Evaluating Privacy Compliance and Usability Risks in Bangladeshi Mobile Applications: A GDPR-Based Sectoral Study

by

MD. Abdullah Al Sadik
23341135
Samia Semu
23341136

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

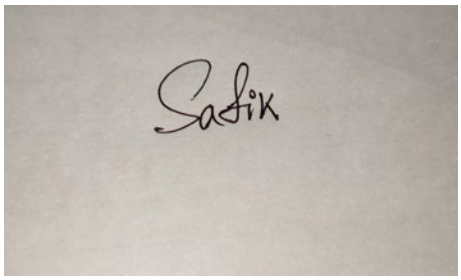
Department of Computer Science and Engineering
Brac University
June 2025

Declaration

It is hereby declared that

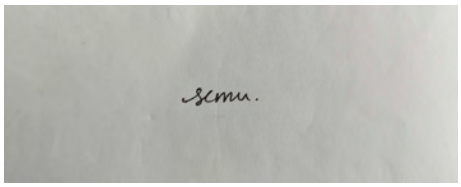
1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:



MD. Abdullah Al Sadik

23341135



Samia Semu

23341136

Approval

The thesis/project titled "Evaluating Privacy Compliance and Usability Risks in Bangladeshi Mobile Applications: A GDPR-Based Sectoral Study" submitted by

1. MD. Abdullah Al Sadik (23341135)
2. Samia Semu (23341136)

Of Spring, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 20, 2025.

Examining Committee:

Supervisor:

Dr. Farida Chowdhury
Professor
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam
Professor
Department
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Associate Professor Chairperson
Department of Computer Science and Engineering
Brac University

Abstract

The ever-growing increase of mobile applications in Bangladesh has intensified concerns over personal data privacy and regulatory compliance. Most of the mobile applications in Bangladesh treat sensitive user data with ambiguous or poor privacy regulation models, which subject customers to the risk of high privacy and security. This research performs an analysis of privacy compliance and usability risks of 32 popular Bangladeshi mobile applications in four major sectors: Healthcare, E-Commerce, Entertainment and mobile financial services on the basis of the European General Data Protection Regulation (GDPR) as one of the global standards, In our study, we focus on exploring how the GDPR compliance of Bangladesh mobile apps is presented in their privacy policies and how the definitions of such statements, whether in the form of explicit obligations or in the form of vague non-committal statements, affect the actual rate of compliance. The sector-based analysis reveals that e-commerce apps and healthcare apps showed a better range of compliance scores in the positive language framing context, whereas negative constructions in all the sectors showed the least compliance scores, which is an indicator of latent privacy risk. This study finds the outcome that identifies sector-specific faults and illustrates the effects of privacy policy wording on privacy usability as well as security usability in less developed nations such as Bangladesh. Regulators and policymakers can use the findings to design actionable insights that can guide decision-makers, app developers, and similar stakeholders interested in improving privacy governance and better protecting the data of digital users in dynamic online environments.

Keywords: Privacy, Privacy Policy, Privacy Policy Violations, GDPR.

Acknowledgement

Firstly, all praise to the Great Allah for whom our thesis have been completed without any major interruption.
Secondly, to our supervisor Dr. Farida Chowdhury ma'm for her kind support and advice in our work. She helped us whenever we needed help.
And finally to our parents, without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

The proliferation of mobile applications across the globe has brought to complicated challenges in the field of individual data protection and regulatory standards. This is of great concern in developing nations like Bangladesh, where the rate of embracing digital services is higher than developing their own legal and regulatory environment. The use of mobile applications at major fronts in several industries like Healthcare, E-Commerce, Entertainment, and Mobile Financial Services (MFS) has become a constant source of collecting, storing, and processing vast amounts of sensitive personal data such as financial information, health records, personal identifications and even behavioral patterns. With the growing popularity of these digital platforms as the intrinsic parts of the everyday life of the Bangladeshi citizenry, the issues related to their protection of privacy, safety of the data, and observance of the regulations are getting more pressing than ever. Since there are no strong laws to protect local data, most Bangladeshi app developers implement loosely framed privacy policy or they seek to match the surface recommendations of international regulations like the European Union General Data Protection Regulation (GDPR). Nonetheless, the inclusion of GDPR-related terms in the privacy policies cannot always suggest that there is real compliance or the meaningful protection of the end-users. (Liu et al.) [4]. Instead, the way privacy obligations are conveyed (by express promises or by indefinite and noncommittal statements) will make the crucial difference as to whether privacy is really complied with, or even usable, in practice. Most studies done on privacy policy analysis have mainly concentrated on keyword-based detection measures to determine compliance levels with the regulations (Raghav et al.) [11]. Although an effective tool, the methods may not sufficiently detect the semantic context of policy statements and their linguistic framing, thus providing erroneous results most of the time, or a misleading evaluation of compliance. This leaves a major research gap in assessing privacy usability or security usability, areas that not only assess the presence of elements of compliance but also the presentation of such elements such that the end-user can take action, understand, and trust them. These limitations triggered the motivation of this study that tries to fill the gap regarding the essentiality of more advanced and context-sensitive approaches to assessing the privacy policies of mobile applications in Bangladesh. This study will seek to introduce a more detailed explanation of how GDPR compliance is further reflected, as well as masked, in app privacy terms by performing a natural language

processing (NLP) analysis of semantic features. In addition, this paper can help regulators, developers and policy formers in the developing world and regions that are highly dynamic in terms of the creation of the digital ecosystem in an attempt to secure its privacy by disposing of useful information about the differences in a sector.

1.2 Research Questions

In a bid to answer the aforementioned challenges and add valuable findings to the usability of privacy and use of security, this study is informed by the following research questions:

- What do the mobile apps of Bangladesh do in terms of GDPR compliance within their privacy policies? How do they develop the compliance within privacy laws, and how does this impact the actual level of privacy compliance?
- Will the compliance method be applicable for various sectors among popular mobile applications in Bangladesh?
- What is the effectiveness of a hybrid approach based on a keyword analysis and contextual interpretation of the language to recognize the unknown privacy risks and enhance the precision of privacy compliance analysis based on mobile application privacy policies of Bangladesh?

1.3 Problem Statement

Despite the growing reliance on mobile applications in Bangladesh, there is limited understanding of how these apps comply with global privacy standards such as the GDPR. The privacy policies have very generic or unspecified contents, which makes it hard on the part of the users to know what happens with their information. It is not clear how app-developing stakeholders consider GDPR compliance to be addressed in their privacy policies and whether the approaches adopted are effective or just on the surface. In addition, existing models of analysis can fail to capture obscure or scenario-specific privacy vulnerabilities. This instigates the necessity of a more specific and scalable mechanism that will be able to not only identify explicit but also implicit privacy compliance problems. As such, this study aims at asking how much Bangladesh-based mobile applications adhere to the GDPR in terms of their privacy policies, how varied the approaches of the various sectors are, and whether a hybrid approach to analysis involving keyword-based and contextual NLP approaches in a combined framework can improve the exactness of the identification of privacy risks and violations.

1.4 Research Objectives

In consideration of the above research questions, the specific objectives of the study are the following:

- Investigate how the issue of GDPR compliance is reflected in the privacy policies of mobile applications in Bangladesh and determine to what extent they maintain compliance in reality.
- Carry out a sector-related assessment of the level of compliance with the GDPR requirements in the Healthcare, E-Commerce, Entertainment, and Mobile Financial Services industries of the Bangladeshi mobile app market.
- Make and test a hybrid framework of analysis that allows mixing fundamental NLP methods with contextual semantic analysis to increase identification of occurrences of privacy-related terms, as well as compliance lapses in privacy statements.

1.5 Methodology in Brief

This research initiated by selecting 32 mobile applications belonging to 4 distinct sectors, the crucial criterion was that the number of downloads that each app had in the background was no less than 100 thousand. After this selection procedure, we identified the 10 dimensions that dealt with GDPR as our privacy compliance benchmarks and consisted of 97 specific keywords. Our data collection phase involved meticulously gathering the Privacy Policies of all selected applications. These policies then underwent text preprocessing including tokenizing and vectorizing them in readiness to be analyzed. We then used the hit-ratio of keywords by app, which measures how many GDPR-consistent terms that the app had on their policies. In order to make our results more precise, we conducted a semantic validation procedure with LLM filtering to ensure the relevancy of the keywords identified. This enabled us to obtain accurate PCS / PRS scores, equivalents of privacy compliance of each of the apps. This approach was further expanded into sector-based analysis to find out trends in terms of various categories of applications, followed by an investigation into demographic implications of compliance. Our findings and conclusions were presented in the end of the whole process. This whole process is summarised in the flowchart (1.1) below.

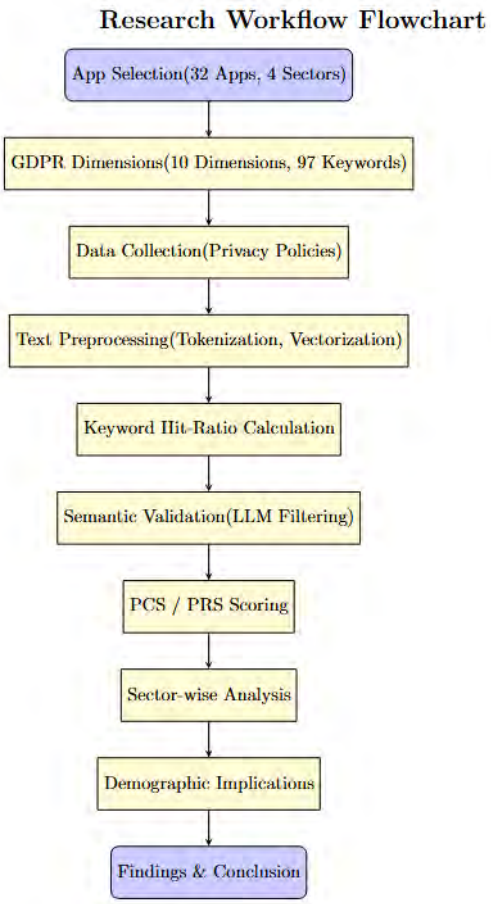


Figure 1.1: Workflow

Chapter 2

Literature Review

2.1 Background

2.1.1 Mobile Applications Privacy Policy

Mobile applications gather and process massive amounts of personal information including financial details, health care, location and user activity. A privacy policy is a legal paper that performs the duty of advising users on the methods of collecting, utilising, sharing, and securing their data. (Schaub et al.) [4] It outlines important aspects as data type, its processing intentions, sharing by third parties, and the users rights. Most of the privacy policies are, however, cumbersome, legalized, and are just foreign to the user. This intricacy poses serious questions to the privacy usability whereby an individual may concur to sign a consent form without effectively knowing what he is signing.

2.1.2 Major Privacy Laws

As a reaction to increasing data privacy concerns, a number of countries brought up privacy regulations. Some of the key frameworks in the world are: General Data Protection Regulation (GDPR) -European Union's comprehensive privacy law focusing on transparency, consent, data minimization, and user rights. California Consumer Privacy Act (CCPA) - Grants California residents control over their personal data. Brazil's LGPD and DPDP Act India: national legislation that relied on the guidelines of the GDPR. Singapore PDPA- Focuses on consent-based data collection and organizational accountability.

2.1.3 The Reason of working with GDPR

In Bangladesh, there is no in-depth data privacy policy that dictates how apps ought to amass, handle, and safeguard user information. The single operational rule is the Digital Security Act (DSA) 2018 that majorly looks under digital crimes, cyberterrorism, and governance of the digital content. The DSA does not have a lot to say about the privacy of personal data, the consent of the individuals using them, data subject rights, or the responsibilities of an organization in case of the misuse of such data. In addition, it is also highly criticized due to the compromised freedom of speech and imprecise general enforcement, as well as political force (Azad et al. [15]. This means that the DSA cannot be used as a useful measure to assess

the privacy policy or guarantee the usability of privacy. GDPR, in turn, presents a universally acceptable and user-friendly data protection system with a broad scope. It also focuses on unequivocal corroborative processing and processing, robust user rights as well as stringent accountability by organizations (Voigt Von dem Bussche) [6]. Most international technological businesses in Bangladesh use GDPR standards informally due to requirements of international business and according to app store policies. Moreover, GDPR principles are becoming a source of inspiration to new privacy legislations across the planet, and as such, it is a reliable and academically sound point of reference(Greenleaf, 2019). Hence, in the present study, GDPR has been adopted as an evaluation framework because it can be used to determine the extent of adherence to the privacy rules of Bangladesh mobile apps against known international privacy laws.

2.2 Related Work

2.2.1 GDPR Compliance Analysis

An Empirical Evaluation of GDPR Compliance Violations in Android mHealth Apps

In this study, Fan et al. [9] investigates the adherence of mobile health applications (mHealth apps) based on Android to the General Data Protection Regulation (GDPR). The authors focus on a very topical topic, the question of the protection of privacy because there is a risk of violating GDPR in the apps related to personal information on health (PHI). Privacy practices that are required by the GDPR involve an end-to-end privacy policy, a consistent collection of data, and security transmission of data. The methodology embraced by the authors is intended to detect and analyze these elements of compliance in mHealth applications, which are especially susceptible considering the particularly sensitive character of such information. Fan suggests an automatic mechanism, HPDROID, that implements the concept of natural language processing (NLP) and static analysis to overcome the semantic gap between the general privacy requirements provided by GDPR and particular implementations in mHealth apps. HPDROID identifies three significantly large categories of breach of GDPR compliance, namely, (1) incomplete privacy policies, (2) inconsistent data collection and (3) insecure data transmission. The tool interrogates 796 mHealth apps and findings show that 23.7% of apps do not exhibit full privacy policy, 77.9% apps have inconsistent data collection measures and most apps do not demonstrate adequate measures of safeguarding data transmission. Its methodological important points include NLP-based classifiers to identify the completeness of privacy policies, data-flow analysis to follow the collection and storage of PHI, and SSL/encryption analysis to find the possible security holes. The authors draw attention to the training of classifiers with labeled ground truth dataset and the proper data flow analysis which helps guarantee the correctness of the results. These findings are frightening: since there is a large number of mHealth applications that are unfavorable with respect to GDPR regulations, specifically the parts covering data transparency, data minimization, and confidentiality. Remarkably, researchers observe that 46 out of 59 applications reviewed are collecting personally identifiable information (PHI) which has not been reported in the privacy policy

and also that a substantial number of applications are utilizing unsecure approaches in sending sensitive information. The study also states that there are partially compliant apps with the standards of encryption with regard to posing additional risks to the privacy of the user. Although the results of this study are encouraging, the authors admit that their strategy has limitations, including the inability to identify every instance of privacy violation because of the complication of apps behavior and encryption in use. Their takeaway is that the possibility of subsequent studies is to leverage dynamic modeling with its traditional techniques in order to maximize HPDROIDs detection range. The study has essential implications of informing about the GDPR-compliant nature of mobile health applications and the prevalence of privacy concerns and provides an automated system to assist the developers in assuring its consideration. Making a conclusion, the work of Fan et al. shows the importance of privacy protection within mHealth applications and informs developers to refer to the tools, such as HPDROID to guarantee the safety of personal information prior to publication. In creating and testing HPDROID the authors have given a good tool to allow both developers and users to detect and correct the privacy breaches before they are too late.

Longitudinal Compliance Analysis of Android Applications with Privacy Policies

In this research, Hashmi et al. [17], performs a longitudinal compliance test of the privacy policies of Android applications of 2008-2016. These authors intend to have a look at the extent to which Android applications report their practices concerning the data collection and sharing, given the fact that there can be inconsistency between the privacy policy and the application behaviors. The research is conducted on 5,240 app releases of 268 Android apps and is conducted by using both static and dynamic analysis to detect any privacy violations as well as personally identifiable information (PII) leakage. The research reveals that there is a disturbing pattern in the increase of violations of the privacy policies of the apps over the years; the newer the app, the more likely they are to lose their privacy policy. In particular, the adherence to privacy policies declined with the drop of 33.16% in 2011 down to only 10.76% in 2016. The research also identifies the tendency towards the growth of leaking PII to third-party services, which are not included in the privacy policies. The main conclusions of the article are as follows: Dissimilar to the rise in data collection practices disclosed in privacy policies, the proportion of apps leaking PII based on disclosure also significantly declined. Using multiple years of data, the authors find that apps leak increasingly more PII to third-party services (advertising and analytics) than first-party services. An important observation made by the study is that the role of privacy policies is not useful when it comes to guaranteeing protection of users because such statements are not always accurate in terms of disclosing real data practices. It should be highlighted that this fact indicates the fact that privacy policies are not successfully conforming to the principle of notice and choice, which should provide a user with clear information about the way data he or she shares will be managed and provide the user with an opportunity to choose or reject sharing. To analyze these privacy breaches, the authors merge the outcomes of the static analysis and dynamic analysis and develop a more detailed picture of

user data processing by applications. Although the presence of privacy policy regulations, including the General Data Protection Regulation (GDPR), has impacted the state of privacy positively, the evidence suggests that the privacy compliance in mobile apps has been getting worse over the years. However, regarding the laws available, it appears that a certain discrepancy between what mobile applications do with the data and what are mentioned in the privacy policies of these apps can be attested. This mismatch presents the requirement of a more solid enforcement strategy and an improved transparency in app privacy engagement.

NLP-Based Automated Compliance Checking of Data Processing Agreements Against GDPR

In their paper, Amaral et al. [7] implement an automated tool to verify the compliance of Data Processing Agreements (DPAs) with General Data Protection Regulation (GDPR) with the help of the Natural Language Processing (NLP). This document is about the difficulty and overtime of manually checking the DPAs compliance of the subordinate legal documents that are related to processing of personal data. Considering the complexity of language and the possible ambiguity of the legal text, manual checking of the adherence of such rules is a difficult process and costly in terms of time. The authors proposed an automation bundle implementing NLP technologies, DERECHA (DPA semantic framework-based Compliance checking Against GDPR), affecting the examination of adherence of the DPAs to the provisions of GDPR. The method begins with identifying the shall descriptions of the GDPR provisions that are then employed as criteria of compliance. These requirements in the system are represented using semantic frames (SF), and the representations of the DPA statements are matched against the already defined representations of the GDPR requirements. The methodology considers five main steps: In cooperation with legal experts, a set of 45 requirements of GDPR compliance is extracted and defined as shall requirements. The text of the DPAs is tokenized, lemmatized, and parsed to be ready to be checked against compliance. The system automatically generates SF representations of each statement that it finds in the DPA. The text in the DPAs is augmented with semantically-related This is a great improvement over baseline performance that involves using off-the-shelf NLP tools, which was less accurate. Another advantage of the approach is that it is efficient. Whereas the manual process of auditing the compliance of a DPA may be time-consuming, DERECHA was able to process a similar-sized DPA in only 2.5 minutes, which is an indication that it can be effectively applied in practice. The system together with the human supervising works best in situations when one wishes to handle edge cases and also in ensuring complete compliance. The paper has therefore brought forth an NLP approach in automation of compliance checking of DPAs against GDPR. The study illustrates a significant reduction of manual work for legal compliance verification to check its validity and to achieve viable results with this approach.

The EU General Data Protection Regulation (GDPR)

Voigt et al. [6], in the given study, introduce a framework supported by AI to automate the procedure of checking the completeness of privacy policies according to the General Data Protection Regulation (GDPR). Privacy policies form the most important document, which notifies the people that use the services about how their data is treated in an organization, outlining the purpose of data-processing and other legal requirements. It is necessary to ensure that these policies comply with the GDPR demands, but the process of manually checking whether they are complete is time-consuming and subject to mistakes. To streamline this activity, the authors suggest the use of an AI technology where Natural Language Processing (NLP) and Machine Learning (ML) will be deployed to find the information components required in the privacy policy. The authors start by proposing a conceptual model and a criteria basis of completeness using GDPR principles. According to this conceptual model, there are 56 detailed metadata types that are to be captured in the comprehensive privacy policy and they include: processing purposes, legal basis, data subject rights, and controller information. These types of metadata reflect several GDPR articles linked to data gathering, data moving circumstances and rights of the client. By way of example, the metadata that indicates the legal basis of the data processing is called the "legal basis", and those that indicate the right to access and rectify or erase personal data are called the "data subject rights". Thereafter, the authors provide 23 criteria of completeness, according to which they dictate the dataset of metadata that is to be contained at various development stages of a product or a site. These criteria are provided in the form of an if-then combination of GDPR requirements and the associated meta-data types. As an example, when a privacy policy states that the personal data is being processed on the basis of user consent, then the privacy policy also has to mention that the users have the right to withdraw consent. There are some criteria that are related to a specific context and can only be applied in certain situations, like in case of data outside the EU. In order to conduct such checks, Amaral et al. suggest a system, named CompAI, which examines privacy policies per two stages: metadata discovery and this stage is followed by completeness assessment. During the initial stage, CompAI will incorporate an NLP-based tool to search the appropriate metadata type across privacy policy documents. The system tokenizes text and identifies named entities and normalizes the vocabulary to allow analysis across multiple policies. The sentences are then transformed into vector representation so that they could be analyzed numerically with word embeddings provided by GloVe. The multi-label classification approach is then used in the system where metadata types identified in sentences are categorized employing a mixture of support vector machine (SVM) classifiers, estimates of similarity, and match-searching definitions. At the second stage, CompAI contrasts the identified metadata with the predetermined completeness standards. It identifies the gaps in the metadata or areas where the vital points are avoided. The system marks the missing metadata at violations or warning, when a violation means that there is obvious non-compliance with GDPR, and warning states that the information is incomplete and needs additional specification. As an illustration, when privacy policies refer to consent but fail to mention a right to withdraw the consent, it is marked as a violation by CompAI. To determine the effectiveness of CompAI, the authors run experiments on a collection of 234 privacy policies in the financial sector, having 19,847 sentences. The results indicate that CompAI generates precision and recall of 92.9 percent and 89.8 percent in completeness checks respec-

tively; this is much better than the conventional keyword-based approach, which is prone to errors. Although these results portend hope, the authors admit that the GDPR compliance cannot be determined solely based on the completeness check as the regulation involves multiple qualitative aspects and interpretations which might need the knowledge of specialists in law. However, CompAI can be discussed as a first-rate tool of compliance, which can filter large quantities of privacy policies in a short time and on a low budget. According to the authors, it is necessary to conduct additional research to support such tools to more legal/regulatory instruments and authority frameworks, eventually enhancing the alignment of automated compliance checking with the legal and regulatory laws.

2.2.2 NLP Techniques in Privacy Policy Analysis

Automated Analysis and Presentation of Privacy Policies Using Deep Learning

Harkous et al., [20] in this work, propose Polisis, a novel framework that aims at automating the process by which the privacy policies are analyzed and presented. To tackle the problem of privacy policy complexity and length the authors touch upon the idea of it becoming the subject of the domination. Privacy policy is most of the time overwhelming to users, researchers, and regulators. The existing gap is what Polisis envisions to fill through a scalable, dynamic, and multi-dimensional query mechanism of privacy policies through the use of deep learning and a privacy-focused language model constructed with more than 130,000 privacy policies. The main contribution of the paper is the creation of a taxonomy of neural-network classifiers that are used to state a hierarchy of privacy policy content into 10 coarse-grained and 122 fine-grained categories. This allows Polisis to annotate the segments of privacy policies in detail and accurately. The system helps two main applications, including the first item which is privacy-icon structured querying, that Polisis is assigned with 88.4 percent of accuracy. The second app, PriBot, is a free-form question-answering system that gives the user the related answers through privacy policies. The top-three results provided by PriBot are found to be 82 percent correct demonstrating the promise of the system in having fairly intuitive interactions with privacy policies. A close inspection is also provided of both applications. The authors employ the dataset of 796 privacy policies and user study consisting of 700 participants to evaluate the usefulness of PriBot. The findings indicate that PriBot can work better than earlier employed term-matching models of retrieval, and on average a user of such a system would be satisfied with 89 percent of the first three answers. The effectiveness of Polisis in auditing privacy compliance is also studied and discrepancies are revealed in the privacy icons allocated by the compliance companies such as TRUSTe. In spite of the successes of this study, it recognizes several limitations such as difficulty in processing adversarial constructed privacy policies, scalability of the approach in more complex legal situations. The authors provide the idea that Polisis could be applied as a supplementary program

to legal professionals, regulators and companies, as helping to automate the process of privacy policy analysis and providing users with better comprehension of privacy practices. Polisis and PriBot prove how deep learning processes can increase the transparency and accessibility of privacy practices that serve the interests of users and regulation authorities.

The Creation and Analysis of a Website Privacy Policy Corpus

The authors in this paper Wilson et al. [4] present and discuss a novel collection of 115 web site-privacy policies, manually labeled with 23,194 fine-grained data practices. The rationale behind the development of this corpus is to enhance the scrutiny of privacy policies, particularly because privacy practices tend to turn out to be intricate and non-comprehensible by users. The reviewers emphasize that despite the importance of privacy policies as legal texts that the users must read and accept, long lists of words and complicated language make them overlooked, even being crucial contents. The intended outcome would make the extraction of the key privacy practices an automated process and assist users and regulators in their comprehension of how their personal information was treated. The article presents the development of such a corpus which was composed of a wide variety of 1,799 websites that were selected in two steps. The initial requirement was choosing the best search results of popular trends, and the second thing that was done was sector-based sampling so that the data presented is not skewed. The final data set will have privacy provisions of websites in 15 business sectors. Such diversity makes the corpus representative of many different privacy practices and it is not biased to any particular industry or type of service. The annotation scheme made specifically to annotate the corpus is extensive, and allows capturing a large number of data practices and is divided into ten categories: first-party data collection, third-party sharing, user choice and control, data retention, data security. All these categories are accompanied by a number of features and they include the nature of the data about a user, the reason for gathering the data, and how a user is able to access or erase his/her data. This degree of specificity matters in realizing how privacy policies distinguish the data practices. To illustrate this scheme of annotation, qualified annotators such as law students (a total of 221) employed via a web-based tool, annotated the privacy policies chosen. The corpus includes salient pieces of information to the schema of privacy policy composition. According to the authors, most common data practices identified in the privacy policies concerned the First Party Collection/Use and Third Party Sharing/Collection, which are the primary issues pertaining to the data collection and sharing that drive the creation of privacy policies. The analysis also indicates that privacy policies tend to be long with a huge amount of data practices yet most of these practices are ambiguous or unclear. In addition to the corpus analysis, the paper provides an activity of automation of privacy policy annotation with machine learning through predicting the category labels of the policy segments on privacy policies. The authors attempt to experiment with such models as logistic regression, support vector machines (SVM), and hidden Markov models (HMM). On the one hand, the models helped to partially automate the annotation process and, on the other hand, they can be used to show the problematic aspects of ambiguous and inconsistent privacy policy lan-

guage. In summary, the authors claim that the corpus can actually be a useful tool both to privacy researchers and regulators of privacy policies. It may be applied to promote natural language processing methods to the privacy policy analysis and the enhancement of user familiarity in privacy practices. The paper also gives weight to form the future research on automating the extraction of privacy practices, and it is the need to develop the tools capable of displaying the privacy practices to people in a manner that is more actionable and understandable.

2.2.3 Privacy Usability and Security Usability

Challenges in Developing Secure Mobile Health Applications — A Systematic Review

Aljedaani et al. [7] in this review conduct a systematic analysis of the problems encountered in the work of developers with regard to ensuring the security of mobile health (mHealth) applications. The development of mHealth applications has yielded great successes as far as reduction of cost and increase of patient awareness are concerned. Nonetheless, all these advantages are shadowed by the growing threats to the security of such sensitive healthcare information. The paper mentions 9 major issues that developers face during the design of secure mHealth apps and also the knowledge, expertise, and resources that can be used to prevent these risks. The analysis uses Systematic Literature Review (SLR) method and chooses 32 pertinent studies out of the range of articles that were published from 2008 to 2020. In thematic analysis, the authors discover such challenges as: absence of security guidelines and regulations, no knowledge of security among the developers, a lack of stakeholder engagement, as well as general negligence of security in the process of developing apps. The review states that the lack of standard security practices and frameworks in mHealth app development is also important because it can be used to reduce the identified threats and risks. Such a gap results in the development of security decisions based on guesses instead of making best practices. Furthermore, the paper also mentions the shortage of resources, time limits, and the absence of the significant number of security tests which are also significant impediments to secure app development. Conceptual framework created by the authors allows one to see the correlation among these challenges, providing an idea of their interrelatedness. As an example, the inability of development teams to have security professionals is revealed to add to the difficulty of limited understanding of security and less focus on security. Summing it up, the review can be considered as a good source of information about the security issues of mHealth app development and included the recommendations to resolve it. Security risks during development of mHealth apps can be greatly mitigated by enhancing the security knowledge of the developers, raising stakeholder take-up in the field, as well as the provision of resources. This paper is an appeal of all researchers as well as practitioners to come up with superior security measures in developing mHealth apps, especially to empower developers through training and guidelines.

Privacy Policies as Decision-Making Tools: An Evaluation of Online Privacy Notices

Jensen and Potts et al. [1] in their paper seek to test the usability of the online privacy policies and their capability to serve as a decision-making tool to the users. As issues of invasion of privacy on the internet grew, privacy policy became common with 77 per cent of the websites publishing a privacy policy. Nonetheless, these policies differ very much depending on the sites and in many cases consider problems that are not congruent with the problems of the users; therefore, not a reliable source of information to the users. The authors review 64 of the existing privacy policies to determine their availability, content, readability, and changes in time and analyze how well current policies address user needs and what can be improved. Among the major issues, the absence of standardization among the policies should be named as it is not always easy to compare and contrast policies properly. This contradiction is further compounded by the complexity and technical jargons contained in the policies which in most cases ends up overwhelming users. The privacy policies, although plainly spelled out in regulated industries, such as the healthcare and finance industries, are difficult to comprehend. The authors also examine the nature of the way privacy policies are displayed on websites. They also discovered that all policies which can be reached through the home page, are usually channeled through inappropriate formatting of links hence difficult to retrieve. Also, policies of certain websites can be divided into several pages which may confuse users and hide the necessary data. Another relevant concern, related to readability, was that the average Flesch Grade Level (FGL) in the policies under analysis was high implying that most of the users, particularly the low-educated ones, would find the content difficult to comprehend. Another finding indicated in the study is that a small proportion of policies can be read by persons with a high school education or below, and it makes the way to informed consent even more complicated. Most policies compel the user to update regularly thus putting an unnecessary burden on the user. The policies do not always have a modification date, and only some policies give notice of any substantial change. Such unresponsiveness in terms of changes gives a situation where users can blindly accept other terms without having any knowledge about them and therefore this might be a cause of privacy violations. In the paper, concerns are raised on the need to make changes in the way these policies are written and communicated with a concern to readability, accessibility, and ease in communicating changes. The authors also raise an issue of improved standardization and communication because, in this way, users will be able to make informed choices regarding their privacy online.

Designing Effective Privacy Notices and Controls

In the present article, Schaub et al. [5] discuss the difficulties that modern privacy notices and controls have with informing the users about their practices and offering them meaningful options. Although the privacy notices have been applied widely, in order to act in conformity with the data protection regulations, the authors have claimed that these privacy notices have been poorly efficient in notifying

users and providing the user with control over his or her privacy preferences. Among the factors that make this ineffective, the paper names the confusion between the legal compliance requirements and user transparency requirements, the absence of meaningful user choice, the exhaustion risk caused by forcing users to read long policies, and the disconnection between privacy notices and user behavior. The authors recommend guidelines on how to design more effective privacy notices and controls. These are the principles to have the privacy notice based on the current context of the user, the provision of the user with actionable choices, and the orientation of the information displayed in a manner comprehensible and user-friendly. The paper highlights that privacy notices should become a part of the flow of interactions that the user performs, and they should not be separate, long documents that the user would not read. The authors recommend not bombarding users with one long notice, but the opposite: brief, more purposeful notices regarding individual transactions or possible functions of the system. In addition, the paper has also emphasized the use of various modalities, which include visual, auditory, and haptic identifications, as the best way of providing privacy information. The authors also suggest the application of just-in-time privacy notices that are indicated at the appropriate moments of user interactions, instead of the use of the generic ones that are indicated at the start of the interaction. The authors evaluate the transactional context and, therefore, claim that it is possible to make shorter, more relevant, and less disruptive privacy notices. Also, incorporating user preferences into the feed of interaction, instead of forcing them to stand out as an independent, yes-or-no choice, can assist users in making even more merited choices regarding their privacy. To sum up, according to Schaub et al., privacy notices should be redesigned to improve the ultimate goal of notifying the user and enabling their control over their data. They feel there should be a change regarding the way businesses and regulators perceive privacy transparency and converting notices to be user-oriented and part of the overall design emanating from the system as a whole. The current paper can be beneficial in designing more useful privacy mechanisms, especially those applying to digital technology, such as mobile applications and IoT devices, to which privacy concerns are more often pertinent.

The effect of online privacy policy on consumer privacy concern and trust

In this research, Wen Wu et al. [19] examines the effect of online privacy policy on consumer privacy concerns and trust and also examine the moderating effect of cross-cultural differences in the relationship. The objective of the search is to comprehend the role played by the contents of privacy policy in affecting the behavior of consumers related to building trust and inclination to share personal information on the internet. The study adopts a model, which is grounded on Privacy-Trust-Behavioral Intention model in order to examine the impact of privacy policies on privacy concern and privacy trust, and the influence of privacy concern and privacy trust on online behavior. The study is built on the survey of 500 individuals, having the equal proportions of Russians and the Taiwanese, with the aim of investigating the cross-cultural implications of reaction to privacy policies. The results demonstrate that the text of privacy policies online has a significant impact on pri-

vacy concern and trust. In particular, such privacy policies that entail the main principles of notice, choice, access, security and enforcement intervene to some appreciable extent on consumer confidence. The research also concludes that the issue of consumer privacy harms trust and inclinations to offer personal information online. Specifically, it was also determined that the existence of clear security practices in privacy policies had the biggest impact related to trust in a positive direction - that is, more willing to trust websites that have indicated what they do to secure personal information. More so, the study reveals that the privacy concern is one of the major factors influencing the inclination of the consumer with regard to sharing personal information to websites. Another question elaborated by the authors is the role of the cultural differences in these relationships and especially power distance. As indicated by the results, power distance moderates the connections between the content of privacy policies and privacy concerns and privacy policies and trust. High power distance culture consumers such as Russia will accredit safer web sites that grant them entry to their personal details and also the safe management of their details over websites of low power distance cultures such as Taiwan. Finally, the paper also implies that in order to increase consumer trust and the willingness to share personal information, companies need to work on the improvements in the transparency of their privacy policies and their security. The implications of the findings also entail the significance of incorporating the cultural differences in the formulation of the privacy policies governing the international markets. The study will be a good guide to businesses involved in e-commerce particularly those doing business in various cultural settings to increase their bonding with consumers by having taken care of issues related to privacy.

Privacy Policies of IoT Devices: Collection and Analysis

Within the framework of this paper, the article by Kuznetsov et al. [10] is devoted to the emergence of the necessity of the accessibility and clarity of privacy policies to Internet of Things (IoT) devices, which are becoming more and more engaged in the collection of personal data. The paper is dedicated to the gathering and study of privacy policies specifically, on IoT devices that provide a novel method of generating a dataset of privacy policies. The authors introduce the issues of comprehensibility of a privacy policy to the user and the necessity of easier to understand privacy policies to allow users to make informed choices regarding their data. The authors reveal the limitations of privacy policies of the present problems and underline that these are mainly the privacy policies of the web pages, but not Internet of Things devices. Because IoT devices, including fitness trackers, smart locks, and smart cameras, capture a large variety of sensitive personal information, including biometric and location information, their privacy policies tend to be very different when compared to those of standard websites. Available data sets, including those obtained on Amazon Alexa and Google Play, do not support finer requirements that the General Data Protection Regulation (GDPR) introduces and, in general, do not fit the data situation that applies to IoT devices. To close these gaps, the authors have devised a new method of gathering privacy policies that specifically take into consideration IoT devices. This method requires using e-commerce websites such as Amazon and Walmart to find the device manufacturers based on which the privacy policies are

extracted. The data collected using this approach includes 592 distinct privacy policies of companies that produce all kinds of IoT devices. The fact that the dataset contains privacy policies, which were elaborated following the adoption of GDPR, makes it important since it allows including them in the latest requirements about the data protection in legislation. The paper also presents the thorough statistical and semantic analysis of the gathered privacy policies. The latter is performed in the statistical and semantic analysis, where the first examines such attributes as the length of documents, the number of paragraphs, etc., whereas the latter incorporates the Latent Dirichlet Allocation (LDA) algorithm to detect various topics concerning the scenarios in which a person might use his/her personal data. The review shows that the majority of IoT privacy policies are concentrated on the first-party data collection and distribution of the data among the third-party, however, generic information is still present in various policies. There were also a few surprising items found by the analysis like allusions to some employee data, something often reflecting the attempts from manufacturers to develop a privacy policy that they can apply to all of their services. The authors compare their data set to other data sets on privacy policies like OPP-115, noting how their data set has a wider variety of items. In some cases, a given IoT privacy approach includes the topics of data protection in terms of the GDPR and the rights of an individual to access his or her personal data and to be notified about any changes in the policy. The given findings emphasize the significance of a specialized dataset of IoT devices so as to enhance the quality of privacy policy analysis methods. Summing up, the paper introduced a new method of gathering and analyzing privacy policies in IoT devices to fill in the gap of the current dataset. The presented tool and data can serve to optimize the models allowing us to automatically check the privacy policies thus making them transparent and understandable to users. Future research directions are related to enhancing the dataset through adopting other languages, to make the process of validating policies automated, and to better detect privacy risks aspects regarding different data usage cases.

2.2.4 Cross-Country or Bangladesh-specific Studies

Digital Security Act in Bangladesh: The Death of Dissent and of Freedom of Expression

In this paper Azad et al. [15] critically analyses the Digital Security Act (DSA) in Bangladesh and how it poses a threat to freedom of expression, especially how it is used to gag dissent. Presented in 2018, the DSA covers the same ground as the Information and Communication Technology Act (ICT Act) that had been criticized due to its perception of censoring free speech. Nevertheless, the author states that the DSA, instead of promoting the idea of freedom of expression, is more restrictive and oppressive legislation which undermines the capacity of citizens to express their views without any restrictions, particularly the ones opposing the government. The paper has commenced by sketching out the historical background of the ICT Act, on which Section 57 used to be infamous in its regard towards suppressing dissenters. After being pressured globally and nationally, the government canceled Section 57

only to substitute it with a longer and more elastic DSA in the form of the latter. The DSA borrows most of its punitive measures such as against the content that damages national image or taints religious beliefs, used in the ICT Act but it has gone further to enact some additional restrictions which restrain political speech, criticism of religion and the right to criticize the historical facts. The author places the DSA in the framework of the international law of human rights, especially in the International Covenant on Civil and Political Rights (ICCPR), which provides freedom of expression. Azad argues that under the leadership of Prime Minister Sheikh Hasina, the DSA goes against the constitutional requirement of Bangladesh and its international commitments in particular due to its broad definitions and vague initiatives that give the government immense powers to censor, rebuke and imprison those who raise their voice against the government. Another idea that Azad analyses in the paper is chilling effects, i.e., citizens remain silent concerning the fear that their voices would become suppressed; it is even worsened by DSA vague and broad language. This, among other provisions, has contributed to the misutilization of the law, especially when they are used against journalists, bloggers, and even activists who are against the government. The overwhelming number of arrests and self-censorship in media, both during the institutionally sensitive periods, such as the 2018 election and COVID-19 pandemic are indicators of the chilling effect of free speech given by the DSA. The paper concludes by making a case to repeal the DSA because of its unconstitutionality as well as its tendency to harm democratic institutions and the freedom of speech. In his article, the author implies that Bangladesh may repeal DSA altogether or revise it in order to meet international human rights standards, so that laws serve the national security and at the same time they would not permit the government to suppress the right to free speech.

(IN)Security of mobile apps in developing countries: A systematic literature review

Diallo et al. [13] in their systematic literature review discuss the situation of mobile applications security studies in developing nations. The authors point at the growing trends of using mobile applications in such areas of serious concern as medicine, budget, and education in these economies. Nonetheless, the issue of mobile app security is critical looking at the threats posed by app weaknesses, malware, and breach of privacy. The aim of this review is to analyze the currently available studies on mobile app security with specific reference to the peculiarities of mobile app security that users and developers of developing countries face. The main results that the authors provide show that there is a lack of studies on the specificities in mobile app security as applied to developing countries. Although vulnerability detection is identified to be of the topmost research, the literature mainly talks about the security issues on mobile financial applications and mostly on mobile banking apps. Most of these apps which deal with sensitive information like financial details and personal details are lucrative targets of cyberattacks. Scientists have developed several methods to find vulnerabilities such as static, dynamic, and hybrid. Nevertheless, what is lacking in these areas is specialized security tools that address the needs and challenges unique to mobile apps security. The paper also explains some

of the challenges peculiar to the developing countries which include poor connectivity, low cost, insecure devices and low degree of computing literacy. All these lead to a high degree of digital security threats. The last point which the authors underline is that, despite the fact that mobile apps security is a subject of huge worldwide research that is done, there still exists a huge gap that needs to be filled by specialized studies related to the problems of developing countries. They propose some future research areas, such as a greater need to make studies of mobile malware detection more comprehensive, create region-specific security solutions, and extend the scope of research to include other app categories that are important, but not limited to financial applications, e.g., in the field of healthcare and education.

Global data privacy laws 2019: 132 national laws and many bills

In their work, Greenleaf et al. [16] propose an automatic framework which implements the privacy policy completeness checks with the help of AI on the basis of GDPR. Privacy policies are very vital legal statements that define how an organization uses the data of its user, why it processes it, and any other legal practices applied. Manual scoping of all these policies, to meet the GDPR requirements, may prove time consuming and error prone. In order to simplify this, Amaral, et. all present an AI-based solution that uses Natural Language Processing (NLP) as well as Machine Learning (ML) to extract necessary pieces of information as specified in privacy policies. The author first came up with a concept model and a set of completeness criteria based on GDPR recommendations. Its conceptual model determines 56 metadata types which must be covered in a comprehensive privacy policy, i.e. the processing purposes, the legal basis, data subject rights, the controller information. All these metadata categories are linked to particular articles of the GDPR addressing the terms of data transfer and collection, as well as users' rights. As an example, the legal basis explains why we should process the data, and data subject rights indicate the rights of people to access or delete their data. On top of this conceptual map, the authors distinguish 23 completeness criteria, according to which metadata types have to be added in a product or a web page at various stages of development. Such criteria are created in terms of the rules of an if-then formula to align GDPR requirements with the metadata types. Taking an example, one of the criteria may be stating that where a policy makes it clear that data processing is done on the basis of consent, the policy also must notify the user that he/she can withdraw consent, at any given time. There are criteria that are contextual, namely, which are only reactive to specific circumstances like when data is moved outside the EU. In order to carry out these checks, Amaral et al. propose the use of the system called CompAI, which is AI-driven. The system has two stages such as identification of metadata and completeness check. During the first step, CompAI determines appropriate metadata in privacy policies. It employs NLP libraries in token creation, identification of named entities as well as generalizing the entities in order to standardize the use of words under various policies. This step also transforms sentences into the form of vectors to allow numerical analogies using word embedding of GloVe. The multiple-label classification approach used by CompAI identifies the presence of metadata types in sentences. It has a wider use of SVM classifiers in broader metadata categories, estimates of similarity to adjust

their classifications, as well as match-searching to extract terms concerning GDPR. Such a structure of classification allows using the full and hierarchical nature of GDPR requirements, but in the second step, the completeness of found metadata with respect to the selected criteria is verified via CompAI. The missing data or those metadata that are not present are flagged as violations (those that show they do not conform to GDPR) or warnings (those that only show partial information that requires further clarification). As an example, when policy references consent without the reference to the right of recalling consent, the policy is flagged by CompAI as a violation. CompAI was tested against a corpus of 234 privacy policies in the financial sector, with 19,847 sentences with metadata types. Policy evaluation was carried out using 20 percent of them. This estimation demonstrated the high level of precision (92.9) and recall (89.8) of CompAI that outperforms conventional methods based on keywords. These results show that AI-based solutions are more precise and less error-card than a keyword search alone. As positive as the results might sound, limitations are admitted by the authors. The completeness check is by no means sufficient to be able to fully comply with the GDPR, given that there are subjective interpretations to be undertaken and legal knowledge to be applied. Amaral et al. observe that CompAI is best used in the first line in cases where large amounts of privacy policies with low costs have to be filtered out quickly. They promote additional studies of intertwining of such tools with the authority compliance frameworks and the application of those tools to other legal instruments.

Design of a Compliance Index for Privacy Policies: A Study of Mobile Wallet and Remittance Services

Valecha et al. [11], present the work on creating a compliance index evaluating the privacy policies of mobile wallet and remittance services (MWR) with particular reference to General Data Protection Regulation (GDPR). With an ever-growing popularity of mobile payment systems, there is a risk of personal data leakage, as they tend to keep the sensitive information of the user. To cope with this, the authors suggest the creation of a methodological way to analyze the adherence of these services to provisions of GDPR by means of a new index of compliance. The authors adopt the methodology to apply the text mining techniques, such as term frequency-inverse document frequency (tf-idf) and topic modeling using the Latent Dirichlet Allocation (LDA) into the process of processing and categorizing the privacy policy documents of MWR services. In this way the major dimensions of the GDPR will be extracted in terms of the policies, which include data protection impact, data subject rights and notification of data breaches. The compliance index will be constructed to calculate two significant scores (emphasis density score and privacy score). Like the emphasis density score indicates how much a policy has stressed the important aspects of GDPR, the privacy score indicates the total level of compliance to the GDPR guidelines. The index of compliance has been calculated and validated by comparing the Cisco Global Readiness Index (GRI) about GDPR compliance which indicates high scores in terms of the country-level privacy practices. To give an example, the level of compliance in a country such as Spain and Germany is high, whereas the compliance in the USA and the UK is average. The comparison assists in the validity of the compliance index in producing an objective, quantifiable

measure of GDPR compliance in case of MWR services. On the whole, this work is a valuable addition to the literature of digital payment system security privacy that presents the useful tool, being able to measure the compliance of privacy policies with GDPR. Besides assisting MWR service providers in assessing their compliance with the law, the compliance index will assist consumers to be more informed when making decisions of what services they will use depending on the extent to which their privacy will be preserved. The results reported in the article demonstrate that there is a necessity of standardization in the privacy practices and that the compliance with GDPR plays an important role to reduce legal risks and develop trust in the digital payment services.

2.2.5 Semantic Analysis & LLM Usage in Digital Privacy

Assessing country-level privacy risk for digital payment systems

In this paper, Akanfe et al. [8], discusses conducting the country level analysis of assessment of privacy risks Digital Payment Systems (DPS), particularly the Mobile Wallet and Remittance (MWR) applications. With the rise of digital payment systems around the world, it has become apparent that privacy is becoming an issue of concern as we have started relying on the system to process massive amounts of personal data. The paper proposes a framework that can be used to measure country-level privacy risks of MWR apps on the basis of their resistance to the General Data Protection Regulation (GDPR). A survey of the extant country-level risk assessment methods is carried out which mostly pay attention to economic, political, and financial risks and give little or no attention to privacy risks. Akanfe et al. even claim that privacy risks have become a major issue especially because they could severely hurt the trust of consumers and international trade. The authors pay attention to MWR apps as such apps are an essential element of DPS and they usually imply cross-border transactions, which introduces more privacy risks. To evaluate the privacy risk level, the authors examine MWR apps privacy policies concerning GDPR ten dimensions, which address the following domains: data processing activities, data transfer to third parties, data security, and data subject rights. They take the GDPR as a compliance benchmark and develop a so-called hit ratio methodology, that is, the proportion of the privacy policies that contain GDPR-related keywords is used to give an app a compliance score. The authors then combine these scores into a country level privacy risk index. Key findings are as follows: By measuring the privacy policies of MWR apps, the authors come up with a country level privacy risk index. There are nations that are good with privacy compliance like the UK, Malaysia, and Canada, but those that have lower privacy compliance rate (more privacy risk) include Nigeria, Saudi Arabia, and Kenya. It is established in the study that most MWR apps do not completely adhere to the privacy needs of GDPR. The hit ratio of GDPR dimensions will be used to come up with findings regarding the existence of compliance gaps in some of the countries, including India and Nigeria that may predispose them to high rates of privacy risk. Alongside with

privacy risk factors, authors encompass general country level risk factors, which include political stability, corruption, and socioeconomic conditions. A more detailed picture of the global risk environment of the country is obtained, with these aspects being incorporated into a wider picture on risk. The research presents a model to assess privacy, as well as overall risks on a country scale. This framework assists policymakers to evaluate the compliance of the regulatory framework on privacy in their country relative to the GDPR and determine areas that privacy practices may be improved. The importance of the study in the field of privacy risk assessment consists in the fact that it presents a new logic of wisdom in how privacy practices in digital payment systems exist in different countries. With the GDPR as a starting point, the framework constitutes an effective tool to detect the gaps in the maintenance of the compliance and guide the policy recommendations. Even though the results have positive outcomes, the authors identify a number of limitations to the study. They note that their framework can be used to reveal interesting insights regarding privacy policy compliance but not the assessment of the compliance with GDPR since certain qualitative preferences might have to be viewed in terms of their legal implications. Also, they argue that it can be a future research area how such tools can be extended to the use of other legal instruments and how the application of privacy risk assessment could be used in an extended scope of digital payment applications. Summing everything up, the study by Akanfe et al. provides an insight into the significance of conducting risk assessment of privacy at the country level, which, in this case, is related to the field of digital payment systems. It is an addition to the existing research on compliance with privacy policy and offers a hands-on guide on how to assess and address privacy risks in the digital payment systems.

Chapter 3

Methodology

In this chapter, we will be talking about how we collected the data required for our research, describing the data in details and how we processed the data to incorporate it into our work. Finally, it covered the calculation process, explaining the necessary formulas.

3.1 Data collection

In the current study, we considered four types of apps: Mobile Wallet and Remittance (MWR) apps, Entertainment apps, E-commerce apps, and Healthcare apps. These types have been chosen due to processing a large volume of sensitive personal data as they are perfect selections in terms of testing the privacy compliance. We specifically focused on apps that are running in Bangladesh, and we only considered apps that have at least 1 million users because we wanted to make sure that we picked popular apps that have a significant number of users and thus have their data processing activities. The privacy policy texts were obtained in the following way: we selected a total number of 32 apps which were directly accessed in the Google Play Store. To measure GDPR compliance, we took a GDPR Keywords Dataset of 97 unique keywords, divided into 10 GDPR dimensions. Such dimensions include fundamental privacy concepts contained in the General Data Protection Regulation (GDPR), like Data Processing Activities (DPA). To test the GDPR compliance, we applied a GDPR Keywords Dataset composed of 97 distinctive keywords. In the construction of a structured dictionary of GDPR keywords, our action was informed by the approach to dimension-based implementation of analysing privacy policies as presented in the earlier study (Akanfe et al.) [11] that relied upon the focus at privacy policies characterised by a number of specific categories of GDPR compliance. Based on this direction, we selected ten keywords under each of the ten fundamental dimensions of GDPR to make our coverage more balanced and representative.Such

dimensions are related to the principles of the General Data Protection Regulation (GDPR) as Data Processing Activities (DPA), Data Subject Rights (DSR), and Data Protection Impact Assessments (PIA).The 10 GDPR dimensions adopted in this research are -DPA – Data Processing Activities PIA – Data Protection Impact Assessment PDD – Data Protection by Default and Design TOM – Technical and Organizational Measures DSR – Data Subject Rights DBN -Data Breach Notification PMS – Data Protection Management System TPD – Third-Party Data TransferCCC – Codes of Conduct and Certification PCI – Data Protection Contact Information. As an illustration, the DPA dimension had keywords, such as, "use," "collect," and "service", whereas the DSR dimension was linked to keywords, such as, "access," "edit," delete, and consent.The table below summarizes the GDPR keywords categorized by their respective dimensions:

Table 3.1: GDPR Dimensions and Their Full Forms

Abbreviation	**Full Form**
DPA	Data Processing Activities
PIA	Data Protection Impact Assessment
PDD	Data Protection by Default and Design
TOM	Technical and Organizational Measures
DSR	Data Subject Rights
DBN	Data Breach Notification
PMS	Data Protection Management System
TPD	Third-Party Data Transfer
CCC	Codes of Conduct and Certification
PCI	Data Protection Contact Information

A set of GDPR-related keywords related to each of the dimensions: We chose them among the sample keywords suggested by Akanfe et al. [8]. As an illustration, the DPA dimension had keywords, such as, "use," "collect," and "service", whereas the DSR dimension was linked to keywords, such as, "access," "edit," "delete," and "consent.". The table below summarizes the GDPR keywords categorized by their respective dimensions:

Table 3.2: Keywords Mapped with Corresponding GDPR Dimensions

GDPR Dimension	**Keywords**
DPA	use, service, download, collect, survey
PIA	request, signal, profile, account, update
PDD	store, delete, record, remove, retention
TOM	firewall, server, technology, track, settings
DSR	disable, choice, agree, option, consent
DBN	communicate, post, decide, notice, breach
PMS	secure, safeguard, encrypt, detect, prevent
TPD	share, third, party, transfer, advertise
CCC	disclose, promote, send, overview, conduct
PCI	contact, feedback, staff, office, file

3.2 Data preprocessing

In order to preprocess the texts of the privacy policies to conduct the evaluation of the GDPR compliance, we created a multi-stage preprocessing pipeline with an adoption of both traditional natural language processing (NLP) algorithms and semantic validation of the texts using a large language model (LLM). We first identified GDPR-related keywords, which were subsequently clustered according to their compliance dimension (e.g., consent and choice, data subject rights), thus obtaining a structured mapping to conduct the analysis category by category. Any keywords were lowercase to make the text matching case insensitive and equivalent, leading to consistent text matching. These documents of privacy policy were then tokenized and vectorized using the scikit-learn library CountVectorizer (Pedregosa et al.,) [3] into a matrix form of token frequencies that could be used to calculate hit ratio. In order to de-noise text and increase precision of keywords, Natural Language Toolkit (NLTK) stopwords were eliminated (Bird et al.,). [18]. In order to investigate further refinements to keyword relevancy and minimize false positives, we applied semantic filtering based on a Large Language Model (LLM) provided by Mistral AI. Once the initial matching on the keywords was carried out, each of those occurrences was then confirmed by asking the LLM to judge whether the context of nearby sentences semantically matched with GDPR compliance meanings. In doing so, the ability of surface-level mentions of GDPR keywords to fail to trigger genuine regulatory compliance was addressed in earlier studies (issues of purely lexical matching), because the pipeline excluded most instances of surface-level mentions due to semantic filtering of LLMs (Zimmeck et al.,). [4]

3.3 Frequency Analysis and Hit Ratio Calculation

As an approach to determining how appropriate mobile application privacy policies are relative to the commonly applied data protection laws like the GDPR, we resorted to keyword-scoring, which has been employed in past research studies (Harkous et al.,) [20]. In particular, we applied the methodology proposed by Liu et al., [4] based on which GDPR-related keywords were used to calculate a hit ratio i.e., a quantitative measure that captures the instances of compliance language in a text of privacy policy. The metric is not new and has been already applied in other related applications such as text retrieval and categorization (Lapata Keller et al.,) [2], and recently, in the context of measuring privacy compliance of mobile wellness apps (Liu et al.,) [9]. To turn this method into practice, we created a GDPR keyword dictionary by using the provided dataset of the Usable Privacy Policy Project (Shvartzshnaider et al.,) [14] and modified/organized it into GDPR dimensions, including transparency, data minimization, consent, limiting the purpose and so on. A custom script was used to match the extracted privacy policy texts with the keywords and it removed stopwords and irrelevant tokens as well. We were able to validate the semantic relevance of the individual instances of the keywords using a large language model (LLM) through Mistral AI with unlike purely syntactic matching. This was necessary to minimize false positives, i.e., cases in which a keyword

occurs but it is not suggestive of the GDPR conformity (cf. Zimmeck et al.,). [4] After the extraction and categorization of the keyword match, we estimated the ratios of hits in each GDPR category in each app. These ratios show how much an app covers a specified area of compliance in the privacy policy compared to the other.

The Hit Ratio (%) is calculated as:

$$\text{Hit Ratio (\%)} = \left(\frac{\text{Number of GDPR Keyword Matches}}{\text{Total Words in Privacy Policy}} \right) \times 100$$

3.4 Privacy Compliance Score Computation

After calculating hit ratios per dimension of GDPR compliance, we obtain a Privacy Compliance Score (PCS) for each app as a sum of the hit ratios of all ten GDPR dimensions. As a proxy measure, this composite score is to be used as an overall measure of the compatibility of the privacy policy of a mobile application with the requirements of GDPR. In the previous research, similar aggregate functions have been used to compute privacy compliance measures at an application level or national scale (Harkous et al.,) [20]. The PCS is computed by determining the arithmetic mean of the individual hit ratios in each of the ten dimensions of GDPR. To measure privacy performance per country, we hypothesize that mobile applications with the same country of operation have a project likeliness or tendency to meet the same regulatory and business trends on privacy associated with their country (Linden et al.,) [12]. Thus, the averages of all the PCS values of the apps belonging to a given country will produce a Country-Level Privacy Compliance Score (CPCS).

Privacy Compliance Score (PCS) Formula:

$$\text{PCS}_{\text{country}} = \frac{1}{10} \sum_{j=1}^{10} \text{HitRatio}_j$$

3.5 Privacy Risk Score Computation

Although the Privacy Compliance Score (PCS) provides an indication of how effectively the privacy policy of an app adheres to the requirements of the GDPR, it is equally worthwhile considering its converse, namely how dangerous this situation is when the requirements of the GDPR are not explicitly indicated. The direct measurement of non-compliance is, however, tricky. This is so because it is not so easy to detect when the language is not in line with the requirements of GDPR or the rules are being overlooked. To address this we took the route of Liu et al.

[9], who employed an alternative measure of privacy risk. They did not attempt to measure non-compliance per se but calculated the Privacy Risk Score (PRS) by flipping the compliance score. And the same concept is what we applied in our study: we summed the PCS of each app and subtracted it from the highest value. This is easy to understand, as the fewer terms of GDPR an app will contain, the greater the associated risk of privacy. We also examined the types of apps that had the lowest PCS values since those types of apps are more prone to introduce higher privacy risks. Such an approach can assist us in making privacy risk estimates under circumstances where law enforcement numbers or monitors are not accessible. The same methods are observed in previous research to measure data privacy through indirect measures. (Zimmeck et al.) [4]

Privacy Risk Score Formula:

$$\text{Privacy Risk Score} = 5 - \text{Privacy Compliance Score}$$

Chapter 4

Findings and Analysis

Chapter 4 gives us the findings and results of our work. We start this chapter by giving an overview of the results we derived, used tables to demonstrate the outcomes and a bar chart to visualize the result at a glance. This chapter ends with a thorough analysis of the outcome.

4.1 Findings and Analysis

Cross-Sector Privacy Risk Analysis in Bangladesh

This chapter presents the results of privacy policy compliance of mobile applications in Bangladesh in four sectors: entertainment, healthcare, e-commerce, and mobile financial services. We evaluate it using a new linguistic framework that we designed to differentiate between the GDPR keywords' usage in positive and negative sentence constructions. Positive structures are the ones that state positively that something is compliant (e.g., we protect user data), whereas negative structures are statements of restriction, omission, or non-commitment (e.g., we are not liable to third-party data breaches). This strategy enabled us to appreciate a more subtle insight into the way apps declare their data protection obligations and user rights. In order to perform such an analysis, we used a large language model (LLM) to find out the frequency and context of GDPR-related keywords within the sentence structures. We asked the LLM to perform sentence-level meaning interpretation in an attempt to answer the question of whether the mentioning of GDPR terminology implied compliance (positive) or evasion (negative). On the basis of this semantic categorization, we determined scores of each of the ten GDPR dimensions, as well as aggregate Privacy Compliance Scores (PCS) and Privacy Risk Scores.

4.1.1 GDPR Compliance Scores of selected Bangladeshi applications

Table 4.1: GDPR Compliance Scores Across Mobile Applications (positive)

Mobile App	CCC	DBN	DPA	DSR	PCI	PDD	PIA	PMS	TOM	TPD
Bkash	0.79	1.64	3.36	0.46	0.26	0.39	0.39	0.59	0.39	0.86
Nagad	0.49	0.79	1.63	0.37	0.18	0.22	0.83	0.12	0.29	0.66
Rocket	0	2.99	4.65	0	1.33	0.33	1.66	0.66	0.66	0.66
BRAC Bank Astha	1.79	1.14	3.01	0.33	0.16	0.33	0.89	0.65	0.08	0.41
EBL Skybanking	0.18	0.87	1.79	0.23	0.17	0.05	1.05	0.41	0.09	0.15
MyCash	0.29	1.86	1.62	0.2	0.98	0.15	0.34	0.69	0.25	0.29
OK Wallet	0.57	0.68	1.89	0.48	0.26	0.24	0.88	0.24	0.42	0.66
Sonali e-Wallet	0.24	1.76	3.83	0.43	0.24	0.12	1.03	1.82	0.43	0.36
SureCash	0.09	1.47	1.29	0.37	0.46	0.55	0.64	0.92	0.74	0.74
TAP	1.07	0.59	2.19	0.42	0.15	0.24	0.71	0.59	0.12	0.36
Upay	0.68	1.22	2.58	0.07	0.14	0	0.95	0.88	0.2	1.02
Bongobd	0.68	0.81	2.89	0.36	0.23	0.23	0.72	0.36	0.27	0.14
Chorki	0.34	0.87	2.18	0.41	0.12	0.15	0.87	0.58	0.27	0.87
iScreen	0.37	1.1	2.92	0.4	0.15	0.27	0.7	0.33	0.37	0.58
Toffee	0.48	1.04	3.14	0.48	0.2	0.3	0.68	0.38	0.28	0.46
Arogga	0.4	1.78	1.88	0.69	0.1	0.4	0.69	0.59	0.1	0.2
Daktarbhai	0	2.25	1.35	0.9	0.9	0	1.35	0.45	0	1.8
DocTime	0.84	1.73	2.35	0.53	0.19	0.43	0.94	0.31	0.19	0.26
Aarong	0.57	1.13	2.55	0.57	0.28	0.28	1.42	0	0	0.85
Ali2BD	0.4	0.97	2.8	0.29	0.23	0.11	0.92	0.46	0.23	0.34
Bangla Shoppers	0.14	0.61	1.97	0.55	0.32	0.61	1.51	0.17	0.2	0.75
Bikroy	0.64	0.97	4.29	0.97	0.97	0.21	0.54	0.43	0.54	1.07
Cartup	0.44	1.05	2.26	0.36	0.08	0.2	1.42	0.49	0.28	0.4
Chaldal	0	0.76	2.02	0.5	0	1.76	2.27	0	0.25	0.25
Daraz	0.44	0.96	2.51	0.38	0.29	0.27	1.07	0.44	0.29	0.77
Othoba	0.68	1.43	2.91	0.8	0.34	0.17	1.14	0.46	0.34	0.46
Paikaree	0.13	2.15	5.64	0.27	0.13	0.13	1.07	0.54	0.13	1.34
Pickaboo	0.25	1.57	3.8	0.15	0.35	0.86	1.52	0.66	0.1	0.91
Rokomari	0.19	1.55	3.67	0.39	0.39	0.58	1.69	0.58	0.48	0.72
Shajgoj	0.41	0.97	2.71	0.31	0.15	0.31	0.97	0.41	0.26	0.41
Shwapno	0.52	1.67	3.66	0.94	0.52	0.31	1.57	0.94	0.73	0.52
The Mall	0.48	1.69	2.17	0	0.96	0.72	1.45	0.48	0	0.48

Table 4.1 displays a table of mobile applications and their "hit ratios" across various dimensions, related to compliance performance. Each row represents a "Mobile App," and the columns (CCC, DBN, DPA, DSR, PCI, PDD, PIA, PMS, TOM, TPD) represent the 10 distinct GDPR dimensions that were analyzed. The numerical values in this table indicate the frequency or extent to which specific GDPR-aligned keywords were found within each app's privacy policy for that particular dimension. A higher hit ratio indicates better compliance.

To determine the best- and worst-performing mobile applications in terms of GDPR-aligned privacy compliance, a total compliance score was computed for each app. This was calculated as the summation of values of all the dimensions. The application that scored a cumulative high, Rocket (12.94), exhibited a wide range of compliance, especially performing particularly well in DPA, DBN, PIA, and TPD and does not have any critical weakness in a particular dimension. Following closely, Shwapno (11.38) and Paikaree (11.33) also showed very high compliance scores. Conversely, Nagad (6.58),TAP (6.44), OK Wallet (6.32) and EBL Skybanking (5.99) recorded the lowest scores, which showed a lack of performance in almost every aspect except occasional excellence. The applications with moderate compliance, including Bkash, Pickaboo, Rokomari, Sonali e-wallet, and Bikroy, got total scores around 9 to 10, which implies that they partially correspond to the privacy standards with a visible lack of improvement potential, particularly in such dimensions as TOM or DSR.

Table 4.2: GDPR Compliance Scores Across Mobile Applications (negative)

Mobile App	CCC	DBN	DPA	DSR	PCI	PDD	PIA	PMS	TOM	TPD
Bkash	0.03	0.05	0.11	0.02	0.01	0.01	0.01	0.02	0.01	0.03
Nagad	0.02	0.03	0.05	0.01	0	0.01	0.03	0	0.01	0.02
Rocket	0	0.1	0.15	0	0.04	0.01	0.06	0.02	0.02	0.02
BRAC Bank Astha	0.06	0.04	0.1	0.01	0	0.01	0.03	0.02	0	0.01
EBL Skybanking	0.01	0.02	0.06	0	0	0	0.03	0.01	0	0
MyCash	0.01	0.04	0.05	0	0.02	0	0.01	0.02	0.01	0.01
OK Wallet	0.02	0.01	0.06	0.01	0.01	0	0.03	0	0.01	0.01
Sonali e-Wallet	0.01	0.04	0.13	0.01	0.01	0	0.02	0.01	0.01	0
SureCash	0	0.03	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.01
TAP	0.04	0.01	0.07	0.01	0	0	0.01	0.02	0	0
Upay	0.02	0.02	0.09	0	0	0	0.02	0.02	0	0.01
Bongobd	0	0	0.05	0.01	0	0.04	0.03	0.02	0	0.01
Chorki	0	0	0.05	0.01	0	0.04	0.03	0.02	0	0.01
iScreen	0	0	0.05	0.01	0	0.04	0.03	0.02	0	0.01
Toffee	0	0	0.05	0.01	0	0.04	0.03	0.02	0	0.01
Arogga	0	0.03	0.01	0.01	0	0	0.01	0	0	0
Daktarbhai	0	0.01	0.01	0	0	0	0	0.01	0	0
DocTime	0	0	0.01	0.01	0	0	0.01	0	0.01	0
Aarong	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Ali2BD	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Bangla Shoppers	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Bikroy	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Cartup	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Chaldal	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Daraz	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Othoba	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Paikaree	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Pickaboo	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Rokomari	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Shajgoj	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
Shwapno	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02
The Mall	0.01	0.03	0.07	0	0	0	0.04	0.02	0.01	0.02

Table 4.2 shows that it quantifies the presence of "negative" words or terms within the privacy policies of various mobile applications. Each row represents a "Mobile App," and the columns (CCC, DBN, DPA, DSR, PCI, PDD, PIA, PMS, TOM, TPD) represent different GDPR-aligned dimensions. The numerical values in the table indicate the cumulative presence or score of these "negative" words for each app within each specific dimension. To determine the best and worst performing mobile applications in terms of privacy compliance, by analyzing the presence of negative words across various dimensions, a total negative word score is computed for each app. Such a score was obtained by adding the measure of each of the dimensions, with a smaller value representing greater compliance (there are fewer negative words used). The application with the lowest overall score was Daktarbhai (0.03), showing the greatest degree of compliance with minimal or absence of negative words in almost all categories. Following closely in terms of high compliance are DocTime

(0.04) and Arogga (0.06), also showing very low overall negative word scores. On the contrary, Rocket (0.42) had the maximum cumulative negative word score and hence a high level of occurrence of negative words in each category especially in DBN and DPA. There is a great number of applications (Nagad (0.20), Aarong (0.20), Daraz (0.20), and numerous others) with moderate ratings, which received responses scores between 0.16 and 0.20. It implies a relatively moderate degree of adherence in this group.

4.1.2 Quantitative Results

The sector-based analysis showed clear differences in the manner by which GDPR compliance is articulated in entertainment, healthcare, e-commerce, and mobile financial services. Our analysis started with sentence-level examination of privacy policy in every sector and was performed based on sentence structure (the presence of negative (compliant-denying, vague, and non-committing) and positive (compliant-affirming) constructions). Based on this structure, this was able to bring out more fine details in the communication of compliance rather than mere existence of keywords.

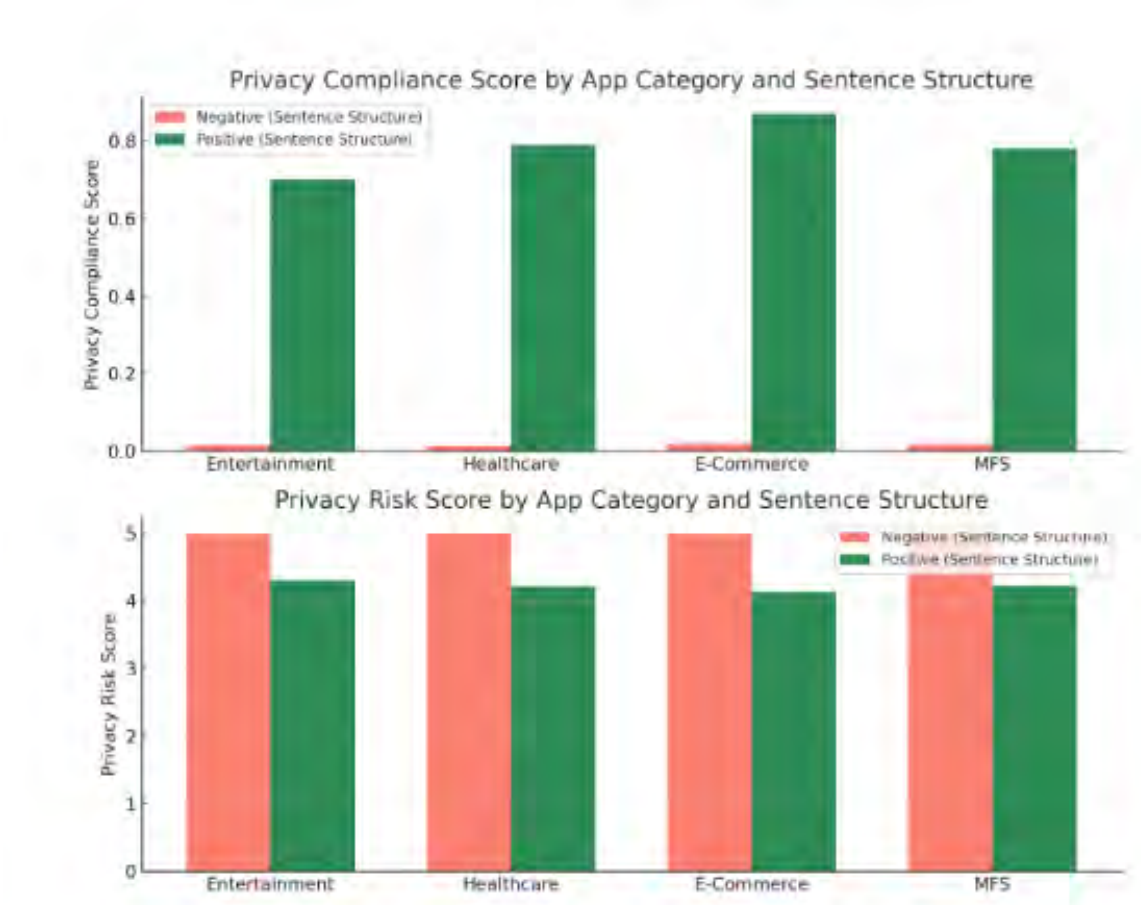


Figure 4.1: Privacy compliance score by app category and sentence structure

This chart 4.1 gives us a demographical state where we see privacy compliance scores by application category and sentence structure. In the x-axis, we plotted the applications and in the y-axis we plotted privacy compliance score. Table 4.3 illustrates this chart with explicit details.

The hit ratio and the formula of compliance posed in 3.3 through 3.5 were used in calculating sectoral Privacy Compliance Scores (PCS) and associated Privacy Risk Scores (PRS).

Table 4.3: Sector-wise PCS and PRS Scores with Key Observations

Sector	**Structure Type**	**PCS (1-5)**	**PRS (1-5)**	**Key Observations**
Entertainment	Negative	0.016	4.984	Very limited GDPR content; no misleading use of terms
	Positive	0.70	4.30	Clear compliance expressions in several dimensions
Healthcare	Negative	0.013	4.987	Extremely sparse GDPR mention; no strategic evasion
	Positive	0.79	4.21	Highest PCS among all sectors
E-Commerce	Negative	0.020	4.980	Slightly higher mentions; still lacking in meaningful commitment
	Positive	0.87	4.13	Most comprehensive and transparent GDPR alignment
MFS	Negative	0.019	4.981	Passive tone; minimal GDPR context without misleading language
	Positive	0.78	4.22	Strong coverage in DPA, DBN, and TOM compliance dimensions

Note: PCS = Privacy Compliance Score; PRS = Privacy Risk Score. Scores scaled between 1 (low) and 5 (high).

In the table 4.3 we see in contrast to the negative PCS scores, that is, scores on sentences denying responsibility or ones avoiding clarity, or vague sentences, all the sectors got very low scores. This implies that apps tend not to use GDPR-related keywords in bad faith and tricky forms when creating disclaimers or evading responsibility. The negative PCS of all sectors ranged between 0.01 and 0.21, whereby the healthcare apps showed the lowest negative PCS, which is 0.013 meaning that they hardly used such words in negative or noncommittal meanings. Then comes the

entertainment sector (0.016), the mobile financial services (0.019), and e-commerce with the biggest negative PCS of 0.020, which is also, nonetheless, very low. These values indicate that, in situations where apps do not want to declare that their app is GDPR compliant, they merely leave the keywords out rather than applying them in a deceptive or contrarian way. That is, a low negative PCS indicates that policies have been written conservatively and truthfully instead of indicating non-compliance. In contrast, positive PCS scores show the frequency of GDPR-keyword-containing statements that are written explicitly and in the positive. In this case there is a broad range of diversity. Its e-commerce sector produced the most positive PCS of 0.87, with such apps being most energetic and direct in how they describe their compliance with GDPR requirements. They particularly did well under such categories as Data Processing Activities (3.07), Privacy Impact Assessment (1.33), and Third-Party Data Sharing (0.66). Healthcare apps were right behind with a PCS of 0.79 as well, showing that there is robust communication on adherence to GDPR, especially in such areas as Data Breach Notification (1.92) and Third-Party Data Sharing (0.75). The MFS sector had a positive PCS of 0.78 as Data Processing Activities (2.53) and Data Breach Notification (1.36) displayed good performance. The entertainment sector had the lowest positive PCS at 0.70, implying that though these apps do publish compliant statements, they may do it less often or less comprehensively than the others.

Sector	**Negative PCS (Low = less negative use)**	**Positive PCS (High = more compliance)**
Healthcare	0.013 (least negative use)	0.79
Entertainment	0.016	0.70 (lowest positive use)
Mobile Financial Services	0.019	0.78
E-Commerce	0.020 (most negative use, still low)	**0.87 (highest positive use)**

Table 4.4: Privacy Compliance Scores (PCS) by Sector

These findings exhibit an obvious structure: apps are likely to mention GDPR whenever they want to align themselves with it, and are unlikely to mention GDPR whenever they would rather absolve any legal registrations. Notably, they will not be using GDPR terminologies in a misleading light; they just remain quiet when they are not assuring to be on the same side. This weary correspondence stems indicates that when apps say they are compliant when they are not, they are not lying, but simply talk about compliance when it suits them, or when it is blatantly obligatory. A high positive score in PCS depicts other areas where transparency and accountability is put in mind and prioritize, in this case in E-Commerce and Healthcare.

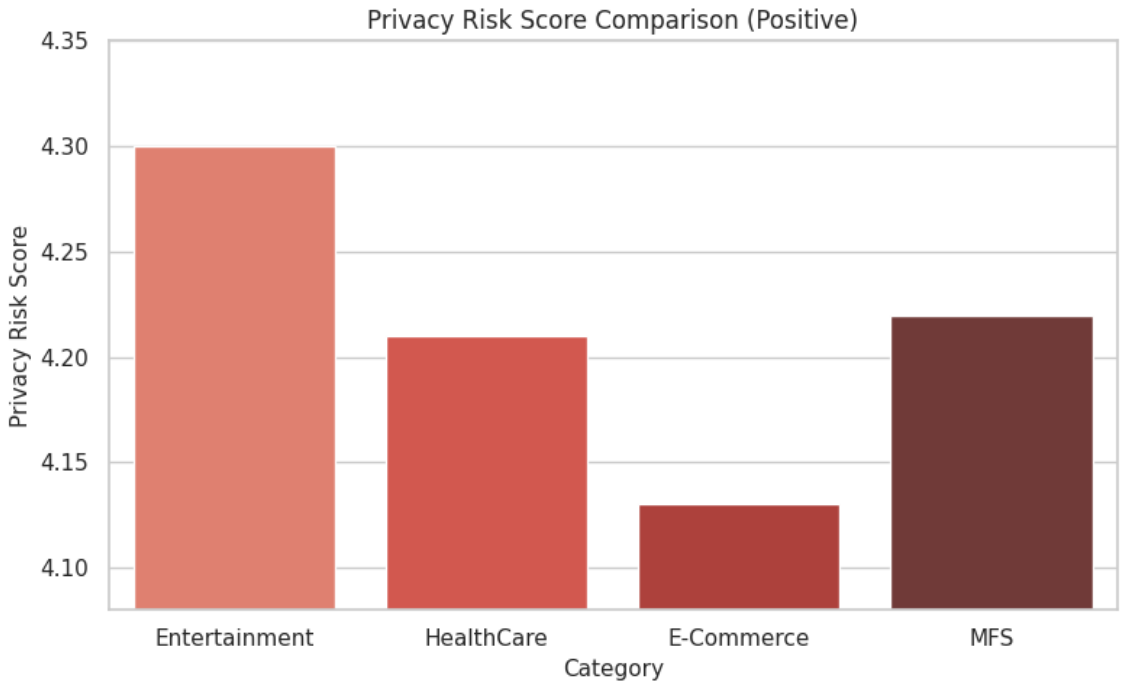


Figure 4.2: Privacy Risk Comparison (Positive)

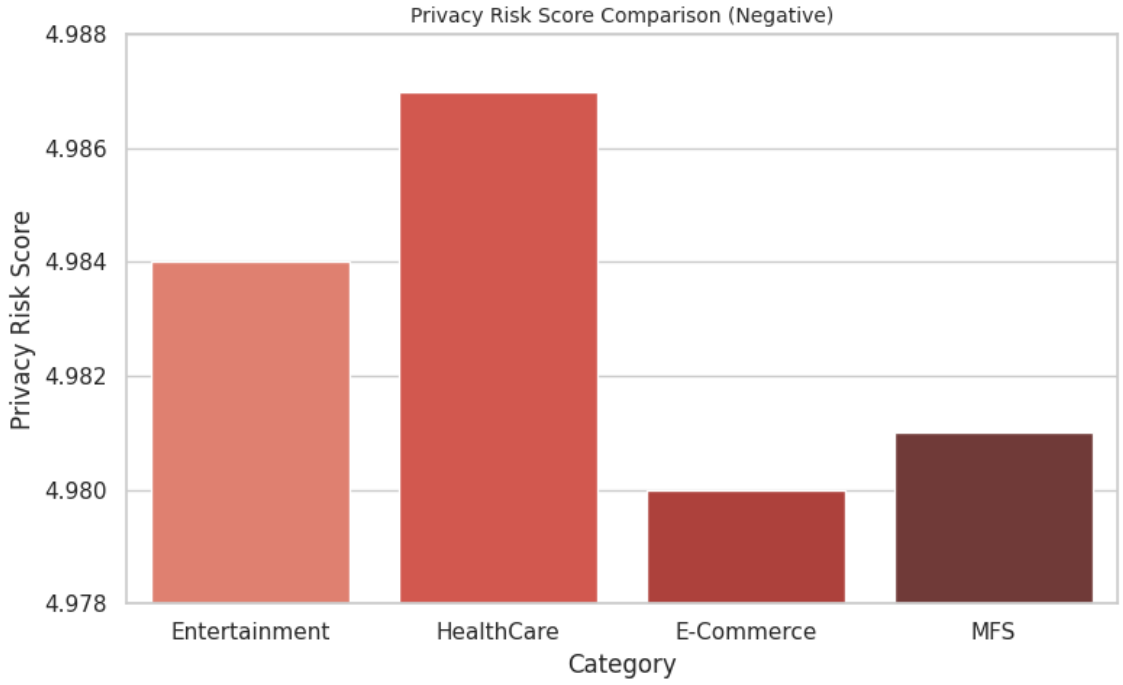


Figure 4.3: Privacy Risk Comparison (Negative)

4.1.3 Analysis

Throughout our analysis, we have been able to provide a strong correlation between sentence-level semantics and GDPR compliances. Policies that had positive framing i.e. had clear and active statements about user rights and data protection responsibilities were always linked with higher scores of Privacy Compliance Score (PCS) and lower risk in privacy. On the contrary, negative construction policies where policies are given in the form of a disclaimer, use of ambiguous language or lack-adaisical statements gave the lowest PCS, and the highest risk scores. It is necessary to mention, though, that it does not imply that instances of GDPR-related terms were employed in a misleading or deceptive manner in negative constructions. Instead, low PCS in negative statements represents that, largely, there were no GDPR keywords in such cases. I.e., apps were likely to leave out all references to GDPR when not claiming compliance. This means that the development team is paranoid and truthful in its communication style and instead of lying about compliance, they do not. Consequently, the presence of GDPR keywords should also be treated as significant, but our data indicates that the way the keywords have been used, whether in a concise and positive way or hidden under vague terms, matters the most. The tone of a policy with GDPR text must not be evasive or take the position of non-commitment. Our solution corresponds to the relatively recent tendency in the domain of privacy studies that focuses on the semantics of the information rather than the mere number of keywords (Zimmeck et al) [4]. Our findings support the fact that framing of language on privacy can play a key role determining perceived and actual compliance- especially when it comes to sensitive environments such as Healthcare and Mobile Financial Services (MFS) where trusting of users is a must.

This was done by including a Large Language Model (LLM), which allowed us to make more accurate interpretations of the meanings of sentences and score along GDPR dimensions in a persistent beyond matching surface-level text associated with individual words. The LLM was able to identify hidden risk as well as latent compliance indicators that the more traditional or search-engine based approaches would fail to detect. It would be of great advantage to integrate the regulatory audit process or app development lifecycle with semantic-aware tools like ours. It would enable the stakeholders to more efficiently realize shortcomings in privacy policies and to facilitate the necessity of more accountable and transparent communication of dealing with the user data.

Chapter 5

Conclusion

5.1 Limitations

Although our research provides important information about the degree of compliance with privacy attributed to the mobile applications in Bangladesh, it has a number of limitations. First, the research only relied on the General Data Protection Regulation (GDPR) as their reference base but also other prominent international privacy regulations such as the California Consumer Privacy Act (CCPA), Brazil's LGPD, and India's DPDP Act. As a result, the analysis may not fully capture jurisdiction-specific legal nuances or allow for comparative insights across global standards. Second, the study comprised analysis of only 32 apps belonging to four categories, which, despite covering the services commonly utilized, are not enough to draw any extended conclusions as they do not represent the full diversity of mobile apps across various sectors and regions. The keyword sample could be expanded to include more diverse and comprehensive GDPR terms to improve the accuracy and depth of compliance analysis. The research primarily focused on analyzing the privacy policies of apps but did not consider user feedback or real-world experiences with data privacy. The Large Language Models (LLMs) have explicitly boosted semantic comprehension; yet, their nature, including latent biases or matters of intended domain-specific misinterpretations, such as those in the law, could undermine the vitality of faulty analysis. Lastly, the selected apps were based on download numbers and popularity, but this could introduce bias, as more popular apps may have better privacy policies than those less commonly used.

5.2 Implications

This research has a number of important practical and policy implications that have a straightforward application in improving digital privacy awareness in Bangladesh.

In a practical sense, the suggested hybrid analytical framework can be considered a scalable and cost-effective system through which application developers, privacy evaluators, and digital service providers can assure in a systematic manner how much GDPR compliance is ingrained into privacy policies. Its usage can also be a motivator to use more transparent, accessible, and regulation-consistent approaches to privacy provisions and, thus, create a culture of accountability and privacy-by-design in development. In addition, the findings have the capacity to raise awareness about digital privacy among the people as they get information on weaknesses occurring in specific sectors, prompting the users to have a better understanding of the decisions they make about their data interactions with mobile applications. At the policy level, the policy implications of the research findings of this study can be applied at the national level and in the case where a country has not elaborated complete laws protecting data. The findings can inform the evidence-based policy-making as it relates to finding solutions and creating the organization-specific regulatory frameworks and privacy policy templates. Taken together, these implications help in the promotion of more privacy-aware implications of the digital ecosystem in Bangladesh and set forth the compelling importance of the ample legal and institutional frameworks entailed to protect personal information on mobile applications.

5.3 Future work

In future, this research could explore several promising directions that can provide solutions to the existing shortcomings and further understanding of privacy compliance in digital ecosystems. To begin with, a broader selection of regulatory frameworks can be introduced into comparative studies, including the emerging digital governance policies of Bangladesh, the Digital Personal Data Protection Act (DPDPA) of India, or the California Consumer Privacy Act (CCPA) to evaluate the degree to which they are compliant in various legal frameworks. Second, increasing the size of such data to cover a wider range of apps, in terms of industry, users, and popularity, may provide evidence of a more general trend and problems in privacy policy compliance within different industries. Third, the integration of policy analysis and technical audit of backend data practices would give a better assessment of the both stated and operational compliance. Lastly, user-centric studies could investigate how individuals in Bangladesh perceive and engage with privacy policies, recognize the usability obstacles, and the informed consent breaches.

5.4 Discussion

This research evaluates privacy policy compliance in Bangladesh's mobile app ecosystem using GDPR as a benchmark, analyzing 32 popular apps across four key sectors. Unlike traditional keyword-based methods, we developed a hybrid NLP-LLM framework incorporating linguistic framing analysis to assess both presence and clarity of compliance statements. Findings reveal most apps include GDPR ter-

minology but frame obligations vaguely, creating compliance gaps and privacy risks. Affirmative statements correlated with higher compliance, while evasive language increased risks, highlighting transparency and usability concerns. The outcome of this research draws attention to the high risk of loss of user privacy in a developing country like Bangladesh where the rate of digital adoption has been escalated much faster than the capacity to develop a strong framework of data protection. It is especially problematic in the nation where digital literacy is not up to standard, legal enforcement mechanisms are still weak, and users might completely fail to recognize and utilize their data rights. This is further increased by the lack of a holistic national privacy law that would subject these users to unregulated data collection, obscure data processing norms and even possible personal data misuse. This paper highlights that there is an urgent need in privacy governance, user education, and usability-minded regulatory guidelines as long as there are still compliance gaps within the privacy policies of many privacy-sensitive digital sites which are badly framed to hide the existence of compliance gaps under-resourced and under-regulated digital environments, such as those in Bangladesh. These findings will help policy formulators and application builders to enhance the control of privacy, particularly in emerging digital markets with evolving legal frameworks.

Bibliography

[1] C. Jensen and C. Potts, "Privacy policies as decision-making tools: An evaluation of online privacy notices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '04, Vienna, Austria: Association for Computing Machinery, 2004, pp. 471–478, ISBN: 1581137028. DOI: 10.1145/985692.985752. [Online]. Available: https://doi.org/10.1145/985692.985752.

[2] M. Lapata and F. Keller, "Web-based models for natural language processing," *ACM Trans. Speech Lang. Process.*, vol. 2, no. 1, 3–es, Feb. 2005, ISSN: 1550-4875. DOI: 10.1145/1075389.1075392. [Online]. Available: https://doi.org/10.1145/1075389.1075392.

[3] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, no. null, pp. 2825–2830, Nov. 2011, ISSN: 1532-4435.

[4] S. Wilson, F. Schaub, A. A. Dara, *et al.*, *The creation and analysis of a website privacy policy corpus*, Aug. 2016. DOI: 10.18653/v1/P16-1126.

[5] F. Schaub, R. Balebako, and L. F. Cranor, "Designing effective privacy notices and controls," *IEEE Internet Computing*, vol. 21, no. 3, pp. 70–77, 2017. DOI: 10.1109/MIC.2017.75.

[6] P. Voigt and A. Bussche, *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Jan. 2017, ISBN: 978-3-319-57958-0. DOI: 10.1007/978-3-319-57959-7.

[7] B. Aljedaani and M. Ali Babar, "Challenges in developing secure mobile health applications: Systematic review (preprint)," *JMIR mHealth and uHealth*, vol. 9, Jul. 2019. DOI: 10.2196/15654.

[8] O. Akanfe, R. Valecha, and H. R. Rao, "Assessing country-level privacy risk for digital payment systems," *Computers Security*, vol. 99, p. 102 065, 2020, ISSN: 0167-4048. DOI: https://doi.org/10.1016/j.cose.2020.102065. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167404820303382.

[9] M. Fan, L. Yu, S. Chen, *et al.*, "An empirical evaluation of gdpr compliance violations in android mhealth apps," in *2020 IEEE 31st International Symposium on Software Reliability Engineering (ISSRE)*, 2020, pp. 253–264. DOI: 10.1109/ISSRE5003.2020.00032.

[10] M. Kuznetsov, E. Novikova, I. Kotenko, and E. Doynikova, "Privacy policies of iot devices: Collection and analysis," *Sensors*, vol. 22, no. 5, 2022, ISSN: 1424-8220. DOI: 10.3390/s22051838. [Online]. Available: https://www.mdpi.com/1424-8220/22/5/1838.

[11] O. Akanfe, R. Valecha, and H. R. Rao, "Design of a compliance index for privacy policies: A study of mobile wallet and remittance services," *IEEE Transactions on Engineering Management*, vol. 70, no. 3, pp. 864–876, 2023. DOI: 10.1109/TEM.2020.3015222.

[12] J. Chen, F. Wang, S. Pang, *et al.*, "A privacy policy text compliance reasoning framework with large language models for healthcare services," *Tsinghua Science and Technology*, vol. 30, no. 4, pp. 1831–1845, 2025. DOI: 10.26599/TST.2024.9010089. [Online]. Available: https://www.sciopen.com/article/10.26599/TST.2024.9010089.

[13] *(In)Security of Mobile Apps in Developing Countries: A Systematic Literature Review — arxiv.org*, https://arxiv.org/abs/2405.05117, [Accessed 18-06-2025].

[14] *Cs.cornell.edu*, https://www.cs.cornell.edu/~shmat/courses/cs5436/shvartzshnaider.pdf, [Accessed 18-06-2025].

[15] *Etd.ceu.edu*, https://www.etd.ceu.edu/2021/azad_ananya.pdf, [Accessed 18-06-2025].

[16] *Global Data Privacy Laws 2019: 132 National Laws & Many Bills — papers.ssrn.com*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3381593, [Accessed 18-06-2025].

[17] *Longitudinal Compliance Analysis of Android Applications with Privacy Policies — arxiv.org*, https://arxiv.org/abs/2106.10035, [Accessed 18-06-2025].

[18] *NLTK Book — nltk.org*, https://www.nltk.org/book/, [Accessed 18-06-2025].

[19] *The effect of online privacy policy on consumer privacy concern and trust — sciencedirect.com*, https://www.sciencedirect.com/science/article/abs/pii/S0747563211002767, [Accessed 18-06-2025].

[20] *Usenix.org*, https://www.usenix.org/system/files/conference/usenixsecurity18/sec18-harkous.pdf, [Accessed 18-06-2025].