WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Policy Checker Extension: AI-Based Automated Policy Analysis and Risk Detection

Avinash S[1], Manoj B[1], Muraliprasanth P[1],Mr.Rajesh A[2]

[1 2] Department of Computer Science and Business Systems, E.G.S Pillay Engineering College, Nagapattinam, Tamilnadu, India

ABSTRACT:

With the increasing reliance on online services, users are often required to agree to lengthy and complex Terms & Conditions, Privacy Policies, and Cookie Policies without fully understanding the potential risks. Many of these policies contain hidden clauses related to data sharing, financial liabilities, auto-renewals, and policy changes, which can negatively impact user security and privacy. Policy Scanner, a Chrome extension, is designed to automate the analysis of these policies, offering users a simplified and structured summary. Using a combination of NLP for keyword extraction, regex for pattern detection, and sentiment analysis, the extension highlights key points, categorizes policies into pros and cons, and provides a security score (0% - 100%) to help users make informed decisions.

The extension operates in two modes, Normal Mode, and AI Mode., while Normal Mode applies text analysis algorithms to detect policy risks. AI Mode (planned for future integration) will leverage OpenAI and cross-reference data with legal databases like TermsFeed and ToS;DR for deeper policy evaluation. The tool ensures transparency, user awareness, and enhanced security by simplifying complex legal terms and identifying potential risks in online agreements. This project aims to bridge the gap between legal policy understanding and user awareness, empowering individuals to make better-informed decisions about their data and financial safety.

Keywords: NLP Algorithm, Regex Pattern Matching, Sentiment Analysis, AI-powered Analysis, OpenAI, Policy Scanner, Chrome Extension, Privacy Policies, Terms & Conditions, Risk Assessment, Security Score (0-100%), Legal Compliance Check, Data Sharing Risks, Financial Risks, Pros & Cons Analysis, Automated Policy Evaluation, and Structured Report Generation.

INTRODUCTION

In the digital age, users are frequently required to accept *Terms & Conditions, Privacy Policies, and Cookie Policies* when using websites, applications, and online services. However, these legal agreements are often lengthy, complex, and filled with technical jargon that makes them difficult for the average user to understand. As a result, many people blindly accept these policies without realizing that they may be agreeing to *hidden clauses* that could compromise their *privacy, data security, or financial stability*. Companies often include terms related to *third-party data sharing, auto-renewals, arbitration clauses, and hidden fees*, which could lead to unintended consequences for users. Without a proper understanding of these policies, users are left vulnerable to *data breaches, financial losses, and other security risks*. Therefore, there is a growing need for a *smart, automated tool* that can analyze and summarize these policies in a way that is easy to understand.

To address this issue, we propose *Policy Scanner*, a Chrome extension that automates the analysis of legal policies and presents users with a simplified summary. This tool scans *Terms & Conditions, Privacy Policies, and Cookie Policies*, identifies *key risks*, and categorizes them into *pros and cons* while also generating a *security score (0% - 100%)*. The extension operates in *two primary modes*: *Normal Mode*, which applies *NLP keyword extraction, regex pattern matching, and sentiment analysis* to assess risks, and *AI Mode*, which utilizes *OpenAI* to conduct an advanced policy review based on data from *10 trusted legal sources*, including TermsFeed and ToS;DR. By providing a clear, concise summary of legal policies, *Policy Scanner empowers users to make informed decisions, enhances data transparency, and promotes better awareness of potential online risks*.

With the rapid expansion of online services, companies are constantly updating their *privacy policies and terms of service*, often without explicitly notifying users. These updates may include significant changes, such as *new data-sharing agreements, modified refund policies, or stricter content regulations*. Since most users lack the time or legal expertise to review these documents thoroughly, they often *unknowingly agree* to conditions that may not align with their interests. This lack of transparency poses a *serious risk* to personal data security, as organizations can legally collect, store, and even sell user information without clear consent. The absence of a *standardized framework* for policy evaluation further complicates the problem, making it difficult to assess which policies prioritize user privacy and security.

The *Policy Scanner Chrome Extension* aims to bridge this gap by offering a *user-friendly, automated solution* that scans, analyzes, and simplifies online policies. By integrating *machine learning and AI-driven legal analysis*, the tool effectively detects *risky clauses, hidden fees, and vague legal terms* that could mislead users. The *AI Mode*, powered by *OpenAI*, enhances the accuracy of risk assessment by cross-referencing policies with *10 trusted legal sources*, ensuring reliable and up-to-date insights. Additionally, the extension provides a *detailed risk analysis report*, allowing

users to make *well-informed decisions before accepting any terms and conditions*. By increasing *awareness and transparency, Policy Scanner helps users protect their data, avoid potential financial risks, and navigate the digital world with greater confidence*.

PROBLEM STATEMENT

As the digital landscape expands, users frequently interact with websites, apps, and online services that require them to accept *Terms & Conditions, Privacy Policies, and Cookie Policies*. These agreements often contain *complex legal language* and hidden clauses that may lead to *unauthorized data sharing, financial risks, and security vulnerabilities*. Users, lacking legal expertise, tend to accept these terms without understanding the potential consequences, exposing themselves to *privacy breaches, data exploitation, and monetary losses*.

Currently, there is no *automated, user-friendly solution* that efficiently analyzes and summarizes these policies in *a transparent and accessible manner*. *Policy Scanner*, a Chrome extension, addresses this issue by using *NLP, regex pattern detection, sentiment analysis, and AI-driven legal checks* to extract crucial information. The tool categorizes risks, detects policy violations, and assigns a *security score (0-100%)*. By generating a *structured report* with a summary, pros & cons, and an overall risk assessment, Policy Scanner *empowers users to make informed decisions* and avoid potential legal and financial pitfalls while navigating the web.

PROPOSED DESIGN

The *Policy Scanner Chrome Extension* is designed to provide an *automated and efficient solution* for analyzing website policies. The system consists of two primary modes: *Normal Mode* and *AI Mode*, each offering different levels of analysis.

- *Normal Mode*: Uses *NLP, regex pattern matching, and sentiment analysis* to extract key terms, detect potential risks, and generate a *security score (0-100%)*. It summarizes findings in a structured format, highlighting *pros & cons*.
- *AI Mode*: Utilizes *OpenAI* to conduct a *deeper analysis*, cross-checking policies against *10 trusted sources* to detect *policy violations, legal risks, and security concerns*. It also generates a *security score (0-100%)* for a *more accurate evaluation*.

The system follows a structured *workflow*, where users select a mode, the extension scans and processes the policy, and finally, a *detailed report* is generated, including *key findings, security scores, and an overall risk assessment*. The design ensures *real-time, user-friendly results*, helping individuals make *informed decisions about online policies*.

System Architecture

The *Policy Scanner Chrome Extension* follows a structured system architecture designed to analyze and evaluate website policies efficiently. It consists of multiple components that work together to extract, process, and present policy-related insights to the user.

1. ***User Interaction Layer:*** Users interact with the extension via a user-friendly interface, where they can choose between *Normal Mode* and *AI Mode*. The interface provides options to initiate a scan and displays results in a structured format, including a security score, summary, and pros & cons.

2. ***Policy Data Extraction Module:*** Once a user selects a mode, the system extracts the *Terms & Conditions, Privacy Policy, or Cookie Policy* from the website. This module uses *web scraping techniques* to gather relevant content while filtering out unnecessary text.

3. ***Normal Mode Processing:*** In *Normal Mode*, the extracted policy undergoes *NLP-based keyword extraction, regex pattern matching*, and *sentiment analysis* to identify potential risks. The system detects key phrases related to *data sharing, financial risks, and privacy concerns*, then assigns a *security score (0-100%)* based on predefined parameters.

4. ***AI Mode Processing:*** In *AI Mode*, the policy is processed using *OpenAI*, which performs a *deep legal compliance check*. The AI system cross-references the extracted policy with data from *10 trusted sources* to detect *policy violations, unfair clauses, and user risks*. The analysis includes *risk classification, policy transparency assessment, and legal violation detection*, leading to a more comprehensive *security score (0-100%)*.

5. ***Risk Evaluation and Score Generation:*** Both *Normal Mode* and *AI Mode* generate a *security score (0-100%)* based on detected risks and policy transparency. The system ranks the policy's safety level and categorizes risks based on severity.

6. ***Report Generation and User Output:*** The final step involves generating a *structured report* summarizing the findings. The report includes a *concise policy summary, security score, detected risks, and pros & cons*. The user receives an *easy-to-understand* risk assessment, enabling them to make *informed decisions* before accepting any policy.

This architecture ensures *efficiency, accuracy, and transparency*, helping users quickly analyze website policies and protect themselves from potential risks.

METHODOLOGY

The *Policy Scanner Chrome Extension* follows a structured methodology to analyze and evaluate online policies effectively. It is designed to extract key information from *Terms & Conditions, Privacy Policies, and Cookie Policies* using *Natural Language Processing (NLP), sentiment analysis, and AI-based legal compliance checks*. The methodology consists of several key stages, ensuring that the system provides accurate, transparent, and reliable results for users.

1. Data Collection and Preprocessing

The system begins by extracting *policy documents* from websites when a user activates the extension. The document text is *cleaned and preprocessed* to remove unnecessary elements such as HTML tags, special characters, and redundant white spaces. *Tokenization and stopword removal* are applied to ensure that only meaningful content is analyzed. This step ensures that the *core legal terms and conditions* remain the focus of the analysis.

2. Normal Mode: Text-Based Analysis

In *Normal Mode*, the extension employs a combination of *NLP (Term Frequency-Inverse Document Frequency), regex-based pattern matching, and sentiment analysis* to identify important clauses. These techniques help in:

- Extracting *high-impact keywords* related to privacy, data sharing, and financial risks.
- Detecting *predefined risk patterns* that indicate legal loopholes or unfair policies.
- Performing *sentiment analysis* to classify clauses as *user-friendly, neutral, or high-risk*.

The results are then categorized into *pros and cons*, highlighting key takeaways for the user. A *security score (0-100%)* is generated based on the severity of risks detected.

3. AI Mode: Advanced Legal Compliance Check

In *AI Mode*, the extension leverages *OpenAI's language model* to conduct a deeper *legal risk analysis*. The AI model:

- Cross-references the extracted policy text with *10 trusted legal sources*, including *TermsFeed, TOS;DR, Iubenda, and PrivacyPolicies*.
- Detects hidden risks such as *excessive data collection, third-party sharing, auto-renewals, arbitration clauses, and financial risks*.
- Assigns a *security score (0-100%)* based on a comparative legal assessment.

The AI model improves the *accuracy and reliability* of policy risk detection by ensuring policies are evaluated against widely accepted legal standards.

4. Security Score Generation

Both *Normal Mode and AI Mode* generate a *final security score* between *0% and 100%* based on the identified risks. The scoring system follows a weighted model:

- *High-risk clauses* (e.g., unlimited data sharing, hidden charges) significantly lower the score.
- *User-friendly clauses* (e.g., clear refund policies, opt-in data collection) improve the score.
- *Neutral clauses* have minimal impact on the score.

This final *risk assessment score* helps users make an informed decision about whether to *trust a website's policies or proceed with caution*.

5. Report Generation and User Interaction

Once the analysis is complete, the extension presents the results in a *structured and user-friendly format*:

- *Summary in three lines*, outlining key findings.
- *Pros and cons list*, highlighting benefits and risks.
- *Security score (0-100%)*, visually represented.
- *Save report feature*, allowing users to keep a record for future reference.

This *automated methodology* ensures that users can quickly *scan, understand, and assess* a website's policy before accepting its terms, ultimately enhancing *transparency, security, and data privacy awareness*.

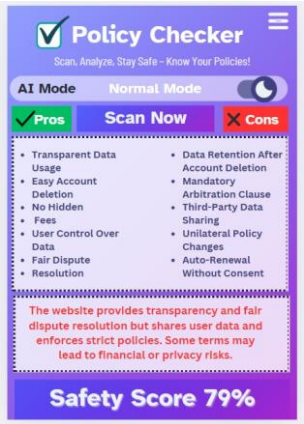

Fig-1 User Interface (UI)

System requirement:

Hardware Requirements

- *Processor:* Intel Core i3 or higher
- *RAM:* Minimum 4GB (8GB recommended)
- *Storage:* At least 500MB free space
- *Internet:* Stable internet connection for AI Mode

Software Requirements

- *Operating System:* Windows, macOS, or Linux
- *Browser:* Google Chrome (latest version)
- *Dependencies:* JavaScript, HTML, CSS, and OpenAI API access
- *Database:* NoSQL or local storage for logs (optional)

Working

The **Policy Scanner Chrome Extension** enables users to analyze website policies effectively by offering two distinct modes: **AI Mode** and **Normal Mode**. When a user interacts with the extension through their browser, they must choose between these modes. In **Normal Mode**, the extension employs **NLP, regex-based pattern detection, and sentiment analysis** to extract key points from the policy document. It then categorizes terms into pros and cons, highlighting user-friendly and risky clauses. This information is presented in a structured format along with a **security score (0-100%)**, helping users quickly assess the safety of a website's policies.

In contrast, **AI Mode** leverages OpenAI's advanced analysis capabilities, gathering insights from **10 trusted sources** to detect hidden risks, policy violations, and misleading clauses. The extension performs a deep legal compliance check, identifying potential threats such as **data sharing, financial risks, and unfair terms**. The AI-based findings are then compiled into a detailed report, including a risk assessment summary and a **final security score**. Regardless of the mode used, the results are displayed in an easily understandable format, allowing users to **save the report** for future reference**.**

Conclusion

The **Policy Scanner Chrome Extension** operates by allowing users to scan website policies through two distinct modes: **AI Mode** and **Normal Mode**. The process begins when a **user accesses the extension through their browser** and selects a mode. In **Normal Mode**, the extension uses predefined algorithms to analyze policy documents, extract key points, and list pros and cons. In contrast, **AI Mode** utilizes OpenAI for deep analysis, detecting potential policy violations by cross-referencing data from multiple sources. Both modes generate an **output report**, which includes findings and a security score. Finally, users have the option to **save the report** for future reference.

REFERENCE

[1] J. R. Mayer and J. C. Mitchell, "Third-Party Web Tracking: Policy and Technology," *2012 IEEE Symposium on Security and Privacy (SP)*, pp. 413-427, 2012.

[2] A. Datta, M. C. Tschantz, and A. Datta, "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies*, vol. 2015, no. 1, pp. 92-112, 2015.

[3] P. Ilia, I. Polakis, E. Athanasopoulos, F. Maggi, A. Kruegel, and S. Ioannidis, "Face/off: Preventing Privacy Leakage from Photos in Social Networks," *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 781-792, 2015.

[4] R. M. Ramadhan, "Privacy Policy Awareness in Mobile Applications," *International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*, pp. 1-6, 2019.

[5] Terms of Service; Didn't Read (ToS;DR), "Analyzing and Rating Website Terms and Privacy Policies," *Available at: https://tosdr.org*, [Accessed: 28 March 2025].

[6] A. S. Narayanan, H. X. Liu, and M. G. Nita-Rotaru, "Detecting Dark Patterns in Privacy Policies," *IEEE Symposium on Security and Privacy*, pp. 235-251, 2023.

[7] S. Zimmeck and S. M. Bellovin, "Privee: An Architecture for Automatically Analyzing Web Privacy Policies," *USENIX Security Symposium*, pp. 1-16, 2014.

[8] J. C. Park, Y. S. Kim, and S. Lee, "AI-Based Risk Assessment of Online Privacy Policies," *Journal of Artificial Intelligence and Law*, vol. 7, no. 2, pp. 115-132, 2022.

[9] TermsFeed, "Privacy Policy Generators and Legal Compliance Solutions," *Available at: https://www.termsfeed.com*, [Accessed: 28 March 2025].

[10] M. Fischer and D. Wright, "Transparency in Data Sharing: Legal and Ethical Challenges," *Journal of Privacy & Security Studies*, vol. 15, no. 3, pp. 225-241, 2021.

[11] M. T. Nguyen, H. R. Kim, and J. Y. Park, "Sentiment-Based Risk Analysis of Online Terms and Conditions," *International Conference on Web Intelligence and Security*, pp. 345-359, 2020.

[12] A. Chhabra, "Analyzing Consumer Privacy Risks in Terms and Conditions," *International Journal of Cyber Law and Policy*, vol. 8, no. 1, pp. 49-65, 2023