many complex points in the policies. Analyzing them will reveal the trends of the market with respect to privacy laws and how new technologies are able to manipulate the laws.

### D. The role of privacy policy on consumers' perceived privacy

This research study is all about the perceived effectiveness of the privacy policies. The study has focused on how the users perceive the privacy policies. There are many policies that do not get attention of the users. Yet, some others are often disliked by the users. The analysis of this perception is the solution that conducted a survey of over 350 participants who use online banking services. The study evaluated the results and it was found that there are different aspects of Fair Information practice principles such as access, notice, security, and enforcement that often are perceived as the most important by the users. The study is not applicable to any particular field and can be extended to other industry such as e-commerce. This study is very closely related to the current research in which the privacy conflict and perception of the users who use e-commerce websites are studied. The privacy policies can be processed using automated tools, analyzed for user perception, and knowledge can be extracted for deeper analysis.

### E. Towards Automatic Classification of Privacy Policy Text

Understanding and analyzing the privacy policy using automated methods is the topic of this research study. The solution is related to how natural language processing can be used to automate and annotate the documents. User interface was also the contribution of this research study in which the solution can be easily visualized and understood by the users. The research analyzed different segments and sentences of the content that had the F1 scores of 0.78. A total of 7000 privacy policies were fed to the system that were classified for various parameters. The study has good results that can be improved bigger dataset. The analysis of many privacy policies can be extended to e-commerce websites. However, the design of the study needs to be changed for higher accuracy.



## F. Toward a Framework for Detecting Privacy Policy Violations in Android Application Code

Mobile applications are on the rise and the cyber-attacks through mobile platforms have also increased in the past few years. The research study proposes a solution in which the privacy policy and the code are compared to check any violations. This is an excellent method to check for privacy violations especially in the mobile platforms because the permissions and all the actions of the applications are embedded in the code. The solution has been implemented and a total of 477 android applications were analyzed. The code in 341 applications were found to be violating the privacy policies. The previously analyzed studies that have NLP mechanism to analyze policies can be combined with this method for efficient analysis. This mechanism can be directly implemented for e-commerce applications. The code of e-commerce mobile applications can be analyzed to check any violations. The legal regulations of privacy policies and laws can also be analyzed using automated methods to check for violations.

## G. Identifying the Provision of Choices in Privacy Policy Text

Privacy policies have become strict and the users are provided with many choices on their rights. This research study is related to a proposed method that analyzes the choice of the users and how they can exercise their right to use the application. This is an advanced method that not only analyzes the text of the policy but also the possible choices of the users.

This method was implemented, and the solution was able to help the users who browse websites and other applications to have a clear visualization of their choices. This method should be a highlight in e-commerce websites and applications that allow the user to select their choice of using or denying the permissions. The solution is an excellent method that fights for the rights of the users to accept or reject only the permissible options such as permission and notifications. This method can be enhanced with machine learning methods and a large dataset for higher accuracy.

## III. DESIGN METHODOLOGY

E-commerce sector has grown into a major industry in the last decade. The market size of ecommerce was \$1336B in 2014. The sector witnessed a continuous growth in the last 6 years and stood at \$3535B in the year 2019. The growth is estimated to be increased at a higher rate and reach \$6542B in the year 2023. E-commerce companies have become major trendsetters in many parts of the world (Global retail e-commerce market size 2014-2023 | Statista, 2020). E-commerce companies interact directly with the users on a large scale, typically in millions. The data generated from the users through websites contain personal information, browsing habits, financial information, details of the identity, family, etc. This information can be used against them for various purposes by the companies. There are strict regulations and laws implemented by the government and regulatory bodies should be followed by every company. However, there are many ways these laws and policies are exploited by the companies using different ways. This project focuses on how the privacy policies bypass the regulations and collect the required data for commercial purposes.

## IV. DATA COLLECTION

The privacy policies of the e-commerce websites are collected from the websites and then analyzed using different tools and approaches available. There are automated tools available online that can generate policies. However, the policies are checked for their relevance, language, perception, and user permissions. The results are based on an approach with different steps used to analyze the privacy policy from its creation to analysis. Simple online tools will be used to generate the policies and compare them with the policies of popular e-commerce websites. This research study is a qualitative analysis. Qualitative analysis has various types of Data Collection Methods. The case study is an important method in which various interviews of the experts in the field ID are collected and analyzed for different parameters. In this current research study, the privacy policies of The E-Commerce companies are collected from online sources. It is not necessary that the case study data collection requires interviews. Instead, the descriptive method of data collection in qualitative analysis requires the data to be collected from various sources in the form of interviews and questionnaires that can be transcribed to understand the results. A questionnaire is another popular method of data collection in the qualitative analysis where a large number of users are involved and their perception of users with respect to a particular technology or field can be analyzed using the responses and statistical methods. The data from all the collected privacy policies can be analyzed for different parameters. There are different components of any privacy policy because it is a written document that has been registered and submitted to the legal authorities for approvals (Obar & Oeldorf-Hirsch, 2020). The document can be studied for its language, complexity terms used, the complexity of the concepts and context, compliance with the regulations, data processing tools and applications, etc. These parameters are studied in the current research study to compare how each company is different in using the privacy policy. The consequences of the privacy policies are also studied by analyzing a few



legal issues and cases that the e-commerce companies have been involved in. The analysis of the privacy policy can also utilize their Technologies such as Natural Language Processing in order to understand the complexity of the language used. Case Method can be employed in order to understand the privacy policy of the major E-Commerce companies Amazon, eBay and Walmart. A specific comparison between the privacy policies of these two companies can clarify the way the data is collected and processed in compliance with all the data processing regulations. The analysis of the privacy policies will reveal several insights and results on how the privacy policies are used in favor of the companies in order to generate and process the user data. The privacy policies are also compliant with all the regulations and privacy laws yet they are able to collect the information and process it as per the need for business intelligence and marketing.

## V. FINDINGS

There are three major e-commerce companies in the world - Amazon, eBay, and Walmart. The privacy policies of these three companies are collected and a new privacy policy for a fictitious company is generated online using automated tools. The average values of readability parameters is compared with the respective values of the online generated template. The comparison is given in table 1 and graphs shown in figure 1. The values show that the readability parameters of the policies of the three companies are also different compared to each other. Amazon is the easiest policy to read followed by Walmart and eBay. Surprisingly, the policy generated online has lower score compared to all the three policies and scores at 33.5 making it the most difficult to read and analyze. The policies of the companies are analyzed and redesigned to have higher readability scored.

Table 1. Comparison of the readability parameters of the e-commerce companies.

	Amazon	eBay	Walmart	Average	Online Generated Privacy policy
	The policy has an average grade level of about 10. It should be easily understood by 15 to 16-year old.	The policy has average grade level of about 12. It should be easily understood by 17 to 18-year old.	The policy has average grade level of about 12. It should be easily understood by 17 to 18-year old.	Average readability of three companies	The policy has an average grade level of about 14. It should be easily understood by 19 to 20-year old.
Flesch Kincaid Reading Ease	52.2	42.8	44.1	46.3666667	33.5
Flesch Kincaid Grade Level	9.1	11.1	10.5	10.2333333	13.2
Gunning Fog Score	11.2	14.1	12.9	12.7333333	15.4

SMOG Index	8.8	10.5	10.3	9.86666667	12.4
Coleman Liau Index	13.7	14.6	15.2	14.5	15.5
Automated Readability Index	8.5	10.5	10.2	9.73333333	12.9
No. of sentences	259	906	411	525.333333	78
No. of words	3228	13875	5598	7567	1445
No. of complex words	571	2960	1284	1605	362
Percent of complex words	17.69%	21.33%	22.94%	0.20653333	25.05%
Average words per sentence	12.68	15.36	13.62	13.8866667	18.53
Average syllables per word	1.6	1.75	1.76	1.70333333	1.83

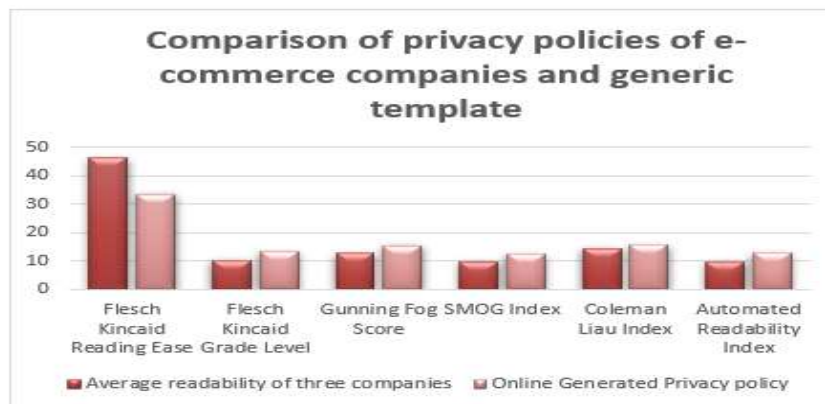


Figure 1. Graphs comparing the readability of policies of e-commerce companies and online template.

## VI. CONCLUSION

Analysis of the privacy policies of e-commerce websites is the focus of this research study. The privacy policies are studied for various parameters and compared with each other. The data collection method for the research study follows the case study method in which a comparison between the privacy policies of Amazon and eBay are studied. The research study can be extended to the analysis of various other privacy policies if an automated system can be used for analysis. The online generator privacy policies can also be included for the comparison as a standard.

## VII. ACKNOWLEDGEMENT

We would like to thank anonymous reviewers for their helpful feedback

## REFERENCES

1. Bennett, C. J., & Raab, C. D. (2017). The governance of privacy: Policy instruments in global perspective. Routledge.
2. Chang, Y., Wong, S. F., Libaque-Saenz, C. F., & Lee, H. (2018). The role of privacy policy on consumers' perceived privacy. *Government Information Quarterly*, 35(3), 445-459.
3. Chua, H. N., Herbrand, A., Wong, S. F., & Chang, Y. (2017). Compliance to personal data protection principles: A study of how organizations frame privacy policy notices. *Telematics and Informatics*, 34(4), 157-170.
4. Global retail e-commerce market size 2014-2023 | Statista. (2020). Retrieved 27 March 2020, from <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales/>
5. Liu, F., Wilson, S., Story, P., Zimmeck, S., & Sadeh, N. (2018). Towards automatic classification of privacy policy text. School of Computer Science Carnegie Mellon University, Pittsburgh, PA, Tech. Rep. CMU-ISR-17-118R and CMULTI-17-010.
6. Obar, J. A., & Oeldorf-Hirsch, A. (2020). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1), 128-147.
7. Paspatis, I., Tsohou, A., & Kokolakis, S. (2018). AppAware: A Model for Privacy Policy Visualization for Mobile Applications.
8. Sathyendra, K. M., Wilson, S., Schaub, F., Zimmeck, S., & Sadeh, N. (2017, September). Identifying the provision of choices in privacy policy text. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 2774-2779).
9. Slavin, R., Wang, X., Hosseini, M. B., Hester, J., Krishnan, R., Bhatia, J., ... & Niu, J. (2016, May). Toward a framework for detecting privacy policy violations in android application code. In *Proceedings of the 38th International Conference on Software Engineering* (pp. 25-36).
10. Such, J. M., & Rovatsos, M. (2016). Privacy policy negotiation in social media. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 11(1), 1-29.
11. Wilson, S., Schaub, F., Dara, A. A., Liu, F., Cherivirala, S., Leon, P. G., ... & Norton, T. B. (2016, August). The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1330-1340).