Journal of Computer Science & Computational Mathematics, *Volume 15, Issue 01, March 2025*
DOI: 10.20967/jcscm.2025.01.004

A Literature Survey: Legal Ease [Privacy Policy Simplification Web Extension]

P. V. Siva Kumar*, D. Nisritha, M. Sreeja, V. Jahnavi, R. N. S. Keerthana, S. Shalini

Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology, Hyderabad, Telangana, India-500090
*Corresponding author email: sivakumar_pv@vnrvjiet.in

***Abstract*:** Legal Ease is a web extension that makes complicated and drawn-out privacy policies of websites, apps, and online services more approachable and intelligible for the typical user. It highlights important details including the kinds of personal data gathered, how it is used, and any possible dangers or privacy issues while distilling these rules into succinct, understandable descriptions.The extension protects users' privacy while assisting them in making informed choices because it does not store or retain any legal papers. Legal Ease enables users to strengthen their awareness of privacy and take control of their digital life.

***Keywords*:** Privacy policies, legalese, user privacy.

1. Introduction

Privacy concerns have grown in importance in today's digital environment, as people disclose more personal information with online platforms. However, privacy policies in the documents that explain how user data is collected, utilized, and protected are sometimes lengthy, complicated and packed with legal terms, making them difficult for the typical user to comprehend. Despite the fact that these rules are essential for safeguarding personal information, most users of digital services are either uninformed of or unable to understand the conditions to which they consent. Legal Ease seeks to solve this issue by improving the accessibility and comprehensibility of privacy policies. The most crucial details, including what data is gathered, how it is used, and any risks involved, are highlighted in these short, easy-to-read summaries of complicated legal regulations. By clearly outlining privacy policies, Legal Ease helps users understand their rights and potential implications of disclosing personal information across various platforms.

Additionally, Legal Ease provides an evaluating tool, letting them make informed judgments about whether platforms correspond with their privacy preferences. Legal Ease's ultimate objective is to enable consumers to take charge of their online privacy, promoting increased openness and confidence between users and the services they utilize.

2. Literature Survey

Julia B. Earp, Annie I. Antón, Lynda Aiman-Smith, and William H. Stufflebeam's [1] study "Examining Internet Privacy Policies Within the Context of User Privacy Values" investigates how well website privacy policies align with user privacy expectations. The study found a significant gap between the protections offered by privacy rules and what people actually value after reviewing nearly 50 websites and surveying more than 1,000 internet users. Notably, users value greater openness in data transfer, knowledge of data usage, and the safe storage of data more than privacy policies, which often emphasize the collection of data and security assurances. This gap suggests that companies should better align their privacy policies with user values to build trust.

The research by Fei Liu [2] examines the problem of using automated analysis to make privacy policies easier to understand. The authors aggregate comparable sections across different privacy rules in an effort to improve readability by applying Natural Language Processing (NLP) techniques, since they realize that users frequently ignore these complex papers. To locate sections of related topics, such as data sharing or cookie use, the study evaluates clustering techniques and a hidden Markov model (HMM). The authors demonstrate where machine learning can closely match user views by contrasting these automated classifications with human assessments.

According to Ruogu Kang, Laura Dabbish, Nathaniel Fruchter, and Sara Kiesler in their paper "My Data Just Goes Everywhere: User Mental Models of the Internet and Implications for Privacy and Security" [3], perceptions of the internet by users shape their decisions in terms of privacy and security. The study's interviews and diagramming exercises revealed that technical participants tend to view the internet as a complex, multi-layered network, whereas non-technical individuals tend to view it as a simple, service-oriented structure. However, there was no direct relationship between these mental models and safer online behavior. Rather, personal experience, trust in established businesses, and overt signals like HTTPS had more influence on privacy practices. According to the report, privacy-preserving technologies that work irrespective of users' technical ability are required.

Using privacy policy analysis, Ashwini Rao et al.'s study "Expecting the Unexpected: Understanding Mismatched Privacy Expectations Online" [4] compares users' expectations and actual privacy practices on well-known websites. The study, which focuses on data collecting, sharing, and deletion, finds discrepancies between user assumptions and revealed practices by surveying 240 people across 16 websites. It demonstrates that, especially with respect to financial and

health data, users often expect less data collection and transfer than actually occurs. Because websites collect and transfer data types users did not expect, this mismatch may lead to privacy concerns. The study suggests that to enhance openness and trust, privacy warnings should address unexpected practices.

"In Limitations of Notice and Choice," Cate, Fred H. [5] points his finger towards the privacy protection these two applications give. The person who signed these agreements, realizing what he was doing, Cate does not argue, but instead because he had no option, which has indeed redressed moral superiority. He has further shown that these are in fact such complex information surveys that, so said, generate only delusion than explanation. Alternatively than such programs, the highlight must be sharper on data usage and schemes for protection. There is really so much need for transparency in this with a particular regard to ethical choices.

The Creation and Analysis of a Website Privacy Policy Corpus by Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N. Cameron Russell, Thomas B. Norton, Eduard Hovy, Joel Reidenberg, and Norman Sadeh [6] created a corpus of 115 website privacy policies, with detailed information about 23,194 data practices annotated. The authors stated the major problem users face in understanding the policies, which usually are lengthy and complex and which serve as concrete legal agreements. To address this issue, the team developed a structured annotation schema that categorized data practices into ten categories, such as "First Party Collection/Use" and "Third Party Sharing/Collection." A team of qualified annotators was engaged to label the policies in a web-based annotation tool, which created a data set full of nuances that could facilitate research in natural language processing and privacy policy analysis.

The paper also further envisions the possibility of automating the annotation task by utilizing machine learning techniques, with results showing that some categories, such as "Do Not Track," can be targeted with a high degree of accuracy.The subsequent section broaches some essential directions for future research, including the need for scaling annotation techniques and establishing user-friendly interfaces for representing privacy guidelines to help end users better understand the rights and choices available to them in connection with data privacy.

Privacy in the digital age: comparing and contrasting individual versus social approaches towards privacy by Marcel Becker [7] writes on a paper related to creating a corpus of website privacy policies containing 115 policies annotated with fine-grained information regarding 23,194 data practices that deals with how these users fail to understand long, complicated policies working as legal documents. To address this problem, the researchers developed an annotation scheme that categorized data practices into ten distinct categories, such as "First Party Collection/Use" and "Third Party Sharing/Collection." Skilled annotators used a web-based tool to label the policies, creating a dataset that supports research in natural language processing and privacy policy analysis.

A Neural Attention Model for Sentence Summarization by Alexander M. Rush, Sumit Chopra, Jason Weston [8] The document discusses the development of a corpus of 115 annotated document privacy policies which detail 23,194 data practices. Discusses text involving a series of illuminated problems related to the deciphering of intricacies of said policies in their ordinary sense as legal agreements. For better understanding, the researchers subsequently developed a structured annotation scheme that put data practices into ten categories, First Party Collection/Use; Second Party Collection/Use; and Third Party Sharing/Collection, among others; trained annotators were lastly used to annotate these web-based tool documents, culminating in the establishment of a dataset intended to support research on natural language processing and privacy policy analysis.

The paper “A Theory of Vagueness and Privacy Risk Perception” by Jaspreet Bhatia et al [9] investigates how consumers’ perceptions of privacy risk are affected by ambiguity in privacy regulations.The authors provide a taxonomy of problematic terms in privacy regulations using empirical text analysis, highlighting the ways in which terms like “may” and “generally” create ambiguity. They demonstrate how more vague rules may undermine user trust and desire to share information. They use factorial vignettes, paired comparison, and content analysis approaches to examine four main areas of ambiguity: conditionality, generalisation,modality and numeric quantifiers. The findings indicate that unclear terms influence users behavior while sharing information and make it more challenging for them to evaluate privacy risks. The authors suggest that policymakers adopt clearer language to improve transparency and user understanding.

Polisis: Deep Learning-Based Automated Privacy Policy Analysis and Presentation [10] The paper “Polisis: Automated Analysis and Presentation of Privacy Policies Using Deep Learning” introduces the automated privacy policy analysis system, Polisis. Polisis uses a large corpus of privacy rules(130,000) to train a language model that is particular to privacy. It employs a hierarchy of neural network classifiers and separates policies into parts to categorize data practices.It accurately labels relevant privacy phrases in policy sections. PriBot, a question answering chatbot, and structured querying for privacy icons are Polisis’s primary features. A user study found that Polisis assigns privacy icons with 88.4% accuracy and answers questions with 82% top-3 response accuracy. PriBot allows users to submit free-form questions, and it will provide relevant policy sections and a confidence level for each answer. Polisis aims to make privacy more transparent.

Pedro G. Leon, Blase Ur,Rebecca Balebako,Lorrie Faith Cranor, Richard Shay and Yang Wang’s study “Why Johnny Can’t Opt Out: A Usability Evaluation of Tools to Limit Online Behavioural Advertising” [11] examines the usability of nine solutions, including opt-out tools, blocking tools, and built-in browser privacy settings, designed to assist users

in managing online behavioural advertising(OBA). In a research with 45 participants, significant usability issues were discovered, including confusing interfaces, technical jargon, insufficient feedback, and ineffective default settings. Many consumers mistook opt-out solutions for totally stopping tracking, whereas browser features like Do Not Track were appreciated but mistrusted. Due to the difficulty of setting up blocking applications like Ghostery and AdBlock Plus, users were unable to achieve the utmost level of anonymity. The survey highlights the necessity of more accessible, user-friendly technologies that help consumers effectively control their privacy.

The importance of client-side security to modern application security is highlighted in the study The Role of Client-Side Protection in Modern Application Security [**?**], which notes that user device vulnerabilities are often overlooked and difficult to address due to the lack of server management.The primary challenges include understanding the behavior of third-party scripts, gaining visibility into them, and putting security measures like Content-Security-Policy headers into place.Threats like JavaScript injection and clickjacking can cause significant data breaches through compromised software supply chains. Effective client-side protection follows PCI DSS 4.0 guidelines and incorporates script validation, continuos monitoring and a zero-trust approach to safeguard user data and maintain compliance.

The paper A Survey on Text Summarization Techniques [12] reviews text summarization techniques in NLP and categorizes them into three categories: extractive, abstractive, and hybrid. Although abstractive techniques can generate summaries that use new language, extractive techniques extract important sentences from the source material. The study measures algorithms such as TextRank and BERT using metrics such as ROUGE and METEOR. Talking about applications in the domains of news media and legal papers, for example, multi-document summarization and user-specific requirements, the study underscores the need to continuously make further progress in this field.

Hui Yang et al. in Speculative Requirements: Automatic Detection of Uncertainty in Natural Language Requirements [13] discusses techniques for finding and removing speculative language from requirements documents by using uncertainty detection in natural language requirements. It uses a two-step process: the first step involves applying a machine learning technique (Conditional Random Fields) to find uncertainty cues within words, and then, with the aid of a rule-based method based on dependency structures, determines the scope of these cues. The study demonstrated the phenomenon of speculative language use that includes auxiliaries and epistemic terms in stakeholder communications through trials on 11 requirements papers. Although the system proved accurate in determining speculative sentences, phrase complexity proved to be challenging to scope.

The article by Abigail See et al. is titled Get To The Point: Summarization with Pointer-Generator Networks, where it assesses developments in sequence-to-sequence models for summarizing lengthy texts using abstractive summarization techniques. [14] presents a hybrid pointer-generator model that successfully copies words from source text while creating new ones by using extractive and abstractive techniques. The model guarantees a thorough summary and minimizes duplication by implementing a coverage method. The study claims improved ROUGE scores, at least two points higher than the previous abstractive approaches when applied to the CNN/Daily Mail dataset. Although this method handles out-of-vocabulary words and improves factual correctness, it has difficulty in generating high degrees of abstraction. The study highlights how summarization methods may be improved for wider applications with additional abstraction capabilities.

In order to increase semantic accuracy and reduce redundancy in summaries, [**?**] proposes the MS-Pointer Network model that integrates multi-head self-attention mechanisms with pointer networks. The experiment with CNN/Daily Mail and Gigaword datasets demonstrates that the MS-Pointer Network outperforms the current state-of-the-art models, such as the Pointer-Generator Network, in terms of ROUGE scores. The model reduces out-of-vocabulary problems and improves semantic feature capture by integrating position embeddings and coverage techniques. The results also show that such sophisticated mechanisms improve abstractive summarization; that is, the coherent semantic accuracy in summaries generated. Lower repetition and enhancing grammatical consistency in more complex models will focus the future of such research.

3. Challenges in Existing System

4. Proposed System

The goal of the state-of-the-art Legal Ease system browser plug-in is to make the privacy policies of websites, apps, and online services less complicated. These rules, which can be complex and full of legalese, make it difficult for people to comprehend how personal information is handled.Legal Ease addresses this issue by improving user accessibility, comprehension, and actionability of privacy regulations through the cutting-edge Machine Learning (ML) and Natural Language Processing(NLP) tools.

4.1 System Architecture

4.2 Algorithms used

A potent ensemble learning method that combines multiple decision trees is called Random Forest.Uses feature randomization and bagging to decrease overfitting and increase accuracy.

Adaptive Boosting, or AdaBoost, is an ensemble technique that improves subpar classifiers iteratively. By focusing on more challenging-to-classify samples, it improves model performance.

Extreme Gradient Boosting or XGBoost, is a very successful gradient boosting implementation. It is well known for its great computational speed, scalability, and anticipated accuracy.

Support Vector Classifier, or SVC, is a data classification technique that finds the optimum hyperplane to divide the

Table 1

Existing System	Purpose	Strengths	Weaknesses
Standard Privacy Policies	Provide detailed legal information on data practices	Comprehensive,directly from the source	Complex language, difficult to understand, time-consuming to read
TOSDR (Terms of Service; Didn't Read)	Summarizes and rates privacy practices for popular websites	Easy-to-understand summaries; provides ratings for quick insights	Limited to popular websites, often lacks depth, may be outdated
DuckDuckGo Privacy Essentials	Blocks trackers and provides basic privacy insights in-browser	Real-time tracker blocking, protects against fingerprinting	Limited insights into full privacy policies; focused primarily on tracker blocking
Polisis	Visualizes privacy policies for easier understanding	Provides unique visual summaries, allows users to see high-level policy insights	Visualization can be complex to interpret, lacks detailed breakdown of legal terms
Guard	Browser extension highlighting important privacy terms in policies	Highlights key terms in real-time, integrated into the browsing experience	Limited to key terms, lacks full policy breakdown; coverage may vary across websites
Privacy Check	Summarizes and analyzes privacy policies for mobile apps	Highlights key privacy risks in mobile app policies	Limited to mobile apps, lacks detailed summaries for web-based policies
Privado	Automates privacy policy management for businesses	Strong data protection insights, automates privacy management for compliance	Primarily enterprise-focused, lacks user-friendly summaries for general consumers
Juro Privacy Simplifier	Simplifies legal documents, including privacy policies	Makes legal language accessible for non-experts, tailored for core privacy terms	Mainly for business use, not widely available for consumer-facing policy comparison

Table 2

Category	TOSDR	DuckDuckGo Privacy Essentials	Polisis	Guard	Privacy Check	Privado	Juro Privacy Simplifier
Public Privacy Summaries	+	-	+	+	+	-	+
Real-Time Highlighting	-	+	-	+	-	-	-
Customizable for User Needs	-	-	-	-	+	+	+
Plain-Language Simplification	+	-	-	+	+	-	+
Comparative Analysis	+	-	-	-	-	-	-
Focus on Mobile App Policies	-	-	-	-	+	-	-
Enterprise Compliance and Automation	-	-	-	-	-	+	+
Cost of Operation	Free	Free	Free	Free	Free	Paid	Paid

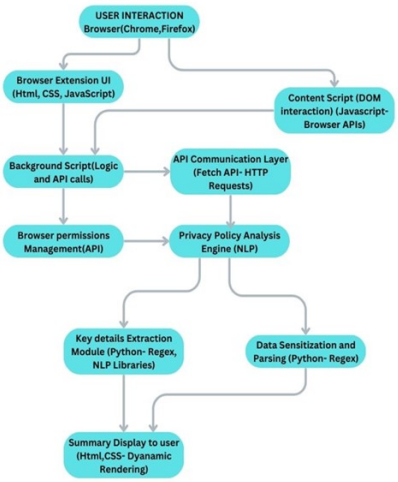

Figure 1. System Architecture

data into classes. It works well with modest to medium-sized datasets and in high-dimensional spaces.

Another sequential model-building ensemble technique is gradient boost. It focuses on minimizing the loss function by iteratively improving weak models.

Predictions from multiple base models, including AdaBoost and Random Forest, are combined in the stacking classifier. A meta-model takes these assumptions into account for better results.

Voting Classifier: Combines predictions from many algorithms by casting hard or soft votes. Hard voting selects the majority class, while soft voting averages probabilities.

5. Expected Results

Legal Ease is intended to make complicated privacy policies easier to understand by providing users with succinct descriptions. The application enables users to rapidly understand important aspects regarding data collection, intended usage, and potential privacy problems by translating legal jargon into simple explanations. This improves accessibility and saves time, allowing users to interact with websites and online services in an informed manner. Additionally, the plugin promotes openness and increases user confidence by highlighting important components like encryption procedures and third-party data exchange. Platforms that put user privacy first can gain more loyalty and trust, and consumers can navigate privacy regulations with less mental strain. One of Legal Ease's most notable features is its privacy-first approach, which guarantees user anonymity by not storing or sending analyzed data. This method maintains data integrity and is consistent with the system's fundamental privacy guarantee. Measurable results, such as less time spent comprehending privacy policies, better decision-making as a result of heightened awareness of data management procedures, and a rise in the use of privacy-focused platforms, can be used to gauge how beneficial the tool is. Legal Ease has the ability to improve user engagement and confidence throughout online ecosystems by facilitating a more transparent and secure digital experience.

6. Conclusion

The literature review highlights a significant gap between consumers' understanding of privacy policies and the purpose these policies serve. While privacy policies are designed to inform users about the collection, usage, and sharing of their personal data, their complex language often deters users from reading or comprehending the full contentAccording to research, this ignorance may result in misinformed consent, putting customers' privacy at danger. Existing initiatives to make privacy information easier to understand, including summarization tools, are helpful, but they frequently fall short of meeting the demands of the typical user. Solutions like Legal Ease can help close this gap by providing succinct and understandable summaries as well as comparative analysis. Legal Ease enhances decision-making and fosters a greater sense of user autonomy and privacy awareness by giving users easily comprehensible information concerning data practices. This solution not only promotes more openness in a world that is becoming more and more data-driven, but it also gives people the ability to make educated decisions, which eventually helps to create a more open and safe digital

environment.

References

[1] J. B. Earp, A. I. Antón, L. Aiman-Smith, and W. H. Stufflebeam, "Examining internet privacy policies within the context of user privacy values," *IEEE Transactions on Engineering Management*, vol. 52, no. 2, pp. 227–237, 2005.

[2] F. Liu, R. Ramanath, N. Sadeh, and N. A. Smith, "A step towards usable privacy policy: Automatic alignment of privacy statements," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 884–894.

[3] R. Kang, L. Dabbish, N. Fruchter, and S. Kiesler, "{"My} data just goes {Everywhere:"} user mental models of the internet and implications for privacy and security," in *Eleventh symposium on usable privacy and security (SOUPS 2015)*, 2015, pp. 39–52.

[4] A. Rao, F. Schaub, N. Sadeh, A. Acquisti, and R. Kang, "Expecting the unexpected: Understanding mismatched privacy expectations online," in *Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, 2016, pp. 77–96.

[5] F. H. Cate, "The limits of notice and choice," *IEEE Security & Privacy*, vol. 8, no. 2, pp. 59–62, 2010.

[6] S. Wilson, F. Schaub, A. A. Dara, F. Liu, S. Cherivirala, P. G. Leon, M. S. Andersen, S. Zimmeck, K. M. Sathyendra, N. C. Russell *et al.*, "The creation and analysis of a website privacy policy corpus," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2016, pp. 1330–1340.

[7] M. Becker, "Privacy in the digital age: comparing and contrasting individual versus social approaches towards privacy," *Ethics and Information Technology*, vol. 21, no. 4, pp. 307–317, 2019.

[8] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," *arXiv preprint arXiv:1509.00685*, 2015.

[9] J. Bhatia, T. D. Breaux, J. R. Reidenberg, and T. B. Norton, "A theory of vagueness and privacy risk perception," in *2016 IEEE 24th International Requirements Engineering Conference (RE)*. IEEE, 2016, pp. 26–35.

[10] H. Harkous, K. Fawaz, R. Lebret, F. Schaub, K. G. Shin, and K. Aberer, "Polisis: Automated analysis and presentation of privacy policies using deep learning," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 531–548.

[11] P. Leon, B. Ur, R. Shay, Y. Wang, R. Balebako, and L. Cranor, "Why johnny can't opt out: a usability evaluation of tools to limit online behavioral advertising," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2012, pp. 589–598.

[12] A. Nenkova and K. McKeown, "A survey of text summarization techniques," *Mining text data*, pp. 43–76, 2012.

[13] H. Yang, A. De Roeck, V. Gervasi, A. Willis, and B. Nuseibeh, "Speculative requirements: Automatic detection of uncertainty in natural language requirements," in *2012 20th IEEE International Requirements Engineering Conference (RE)*. IEEE, 2012, pp. 11–20.

[14] A. See, P. J. Liu, and C. D. Manning, "Get to the point: Summarization with pointer-generator networks," *arXiv preprint arXiv:1704.04368*, 2017.