Compliance Framework for Personal Data Protection Law Standards

Norah Nasser Alkhamsi , Sultan Saud Alqahtani
Computer & Information Sciences College, Al Imam Mohammad Ibn Saud Islamic University, Riyadh, Saudi Arabia

***Abstract*—Personal data protection laws are crucial for protecting individual privacy in a data-driven world. To this end, the Kingdom of Saudi Arabia has published the Personal Data Protection Law (PDPL), which aims to empower individuals to manage and control their personal information more securely and effectively. However, data management ecosystems that process such data face challenges directly applying PDPL due to difficulties translating legal provisions into a technological context. Furthermore, non-compliance with PDPL can result in financial, legal, and reputational risks. To address these challenges, this paper developed an approach for legal compliance with PDPL through a framework that analyses and translates legal terms into measurable data management standards. The framework guides data management ecosystems in implementing and complying with PDPL requirements and covers all integral parts of data management. To demonstrate the practical application of this approach, a case study utilized two advanced deep learning models, MARBERTv2 and AraELECTRA, to enhance privacy policy adherence in Saudi Arabian websites with PDPL requirements. The results are highly promising, with MARBERTv2 achieving a micro-average F1-score of 93.32% and AraELECTRA delivering solid performance at 92.46%. This underscores the effectiveness of deep learning models in facilitating PDPL compliance.**

***Keywords*—Personal data protection law (PDPL); framework; data management; data protection; privacy policy**

I. Introduction

The existence of personal data protection laws has significant benefits in protecting and governing individual privacy and empowering them with the ability to have a clear vision of their data in a data-driven world where sharing such data has become common and essential to benefit from the services provided in all fields, such as financial, health, etc. On the other hand, the implication of non-compliance with such regulations leads to catastrophic consequences such as financial loss of the issued penalties and breaches lawsuits along with reputational damage. Furthermore, applying governmental regulations is a challenging mission from a data management perspective, as the major obstacle is how to comply with Saudi Arabia's Personal Data Protection law (PDPL) [1], a legally written document in technological environments. For that, this paper aims to develop a solution to the legal compliance problem with PDPL by adopting a framework that illustrates legal terms into technologically measurable standards to guide the organization to implement and comply with PDPL requirements.

As technology advances, various activities rely on personal data, which comprises any information that may potentially lead to identifying an individual. This raises concerns for the privacy of individuals regarding the proper usage and protection of their data. In that concern, many countries have put in place specific laws and regulations for privacy and data protection, such as the General Data Protection Regulation (GDPR) [2] for the European Union and the PDPL, which is the first personal data protection law in Saudi Arabia [1]. This illustrated the standard of data privacy and protection requirements regarding individual data, which is an integral part of governing data privacy. These regulations aim to empower individuals with certain rights (i.e., data subject rights) to manage and control their personal information more securely and effectively and grant people the right to be informed of all operations that are carried out on their data, including collection, processing, and other activities, as well as the right to access, obtain, correct and delete these data. The law was issued in September 2021 and was enacted on 14 September 2023 [1].

The law will be applied to any organization processing personal data related to individuals in Saudi Arabia (even if the processor is an entity present abroad) by any means and lays out penalties in case of non-compliance with the PDPL. Failure to adhere to the regulation requirements can pose substantial financial, legal, and reputational hazards for companies. Also, ensuring compliance with regulations and implementing measures to meet their demands, particularly within technological systems, can prove a pivotal and challenging task for all organizations. These regulations specify what needs to be done without providing explicit guidance on how to accomplish it. Therefore, comprehending and applying legal requirements to an organization is often far from straightforward. This difficulty arises from the numerous ambiguities, cross-references, and domain-specific definitions present in these regulations, which may be quite complex to grasp for individuals without a legal background [3 - 5].

Practitioners and data engineers in the data management community will play a significant role in implementing the compliance requirements as they work directly with the data. Moreover, the absence of the resources and guidance that translate regulation requirements into applicable concepts that could be implemented would make the mission more difficult for data management as it has been addressed by previous research in complying with governmental regulations such as GDPR, along with other obstacles such as a lack of awareness of the upcoming changes and requirements that the law will impose [5]. To overcome these challenges, our approach aims to support organizations in implementing and complying with

PDPL requirements and automating the process. Overall, our paper made the following contributions:

- Determine and analyze the PDPL provisions that are to be translated into organizational and technical standards.
- Develop a framework to streamline the implementation and compliance with PDPL requirements through the analysis and interpretation of regulatory norms into practical organizational and technological strategies.
- A case study demonstrating the use of deep learning classifiers to aid in the compliance of privacy policies with PDPL requirements.

The paper is structured as follows in Section II. We introduce the background information of our research. Next, we present related work in Section III. We then describe our framework in Section IV. Following this, we have a qualitative evaluation of the framework in Section V. After that, a use case scenario in the PDPL privacy policy compliance will be presented in Section VI. Finally, Section VII contains our paper's future directions and conclusions.

II. Background

This section will discuss the pertinent legislation, critical discoveries in this field, and contemporary publications that address the topic.

A. Legislation of Personal Data

PDPL (Personal Data Protection Law) in Saudi Arabia and GDPR (General Data Protection Regulation) in the European Union will be discussed as data protection regulations that apply to the processing of personal data.

1) The personal data protection law (PDPL): Due to the importance of ensuring the privacy of individuals, many countries have introduced laws and legislation that govern the use of personal data to ensure the privacy of individuals and provide the proper protection, such as The Personal Data Protection Law (PDPL) in Saudi Arabia [1]. It was issued by Royal Decree M/19 of 9/2/1443H (16 September 2021), approving Resolution No. 98 dated 7/2/1443H (14 September 2021). which is the first data protection standalone law that governs the use and process of Saudi resident's data by any entities (including public or private) and for the entities outside Saudi Arabia that process residents' data, also including data of a deceased person or their family members, and excludes information used for household or personal proposes.

PDPL defines two types of data personal data, which is "Every statement - whatever its source or form - that would lead to the individual being specifically identified, or make it possible to identify him directly or indirectly, including name, personal identification number, addresses, contact numbers, license numbers, records, and personal property, bank account and credit card numbers, still or moving photos of the individual, and other data of a personal nature." And sensitive data as a part of the personal data which is "Every personal statement that includes a reference to an individual's ethnic or tribal origin, religious, intellectual or political belief, or indicates his membership in civil associations or institutions. As well as criminal and security data, bio-identifying data, genetic data, credit data, health data, location data, and data indicating an individual is unknown to one or both parents" [1]. The Objective of the law is to provide proper protection for individuals' privacy and prevent abuse of any personal data by granting all rights to individuals related to the control of their data. PDPL contains several definitions that must be considered by any entity [6], such as Data Subject, which is defined as "an individual to whom the personal data belongs, his representative, or whoever has legal guardianship over him," a Data Controller that is "any Public Entity, a natural person or private legal person that specifies the purpose and manner of Processing Personal Data, whether the Data is processed by that Controller or by the Processor" [6].

A Data Processor which defined as "Any Public Entity, a natural person or private legal person that processes Personal Data for the benefit and on behalf of the Controller" [6], and a Privacy policy that must include the purpose of collection, the content of the personal data to be collected, the method of collection and storing, how to process and destroy also for the owner's rights with data and how to practice which support the transparency between individuals and any entities work with personal data, another principle is the Purpose limitation dictates the process of personal data is only for the purpose collected. Also, the main principle is the consent of the data owner to carry on the data processing. The implementation of the law will be supervised by The Saudi Data & Artificial Intelligence Authority (SDAIA).

2) The general data protection regulation (GDPR): Another analogous law to PDPL is the General Data Protection Regulation (GDPR) [2]. It is a regulation of the European Union that came into effect on May 25, 2018, and applies to all associations that process the particular data of EU citizens, anyhow of where the association is located in the world. The GDPR aims to strengthen the protection of particular data, giving EU citizens more control over their particular information and mandating that companies handle this data in a biddable and transparent manner. And contains 99 articles that introduce some of the crucial points, including.

- Every European citizen is entitled to eight rights: the right to be informed, access, rectification, erasure, restriction of processing, data portability, avoiding automated decision-making, and object.
- Unequivocal consent from the data owner before collecting and using the data must be assured.
- Appoint a Data Protection Officer (DPO) responsible for every manner in protecting particular data.
- Data breach announcement to authorities and individuals.

B. Privacy Frameworks

This section discusses best-practice privacy frameworks that are built on risk-based approaches to provide businesses

with standards and guidelines for protecting personal data during processing.

1) NIST privacy framework: The National Institute of Standards and Technology (NIST) has established the NIST Privacy Framework as a tool to help in developing services and products innovatively by managing the privacy risks regarding the processing of related personal data, which works as a guideline for organizations to build a privacy program. The framework consists of three components. The "Core" is the first part of the privacy framework. To better manage privacy risks throughout the entire enterprise, the Core is made up of a table of Functions, Categories, and Subcategories that describe certain privacy operations and results. The second component is the profile, which represents the organization's current and desired activities based on the assessment conducted of the core activities on the organization's privacy program. The third component of the Privacy Framework is called the Implementation Tiers which have a view of current privacy risk management practices in the organization to determine the requirements that need to be met that are identified in the profile component [7].

2) ISO/IEC 27701:2019: ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) introduced ISO 27701 as an extension of ISO/IEC 27001, which includes additional controls and privacy requirements to guide organizations in implementing and improving the privacy information management system (PIMS) for providing the proper protection of the personal data [8].

C. Data Protection Regulations and Best Practices Privacy Frameworks Comparison

The main commonality between Data protection regulations and Best Practices Privacy Frameworks is the scope intended to protect the processes of personal data, while the difference is that the regulations have been issued by governments, which means that they must be complied with to avoid non-compliance penalties. Also, the main goal of these regulations is empowering individuals with the right to have control over their data.

On the other hand, Best Practices Privacy Frameworks offer what could be described as the "best to be followed" which means no fines regarding the non-compliance also, the guidelines presented are in a high level of abstraction which will help to build off an effective privacy program in processing data for institutions to gain the trust of relevant individuals.

In addition, and through our reviewing process, these best practices tools could be considered as assisting tools, but not as the main ones for the compliance process with the regulations due to what has been mentioned earlier that these standards work in more general approaches, unlike regulations which are written legal-specific instruction documents that must be followed to ensure the compliance with, and that what has been address by our proposed study.

D. Deep Learning

Deep learning is a branch of intelligence (AI) that falls under the umbrella of machine learning. Its main objective is to enable machines to imitate behavior by utilizing neural networks, also known as deep neural networks (DNNs), for solving complex problems. What sets it apart is that these networks consist of layers of interconnected nodes, allowing them to learn representations of data.

In a network, each layer performs a transformation on the input data, which is then passed on to the next layer. The final layer produces the desired output generated by the network. The remarkable aspect is that these transformations are learned automatically from training data, eliminating the need for feature engineering.

The ability to learn from amounts of data has resulted in significant advancements across various fields, such as computer vision, natural language processing, and speech recognition [9, 10]. Furthermore, numerous models have demonstrated their efficacy in ensuring compliance with the regulation process, as we discuss in our paper, such as checking compliance in privacy policies with GDPR using the Transformers model [11 - 13] and ensuring privacy by applying de-identification techniques on patient images [14].

1) The transformer architecture: Transformers are deep learning models that were developed in 2017 by researchers at Google [15]. They have had a significant breakthrough in the field of natural language processing (NLP) in recent years in different tasks such as language translation, question-answering, and generating human-level text. Before Transformers, models like RNNs [16] and LSTMs [17] struggled with long-range dependencies and parallel processing. The Transformer model addressed these issues through its innovative use of the self-attention mechanism. This mechanism enables the input to interact with each other and understand the context around it through mathematical equations.

a) MARBERTv2: In the evolving field of natural language processing (NLP), several models have been developed to overcome the challenges present in the Arabic language due to its diverse dialects and the combination of Modern Standard Arabic (MSA) with Dialectal Arabic (DA). In order to deal with these complexities, models need to be able to process Arabic as it appears in its various forms. In this domain, MARBERTv2 and its predecessor MARBERT [18] offer enhanced capabilities for Arabic NLP, building upon the innovative BERT (Bidirectional Encoder Representations from Transformers) [19] framework to provide enhanced capabilities for processing Arabic text.

The significant enhancement in MARBERTv2 is extending the sequence length to 512 tokens, compared to the original MARBERT's 128. This adjustment allows MARBERTv2 to encapsulate more extensive text fragments, improving its ability to comprehend and process complex queries and documents in Arabic.

b) AraELECTRA: AraELECTRA is an advancement in Arabic language representation [20], building on the Efficiently Learning an Encoder that Classifies Token Replacements Accurately (ELECTRA) framework [21]. Unlike the approach taken by previous Arabic language models, which primarily relied on masked language modeling for pre-training, AraELECTRA introduces a novel methodology by pre-training a discriminator model. This model is trained to distinguish between valid input tokens and corrupted tokens replaced by a generator network, leading to a more sample-efficient pre-training task. AraELECTRA was pre-trained using the replaced token detection (RTD) objective on large Arabic text corpora. It has been evaluated across multiple Arabic NLP tasks, including reading comprehension, sentiment analysis, and named identity recognition. The results showed that AraELECTRA outperforms some of the current state-of-the-art Arabic language representation models in performance, even with smaller model sizes and given the same pre-training data.

III. Related Work

In the current absence of research on PDPL, this section discusses GDPR compliance as it is the most relevant area of study to our topic. Several researchers proposed different approaches to compliance with GDPR. All researchers attempted to address the challenge of translating the legal requirements into a technical context through the implementation of different mechanisms.

Labadie et al. [4] discussed that organizations struggle to implement GDPR requirements due to a lack of understanding between legal regulations and data management. A capability model was proposed to act as an abstraction layer between regulatory guidelines and compliance requirements. It defines organizational and system capabilities to comply with EU-GDPR. The model helps companies develop approaches to achieve compliance. However, the model does not cover all GDPR requirements, such as the subject's access rights.

Brodin et al. [22] presented a comprehensive framework to support small and medium-sized enterprises (SMEs) in complying with GDPR. The framework comprises three phases: analysis, design, and implementation, and it involves defining personal data, developing policies, and assigning roles to ensure adherence. The framework presented a more abstract level with not much clarification details on how to implement these steps. In addition, most GDPR requirements, such as the security requirements, have not been included.

Rivera et al. [23] proposed GuideMe, a six-step approach to map legal provisions to privacy controls to help elect an applicable solution that could be implemented in software systems for GDPR compliance. It includes a data audit, gap analysis, solution selection, plan review, implementation, and evaluation. The approach is structured to be adoptable by any organization. Yet, they validated and focused on only two GDPR articles (Articles 5 and 25) in the software systems.

L. Piras et al. [24] proposed the DEFeND platform to help organizations comply with GDPR. It integrates various tools and solutions for comprehensive monitoring and control of compliance processes from a single channel and enables users to exercise their data processing rights. However, no platform implementation is mentioned to measure its effectiveness.

Other research focused on a specific aspect of GDPR, such as the privacy policy by El Hamdani et al. [13], who proposed an automatic compliance check for GDPR in privacy policies using machine learning models such as XLNet, T5, and CNN, along with a rule-based approach. The compliance process consists of three main components: (1) extracting and classifying data practices from a privacy policy using machine learning models, (2) encoding Articles 13 and 14 of the GDPR, and (3) assessing the existence of mandatory information using a rule-based mechanism.

Previous research on GDPR compliance has yielded a substantial number of proposed methods and approaches. Some of these methods have focused on specific aspects of the law, whereas others have presented more comprehensive approaches for organizations to implement. However, none of these researchers have achieved a high level of maturity in covering the essential aspects of the regulations or provided clear guidelines that are universally applicable within technology communities in organizations. This underscores the need for a mechanism that comprehensively addresses the essential requirements of the regulation and serves as a roadmap in the compliance process. Furthermore, structuring this mechanism at a level familiar to those immersed in the technological environment will greatly facilitate the compliance process. To the best of our knowledge, there is no framework available to check and assist in the application of PDPL requirements.

IV. PDPL Compliance Framework

Developing suitable methods and techniques for addressing governmental regulations within the data management ecosystem is crucial to ensuring compliance with the requirements set forth. This compliance is necessary to mitigate potential risks, including legal penalties, as the PDPL exemplifies.

The proposed PDPL Compliance framework holds significant importance. Its primary role is to aid in assessing the current state of regulatory compliance and to serve as a guide for achieving the foundational level of PDPL compliance. This will be accomplished by implementing the technical and organizational aspects outlined in the regulation, with a focus on breaking down their interconnected components. Since the framework's core revolves around the PDPL, a legal document composed in plain language, several phases are required to construct the framework and carry out the process of translating the legal provisions into a technical context.

A. Framework Construction Phases

Our proposed framework comprises a series of phases designed for constructing the framework, as depicted in Fig. 1. We analyze the PDPL provisions in the initial phase to extract its core principles. Moving on to the second phase, we translate and map these principles to the relevant data management requirements that are applicable in technological environments.

The requirements are then thoroughly reviewed and formulated to shape the final phase, thus structuring the framework.

Fig. 1. PDPL development processes.

A detailed explanation of each phase and its role in advancing the development of the framework is presented as follows:

Phase One - Analyze PDPL provision: Due to the complexity and ambiguous nature of challenges in legal documents, a consultation with a legal expert and an in-depth analysis of the PDPL is conducted to identify, first of all, the objective of the law, which is ensuring individual privacy and protection of their data through enforcement of principles that stipulate the procedures to be followed by entities that process personal information.

Secondly, the principles that are related to the regulation, such as the Data Subject, Data Controller, and Data Processor, which have been described earlier in the Background section, and finally, identifying all the articles that would pose specific requirements on systems, for that concern the articles that are not related to data processing processes such as penalties and Competent Authority responsibilities has been excluded from the analysis process.

Some of the extracted articles stated the legal requirement clearly and straightforwardly to be articulated to the corresponding technological and organizational context. For instance, Article 12 outlined all the essential points that must be specified in the privacy policy and Article. 30 (2) Stated appointing a Data Protection Officer (DPO) to implement PDPL provisions.

However, most articles lack a direct description, requiring interpretation to facilitate the compliance process in data management systems. The legal and technical requirements will be mapped in phase two.

Phase Two - Map the Legal and Data Management Requirements: As the primary domain of the PDPL law is the individual's privacy, incorporating knowledge of that area has been included via a variety of sources, such as the best practices standards ISO 27701[8] and NIST Privacy Framework [7] that be published to help in protecting the privacy of processing personal data in an organization through laying out the guideline to be followed to achieve the privacy goals also, the previous studies of implementing GDPR compliance have been extensively reviewed to participate in the process of extracting and translating legal requirements into measurements that can be applied to data management, and our analysis revealed sixteen main requirements that are listed with the corresponding articles.

Phase Three - Develop the Framework: The requirement outcomes from the previous steps have been reviewed and formulated into two main modules for structuring the framework: organizational and technical controls, which will provide a clear vision for the data management principles that are responsible for enforcing compliance with the law and help identify the roles and responsibilities of each requirement, the framework architecture depicted in Fig. 2.

Each component is called a control and is broken into a more specific measurement called a sub-control. For instance, in the Organizational category, the sub-control of DPO control stipulates the necessity of appointing a Data Protection Officer responsible for PDPL provisions implementation as the regulation specified in Art. 30(2). Another component, Notifications, represents all the processes and procedures required to notify all related parties about the relevant principles stated in the regulation to ensure transparency between the data controller and relevant parties. Any related data party is notified when data is amended under the first sub-control, such as when a correction is made per Art. 17(1). However, for the second sub-control, as it corresponds with Art. 20, a notification process must be adopted in case of a breach for both the Competent Authority and the data subject. The last sub-control is specific to credit cards for disclosure requests, and a process must be implemented to notify the personal data owner, as stated in Article 24(2).

Data Security control in the technical category contains several sub-controls that represent more details of the necessary technical measurement for protecting the data to provide a clarification of what has been stated on.

Technical measures in Article 19: "The Controller shall take all the necessary organizational, administrative and technical measures to safeguard Personal Data," In the first two sub-controls, we see that it is essential to implement safeguards for ensuring confidentiality by protecting data using proper encryption methods and data loss prevention (DLP) techniques. As for the third sub-control, it is imperative that data integrity is ensured by implementing an integrity checksum mechanism, such as hashing, and for the last sub-control, it is necessary to complete all the backup and recovery operations to ensure the data's availability.

B. Framework Components

The proposed framework is designed to be adaptable and expandable for managing regulatory changes and the addition of any future components that may be added to provide privacy and data protection.

The first module of the framework is Organizational controls, which are the strategic processes implemented to ensure the protection of personal data from a managerial point of view to enforce compliance. It consists of eight organizational control components that are documented as follows:

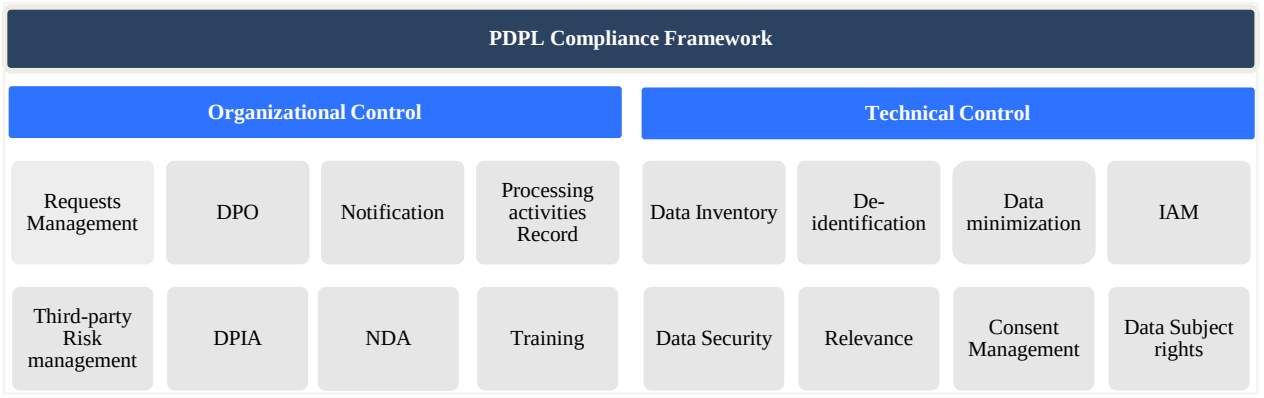

Fig. 2. PDPL compliance framework.

1) Data protection officer (DPO): Assign the Data Protection Officer to ensure that all personal data processing activities comply with the relevant data protection laws and regulations through governing and implementing PDPL provisions and policies.

2) Requests management: Provide mechanisms that must be adhered to give a response channel to data subject requests regarding their rights.

3) A non-disclosure agreement (NDA): A non-disclosure agreement with any related party regarding processing recorded personal data is documented (e.g., signing an NDA document for employees).

4) Notifications: Processes and procedures to notify the relevant principles stated in the regulation, such as Data Subject or Authority for the necessary conditions, for instance, security breaches.

- Notification for personal data amendment to any related parties that process such data, for example, by email.
- Procedure for the Competent Authority notification of a privacy breach or event within 72 hours and for the data subject' in case of harm.
- Procedure for notifying the data subject in Credit data disclosure requests by any party.

5) Processing activities record: To ensure that Authority requests are adequately documented, it is necessary to include specific information in the record of processing activities. Which should consist of the controller's contact details, the purpose of the processing, a description of the data subjects, any other entities that personal data has been or will be disclosed to, whether personal data has been or will be transferred outside the Kingdom or disclosed to an entity outside the Kingdom, and the retention period.

6) Data protection impact assessment (DPIA): Data Protection Impact Assessment, which is the process of addressing privacy and data protection risks in processing personal data to provide proper assurance in mitigating risks and providing protection, must be included for processing activities relating to any product or service offered to the public, according to the nature of the processing carried out by the controller.

7) Third-party risk management: A critical aspect of protecting personal data is to evaluate and assess third-party entities who handle it on behalf of the organization. This involves conducting due diligence on data processors to ensure they have sufficient data protection guarantees, such as using a privacy and security checklist, before allowing them to process the data.

8) Training: Implement a Personal Data Protection training program to foster a culture of safeguarding personal information. By providing employees with the tools and knowledge to handle such data properly, this program can increase awareness levels and promote responsible practices in data protection.

The Technical controls comprise the second module of the framework, containing eight controls, and are responsible for protecting and preventing personal data from being compromised. These controls are documented below.

9) Data inventory: An inventory that contains any assets or processes related to processing personal data, such as the data itself, location, action, or purposes, etc.

- Personal data and sensitive data elements are specified and categorized (e.g., Health data).
- The Data Subject ("data owner") is specified and connected to the data.
- Processing actions on personal data are defined and mapped to the data (e.g., collect, store).
- Systems that Process personal data and purposes of processing are identified and mapped to each data.
- Data Processing Locations are specified and mapped to the data (e.g., geographic location, Cloud).
- Retention periods of data are defined and mapped to the data.

10) Consent management: The procedures to provide transparency regarding Data Subject consent through

implementing a precise mechanism for consent and consent withdrawal.

- Clear procedures must be established to obtain valid consent from data subjects for processing their data. These procedures include opt-in consent checkboxes or buttons, signed consent forms, etc.
- Procedures for consent withdrawal are implemented clearly to the data subject, such as unsubscribe links.
- Procedure for explicit collection consent, changing of collection purpose, disclosure or publication of Credit Data obtained along with the Credit Information Law depicts [25].

11) Data subject rights: The process enables Data subjects to practice the rights dictated in the law regarding their data, such as access, deletion, modification, etc.

- The Privacy Policy outlines the details of the collected personal data, including the purpose, method, storage, and processing, how the Personal Data shall be destroyed, the data subject's rights, the legal basis for data collection, the data controller's identity, entities to whom Personal Data will be disclosed and the Consequences and risks of not gathering Personal Data is adopted and available in clear text to Data Subject.
- A mechanism for the data subject to access personal data. (e.g., preview data on a Web page)
- A mechanism for the data subject to request (obtain, correct, or delete) personal data. (e.g., Requests Web page).

12) Data minimization: Limit the minimum amount of personal data to the purpose of the collection process.

- After fulfilling collection purposes, personal data is destroyed.
- Personal Data out of Data Controller business purpose scope destroyed.

13) Relevance: Data control and audit processes to ensure accuracy and relevance to processing purposes as Audit/log records are implemented and reviewed to incorporate the principle of Purpose limitation and data accuracy and comply with the privacy policy (e.g., limiting processing to collection purposes only).

14) De-identification: The process of discarding any data directly related to the identity of a particular individual is applied to retain, collect, or process Personal Data without consent, such as implementing the Data Masking process.

15) Identity and access management: Access to Personal Data is restricted to authorized individuals, processes, and devices.

- Identities and credentials are issued, managed, verified, revoked, and audited for authorized individuals, processes, and devices.
- Remote access is properly managed.
- Access permissions and authorizations are managed, incorporating the principles of least privilege and separation of duties, such as implementing a Role-based access control mechanism.

16) Data security: The necessary safeguard to ensure personal data's confidentiality, integrity, and availability.

- Encryption methods are used to protect personal data (e.g., Database Encryption, TLS)
- Data loss prevention (DLP) techniques protect data from loss.
- Integrity-checking mechanisms verify Personal Data integrity (e.g., hashing, logging).
- Implementation of the Backup and Recovery process.

Through the implementation of the framework components, organizations can ensure compliance with the PDPL requirements. The framework provides a comprehensive assessment of the current state of regulatory compliance and offers guidance to organizations to achieve a baseline of PDPL compliance. Additionally, the framework assists data management in identifying the roles and responsibilities of each framework component, which is essential for effective data management. The PDPL compliance framework is a valuable tool for organizations seeking to maintain compliance with regulatory requirements and ensure the protection of personal data.

V. Evaluation

Throughout the creation of the PDPL framework, great emphasis was placed on the crucial role of individuals in its implementation, considering them as the main element in any compliance process. To this end, we employ semi-structured interviews with professionals in data management, governance, and privacy engineering to answer the research questions. Inspired by the structured approach to qualitative research as outlined by Kallio et al. [26]. We have conducted an in-depth analysis of the literature review and a comprehensive examination of the PDPL and GDPR, and in incorporating best practices such as ISO/IEC 27701:2019 [8] and NIST Privacy Framework [7] standards. From these studies, we developed a set of criteria named “Framework Assessment Criteria,” which are clarity, applicability, usability, comprehensiveness, adaptability, accountability, and continuous improvement to evaluate the effectiveness of the framework.

A. Participant Selection

1) Scope and criteria: The research is centered on experts who work in data management, governance, and privacy. They play a vital role in enforcing and implementing PDPL compliance frameworks. These professionals are in a unique position to offer valuable feedback on the effectiveness of the framework, any challenges they face, and how the framework can be improved.

2) Sample size: For this study, a sample size of six participants was chosen due to the qualitative research method's focus on depth over breadth. This allows for thorough and nuanced insights into the application and impact

of the PDPL compliance framework while remaining manageable for detailed analysis.

3) Selection process: Participants for the study were chosen through a systematic review of LinkedIn profiles. This allowed us to identify professionals who had relevant experience and expertise in data management, governance, and privacy. Our selection criteria included the following:

a) Demonstrated expertise in data management, governance, and privacy practices.

b) Working on Saudi Arabia.

c) Involvement in PDPL-Related Projects or Any Relevant Regulation.

By following this process, we aimed to ensure that we selected individuals who could offer valuable perspectives on evaluating the PDPL compliance framework. Table I illustrates the overview of participant informants.

TABLE I. PARTICIPANT OVERVIEW INFORMATION

Code	***Job Title/Position***	***Years of experience***
P-01	Senior privacy consultant	5
P-02	Chief Information Security Officer	9
P-03	Data protection and privacy supervisor	6
P-04	Data Governance Advisor	8
P-05	Data Protection Manager & DPO	7
P-06	DPO	1

B. Data Collection

A pilot interview with a data privacy specialist was conducted to verify the initial interview guide developed. The pilot interviews demonstrated the need to present more questions about the current state of the PDPL compliance process that the participants do and the challenges they face as the law has taken place to help compare what has been applied in the workplace. And our proposed framework. For that, two questions were added to the opening questions: "Can you describe the process your organization follows to ensure compliance with the PDPL or relevant data protection regulations?" and "Do you face challenges applying these regulations? What are they?". By adding these questions, we will provide a comprehensive overview of PDPL compliance practices in real-world settings and assess how our proposed framework might enhance these practices.

1) Interview process: The interviews with participants were conducted online utilizing Microsoft Teams [27], where both sides mutually agreed upon the interview time, and the discussion was conducted in Arabic and English. At the start of each session, the study's aims were introduced, followed by a guided discussion that allowed for a detailed exploration of the goal and components of the PDPL framework. The interviews were audio-recorded with prior consent from the participant, anonymized, and stored on a secure drive before being destroyed post-transcription, and the interview duration ranged from 45 to 90 minutes.

C. Data Analysis

For data analysis, a Thematic analysis (TA) approach was followed, which is a qualitative research method used to identify, analyze, and report patterns or themes within data [28].

This approach was followed to evaluate the PDPL compliance framework's effectiveness in applying and facilitating the compliance process through analyzing the interview scripts to identify and outline the ability to provide significant insights into the facilitative role of compliance frameworks in aligning organizational practices with PDPL provisions.

A detailed evaluation of the framework's applicability in practice employs both inductive and deductive methods.

The interviews were reviewed and analyzed for patterns and themes following the six-phase thematic analysis process detailed by Braun and Clarke [28]. The six-phase process of thematic analysis is widely employed for the study of qualitative data. A systematic and adaptable approach is provided for the identification, analysis, and reporting of themes within a dataset. The six phases are outlined as follows:

1) Familiarization with the data: In this phase, immersion in the data is undertaken to become acquainted with its content. Transcripts, notes, or other qualitative materials are read and re-read to gain a comprehensive understanding of the meaning and patterns. The interview transcripts were imported into MAXQDA 2024[29]. Moreover, two stages of the analysis process were defined in this phase.

- Stage -I: This stage focused on a general inquiry regarding the current state of compliance with the PDPL in the organization, where the inductive approach was followed.
- Stage -II: Following the inductive insights gained from Stage-I, a deductive approach was followed and grounded in seven categories pertaining to the evaluation of the framework in Stage-II. These seven categories were identified based on pre-defined “Framework Assessment Criteria” to assess the effectiveness and capability of the PDPL compliance framework to facilitate the compliance process.

2) Generating initial codes: The hybrid approach that has been followed structured the generating of the initial codes based on the two stages. In Stage -I, the codes were developed through meticulous line-by-line reading of the interview transcripts, while in Stage – II, the data were coded based on their relevance to the pre-defined Assessment Criteria.

3) Searching for themes: Patterns and clusters of codes were identified and categorized into themes, creating a thematic map containing the initial themes and codes correlated.

4) Reviewing themes: There is a two-level creating this phase. The first level is reviewing the themes, sub-themes, and code to ensure consistency and logical connection among the extracted data.

For the second level, a similar process followed, but a comparison will be made for the entire data set to ensure the validity of the extracted data to the main analysis goal.

5) Defining and naming themes: After refining the themes, the names were clearly defined, and a clear and concise description for each theme was created, ensuring an accurate representation of the underlying data. The themes are outlined in the mind map in Fig. 3.

6) Writing the report: The final phase involves the writing of the report, where the findings of the thematic analysis are presented. This includes providing a clear account of the research question, the analytic process, and the identified themes. Illustrative quotations and examples from the data are often included to support each theme.

D. Findings and Discussion

1) Stage -I:

a) Theme 1: Organizational compliance process: Compliance with the regulation process is the approaches and standards that have been followed to adhere to regulation, and as it is not a new concept for organizations, as a comprehensive set of regulations in Saudi Arabia has been imposed, including business conduct and labor laws, as well as data protection and cybersecurity. All the participant were familiar with PDPL and has been involved in the compliance with the regulation along with other regulation such the financial sectors that follows SAMA [30] by P-02 and P-06, for the PDPL the regulation is supervised by SDAIA for that all the participant follows and uses procedures and tool that been presented some of them also created their own privacy system before even the regulation took place following the international standards as P-04 outlined 'We established our privacy department in 2018 before the law was published, which was in 2021, we followed the international standards regarding privacy', yet it is been emphasized that the compliance process must start with robust system for the entire organization, starting with top management support and understanding as P-01 and P-06 stressed 'where I think that it's so important for information to be, on privacy to be understood at top level, from there that then cascades down operational level', 'Very necessary that we bring our, higher management on board because this program, You cannot run this program without the buy-in of the higher management', along with structuring a Data Protection Governance Program that responsible for all the related strategic and operational process to be involved in, also P-04 and P-05 stipulated the necessity for an internal awareness campaigns to clarify the objectives of the PDPL and processes that must be followed to adhere to this regulation in order to facilitate the compliance process as it relates to all departments and parties in the organization such as employees .

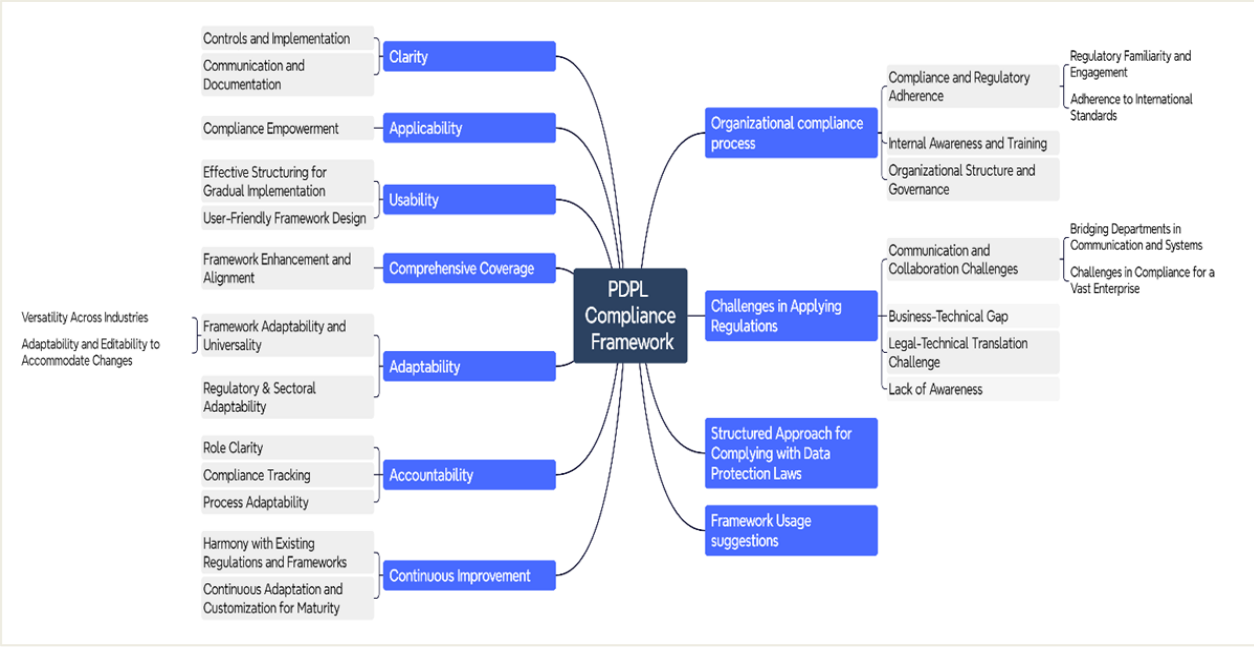

Fig. 3. Compliance process themes.

b) Theme *2: Challenges in applying regulations*: Applying such regulation is not an easy task for the organization due to it is connectivity to all parties within the organization, such as departments or even external parties, such as subjects or clients. All the participants agreed on having challenges in applying it, such awareness as most of the participants stressed that one of the major challenges is the lack of awareness of the law itself for employees and subjects who share the information without proper consideration of what the regulation stated, another challenge in The implementation process in some entities arise from a mismatch between the business and technical requirements as the case for P-03 'Yes, The pinpoint of challenges is the mix in the implementation between business requirements and technical requirements in some entities.' The size of the

organization and its clients play an essential role in complying with regulatory requirements. The larger the size of the company and the number of clients, the more complex and challenging the process becomes because it requires more significant effort in internal communication in the organization and communication with external parties such as clients, in addition to efforts to balance customer protection and protecting of the organizations' revenues as well. Also, struggling to translate legal jargon for technical understanding is a challenge that has been mentioned by P-04 and P-05 as it is necessary to keep all the requirements clear to implement for the related teams and how to overcome this challenge according to P-04 and P-05 'We read and analyzed it and created a control framework derived from the law that translates the regulation language into an understandable language in the company, to facilitate the compliance in the department, and it is similar to the framework you suggested,' 'The technical department was confused about what should be done, making us work with them through steps to clarify the requirements.'

c) Theme 3: Structured approach for complying with data protection laws: It is crucial to comply with the PDPL. However, attempting to achieve compliance in a disorganized, disjointed, or step-by-step manner is counterproductive and unlikely to meet the rigorous standards set by the regulation or the specified compliance deadlines as mentioned by P-04 'if I take one article of the law and work to apply it and then shift to another one, I will be distracted, unlike if I have a structured approach for me as responsible for applying it and for the other related departments, which will ensure to comply with the PDPL before the deadline and provide the top management an organized view on the compliance process state.' Also, all the participants agreed and asserted having a structured approach to compliance is crucial as it provides clear guidance to all who are involved in the compliance process, including top management, departments, etc.

d) Theme 4: Suggestions regarding the use of the compliance framework: Soliciting Suggestions Regarding the Use of the Compliance Framework play a pivotal role in evaluating and enhancing the framework's effectiveness. This approach fosters active engagement with participants, encouraging them to share their experiences and insights. Such interaction enriches the framework with diverse perspectives and prompts a reflective process aimed at continual refinement and adaptation. By integrating feedback from those directly impacted by the framework, we ensure its relevance, practicality, and efficiency in addressing current and future data protection challenges. For that, several suggestions were presented by that participant, as some were related to the framework structure, which would help easily to add more controls, such as adding the policies related to applying the data protection controls as a responsibility of the DPO, or adding tooling and data-sharing control; as P-03 stressed, "Adding a component called tooling, which contains all necessary tools such as consent management, metadata management, etc. and Adding a data-sharing component" will help to have a clearer path in implementation process while others suggested linguistic refinement of some of the terminologies has been mentioned in the Stage - II themes to reflect the marketplace terms.

2) Stage - II:

a) Theme 5: Clarity: Clarity of the framework measures the quality of the terminology being clear, understandable, and free from ambiguity, and it has emerged as a crucial aspect of the compliance framework. Participants highlighted the importance of precise language and terminology in the framework to ensure straightforward interpretation and application for Specialized and Non-Specialized Audiences as it has been applied, as many participants agreed on the benefit of diving the controls into organizational and technical as P-04 stressed, "It is very clear, especially dividing the controls into two levels, organizational and technical controls, which will provide a great benefit for companies to separate the focus of the control types" also for P-06 who described the necessary requirements of a framework would be "A framework needs to be short, concise, and robust" to make it easy to understand and pinpoint the core objectives. While the framework was generally perceived as clear and straightforward, there were suggestions for linguistic refinement in the privacy domain. For instance, replacing the control "processing activities record" with "Record of processing activities (ROPA)" and "Relevance" with "Data Monitoring".

b) Theme 6: Applicability: The applicability focused on the framework's practical guidance for different industries and data management activities to be able to apply the framework component in their environment. Participants noted that the framework provided valuable insights into PDPL controls as the framework emphasized the essential aspect of the law and structured to be applicable by the data management community as P-05 stressed, "I think it's really good. The reason being is because you've really touched upon the core fundamentals" and P-06 outlined regarding the using of the framework in the compliance process with the law "The chance is very big because here in the framework, the scope is identified, and the main points are clear to start the compliance process from, compared to the main law and regulations, which could be interconnected and have multiple exceptions.", making it suitable for implementation across various organizational contexts. However, there were suggestions to enhance the framework with additional components, such as tools for consent management and metadata management, to further improve its applicability and add a new main control for security measures.

Applicability plays a crucial role in any framework, as it specifies the capability of an organization to adapt and apply the components of the framework efficiently. The value of a framework lies not just in its theoretical underpinnings but, more importantly, in its practical application across diverse organizational contexts and challenges.

c) Theme 7: Usability: Usability emerged as a key theme as it represents the user-friendliness level of the framework to be understandable to the targeted audience, with most of the participants expressing satisfaction with the

framework's user-friendly level. The division of the framework into organizational and technical components was appreciated for its clarity and ease of implementation. Also, the clarity and simplicity of the language used, as highlighted by participant P-06: "The language use is perfect, because there's not so much legal jargon, but there's enough to be understood on what is required." stressed the significance of ensuring information is easy to understand and process. It is crucial to avoid using complicated legal terms as this delicate balance appears to have been successfully achieved in the presented framework. Suggestions for minor changes in terminology were made to enhance usability further. For instance, participants recommended changing some of the control names to be aligned with the data management common terminology language, such as "organizational control" into "business control,"

The role of usability in developing the framework is to produce a guideline that is easily followed and implemented by the users without meeting difficulties. Usability will enable the process to be adopted smoothly, making it easier for the organization to apply the framework most effectively.

d) Theme 8: Comprehensive coverage: Comprehensive Coverage refers to the ability to encompass all the necessary fundamental concepts of the PDPL regulations in the framework to ensure reach out to a high level of maturity in data protection practices, which in turn facilitates the compliance process. All Participants highlighted the importance of comprehensive coverage within the framework to address all aspects of compliance effectively and agreed that the framework covered a comprehensive and essential range of regulations such as consent, consent Withdrawal, and data subject rights, etc. Moreover, it guides in structuring the data management office (DMO) that is responsible for applying the PDPL regulations as P-03 and P-04 outlined: "It covered a wide range of regulations such as the subject rights, data inventory, and consent. Also, having sub-controls that define the process that must be implemented made it very mature.", "This framework provides the overall view of the compliance process with PDPL, which makes it able to structure the data management office (DMO) based on it.". However, there were suggestions to include additional components, such as adding the policies related to applying the data protection controls as a responsibility of the DPO, as mentioned by P-05: "The framework is very comprehensive, and it covers a wide range of technical parts, such as data subject rights, data minimization, and others, this is also the same for the organizational components for improvement in the DPO component. Creating the required policies must be mentioned as one of the DPO responsibilities, "while others suggested including data transfer outside the Kingdom of Saudi Arabia in the framework.

Overall, feedback from the participants emphasized that the framework reached a high level of maturity in encompassing the crucial requirements of the PDPL regulatory that are related to the compliance process within a data management system and the ability to be enhanced and cover a broader scope.

e) Theme 9: Adaptability: The adaptability theme emphasized the framework's flexibility in responding to changes in data protection regulations and organizational requirements, along with being designed to be applicable across different industries, which significantly impacts the framework's effectiveness level and long-term viability. The participants noted that the framework was adaptable and editable, making it easy to incorporate new changes and updates, for instance, in technologies, etc. This ease is attributed to the structured approach that has been followed to build the framework, characterized by the methodological division of the organizational and technical controls. Such a division aids in seamlessly integrating changes related to sector-specific and regulatory requirements, as mentioned by P-03: "It would be adaptable to changes, as the changes in the regulation will be minor and sectorial such as in health data or credit data and these changes could be added to the framework as controls or domains."

f) Theme 10: Accountability: Accountability is a crucial aspect of any framework in any data protection system. It involves assigning clear roles and responsibilities to every party involved in the implementation of the framework to aid and track compliance. Based on robust feedback received from participants, it is evident that the framework plays a pivotal role in ensuring accountability when applying the PDPL requirements. The controls structured within the framework make it significantly easier to assign roles and responsibilities, and one such example is the implementation of RACI matrices for each control. As noted by P-04, "the framework controls made it easy to assign a RACI matrix to each control, which ensures that each employee's responsibility is identified." Moreover, Participant P-05 highlighted the framework's adaptability in incorporating additional details like timelines and procedural steps. This adaptability not only advances accountability but also streamlines the application of PDPL requirements, demonstrating the framework's effectiveness in fostering a robust data protection environment.

g) Theme 11: Continuous improvement: Continuous improvement is a vital component of any framework in data management to accommodate the fast changes that appear in the surrounding community, such as the development of technology, evolving changes in rules and regulations, and the ability to keep compliance process with regulations. Participants noted that the framework provided a foundation for ongoing improvement and gave the opportunity to encompass regulatory and business process improvements. Furthermore, one of the participants highlighted the ability of the framework to harmonize with other regulations and best practices such as National Cybersecurity Authority (NCA) regulation, NIST and ISO, etc. This capability ensures that organizations can not only comply with current standards but also remain poised to incorporate future developments in data protection and any other domains.

The evaluation process was conducted to assess the framework's effectiveness in facilitating the compliance process for organizations with PDPL provisions. According to

the findings from the analysis process, the PDPL compliance framework has been recognized overwhelmingly by all participants as a robust and effective instrument crafted with a focus on practicality and ease of implementation as it has been managed to be structured firmly, which it has been able to translate the ambiguity of the Legal Jargon into clear and understandable terminology to be implemented efficiently. The division of the controls into two main modules, organizational and technical controls, has provided a clear vision for the data management principles responsible for enforcing compliance with the law. Also, the applicability and usability provided a versatile, straightforward application and user-friendly design guidance tailored to diverse industrial needs and data management activities. Moreover, the framework was built to be flexible and adapt to ongoing technological and regulatory developments, which is a critical advancement. This adaptability ensures that the framework is capable of addressing future challenges and evolving data protection standards. Additionally, its comprehensive scope, covering all essential PDPL requirements, sets a high benchmark for data protection maturity. This extensive coverage guarantees that organizations employing the framework can achieve and maintain advanced data security and compliance.

While the findings are predominantly positive, we must acknowledge the constructive feedback and suggestions for improvement identified through our analysis.

In conclusion, the PDPL compliance framework emerges from this analysis as a vital instrument for organizations to streamline the implementation and compliance process of data protection regulation.

VI. Case Study – Privacy Policies Compliance

The core objective of PDPL is to empower users with control over their personal information. This legislation encompasses a range of rights, including the right to access a privacy policy. This legal document serves as a guide to the processes that an organization or company has established to manage the personal data of its users. Additionally, it is crucial for companies to instill confidence in their users by assuring them of the security of their personal information. Without a well-crafted privacy policy, companies run the risk of damaging their reputation and losing the trust of their customers. In that matter, several studies have been conducted to check compliance with privacy policies on regulations such as GDPR using deep-learning methods such as Transformers [15], which have demonstrated their efficacy in assessing the compliance of privacy policies with laws such as GDPR [11 - 13].

However, in this case study, we choose to implement one component (i.e., privacy policy from data subject right in the technical controls module) from the proposed PDPL framework as proof of concept. To the best of our knowledge, this is the first case study to automate and implement the PDPL framework in Saudi website data in terms of analyzing privacy policies.

A. Privacy Policies Compliance Approach

Deep learning models were developed and built to evaluate the adherence of websites in Saudi Arabia to PDPL standards in their privacy policies. Various models from the Transformers family have been utilized for multi-class classification purposes. These models are pre-trained in Arabic domains to achieve superior results in our area of focus. The models employed include MARBERTv2 and ARAELECTRA, a set of models that were developed to handle natural language processing tasks for the Arabic language. They are based on the transformer architecture, which has been widely used in deep learning for various NLP tasks.

The classifiers were fine-tuned on the Saudi Privacy Policy Dataset [31]. The process was carried out in Google Colab [32], where users can write and execute Python code in a collaborative and interactive environment without installing any software and running it in the cloud using eight epochs. The dataset was randomly split into the following subsets:

- Training set: 80% of the data (3710 of a total of 4638 text lines)
- Testing set: 20% of the data (928 of a total of 4638 text lines)

B. Dataset

The dataset used for this study is the Saudi Privacy Policy Dataset [31], which comprises a collection of privacy policies from 1,000 websites representing diverse sectors in Saudi Arabia, including healthcare, education, finance, government, e-commerce, and other industries.

TABLE II. Dataset Statistics

No. Files	1000
No. Tokens	775,370
No. Text Lines	4,638

The corpus statistics are shown in Table II, which contains more than 4K lines of text and 775K tokens, with a corpus size of 8,353 KB. The feature annotations are based on ten high-level categories derived from the Personal Data Protection Law (PDPL). They are numbered from 1 to 10, further branching into 21 specific content categories, as shown in Fig. 4. The PDPL category distributions among the datasets are shown in Fig. 5.

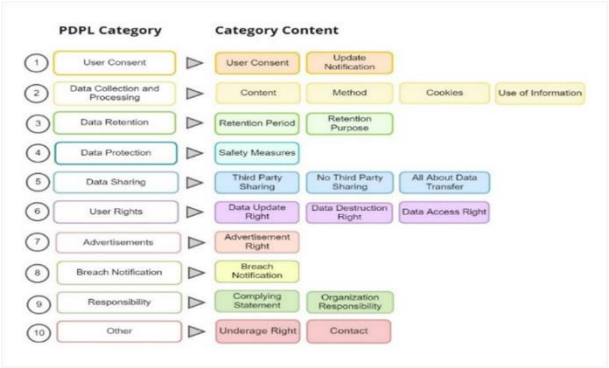

Fig. 4. PDPL annotation category [31].

The PDPL categories were considered in the classification process, and details of the categories and their correspondences with PDPL clauses are explained in the following:

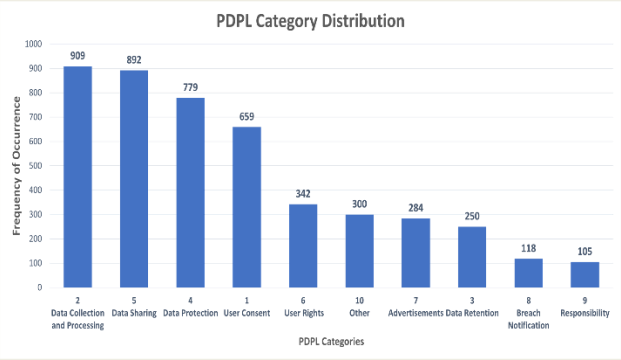

Fig. 5. PDPL category distribution.

- User Consent: The user's personal information cannot be processed without consent, except for essential services or legal purposes. The controller must inform any other party if data is modified [PDPL Art. 5(1)].
- Data Collection and Processing: The data controller defines the collection's purpose, method, and content. The user is informed of the collector's identity except for security reasons. Data is only used for the initial collection purpose [PDPL Art. 12)].
- Data Retention: The controller must delete personal data when its purpose is fulfilled unless certain cases allow data retention by the controller [PDPL Art. 12].
- Data Protection: Personal data storage and transfer must be secure, and controllers must protect user data during this process [PDPL Art. 19].
- Data Sharing: The controller is prohibited from sharing, transferring, or disclosing personal data except in special cases, and access to the data is strictly limited to these special instances [PDPL Art. 13(4)].
- User Rights: An individual's rights include obtaining a copy of the data collected by the controller, requesting its destruction when no longer needed, and rectifying any inaccuracies [PDPL Art. 4 (5-3),12].
- Advertisements: The controller must not send promotional or educational materials to the user's personal addresses without their consent. If approved, a mechanism must be in place for the user to opt-out [PDPL Art. 25].
- Breach Notification: The controller must promptly notify the competent authority and data owner upon becoming aware of any personal data leakages, corruptions, or unauthorized access [PDPL Art. 20].
- Responsibility: Organizations must comply with PDPL for accountable and secure data processing [PDPL Art. 19].
- Other: The controller's contact details information and rights related to underage [PDPL Art. 13(3)].

C. Evaluation Metrics

To make the evaluation results objective, we use in our study a group of performance metrics applied by several studies in natural language processing studies. These metrics include Recall, Precision, and F1-score. They are commonly used by many studies in machine learning and deep learning [12, 13],[33]. In our study, there are three possible outcomes of the classification results:

TP (True Positives): Instances that belong to the "PDPL category" and are correctly predicted as such.

FP (False Positives): Instances that do not belong to the "PDPL category" but were incorrectly predicted as such.

FN (False Negatives): Instances that belong to the "PDPL category" but were incorrectly predicted as a different class.

Based on the possible outcomes, the performance metrics Recall, Precision, and F1-score can be calculated as follows:

$$Micro\ Precision = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FP_i)} \quad (1)$$

$$Micro\ Recall = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FN_i)} \quad (2)$$

$$Micro\ F1 = 2.\frac{Micro\ Precision\ .Micro\ Recall}{Micro\ Precision + Micro\ Recall} \quad (3)$$

It is important to note that among these three-performance metrics, the higher the Recall, Precision, and F1-score, the better model performance, while the lower the value, the worse.

1) Results and discussion: In this section, we present the results of our case study to examine and compare the performance of two models used in our experiment, namely, MARBERTv2 and AraELECTRA. Both models were pre-trained on Arabic domains to achieve a better result with the Saudi Privacy Policy Dataset that we used. During the training process, we recorded the precision, recall, and F-measure metrics achieved for each class. The dataset has ten classes, and the results for each class are shown in Table III, along with the micro-average.

The results indicate that deep learning models can serve as effective tools for detecting and classifying privacy policies. Additionally, they can aid in measuring compliance with Personal Data Protection Law (PDPL) requirements.

Looking at the precision, we can observe that both models achieved values consistently above 83% for all classes. MARBERTv2 achieved values ranging from 87.50 % to 95.43%, while AraELECTRA achieved values ranging from 83.33% to 95.83%. This indicates that both models perform well in making positive predictions of the PDPL categories.

For the recall, MARBERTv2 achieved consistently higher recall values, ranging from 86.21% to 97.89%, compared to AraELECTRA, which ranged from 63.33% to 97.37%. This means that MARBERTv2 captured a substantial portion of the actual instances for these classes.

TABLE III. PRECISION, RECALL, AND F1-SCORE OF MARBERTV2 AND ARAELECTRA

	Model	MARBERTv2			AraELECTRA		
#	Class	Precision	Recall	F1- score	Precision	Recall	F1- score
1	User Consent	93.46%	86.21%	89.69%	**93.75%**	**93.02%**	**93.39%**
2	Data Collection and Processing	92.59%	**97.77%**	95.11%	**94.57%**	97.21%	**95.87%**
3	Data Retention	**89.29%**	**90.91%**	**90.09%**	85.00%	87.17%	86.08%
4	Data Protection	**95.43%**	92.99%	**94.19%**	89.68%	**93.91%**	91.75%
5	Data Sharing	**94.42%**	**97.89%**	**96.12%**	93.43%	97.37%	95.36%
6	User Rights	89.83%	89.83%	89.83%	**91.67%**	**91.67%**	**91.67%**
7	Advertisements	**95.00%**	**86.36%**	**90.48%**	92.31%	82.19%	86.95%
8	Breach Notification	89.29%	**89.29%**	**89.29%**	**95.00%**	63.33%	76.00%
9	Responsibility	**87.50%**	**93.33%**	**90.32%**	83.33%	83.33%	83.33%
10	Other	95.24%	**95.24%**	**95.24%**	**95.83%**	92.00%	93.88%
Micro Avg		**93.32%**	**93.32%**	**93.32%**	92.46%	92.46%	92.46%

Regarding the F1-Score, the two models display impressive scores, surpassing 90% for most classes. However, MARBERTv2 has slightly higher scores. MARBERTv2 and AraELECTRA achieved an overall micro-average F1-score of 93.32% and 92.46%, respectively.

Overall, both models are reliable and efficient for classifying into PDPL categories, but MARBERTv2 outperforms AraELECTRA by a small margin. These results are significant because they demonstrate the potential of using pre-trained models for Arabic text classification, specifically in the domain of privacy policies.

VII. CONCLUSION AND FUTURE DIRECTIONS

Complying with governmental regulations is a crucial mission. The Saudi Arabia Personal Data Protection Law regulates the use of individual personal data to ensure privacy and empower individuals to have control over their data. However, as these regulations are written in a clear legal format, complying with them has become an obstacle for the data management community. Therefore, this paper addresses the problem and proposes a comprehensive framework to help organizations implement PDPL requirements and comply with them by illustrating a clear roadmap on how to comply with the rules by analyzing and translating the normative aspects of the regulation into applicable organizational and technological standards. Moreover, we conducted a case study that utilized deep learning classifiers to enhance privacy policy compliance with PDPL requirements.

To move forward, we will apply the PDPL compliance framework within actual organizational environments. This practical application will enable a more detailed assessment of the framework's effectiveness. Testing the framework in a variety of real-life settings will also offer insights into its adaptability across different industries and organizational sizes, further refining its utility and impact. We will also incorporate advanced technologies to automate the framework and improve the efficiency of data privacy governance.

REFERENCES

[1] "Saudi Arabia Personal Data Protection Law." https://laws.boe.gov.sa/boelaws/laws/lawdetails/b7cfae89-828e-4994-b167-adaa00e37188/1. Accessed 7 Mar 2024

[2] "General Data Protection RegulationGDPR." https://gdpr-info.eu/. Accessed 7 Aug 2024

[3] P. N. Otto and A. I. Anton, "Addressing legal requirements in requirements engineering," in Proc. 15th IEEE Int. Requirements Engineering Conf. (RE 2007), Delhi, India, 2007, pp. 5–14. IEEE, New York. https://doi.org/10.1109/RE.2007.65.

[4] C. Labadie and C. Legner, "Understanding data protection regulations from a data management perspective: A capability-based approach to EU-GDPR." https://aisel.aisnet.org/wi2019/track11/papers/3/. Accessed 22 Apr 2024.

[5] S. Sirur, J. R. C. Nurse, and H. Webb, "Are we there yet?: Understanding the challenges faced in complying with the General Data Protection Regulation (GDPR)," in Proc. 2nd Int. Workshop on Multimedia Privacy and Security, Toronto, Canada, 2018, pp. 88–95. ACM, New York. https://doi.org/10.1145/3267357.3267368.

[6] "Saudi Arabia's Personal Data Protection Law (PDPL)." CookieYes. https://www.cookieyes.com/blog/saudi-arabia-personal-data-protection-law/. Accessed 7 Apr 2024

[7] "NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management." NIST. https://www.nist.gov/system/files/documents/2020/01/16/NIST%20Privacy%20Framework_V1.0.pdf. Accessed 7 Apr 2024

[8] "ISO/IEC 27701:2019." https://www.iso.org/standard/71670.html. Accessed 7 Apr 2024

[9] "What is deep learning?" https://www.ibm.com/topics/deep-learning. Accessed 13 Apr 2024

[10] LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature 521, 436–444 (2015). https://doi.org/10.1038/nature14539

[11] A. Qamar, T. Javed, and M. O. Beg, "Detecting compliance of privacy policies with data protection laws," 2021. https://doi.org/10.48550/ARXIV.2102.12362.

[12] S. Liu, B. Zhao, R. Guo, G. Meng, F. Zhang, and M. Zhang, "Have you been properly notified? Automatic compliance analysis of privacy policy text with GDPR Article 13," in Proc. Web Conf. 2021, Ljubljana, Slovenia, 2021, pp. 2154–2164. ACM, New York. https://doi.org/10.1145/3442381.3450022.

[13] R. E. Hamdani, M. Mustapha, D. R. Amariles, A. Troussel, S. Meeùs, and K. Krasnashchok, "A combined rule-based and machine learning approach for automated GDPR compliance checking," in Proc. Eighteenth Int. Conf. Artif. Intell. Law, São Paulo, Brazil, 2021, pp. 40–49. ACM, New York. https://doi.org/10.1145/3462757.3466081.

[14] Y. U. Jeong, S. Yoo, Y.-H. Kim, and W. H. Shim, "De-identification of facial features in magnetic resonance images: Software development using deep learning technology," J Med Internet Res, vol. 22, e22739, 2020. https://doi.org/10.2196/22739.

[15] A. Vaswani et al., "Attention is all you need," 2017. https://doi.org/10.48550/ARXIV.1706.03762.

[16] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," Nature, vol. 323, pp. 533–536, 1986. https://doi.org/10.1038/323533a0.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, pp. 1735–1780, 1997. https://doi.org/10.1162/neco.1997.9.8.1735.

[18] M. Abdul-Mageed, A. Elmadany, and E. M. B. Nagoudi, "ARBERT & MARBERT: Deep bidirectional transformers for Arabic," arXiv, 2021. https://arxiv.org/abs/2101.01785. Accessed 28 Apr 2024.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv, 2019. https://arxiv.org/abs/1810.04805. Accessed 28 Apr 2024.

[20] W. Antoun, F. Baly, and H. Hajj, "AraELECTRA: Pre-training text discriminators for Arabic language understanding," arXiv, 2021. https://arxiv.org/abs/2012.15516. Accessed 29 Apr 2024.

[21] K. Clark, M.-T. Luong, Q. V. Le, and C. D. Manning, "ELECTRA: Pre-training text encoders as discriminators rather than generators," 2020. https://doi.org/10.48550/ARXIV.2003.10555.

[22] M. Brodin, "A framework for GDPR compliance for small- and medium-sized enterprises," Eur J Secur Res, vol. 4, pp. 243–264, 2019. https://doi.org/10.1007/s41125-019-00042-z.

[23] V. Ayala-Rivera and L. Pasquale, "The grace period has ended: An approach to operationalize GDPR requirements," in Proc. 2018 IEEE 26th Int. Requirements Engineering Conf. (RE), Banff, AB, 2018, pp. 136–146. IEEE, New York. https://doi.org/10.1109/RE.2018.00023.

[24] L. Piras et al., "DEFeND architecture: A privacy by design platform for GDPR compliance," in G. T. Rado and H. Suhl, Eds., Trust, Privacy and Security in Digital Business, Lecture Notes in Computer Science, vol. 11711. Cham: Springer, 2019, pp. 78–93. https://doi.org/10.1007/978-3-030-27813-7_6.

[25] "Credit Information Law." https://laws.boe.gov.sa/BoeLaws/Laws/LawDetails/63dc01a6-fc5c-4600-9171-a9a700f2d222/2. Accessed 30 Apr 2024

[26] H. Kallio, A. Pietilä, M. Johnson, and M. Kangasniemi, "Systematic methodological review: Developing a framework for a qualitative semi-structured interview guide," J Adv Nurs, vol. 72, pp. 2954–2965, 2016. https://doi.org/10.1111/jan.13031.

[27] "Microsoft Teams." https://www.microsoft.com/en-us/microsoft-teams/group-chat-software. Accessed 2 May 2024

[28] V. Braun and V. Clarke, "Using thematic analysis in psychology," Qual Res Psychol, vol. 3, pp. 77–101, 2006. https://doi.org/10.1191/1478088706qp063oa.

[29] "MAXQDA." https://www.maxqda.com/. Accessed 2 May 2024

[30] "SAMA-Banking Rules and Regulations." https://www.sama.gov.sa/en-us/laws/pages/bankingrulesandregulations.aspx. Accessed 2 May 2024

[31] H. Al-Khalifa, M. Mashaabi, G. Al-Yahya, and R. Alnashwan, "The Saudi privacy policy dataset." https://www.researchgate.net/publication/369854910_The_Saudi_Privacy_Policy_Dataset. Accessed 30 Apr 2024.

[32] "Colaboratory." https://colab.research.google.com/. Accessed 2 May 2024

[33] Y. Ling, K. Wang, G. Bai, H. Wang, and J. S. Dong, "Are they toeing the line? Diagnosing privacy compliance violations among browser extensions," in Proc. 37th IEEE/ACM Int. Conf. Automated Software Engineering, Rochester, MI, USA, 2022, pp.