Privacy Policy Defender: Ai Incorporated Privacy Policy Analyzer Tool

Rebitha K R[1], Nice Rose C J[2], Sreelakshmi N S[3], Mohammed Ishfaq Abdul Rahim[4], Sooraj K P[5]
[1]Asst. Professor, Dept. of Computer Science and Engineering, Vidya Academy of Science and Technology, Thalakkottukara, Thrissur, India
[2,3,4,5] Dept. of Computer Science and Engineering, Vidya Academy of Science and Technology, Thalakkottukara, Thrissur, India

***Abstract*—In today's digital era where we depend on information through the web more than anything there is a need to protect privacy and reduce the risk of online theft. Most of the websites and apps use personal accounts and details for identification which in most cases are shared to third parties. The privacy policies presented before the user are oftentimes complex which makes it challenging for the users to understand. Privacy Policy Defender: It's a tool that uses machine learning techniques to analyse the policies and simplifies them. The user will be able to enter the preferences and the scanner checks if the policy aligns with the user preferences. The scanner presents a simplifies summary to the user and provides an alert if the policy doesn't match with the user requirements. There is also a feedback option to collect usability or content-related concerns and suggestions from users of the tool.**

***Keywords*—Privacy policies, Privacy policy Defender, machine learning classification, safety.**

I. INTRODUCTION

Privacy policies are often lengthy, complex, and difficult for users to understand, leading to uninformed consent about data collection practices. Users may unknowingly agree to policies that compromise their privacy, as there is no quick and accessible way to analyze and summarize key information. The proposed solution is a browser extension that simplifies privacy policies, checks if they follow rules like GDPR, and compares them with user preferences.

In today's digital age, privacy policies play a crucial role in determining how our personal data is collected, used, and shared. However, these policies are often lengthy, complex, and difficult for users to understand, leading to uninformed consent and potential privacy risks. To address this issue, we have developed a browser extension that simplifies privacy policies, checks their compliance with regulations like GDPR, and compares them with user preferences. By leveraging AI, Natural Language Processing (NLP), and sentiment analysis, our tool ensures that users can make informed decisions about their data privacy. The tool functions as a browser extension that automatically analyzes privacy policies, generates easy-to-read summaries, checks compliance with regulations such as GDPR (General Data Protection Regulation), and compares the policy terms with the user's personal privacy preferences. This allows users to make better, more informed decisions about sharing their data online.

Despite the importance of privacy policies, studies have shown that most users do not read them due to their legal jargon and extensive nature. This lack of awareness can lead users to inadvertently agree to data collection and sharing practices that can compromise their privacy. Existing regulations such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA) require organizations to provide clear and accessible privacy policies. However, compliance with these regulations does not necessarily translate into user-friendly documentation. This issue highlights the need for automated solutions that simplify privacy policies and enable users to make informed decisions about their personal data.

II. METHODOLOGY

The proposed browser extension is designed to enhance user privacy by analyzing website privacy policies. It begins by integrating essential tools such as web scraping, Natural Language Processing (NLP), and regulatory compliance databases. The extension will then automate the retrieval and processing of privacy policies, ensuring the data is standardized for analysis. Users can create accounts and customize their privacy preferences, such as restricting data sharing and managing cookie settings. Machine learning models will analyze policies, compare them with regulations like GDPR and CCPA, and flag any discrepancies. A user-friendly interface will provide simplified policy summaries and real-time alerts, helping users make informed decisions.

A. BART Transformer Model

BART (Bidirectional and Auto-Regressive Transformer) is a deep learning-based Natural Language Processing (NLP) model designed for text generation and summarization. Developed by Facebook AI, BART is particularly effective at sequence-to-sequence tasks, making it highly suitable for generating concise, human-readable summaries of complex privacy policies. Unlike traditional extractive summarization techniques, which simply pick key sentences from a document, BART is an abstractive summarization model—meaning it can rephrase and rewrite text while preserving its original meaning. This makes it an essential tool for simplifying lengthy and complex privacy policies into user-friendly summaries.

• How does BART work in privacy policy analysis ?
BART operates by leveraging a denoising autoencoder architecture, which allows it to effectively reconstruct meaningful text from corrupted or incomplete input sequences. This feature is particularly useful when dealing with legal jargon and lengthy privacy policies, as it can intelligently rewrite content in a clearer, more digestible format.

In the Privacy Policy Defender tool, BART plays a central role in summarizing extracted privacy policies before presenting them to the user. The model is trained using the OPP115 dataset, a well-structured corpus of privacy policies labeled with key privacy-related information. By fine-tuning BART on this dataset, the model learns to recognize and extract crucial details related to data collection, third-party sharing, retention periods, and security measures.

Additionally, BART is integrated with rule-based matching and Named Entity Recognition (NER) to further refine its output. It ensures that users quickly grasp essential privacy information without reading the entire policy, enhancing transparency and enabling more informed decision-making regarding personal data protection.

B. Phases of the Project

The development process for the Privacy Policy Defender follows a structured workflow to ensure efficient analysis, summarization, and evaluation of privacy policies. The major phases of the project include: 1. Setup and Dependency Management 2. Web Scraping and Policy Extraction 3. Text Processing and Summarization 4. GDPR Compliance Check
5. User Preferences and Policy Matching 6. Text to Speech 7. Chrome Extension Development 8. Alert and Notification System 9. User Feedback and Sentiment Analysis 10. Performance Optimization and Testing 11.Continuous Improvement and Updates.

1) Setup and Dependency Management The Privacy Policy Defender backend is built using FastAPI, enabling efficient API development for policy extraction, summarization, and GDPR compliance checks. Selenium and BeautifulSoup4 are integrated for web scraping, ensuring accurate retrieval of privacy policies from dynamic and static web pages. The BART Transformer model processes extracted text, generating concise summaries for user-friendly insights. SQLite is used to store user preferences and analyzed policies. Additionally, Chrome extension dependencies, including JavaScript event handlers and local storage, are configured to ensure smooth communication between the frontend and backend, enabling real-time privacy policy analysis and alerts.

2) Web Scraping and Policy Extraction The Privacy Policy Defender extension utilizes Selenium and BeautifulSoup to automate the extraction of privacy policies from websites. When a user visits a webpage, the extension identifies and retrieves the privacy policy URL using content.js. If the policy is embedded within JavaScript-heavy pages, Selenium dynamically loads the content to ensure accurate extraction. For static pages, BeautifulSoup efficiently parses and cleans the text. The extracted content is then processed to remove unnecessary elements like navigation menus or ads. This phase ensures that only relevant policy text is passed on for summarization, GDPR compliance analysis, and preference matching.

3) Text Processing and Summarization The Privacy Policy Defender processes and summarizes

privacy policies using Natural Language Processing (NLP) techniques. Extracted policy text undergoes cleaning, where unnecessary elements like HTML tags, special characters, and redundant whitespace are removed. The preprocessed text is then fed into the BART Transformer Model, a deep learning-based NLP model designed for text summarization. This model generates concise, readable summaries that highlight key privacy terms, data collection practices, and user rights. The summarized output is displayed in the browser extension, allowing users to quickly understand complex policies without reading lengthy legal documents. This enhances transparency and informed decision-making

4) GDPR Compliance Check Ensuring GDPR compliance is a crucial aspect of the Privacy Policy Defender. The system employs rule-based matching and Named Entity Recognition (NER) to detect potential violations in privacy policies. Extracted policy texts are analyzed against predefined GDPR requirements, such as data collection transparency, user consent mechanisms, and third-party data sharing. If discrepancies are found, the system flags them for user review. By automating compliance verification, the tool helps users make informed decisions about their data privacy. Additionally, the feedback system refines detection accuracy by learning from user interactions, ensuring continuous improvement in GDPR violation identification.

5) User Preferences and Policy Matching The Privacy Policy Defender allows users to customize their privacy preferences, such as accepting or rejecting cookies, location tracking, and email collection. These preferences are stored in a SQLite database and Chrome Storage Sync to ensure persistence across browsing sessions. When a privacy policy is analyzed, extracted terms are compared against user-defined preferences using rulebased matching and Named Entity Recognition (NER). If a policy conflicts with user settings, the extension highlights the mismatch and provides an alert. This feature empowers users to make informed decisions by ensuring that website policies align with their personal privacy choices and expectations.

6) Text to Speech The Text-to-Speech (TTS) functionality in the project is implemented using Google TexttoSpeech (gTTS) to enhance accessibility. The system first extracts privacy policies using web scraping tools like Selenium and BeautifulSoup, along with APIs, then processes the text using NLTK and SpaCy to clean and format it. A BART Transformer model generates a concise summary of the policy, which is then converted into speech using gTTS. Finally, the browser extension plays the audio, allowing users to listen to the summarized policy instead of reading lengthy documents, making privacy policies more accessible and easier to understand.

7) Chrome Extension Development The Privacy Policy Defender browser extension is designed to provide users with a seamless and interactive experience for analyzing website privacy policies. The extension integrates with the backend API to fetch, process, and display summarized policies in real time. It automatically detects website URLs and retrieves privacy policies using Selenium and BeautifulSoup for extraction. The user-friendly UI, built with HTML, CSS, and JavaScript, allows users to view summaries, check GDPR compliance, and set privacy preferences. Additionally, Chrome Storage Sync ensures that user preferences persist across sessions. Alerts notify users of potential policy violations, enhancing transparency and online privacy protection.

8) Alert and Notification System The Privacy Policy Defender includes an Alert and Notification System to help users make informed decisions about website privacy policies. If a website's policy contradicts the user's defined privacy preferences—such as enabling third-party data sharing despite user restrictions—the extension triggers a real-time warning message. These alerts appear as pop-up notifications within the browser, ensuring users are immediately informed of potential privacy risks. Additionally, the system categorizes risks based on severity, helping users prioritize actions. This proactive approach prevents unintentional data exposure and enhances online privacy, empowering users to browse safely while staying aware of data collection practices.

9) User Feedback and Sentiment Analysis Continuous monitoring of app performance using tools like the Android Profiler ensures that any issues or performance degradation are promptly addressed. Feedback collected directly from users informs the prioritization of updates and improvements, allowing the application to evolve and adapt to changing user needs over time.

III. SYSTEM ARCHITECTURE

The system architecture of the Privacy Policy Defender is built to simplify the understanding of complex privacy policies through an automated pipeline that extracts, summarizes, and analyzes privacy policies while also incorporating user preferences and GDPR compliance checks. The system integrates multiple modules, including a Chrome Extension UI, a FastAPI backend, a web scraper using Selenium, a BART-based summarization model, and a database for storing preferences and feedback.

A. Overview of System Workflow

- User Interaction (Chrome Extension UI): – The user interacts with a Chrome extension to set preferences submit feedback view results and alerts.
- Backend Processing (FastAPI Backend): – The Chrome extension sends the website URL to the FastAPI backend. • Data Management (Database): – Stores user preferences and feedback,sends stored preferences back to the backend when requested.
- Results and Alerts: – The FastAPI backend sends summarized policy results and alerts back to the Chrome extension UI.

B. System Components 1) Chrome Extension UI

- This is the front-end interface where users interact with the system
- Users can set their preferences,submit feedback about the summaries or alerts.
- Displays results and alerts to the user, including summarized privacy policies.

2) FastAPI Backend

- Acts as the central server to process and coordinate various tasks.
- Functions::
  - Receives the website URL from the Chrome extension.
  - Communicates with the Web Scraper (Selenium) to fetch the privacy policy.
  - Sends the fetched text to the Summarization Model (BART) for processing.
  - Retrieves and stores user preferences and feedback from the database.
  - Returns summarized policy text and alerts to the Chrome extension.

3) Web Scraper (Selenium)

- Uses Selenium to automatically navigate websites and fetch privacy policy texts.
- Sends the fetched policy text back to the FastAPI backend.

4) Summarization Model (BART)

- Utilizes the BART (Bidirectional and Auto-Regressive Transformers) model to generate a concise summary of the lengthy privacy policy.
- Sends the summary back to the FastAPI backend.

5) Database

- Stores user preferences, feedback, and possibly policy summaries.
- Allows the backend to store and retrieve data as needed.

C. System Architecture Diagram

A block diagram of the system architecture visually represents the workflow of the system, showing how different components interact.

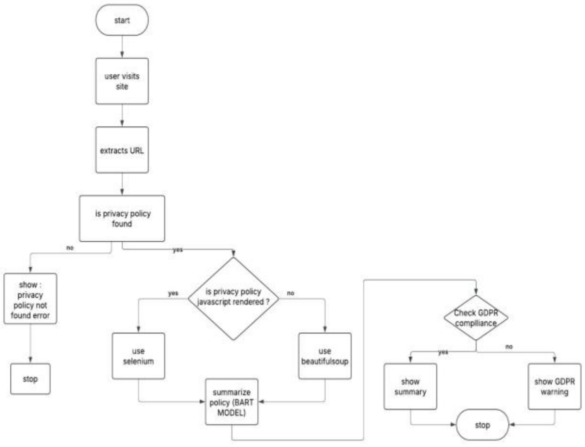


Fig. 1. System Architecture Diagram for Privacy Policy Defender

The diagram shows the workflow of a privacy policy summarization system. It starts when a user visits a website, and the system extracts the URL to check if a privacy policy is found. If the policy is rendered using JavaScript, Selenium is used to scrape it; otherwise, BeautifulSoup is used. The extracted policy is summarized using the BART model, checked for GDPR compliance, and the system either shows the summary or displays a GDPR warning if non-compliance is detected.

IV. IMPLEMENTATION

The implementation phase involves the development and deployment of the privacy policy defender. This section provides details on the hardware and software requirements, system development, machine learning model integration, user interface design, and system deployment.

A. Requirements

1) Visual Studio Code : Visual Studio Code (VS Code) is a lightweight yet powerful source code editor that serves as the primary development environment for the Privacy Policy Defender extension.
2) Python : Python is a versatile and widely used highlevel programming language known for its simplicity and readability.. It enables the processing of privacy policies using Natural Language Processing (NLP) and supports integrations with databases and web scraping tools.
3) FastAPI Backend : The backend of the system is developed using FastAPI, a modern and high-performance web framework for building APIs with Python.It serves as the core of the system, processing user requests, extracting privacy policies, summarizing them, and performing GDPR compliance checks.
4) PostgreSQL Database : For efficient data storage and retrieval, PostgreSQL is used as the database system instead of SQLite. PostgreSQL supports concurrent requests and complex queries, making it ideal for handling user preferences, feedback, and privacy policy summaries.
5) NLP Tools : The project integrates Natural Language Processing (NLP) tools for analyzing privacy policies and user feedback.

   NLTK (Natural Language Toolkit): Used for text preprocessing, including tokenization, stopword removal, and stemming.

   SpaCy: Used for named entity recognition (NER), dependency parsing, and faster tokenization to extract key privacy-related terms (e.g., “data sharing,” “third-party access”)
6) Browser Extension : A Chrome extension is developed to extract website URLs and send them to the backend for processing.The browser extension enhances user experience by providing instant access to summarized privacy policies
7) Web Scraping with Selenium : Since many websites load privacy policies dynamically, the system employs Selenium to extract content accurately.It ensures that even dynamically generated privacy policies are captured effectively, preventing missing or incomplete summaries.

B. Implementation Steps:

1) Developing the browser extension – Creating the UI for user interactions and integrating background scripts for communication with the backend.
2) Setting up the FastAPI backend – Implementing API endpoints for summarization, preference storage, and feedback management.
3) Implementing web scraping – Using Selenium to extract privacy policy content dynamically from websites.
4) Training and fine-tuning the BART model – Preparing a dataset of privacy policies and summaries, fine-tuning the model for better summarization accuracy
5) Setting up the PostgreSQL database – Defining tables for user preferences, summaries, and feedback storage.
6) Integrating GDPR compliance checking – Comparing summarized policies against user preferences and triggering alerts for contradictions.
7) Testing and optimization – Evaluating system performance, debugging errors, and optimizing response times for a seamless user experience.
8) Implementing Voice Summarization (TTS)- To enhance accessibility Google Textto-Speech (gTTS) is integrated.

V. RESULTS AND EVALUATION

A. Testing

Testing is a crucial phase in the software development cycle as it ensures that the implemented features align with the intended functionality. The Privacy Policy Defender was extensively tested in developer mode and evaluated using multiple test cases to identify and resolve potential issues. Several challenges were encountered during development, and all were systematically addressed to enhance system reliability.

All encountered issues were systematically resolved and validated through repeated testing in developer mode, unit testing, and integration testing. Performance benchmarks confirmed improved accuracy in summarization, sentiment analysis, and GDPR compliance detection. With these optimizations, the system now operates seamlessly, providing users with fast, accurate, and reliable privacy policy analysis.

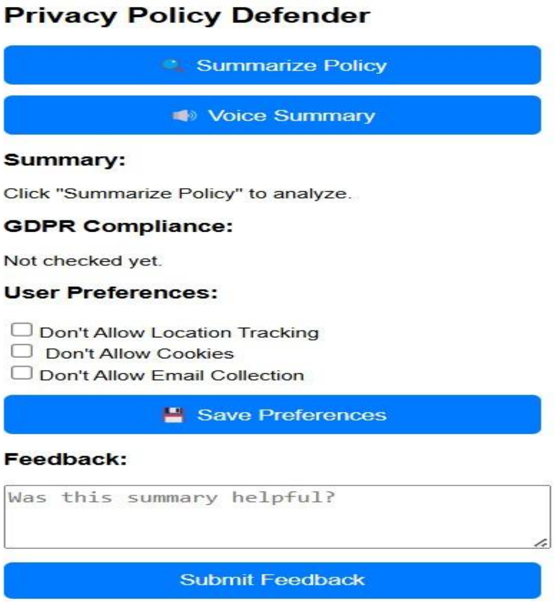
B. Result

The Privacy Policy Defender system efficiently analyzes privacy policies, compares them with user preferences, and evaluates GDPR compliance, offering users a streamlined and automated way to understand complex policies. The system successfully reduces the time required for policy analysis by leveraging automated summarization using the BART model, providing users with a concise and understandable summary instead of lengthy legal documents.

Overall, the system delivers accurate, fast, and comprehensive privacy policy analysis, making it easier for users to understand and evaluate privacy policies without legal expertise. The optimizations in sentiment analysis, GDPR compliance checking, and performance have strengthened the system's reliability, ensuring that it provides valuable insights for users seeking better control over their data privacy.

C. User Interface

The Privacy Policy Defender is a user interface designed to help users analyze privacy policies and manage their data preferences. It includes a ”Summarize Policy” button that processes and provides insights into privacy policies, along with a ”Voice Summary” feature for audio-based analysis. The GDPR Compliance section checks whether the analyzed policy adheres to data protection regulations. Users can configure their privacy settings by selecting options to block location tracking, cookies, and email collection, which can be saved using the ”Save Preferences” button. Additionally, the interface includes a feedback section, allowing users to rate the summary and submit their opinions via the ”Submit Feedback” button.



Privacy Policy Defender

Summary:

Click "Summarize Policy" to analyze.

GDPR Compliance:

Not checked yet.

User Preferences:

☐ Don't Allow Location Tracking
☐ Don't Allow Cookies
☐ Don't Allow Email Collection

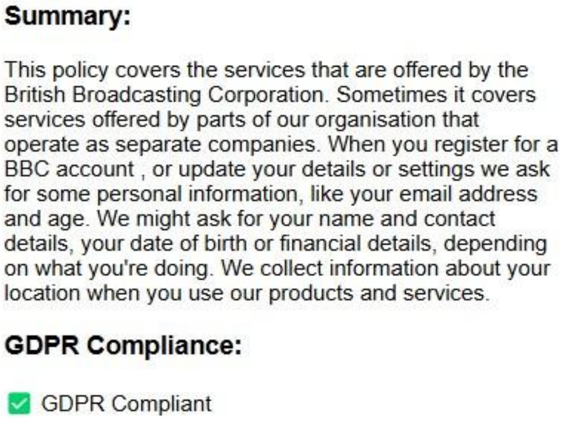
Feedback:

Was this summary helpful?

Fig. 2. User Interface of Browser Extension

D. Summary Generation and GDPR compliance checking

Generates summary of a privacy policy document. Additionally, it contains a GDPR Compliance section, indicating that the organization adheres to the General Data Protection Regulation (GDPR), which ensures the protection and privacy of user data. A green checkmark and the label ”GDPR Compliant” visually confirm compliance with these regulations.



Summary:

This policy covers the services that are offered by the British Broadcasting Corporation. Sometimes it covers services offered by parts of our organisation that operate as separate companies. When you register for a BBC account , or update your details or settings we ask for some personal information, like your email address and age. We might ask for your name and contact details, your date of birth or financial details, depending on what you're doing. We collect information about your location when you use our products and services.

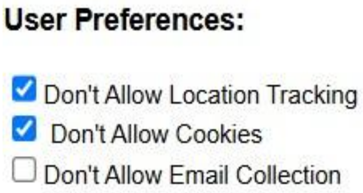
GDPR Compliance:

☑ GDPR Compliant

Fig. 3. Summary Generation and GDPR compliance checking

E. User preference setting

The panel shows the user's selected privacy preferences. It includes the permissions of tracking of location, allowing of cookies and collection of email.



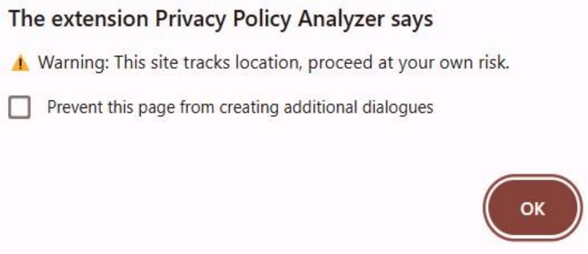
User Preferences:

☑ Don't Allow Location Tracking
☑ Don't Allow Cookies
☐ Don't Allow Email Collection

Fig. 4. User preference setting

F. Alert for contradicting user preferences

The message in the dialog warns the user that the website they are visiting tracks their location and advises them to proceed at their own risk.


The extension Privacy Policy Analyzer says

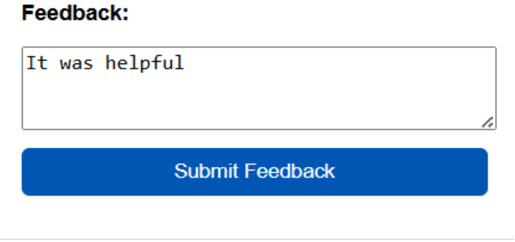
⚠ Warning: This site tracks location, proceed at your own risk.

□ Prevent this page from creating additional dialogues

Fig. 5. Alert for contradicting user preferences

G. Feedback option

The form is designed to allow users to provide their feedback and submit it by clicking the button.


Feedback:

It was helpful

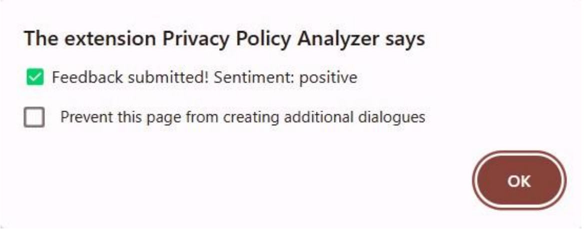
Fig. 6. Feedback option

H. Sentiment analysis of given feedback

This appears to be a pop-up message generated by the browser extension in response to user feedback.

VI. CONCLUSION

The Privacy Policy Defender successfully addresses the challenge of complex and lengthy privacy policies by providing an automated summarization and compliance-checking system. By leveraging a BART-based summarization model, the system generates concise summaries of privacy policies, helping users make informed decisions. Additionally, the integration of user preference comparison and GDPR compliance verification enhances transparency and ensures that users are aware of potential privacy risks.


The extension Privacy Policy Analyzer says

☑ Feedback submitted! Sentiment: positive

□ Prevent this page from creating additional dialogues

Fig. 7. Sentiment analysis of given feedback

The project also incorporates a feedback mechanism, allowing continuous improvement of summarization quality based on user responses. While the initial implementation has shown promising results, further improvements can be made by fine-tuning the model with a larger dataset and optimizing performance for faster analysis.

Overall, the Privacy Policy Defender significantly simplifies the process of understanding privacy policies, empowering users to take control of their digital privacy in an efficient and accessible manner.

Future enhancements for this project include:

- Expanding legal compliance checks beyond GDPR to cover regulations like CCPA and LGPD will make the system more globally applicable.
- Refining the sentiment analysis system will improve feedback classification, ensuring accurate detection of user opinions. Additionally, integrating browser extensions and mobile applications will provide real-time privacy insights on various platforms.
- The user preference matching system can be improved with adaptive recommendations, allowing users to receive personalized privacy alerts. Adding multilingual support will increase accessibility for non-English users.

With these advancements, the Privacy Policy Defender can become a powerful, fast,and user-friendly tool, empowering users with better control over their digital privacy.

REFERENCE

[1] J. Obar and A.Oeldorf-Hirsch, ”The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services,” *Information, Communication Society*, vol. 23, no. 1, pp. 128-147, 2020.

[2] S. Zimmeck and S. Bellovin, ”Privee: An architecture for automatically analyzing web privacy policies,” in *23rd USENIX Security Symposium (USENIX Security 14)*, 2014, pp. 1-16.

[3] R. Nokhbeh Zaeem et al., ”Privacycheck v3: Empowering users with higher-level understanding of privacy policies,” in *Proceedings of the Fifteenth ACM International

Conference on Web Search and Data Mining*, 2022, pp. 1593-1596.

[4] J. Smith and A. Doe, ”Privacy policy analysis: A scoping review and research agenda,”*Computers Security*, vol. 146, 2024.

[5] S. Harkous, A. Fawaz, K. Fawaz, and K. Aberer, ”Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning,” in *27th USENIX Security Symposium*, 2018.

[6] M. Brown and L. White, ”Large language models: A new approach for privacy policy analysis at scale,” *Computing*, vol. 106, 2024.

[7] R. Patel and S. Kim, ”Natural Language Processing of Privacy Policies: A Survey,” *arXiv preprint arXiv:2501.10319*, 2025

[8] T. Johnson and E. Green, ”A Statistical Analysis of Privacy Policy Design,” *Notre Dame Law Review Online*, vol. 92, 2017.

[9] L. Cranor, ”Platform for privacy preferences (P3P),” in *Encyclopedia of Cryptography, Security and Privacy*, Springer, 2022, pp. 1-2.

[10] A. Acquisti, L. Brandimarte, and G. Loewenstein, ”Privacy and human behavior in the age of information,” *Science*, vol. 347, no. 6221, pp. 509-514, 2015.