



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

Leveraging BERT for Enhanced GDPR Compliance: A Novel Approach to Privacy Policy Analysis

Presenter's Name : Sanat Paranjape

Course : MSc Computer Science – Intelligent Systems

Student. ID : 23347997

Date 23/07/24

Overview

- Introduction
- Background
- Literature Survey
- Design
- Implementation Issues
- Demo of Project
- Evaluation
- Conclusions

Introduction

GDPR Overview

- The General Data Protection Regulation (GDPR) is a robust EU regulation safeguarding the personal data and privacy of individuals.
- It enforces clear guidelines on how organizations collect, use, and protect personal data.

Compliance Challenges

- Organizations struggle to interpret and implement the intricate GDPR requirements within their privacy policies.
- It enforces clear guidelines on how organizations collect, use, and protect personal data.

Motivation for the Privacy Policy Analyzer

- The need for an accessible, automated tool to simplify GDPR compliance checks.
- Aim to enhance the transparency and readability of privacy policies for both organizations and users

Background

- **The Rise of Data Privacy Concerns**

- With the growing volume of personal data collected online, robust regulations are crucial to safeguard user privacy. The General Data Protection Regulation (GDPR) implemented by the EU mandates clear guidelines for how organizations handle user data. However, ensuring compliance presents challenges for both individuals and organizations.

- **Challenges with Privacy Policies**

- **Technical Language & Inconsistent Format:** Often lengthy and filled with legal jargon, privacy policies are difficult for most users to understand. The lack of standardized formatting further complicates comparisons between organizations and creates ambiguity regarding GDPR adherence.
- **Impact on User Experience:** This complexity impedes users' ability to exercise their data rights under GDPR (e.g., right to access, right to erasure). Studies show a low user comprehension rate of about 33% of data practices outlined in privacy policies .

- **Challenges for Organizations**

- **Resource-Intensive Analysis:** Manually analyzing privacy policies for GDPR compliance is time-consuming and prone to human error.
- **Risk of Non-Compliance Penalties:** Organizations face significant penalties for non-compliance with GDPR regulations.

Literature Survey

A BERT-based Empirical Study of Privacy Policies' Compliance with GDPR

Conference: IEEE International Conference

Abstract: Used a BERT-based model to analyze GDPR compliance in privacy policies from 5G MNOs. Found 51% of companies showing strong adherence to GDPR but noted issues with policy readability and the need for clearer documentation.

Content Analysis of Privacy Policies Before and After GDPR

Conference: IEEE International Conference

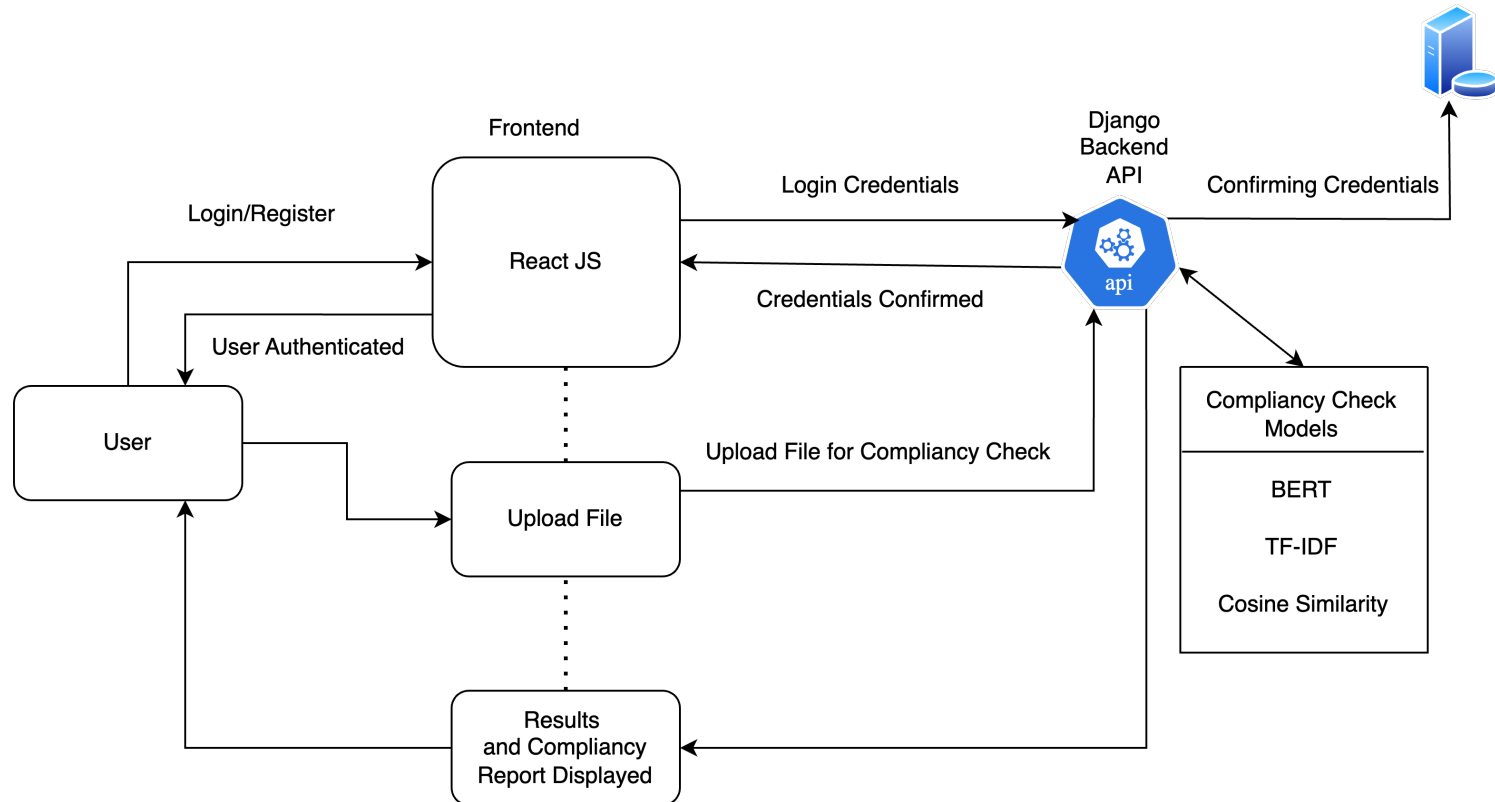
Abstract: Introduced a framework for evaluating GDPR compliance using text feature analysis, topic coverage, and content analysis. The study revealed that while GDPR has led to more comprehensive privacy policies, many still fall short of full compliance. The framework utilized NLP and topic modeling to assess changes in privacy policies before and after GDPR, highlighting areas for further improvement.

Detecting Privacy Policies Violations Using Natural Language Inference (NLI)

Conference: IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), 2022

Abstract: Developed an approach using Natural Language Inference (NLI) to detect violations of GDPR in privacy policy texts. The study presented models using Stanford Natural Language Inference (SNLI) and Multi-Genre Natural Language Inference (MultiNLI) datasets to compare privacy policies against GDPR requirements. The proposed models achieved up to 76% accuracy in detecting deviations from GDPR guidelines.

Design



Implementation Issues

- **Data Extraction Accuracy:**
 - Extracting text from privacy policies, especially PDFs, can be difficult due to complex layouts, embedded images, and non-textual content (charts).
- **Scalability and Performance:**
 - Processing large and complex privacy policies can be computationally expensive, leading to slow analysis times, especially when dealing with many documents.
- **User Role Management:**
 - The application needs to differentiate between different user roles and manage their access for data security and controlled interaction.



Demo of Project

GDPR Insight [Home](#) [Upload](#) [Check Policies](#) [Logout](#)

Welcome to GDPR Insight

Understanding GDPR: A Comprehensive Overview

The General Data Protection Regulation (GDPR) is a landmark regulation that governs data protection and privacy within the European Union (EU) and the European Economic Area (EEA). Enacted in May 2018, GDPR represents a significant shift in data protection law, aiming to give individuals greater control over their personal data while harmonizing data protection rules across Europe.

Why GDPR Matters

GDPR is crucial because it:

- **Empowers Individuals:** GDPR provides individuals with robust rights over their personal data, including the right to access, rectify, and delete their data.
- **Enhances Transparency:** Organizations are required to be transparent about their data processing activities, improving trust and accountability.
- **Strengthens Data Security:** GDPR mandates stringent security measures to protect personal data from breaches and misuse.
- **Regulates Data Transfers:** It addresses the transfer of data outside the EU/EEA, ensuring that personal data is protected internationally.

Core Principles of GDPR

GDPR is underpinned by key principles that organizations must adhere to:

- **Transparency:** Clearly inform individuals about how their data is collected, used, and

Demo of Project

GDPR Insight [Home](#) [Upload](#) [Check Policies](#) [Logout](#)

Upload Privacy Policy

Choose file

usenix.org.md

Upload

The policy is not GDPR compliant.

File Preview

```
> *The following text is extracted and transformed from the usenix.org privacy policy that was archive
# USENIX Privacy Policy | USENIX

USENIX respects your privacy and does not rent, trade, or share your contact information with any othe
This Privacy Policy describes how we collect and use your information, and how you can exercise your p

## Personal Information You Provide to Us and Our Data Processors

### USENIX Membership and Account Creation

When you create an account on the usenix.org website, you provide your name, email address, job title,
We use this information to set up your online account, process your registration, and provide updates on online and
```

Demo of Project

Check Similarity between policies

Choose file retailmenot.com.md

Choose file googleenterprise.com.md

Upload and Check

Files processed successfully!

Similarity Score: 0.47

Top Similar Texts:

Section 7 (File 1) and Section 147 (File 2)

File 1: Personal Information is information that can be used to identify, locate, or contact an individual. It also includes other information that may be associated with Personal Information. We collect the following types of Personal Information:

File 2: ### Personal information

Similarity: 0.77

Demo of Project

GDPR Insight [Home](#) [Upload](#) [Check Policies](#) [Check Compliance with BERT](#) [Logout](#)

Check GDPR Compliance with BERT

Select Privacy Policy File

Choose file

googleenterprise.com.md

Check Compliance

File Preview

```
http://www.google.com/search?q=cars -
Firefox 1.0.7; Windows NT 5.1 -
740674ce2123e969
```

- * `123.45.67.89` is the Internet Protocol address assigned to the user by the user's ISP. Depending on the user's service, a different address may be as
- * `25/Mar/2003 10:15:32` is the date and time of the query.
- * `http://www.google.com/search?q=cars` is the requested URL, including the search query.
- * `Firefox 1.0.7; Windows NT 5.1` is the browser and operating system being used.
- * `740674ce2123a969` is the unique cookie ID assigned to this particular computer the first time it visited Google. (Cookies can be deleted by users. If

Unique identifiers

A unique identifier is a string of characters that can be used to uniquely identify a browser, app, or device. Different identifiers vary in how permanent

Unique identifiers can be used for various purposes, including security and fraud detection, syncing services such as your email inbox, remembering your p

On other platforms besides browsers, unique identifiers are used to recognize a specific device or app on that device. For example, a unique identifier cu

Evaluation

Compliance Criterias :

1. Data Processing Purposes : Ensures the privacy policy clearly states why personal data is being processed.
2. Data Subject Rights Informs users about their GDPR rights, such as access and deletion of data.
3. Data Protection Officer: Mentions the appointment and contact details of the Data Protection Officer.
4. Lawful Basis for Processing: Outlines the legal grounds for processing personal data.
5. Data Retention Period: Specifies how long personal data will be retained.
6. Data Transfer: Describes any international transfers of personal data and their safeguards.
7. Security Measures: Details the security measures to protect personal data.
8. Cookies and Tracking: Provides information on the use of cookies and tracking technologies.
9. Children's Privacy: Includes provisions for protecting the personal data of children.
10. Data Breach Notification: Outlines the procedures for notifying data breaches to users and authorities.
11. Automated Decision-Making: Informs users about automated decision-making processes and their impact.

Evaluations

Compliance Results

Criterion	Similarity	Compliant
DATA PROCESSING PURPOSES	0.73	Yes
DATA SUBJECT RIGHTS	0.75	Yes
DATA PROTECTION OFFICER	0.72	Yes
LAWFUL BASIS FOR PROCESSING	0.70	Yes
DATA RETENTION PERIOD	0.72	Yes
DATA TRANSFER	0.73	Yes
SECURITY MEASURES	0.73	Yes
COOKIES AND TRACKING	0.64	Yes
CHILDREN'S PRIVACY	0.73	Yes
DATA BREACH NOTIFICATION	0.74	Yes
AUTOMATED DECISION-MAKING	0.61	Yes

Overall Compliance: Yes

Conclusions

- **Impact of BERT on GDPR Compliance:** The implementation of BERT in analyzing privacy policies has shown promising results in assessing compliance with GDPR. Its ability to process and understand complex legal language allows for a more accurate evaluation of whether policies meet regulatory standards.
- **Enhanced Readability and Understanding:** BERT's capabilities can be leveraged to improve the readability of privacy policies. By simplifying the language used in these documents, BERT can help organizations communicate their data practices more effectively, thus enhancing user comprehension and engagement.
- **Future Directions:** Continued research and development of BERT-based models can lead to more sophisticated tools for compliance analysis. These advancements could include real-time monitoring of policy updates and automated alerts for non-compliance, further supporting organizations in their GDPR obligations.



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

Questions ?



Trinity College Dublin

Coláiste na Tríonóide, Baile Átha Cliath

The University of Dublin

Thank You

