Browser API for Managing and Recording Web Consent

Thilak Ramanie Shanmugasundaram
MSc in Computing
Dublin City University
Dublin, Ireland
thilak.shanmugasundaram2@mail.dcu.ie

Apurva Ravikiran Shirbhate
MSc in Computing
Dublin City University
Dublin, Ireland
apurva.shirbhate2@mail.dcu.ie

***Abstract*—Web consent refers to obtaining user permission online for data collection, usage, and sharing, ensuring compliance with privacy policies and legal requirements. Dark patterns frequently occur when users are tricked into accepting terms without proper recognition, When users clear cookies, the proof of consent acceptance is removed from the client side, leaving no way to track their consent. Solving these problems will require concerted effort for the consent process to be improved, increased transparency, and easy ways for users to manage their preferences across several websites, thus improving control and overall data visibility. Thus, this research focuses on deploying a browser API for handling and monitoring consent, tracking data usage, and revocation. Further, the API contains Smart Cookie Recommendations, which target the cookie pre-choices based on the user's likelihood and previous consent, improving the user experience and increasing the consent level. Moreover, supporting cross-platform consent synchronization, the research aims to provide a single experience to handle consent across multiple platforms – natural for a user. The results verify that using the Browser API for web consent management is effective in providing a solid solution against dark patterns and short-lived consents while at the same time preserving the user-oriented nature of the online context.**

***Index Terms*—User awareness, Web consent banners, User consent tracking, Browser extension, Informed decision-making, Privacy practices, User behaviour analysis, Smart cookie recommendation**

I. Introduction

In the contemporary world of digital use, privacy and information security have become critical issues for users and institutions. Using intervention-connected devices and complex internet services, the gaining and analyzing individual data has risen significantly. While helping to improve user experiences and business revenues, this data is privacy-sensitive if not processed appropriately. Thus, legislation like the General Data Protection Regulation (GDPR) of the European Union or the California Consumer Privacy Act (CCPA) of the United States has been developed to protect people's rights to privacy and guarantee that the information is processed by the companies lawfully. However, there are still problems regulating consent in the online environment. Inconsistencies within the GDPR, especially in cookie consent—underline the need for more explicit guidelines with respect to user monitoring and actual consent. For instance, different meanings associated with "informed consent" create confusion and uneven enforcement across different jurisdictions [1]. Various studies show that people need to properly comprehend the consent requests they interact with, thus making irrational consent choices and likely causing their data to be misused [2]. These problems are further complicated by requirements that would very clearly identify the service provider, the categories of personal data to be collected, the purposes for which the data will be processed, and the categories of recipients to which it will be transmitted [3].

In this regard, much has been invested in the research and development of consent management due to the increasingly strict global regulations regarding data privacy. The literature [2] reveals that different authors noted difficulties and issues when adopting proper consent management strategies. For instance, some studies on the GDPR indicate a need for uniformity in handling GDPR, especially regarding cookie consent notification. Hence, there is a need for better standards and friendly ways of dealing with the same [1]. Moreover, dark patterns, the strategies for making users initiate actions they do not want to undertake, also affect the consent mechanisms and limit the users' autonomy and trust [4]. The study [5] reveals that only 11.8% of these consent popups comply with the GDPR, possibly leading users to make uninformed decisions.

Research Objective: This research investigates user interactions with web consent banners. This study will try to understand users' awareness of the types of consent they provide, identify the most important factors that influence their consent decisions, and assess the impact of saving consent choices in an extension's dashboard on user trust. Furthermore, by examining previous user choices and utilizing incremental learning strategies to adjust to shifting user preferences and behaviours, we will develop and improve a machine learning-based recommendation system that precisely forecasts consumer preferences about cookie acceptance on websites.

The primary objective of this dissertation is to construct and appraise a comprehensive Browser API for web consent management. This API is designed to enhance user control over personal information, improve transparency in consent processes, and ensure compliance with international privacy laws. The research will involve the API's design, implementation, and deployment, its integration with smart cookie recommendation systems, and the evaluation of its effectiveness through experimentation. This recommendation system enhances user consent by offering consent options based on user behaviour and interests [6]. The extension was tested and deployed locally to users' machines to evaluate the research. Feedback and sustainability surveys were conducted, with most participants reporting improved transparency of data and greater user control over their choices compared to being tricked by dark patterns, one of the prevailing problems in the world of web consents [7].

Understanding the details behind online consent forms is crucial in today's world. In fact, consent has emerged as one of the main elements of data privacy and protection in the digital domain. An API like this would provide customers with much better control over their data, ease consent management for website owners, and enable compliance with strict privacy regulations.

II. Related Work

In this section, we present the state of the art in the more general areas of consent, its handling, and the representation of related information in online digital services.

A. Machine-Readable Representations :

Making consent records useful in digital settings must be machine-readable [8]. This allows digital tools and agents (like web browsers) to use and work with them. [9] The Data Privacy Vocabulary (DPV) developed by the W3C Data Privacy Vocabulary and Controls Community Group reflects a broad agreement among experts. The goal of the DPV is to standardise how consent and related information are represented, ensuring interoperability and compliance. It offers a structured, machine-readable format that helps manage consent records consistently and transparently across various platforms and jurisdictions.

B. Use of Consent Receipts

Consent receipts act as documented records of consent interactions, helping to ensure transparency and accountability in data processing activities [10].

1) Conceptual Framework and Implementations: Several Consent Management Platforms (CMPs), like ConsentEye and Signature, have adopted consent receipts to boost transparency and user trust. These platforms document user consent interactions and make this information easily accessible to users and auditors. [11] OneTrust LLC has developed a consent receipt management system highlighting the industry's move towards structured consent documentation.

2) Challenges and Requirements : Enhancing Accessibility and Usability :

Ensuring that consent receipts are accessible to users, and not just service providers, is essential for building trust and improving the consent process. Users need to easily access, review, and manage their consent receipts. This means designing user-friendly interfaces and systems allowing users to view their consent history, understand the details of each interaction, and withdraw or modify their consent as needed [12]. Providing users with control over their consent receipts helps boost their trust in the data processing practices of service providers [13]. However, creating such accessible and user-centric systems requires significant design and implementation efforts, particularly to ensure that these systems are intuitive and easy to use for individuals with varying levels of technical expertise [14].

C. Standards for Consent:

The Kantara Initiative's specification offers a starting point but lacks information on recent consent mechanisms, such as GDPR compliance. The specification must be updated to include new regulatory requirements and technological advancements [8]. ISO/IEC 29184:2020 outlines requirements for online privacy notices and consent, with implications for consent receipt design and utilisation [10].

D. Privacy Preference Signalling:

Privacy preference signals have evolved significantly over the years. Initial efforts, such as the "Do Not Track" (DNT) signal [15] and the Platform for Privacy Preferences Project (P3P), aimed to provide users with mechanisms to express their data preferences easily. These signals aim to provide a standardized way for users to communicate their privacy preferences and for service providers to respect them. This allows users to send a signal from their browser indicating their desire to opt out of data sales and sharing. This approach aims to simplify the process for users and ensure compliance from service providers.

E. Automated Consent Management and Revocation:

The studies [16] [17] address exactly how to make a cookie consent mechanism compliant under the GDPR. It proposes an automated framework using web scraping, machine learning, and natural language processing to identify cases of GDPR violation. [18] Consent-O-Matic is a browser extension designed to automatically handle consent pop-ups based on user preferences. Consent-O-Matic aims to simplify user interactions with consent banners by automating the consent process, ensuring that user preferences are consistently applied across different websites.

However, challenges in automation arise due to the complex nature of consent providers and the storage of consent options. Although the extension stores the number of clicks and websites visited, crucial information such as the consent acceptance status, types of consent accepted, and corresponding vendors are not stored. Consequently, Consent-O-Matic provides detailed steps for customising the framework, aiming to enhance its automation framework to handle much more complexities better. The implementations so far have paved the way for accepting consent automatically. Still, regarding the right to know about the data being used, this implementation is incomplete [4].

F. *Data Analytics for Personalized Consent Management:*

Machine learning and data analytics in consent management have colossal potential for enhancing user experience. Technologies in this space make modifying privacy settings more accessible and intuitive for users by predicting their preferences and speeding up the consent process.

1) Foundations of User Consent and Privacy Choices: Generalizable Active Privacy Choice: Proposal of a GUI for the Regulation of World Privacy This paper [19] discusses the user interface design concept for privacy decision-making, which could be useful in improving the application of scenario-based questions in determining initial user preference. The knowledge gained in this investigation can help fine-tune the visualization of these scenarios so that they are as unambiguous as possible to promote data entry accuracy.

2) Reinforcement Learning in Recommendation Systems: This survey [20] outlines Reinforcement learning applications with special emphasis on recommendation systems, describing different reinforcement learning approaches and their interaction with users' preferences. Reinforcement Learning is described as retraining SGD Classifier-based recommendations based on new occurrences of user interactions.

This paper [6] specifically covers Reinforcement Learning in online contexts and outlines data analysis and modification procedures that may improve the interactivity of the recommendatory system.

3) Hybrid and Real-Time Learning Approaches: This research [21] also discusses integrated models, which are formed of two or more algorithms and enhance performance when dealing with dynamic databases. While the system does not currently incorporate matrix factorisation, these methods could serve as ideas for future improvements.

4) Behavioral Dynamics and User Modeling: The research [22] lays the central groundwork for further study of user behaviour, thus making it exceedingly useful for theorizing systems such as the Consent-O-Matic. The paper also describes techniques for identifying patterns of users' behaviour over the course of time, which correlates with how our system monitors users' decisions related to cookie consent on various websites. It covers the topic of dynamic user profiling, with the help of the user models being updated when more information is collected. This is also evident in the kind of classifier you have used; the SGDClassifier improves its prediction with more user feedback.

5) Incremental and Online Learning Techniques: This paper [23] describes techniques used in incremental learning scenarios where the model's parameters are updated progressively with fresh data. It offers a foundation for those approaches and measures and applies these techniques when processing the new user data to maintain the model efficiently.

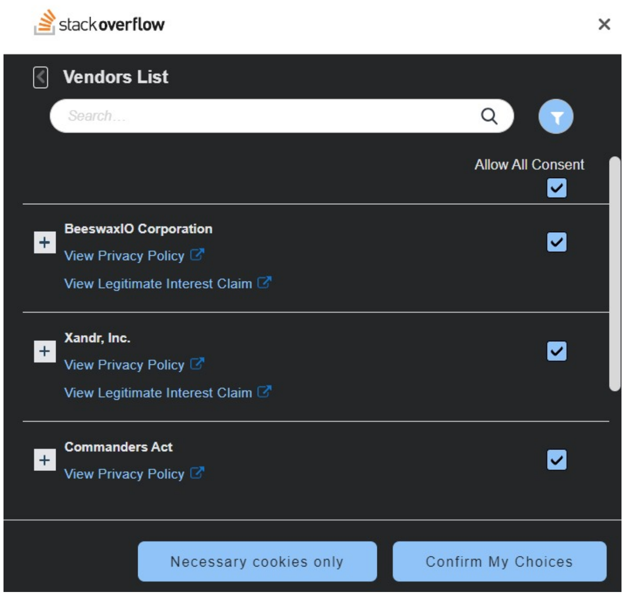
The study [24] discusses the uses of Stochastic Gradient Descent (SGD) as an optimizer for the neural networks regarding learning more complex functions mapping from the datasets. In the references made in our work, the SGDClassifier is applied to manage the dynamics of consent decisions made in website cookies. While our model does not work directly with neural networks, the principles of SGD described in the given paper are essential. That SGD offers the possibility to adjust model weights stepwise is beneficial for the idea of online learning as the data arrives continuously, like in the case of the users' interactions with the cookie consent banners.

III. Understanding Modern Consent Management Platforms (CMPs):

Consequently, in data protection, CMPs play a crucial role in explaining how websites collect and process users' consent to data processing operations. These third-party service providers are employed on websites to ask visitors' permission to read and write cookies and handle their information. Because of the frequently automatic placement of advertisements, many CMPs also have scanning capabilities for new purposes or new vendors seeking consent from the users. Notably, in Q3 2020, 58% of the 10,000 most popular websites in the UK utilized one of five prominent CMPs, Including QuantCast, OneTrust, TrustArc, Cookiebot, and Crownpeak.

From a user interface perspective, CMPs are generally structured into three sections: The first option entails the display of a landing page where the consumer is presented with general consent to the processing of his or her data; the second option involves the presentation of a second page containing the details of the processing of the consumers' data for various purposes; the third option exposes a third that lists some third parties involved in the processing of consumers' data.

A. *Vendor Lists and Consent Options:*



Allow All Consent ☑

BeeswaxIO Corporation ☑
View Privacy Policy
View Legitimate Interest Claim

Xandr, Inc. ☑
View Privacy Policy
View Legitimate Interest Claim

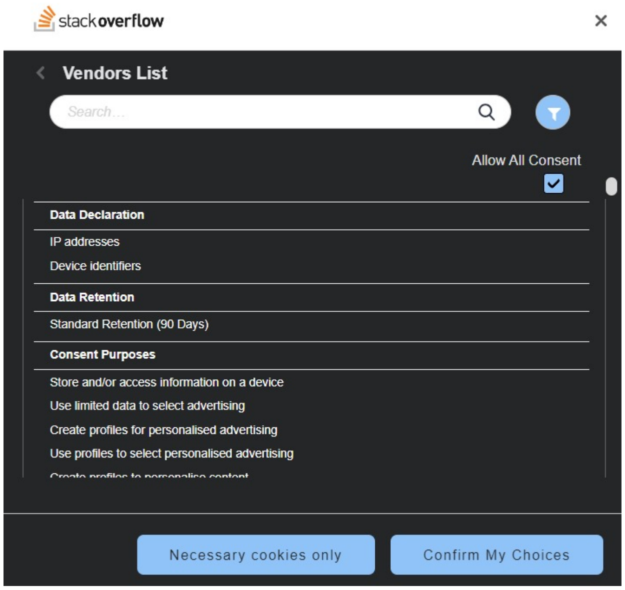
Commanders Act ☑
View Privacy Policy

Fig. 1. OneTrust vendors list

The consent shown in Figure 1 displays a vendor list with checkboxes for consent options when selecting the Accept all cookies option. The veteran "Allow All Consent" checkbox can be named one of the unquestionable, as it is usually pre-ticked. This form of nudging pressurizes users into agreeing with every vendor listed and processing their data.

B. Nested Consent Purposes:

Another example of the hidden essential information is in the image 2, where other details regarding the declaration of data, its retention and consent purposes are hidden in the foldable sub-sections. Choosing the design in this way imposes on the users some extra clicks that have to be made to ensure performance with lots of options and details: as pointed out above, users' choice is therefore not motivated to be implemented with a concern of its analysis and the reflection on the choice made.



Allow All Consent ☑

Data Declaration

IP addresses

Device identifiers

Data Retention

Standard Retention (90 Days)

Consent Purposes

Store and/or access information on a device

Use limited data to select advertising

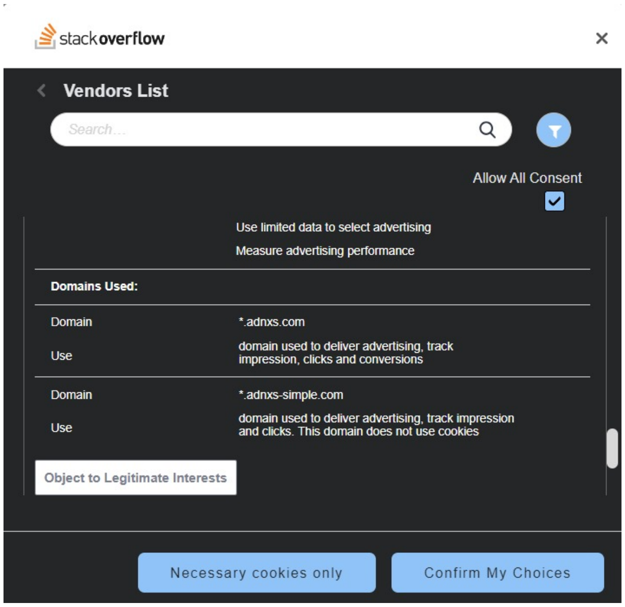
Create profiles for personalised advertising

Use profiles to select personalised advertising

Fig. 2. OneTrust consent purposes

C. Legitimate Interest Claims:

The specific and concrete domains belong to the legitimate interest for data processing to individual banners. This section contains an "Object to Legitimate Interests" button. However, it is considerably less visible than the previous one and needs to be clicked twice to activate it, probably cancelling the chances for effective interaction with the privacy elements.



Allow All Consent ☑

Use limited data to select advertising

Measure advertising performance

Domains Used:

Domain	*.adnxs.com
Use	domain used to deliver advertising, track impression, clicks and conversions
Domain	*.adnxs-simple.com
Use	domain used to deliver advertising, track impression and clicks. This domain does not use cookies

Fig. 3. OneTrust legitimate interest

IV. Methodology:

A. Requirements and Design Implementation:

1) The Framework: The primary requirement was to ensure the system could seamlessly interact with popular Consent Management Platforms (CMPs) like OneTrust. We chose Consent-O-Matic for its open-source nature, which provided a flexible foundation [17]. Leveraging Consent-O-Matic, we extended its capabilities to handle dynamic content and various implementations of OneTrust's CMP. OneTrust's pop-up content loads dynamically, posing a challenge for automated interactions. To address this, our system continuously monitors the network interaction intervals to ensure effective data retrieval.

2) Receipt Storage: We chose local storage for consent preferences and options over cookies for a seamless user experience. This is so because cookies are often cleared by the user or through the browser settings, which would remove the consent preferences. Local storage will help our system remember user choices and dynamically update the consent receipts on each website, even when cookies are cleared. This approach gives a more persistent and reliable way to manage consent data, ensuring user preferences are remembered across browsing sessions and making the system's usability much easier.

3) Recommendation System: The system intends to provide consent decision recommendations to users as they engage with website consent banners. To enhance its efficiency, it has the functionality of incremental learning, meaning that it can learn new preferences from the user without being trained all over. Furthermore, the system is designed to grow with user data volumes in an efficient manner, guaranteeing accuracy and excellent performance.SGDClassifier and TfidfVectorizer are two design decisions that were made in response to the need for a dependable, scalable system that could quickly provide customised suggestions. The dynamic nature of web interactions, where user preferences may change, and new data is always becoming available, is why these technologies were chosen.

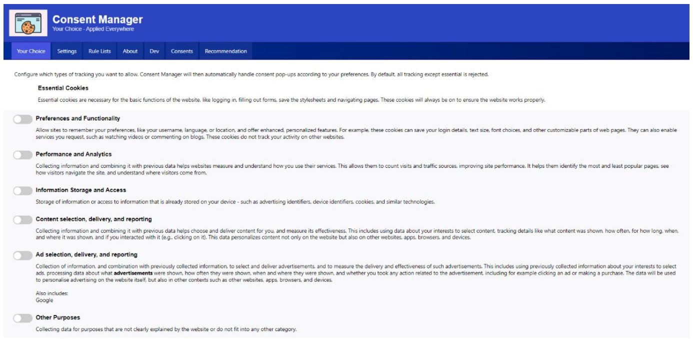
B. Extension Implementation:

The implementation is mostly concentrated on the OneTrust CMP. We undertook a reverse engineering process to sort through the DOM and the HTTP requests that OneTrust employed. It involved in-depth analysis of the available client documentation and drawing out detailed information about their implementations [25]. Customer-facing user experience masks technical subtleties in OneTrust's implementation, making interoperability particularly tricky.

One major challenge identified was the difference between dynamic and static HTML in OneTrust. The content of OneTrust pop-ups loads dynamically as the user interacts with the interface. This necessitates a mechanism in our system to continuously check for the presence of elements on the page and manage the intervals between these checks, thus limiting how quickly interactions with the interface can be automated. For instance, OneTrust often retrieves vendor information through remote calls with potential delays, which can significantly affect automation speed.

1) Targeting and Identifying Elements: Accurately targeting elements for interoperability with OneTrust requires dependable semantic information. OneTrust, however, often uses the same class names for different purposes, making it hard to differentiate between them. Some versions of OneTrust also mislabel optional items as necessary or use auto-generated classes, all hampering consistent targeting. In the absence of semantic information, our system has to fall back to navigation in the DOM tree, which is less reliable since any changes in structure may break the detection mechanism. The approach works fine for most OneTrust websites but only for some other websites, especially SPA(Single Page Applications). Yet, modifying or manually updating the rules schema worked for other web applications.

2) Automated Actions: The extension's actions will be comparable in interaction with the OneTrust interface; for example, clicking buttons or toggling checkboxes is performed based on the user's consent preferences. The user can set such preferences within the extension's settings page, which contains six purposes for data processing identified from a survey of CMPs used by popular websites. These designs help prevent user manipulation via dark patterns since they are, by design, transparent and automated in consent management. Also, the settings have a skipSubmit button that allows users to cross-check their choices. The options toggle without submitting the action, thereby allowing users, after viewing the results, to manually review and submit their preferences if they want to, improving trust and accuracy in consent management.



Configure which types of tracking you want to allow. Consent Manager will then automatically handle consent pop-ups according to your preferences. By default, all tracking except essential is rejected.

Essential Cookies

Essential cookies are necessary for the basic functions of the website, like logging in, filling out forms, save the stylesheets and navigating pages. These cookies will always be on to ensure the website works properly.

○ **Preferences and Functionality**

Allow sites to remember your preferences, like your username, language, or location, and offer enhanced, personalized features. For example, these cookies can save your login details, text size, font choices, and other customizable parts of web pages. They can also enable services you request, such as watching videos or commenting on blogs. These cookies do not track your activity on other websites.

○ **Performance and Analytics**

Collecting information and combining it with previous data helps websites measure and understand how you use their services. This allows them to count visits and traffic sources, improving site performance. It helps them identify the most and least popular pages, see how visitors navigate the site, and understand where visitors come from.

○ **Information Storage and Access**

Storage of information or access to information that is already stored on your device - such as advertising identifiers, device identifiers, cookies, and similar technologies.

○ **Content selection, delivery, and reporting**

Collecting information and combining it with previous data helps choose and deliver content for you, and measure its effectiveness. This includes using data about your interests to select content, tracking details like what content was shown, how often, for how long, when, and where it was shown, and if you interacted with it (e.g., clicking on it). This data personalizes content not only on the website but also on other websites, apps, browsers, and devices.

○ **Ad selection, delivery, and reporting**

Collection of information, and combination with previously collected information, to select and deliver advertisements, and to measure the delivery and effectiveness of such advertisements. This includes using previously collected information about your interests to select ads, processing data about what **advertisements** were shown, how often they were shown, when and where they were shown, and whether you took any action related to the advertisement, including for example clicking an ad or making a purchase. The data will be used to personalise advertising on the website itself, but also in other contexts such as other websites, apps, browsers, and devices.

Also includes:
Google

○ **Other Purposes**

Collecting data for purposes that are not clearly explained by the website or do not fit into any other category.

Fig. 4. Consent preferences

3) Enhanced Interoperability with OneTrust: One specific enhancement focuses on OneTrust CMPs. Our system captures the request sent when accepting or rejecting consent and stores this information in a machine-readable format. The consent decisions are displayed on a dashboard, and steps for revocation are clearly outlined for each website. This information is formatted according to ISO/IEC 29184:2020 standards, ensuring compliance with international guidelines.

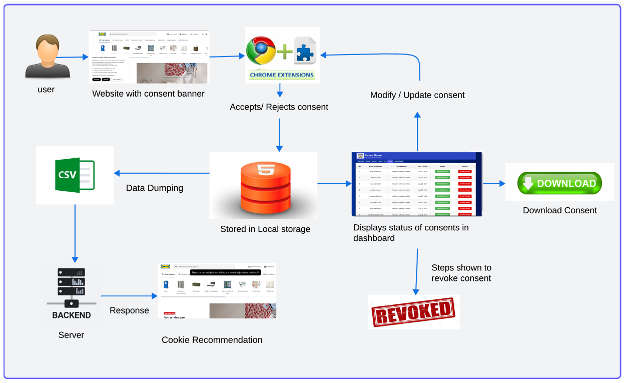
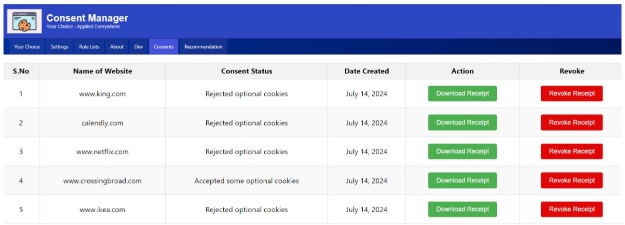


Fig. 5. Cookie management system

4) Local Storage and Dynamic Receipts: To provide users a seamless experience, we store their consent preferences and options in local storage. This approach ensures the system remembers user choices and dynamically updates consent receipts for each website visited, as depicted in the image below.



S.No	Name of Website	Consent Status	Date Created	Action	Revoke
1	www.king.com	Rejected optional cookies	July 14, 2024	Download Receipt	Revoke Receipt
2	calendly.com	Rejected optional cookies	July 14, 2024	Download Receipt	Revoke Receipt
3	www.netflix.com	Rejected optional cookies	July 14, 2024	Download Receipt	Revoke Receipt
4	www.crossingbroad.com	Accepted some optional cookies	July 14, 2024	Download Receipt	Revoke Receipt
5	www.ikea.com	Rejected optional cookies	July 14, 2024	Download Receipt	Revoke Receipt

Fig. 6. User dashboard

When users modify their consent options on a website, the system immediately reflects these changes, generating a new receipt in a machine-readable format that complies with ISO/IEC 29184:2020. Clicking on the download button in the dashboard provides the user with the receipt formatted below.

```
{
  "@type": "dpv:ConsentRecord",
  "dct:conformsTo": "https://w3id.org/dpv/schema/iso-27560"
    ,
  "dct:created": "2024-07-13T19:32:30.216Z",
  "dct:identifier": "e87740d0-06f5-4e12-b06a-b86deb8f8c5f",
  "dct:modified": "2024-07-13T19:41:44.959Z",
  "dct:valid": "2025-07-13T19:32:30.216Z",
  "dpv:hasConsentStatus": [
    {
      "@type": "dpv:ConsentNotGiven",
      "dpv:hasDuration": {
        "@type": "dpv:TemporalDuration",
        "rdf:value": "P12M"
      }
    }
  ],
  "dpv:hasEntity": [
    {
      "@id": "",
      "@type": "dpv:DataController",
      "dpv:hasName": "Netflix International B.V.",
      "schema:contactPoint": {
        "@type": "schema:ContactPoint",
        "schema:address": {
          "@type": "",
          "schema:streetAddress": "Karperstraat 8-101075 KZ
               Amsterdam, the Netherlands"
        },
        "schema:contactType": "Customer Service",
        "schema:email": "contact@",
        "schema:telephone": "62266519",
        "url": "https://help.netflix.com/en/node/134094"
      },
      "schema:url": ""
    }
  ],
  "dpv:hasProcess": [
    {
      "dpv:hasPersonalData": [
        {
          "@type": "",
          "dct:description": ""
        }
      ],
      "dpv:hasPurpose": {
        "Essential Cookies": true,
        "Functional Cookies": false,
        "Performance Cookies": true,
        "Targeting Cookies": false
      },
      "dpv:hasRecipient": [
        {
          "@id": "",
          "@type": "dpv:Recipient",
          "dpv:hasLocation": "loc:EU",
          "Vendors": [
            "spotify",
            "tiktok",
            "doubleclick",
            "pingdom",
            "instagram",
            "google",
            "bing",
            "facebook"
          ]
        }
      ],
      "dpv:hasRight": [
        {
          "@type": "dpv:DataSubjectRight",
          "dct:title": "Right to Withdraw Consent",
          "dpv:hasApplicableLaw": "dpv-gdpr:GDPR"
        }
      ]
    }
  ]
}
```

- **@type: dpv:ConsentRecord:** This indicates that the record is a consent record, compliant with the Data Privacy Vocabulary (DPV) standards.
- **dct:conformsTo:** This field specifies the URL to the standard the record conforms to, in this case, ISO/IEC 29184:2020.
- **dct:created and dct:modified:** These fields provide timestamps for when the consent record was created and last modified, respectively.
- **dct:identifier:** A unique identifier for the consent record.
- **dct:valid:** This indicates the validity period of the consent record.
- **dpv:hasConsentStatus:** This field describes the consent status (e.g., consent given or not given) and its duration.
- **dpv:hasEntity:** Details about the data controller, including the organization's name, contact information, and address.
- **dpv:hasProcess:** This section provides information on the data processing activities, types of personal data involved, the purpose of data processing, recipients of the data, and the data subject's rights.

C. Personalised Cookie Consent Recommendations

Unlike others, our system tends to work based on cookie consent, users' past choices, and the appearance of suitable notifications. The base of the developed system is constructed on the machine learning model, namely the SGD (Stochastic Gradient Descent) Classifier. This model analyzes the user's behaviour and determines the probable responses of the client towards cookies within diverse websites. The SGD Classifier is well suited for this purpose given that the method can deal with large-scale data and update this information quickly from new information obtained from activities such as web browsing, where preferences frequently change.

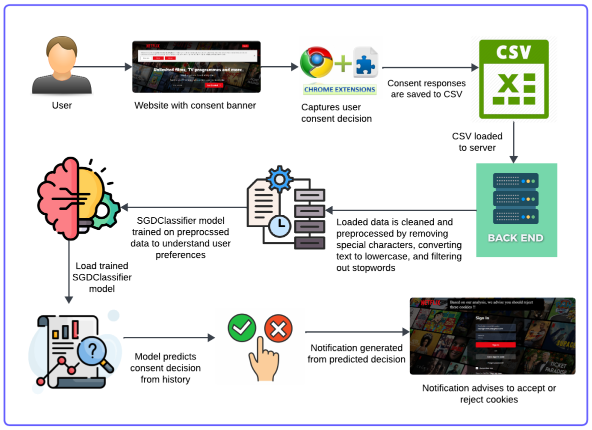

User
Website with consent banner
CHROME EXTENSIONS
Captures user consent decision
Consent responses are saved to CSV
CSV
CSV loaded to server
BACK END
Loaded data is cleaned and preprocessed by removing special characters, converting text to lowercase, and filtering out stopwords
SGDClassifier model trained on preprocssed data to understand user preferences
Load trained SGDClassifier model
Model predicts consent decision from history
Notification generated from predicted decision
Notification advises to accept or reject cookies

Fig. 7. Cookie Consent Recommendation System

1) Data Collection : The data collection approach that forms the structure of the suggested cookie consent recommendation system involves logging users' interactions with consent banners across different websites. This data collection is facilitated through a browser extension designed specifically to:

a) Dynamic Data Capture:

- Capture User Decisions: Every time users make a decision regarding the consent banner (either to accept or decline), it is counted. Moreover, these decisions vary based on the type of cookie and the site, which equips analysts with a great deal of contextual information.
- Data Storage: This selection is written in the CSV file, including a few additional data: cookie description and the website it is connected with. The most critical choice is the CSV file format, which is used by the program and is delivered to the machine learning model; thus, the data should be updated as much as possible to cover the latest user activities.
- Real-Time Updating: In the case of the data file, new records are always created in an ever-expanding list due to more user engagements demanding more consent banners. This feature is very beneficial in ensuring that the system is always informed of the latest behaviours from the users. Thus, the availability of enough information mars the growth of new trends and tastes among the users.
- Initial User Interaction: When giving recommendations to employ the preference modelling to new users without prior history in the system, to estimate new users' preferences and build heuristics, the authors have designed 23 scenario-based questions. When they first load the extension, these appear to the user, and they are supposed to select their consent to share different data. This initial interaction assists the system in offering recommendations for items in the user's profile genre, even when a user history element is yet to be developed. Users' responses to these questions are recorded and stored in the same CSV file as their consent decisions. These responses are viewed as separate features in the dataset; moreover, they reflect more detailed information about users' preferences than common acceptance or rejection.

2) Machine Learning System: The central component of the recommendation system is the SGDClassifier, which is used in conjunction with Reinforcement Learning (RL) techniques to consistently adjust the user's consent patterns. Text data is preprocessed with the help of TfidfVectorizer, which allows the data to be implemented efficiently and accurately in the SGDClassifier.

a) Data Preparation:

- Text Vectorization with TfidfVectorizer: For text data obtained from user interactions and those scenario-based questions, TfidfVectorizer is used to transform the data into a format that can be used in machine learning. they are TF-IDF (term frequency-inverse document frequency) scores that indicate the importance of each word in the dataset, which is useful for the model input. TfidfVectorizer makes the model more sensitive to the text's importance and the decision's context, critical for analysing the subjects' decision-making processes in diverse and intricate consent-related contexts. This method preserves the number of the term occurrences and their importance within the collection. It helps to get more profound insight into users' preferences and increases the accuracy of the system prediction.
- Data Cleaning and Engineering: Cleaning and engineering are performed on the data after vectorization so that the categorical form of the data is compatible with training. It is beneficial in guaranteeing that all the data inputs are channelled through the right format, which is required in the next phases of the model training.
- Dataset Splitting: The processed data set split into training and testing sets enables the model to be 'tested' and helps avoid overfitting.

b) Model Training:

- SGDClassifier Configuration: The SGDClassifier is recommended for this application because of its suitability in large-scale streaming data and speed in online learning scenarios. It provides accurate, immediate filters which enhance the experience but do not encroach on the user's private sphere. The SGDClassifier is fitted with the 'log_loss' loss function since the pipeline is built to solve binary classification problems. This loss function is suitable for constructing a probability 'feel' that quantifies the probabilities of the outcomes, which is appropriate for binary choices such as the accept or reject decisions in cookie consents.
- Online Learning: The SGDClassifier class's online learning property is utilized by training incrementally on the dataset. The fact that each instance is misclassified is utilized to adjust the weights to the model to improve the classifier's learning with time as it is exposed to new or different data. This feature makes the SGDClassifier well-placed in content, receiving a continuous feed of new data like real-time user interaction.

c) Incremental Learning:

- Continuous Updates with TfidfVectorizer: With the new text data obtained, the TfidfVectorizer transforms it into numeric data, which the SGDClassifier consumes. This structure allows for the model's gradual development, where its estimations of user preferences are improved with new data, as the model doesn't have to train from scratch. This configuration allows for continuous model improvement without requiring whole retraining, which increases system scalability and memory efficiency.
- Adaptation to New Patterns: TfidfVectorizer and SGDClassifier work together to enable real-time learning and adaptability to change user behaviour patterns and external changes, including varying consent choices over

time or across websites.

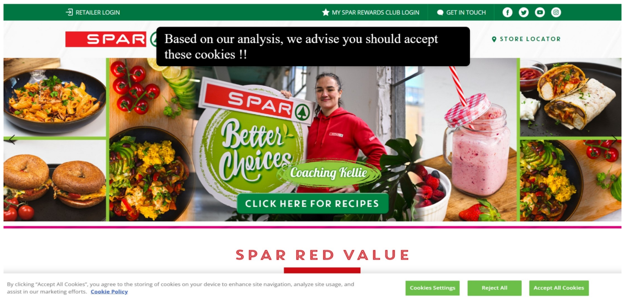
3) Prediction System: The trained model is used to predict user consent in real-time applications using the following model.

a) User Visits New Website: There is the activation of a prediction mechanism when a user, for the first time, signs into a new site is put in place.

b) Model Prediction: The SGD model predicts the user's likely consent based on historical data and the current context. Predictions are made in real-time, providing users with immediate recommendations on whether to accept or reject cookies.

4) Notification System: The efficacy of the recommendation system is also reflected in how it communicates with the user:

a) User-Friendly Notifications: Using the trained model, the recommendation system notifies websites about their cookie settings in real-time. Upon accessing a website, users are presented with a brief and easy-to-understand message advising them whether to accept cookies based on their prior choices. This automated guidance improves users' surfing experiences while honouring their privacy preferences by assisting them in making well-informed decisions more rapidly.



Based on our analysis, we advise you should accept these cookies !!

By clicking "Accept All Cookies", you agree to the storing of cookies on your device to enhance site navigation, analyze site usage, and assist in our marketing efforts. Cookie Policy

Fig. 8. This figure shows a consent banner on the SPAR website, displaying a notification that recommends accepting cookies based on a model trained on previous user data.

V. Evaluation:

The built Browser API and its integrated recommendation system were tested in the real world with a chosen set of 24 users to thoroughly assess the system's functionality and performance. During normal surfing circumstances, this user group interacted with the extension, offering useful insights regarding the system's responsiveness and UI design.

Following deployment, three Google Forms were distributed to collect structured input as part of an extension of the assessment:

a) System Usability Scale (SUS): This form Focused on the new consent management tools' simplicity of use and learning curve to statistically assess the extension's usefulness. The generated extension's usability was evaluated using the SUS, which yielded a median score of 70 and an average score of 71.48, suggesting reasonable usability. Most consumers gave satisfactory ratings for the usability and functionality of the product, with most falling between 50 and 80. The add-on addresses a crucial requirement for cookie consent management, as indicated by its aim for frequent usage. Feedback did, however, point out certain areas that needed work, such as making the UI more user-friendly and improving consistency across various scenarios. A few customers also mentioned that they could want technical support, emphasizing the need for improved user manuals and support materials. Though there is still an opportunity to improve user experience and usability, the addon is generally well-received.

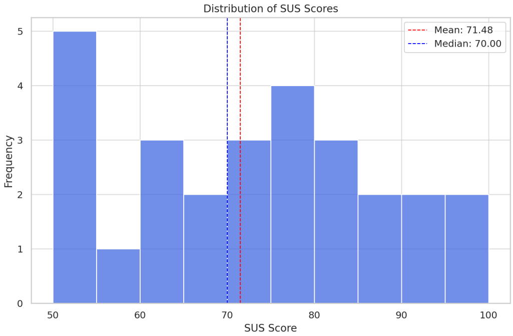


Fig. 9. Visual Analysis of SUS Score Distribution

b) Feedback Survey: Gathered qualitative data on user satisfaction and targeted areas for development, letting users share their thoughts on the usefulness of the product and any difficulties they had. The feedback reveals high satisfaction with functionality, transparency, and ease of managing consent settings, suggesting our design effectively counters dark patterns by supporting informed user choices. One user's concern about data collection highlights the need for clearer communication to prevent any perception of manipulative practices. Users value personalized recommendations and the user-friendly interface, indicating that these features enhance their ability to make autonomous consent decisions.

c) Data Contribution Form: We requested voluntarily from users information about their experiences handling data, which aided in determining whether privacy laws were being followed and how well the consent alternatives offered worked. Feedback shows that users are generally at ease with device data collection and browsing history collection since they consider these less invasive. But there's a noticeable reluctance to divulge more private information, like location or demographics, highlighting the importance of strong data protection and openness. Users voice concerns about security and data management, even though they regularly check their consent settings and trust the extension. The feedback indicates areas needing improvement to increase user engagement and experience, such as inconsistent consent banner management and a desire for more comprehensive, interactive consent alternatives.

Product Evaluation:
In testing, bugs were raised by the users, where it failed

to recognize consent banners by TrustArc and CookieYes. Further, it was unable to get the list of vendors due to the high level of complexity in the DOM structure that required recursive scraping. These kinks should be sorted for improvement in terms of its reliability and comprehensiveness.

Other potential bugs include inconsistent behaviour across different browsers, such as discrepancies in how consent banners are detected in Chrome versus Firefox, where the extension occasionally froze or became unresponsive when dealing with particularly complex or heavily scripted websites. Addressing these issues is crucial for enhancing the reliability and comprehensiveness of the extension. The overall assessment thus comes to the conclusion that users welcome the extension's good functionality and transparency.

Various quantitative and qualitative criteria were used to evaluate the cookie consent suggestion system's accuracy, user happiness, and adaptability. The main area of focus was the system's capacity to accurately forecast user consent decisions based on inputs from scenarios and previous data.

Accuracy and Performance Metrics: The SGD Classifier's predictive performance was assessed using common classification metrics, including accuracy, precision, recall, and F1-score. These metrics were computed using a test dataset divided from the original data gathered by the browser extension. In addition, user opinions were gathered to assess the notification system. Users said the alerts let them make rapid decisions without interfering too much with their browsing, and they were helpful and non-intrusive. Most users confirmed the system's usefulness in practical applications by praising its transparency and ability to modify privacy settings.

VI. Conclusion:

The developed extension has thus significantly increased users' control over consent regarding cookies. Functions of legal standards have been independently affirmed by the users, and aspects such as functionality, transparency, and efficiency have been highlighted. To further enhance the performance, the development requires refining the interface, asserting users' privacy, and maintaining high quality irrespective of the context. Future enhancements will try how to deal with nested DOM structures in a better way to include more specific consent details like vendors, the data controller, etc., to make the consent receipts more accurate. The user ratings, aimed at comparing the stability of reactions in the automatic and the manual consent responses, will offer insights into the further debate on the effectiveness of the signals collected by the browser-based consents under the ePrivacy Regulation. The plan also involves ascertaining the user-data frequency patterns and using state-of-the-art machine learning techniques such as feedback-integrated real-time learning to enhance the extension's performance and conformity [25]. Furthermore, a dedicated server for securely storing consent data can be developed, and a fast-processing solution can be implemented for user withdrawal requests to Data Controllers, thereby increasing transparency.

This research successfully constructed a recommendation system utilizing an SGD Classifier to improve user control over website cookie consent. The system adjusts to user preferences based on previous choices and scenario-based questions for new users to provide individualized and responsive recommendations. The system's capacity for incremental learning enables it to function well in the ever-changing context of online browsing, simplifying and making the cookie consent procedure less obtrusive and more user-friendly.

A crucial improvement for this project's future iterations might be creating a default model based on collected information from 400–500 users to identify common trends in users' choices about cookie consent. This model would act as a baseline for new users or those with little interaction history and offer basic recommendations until individualized data becomes available [3]. Additional investigation may delve into sophisticated machine learning techniques like ensemble or deep learning to enhance the system's forecast accuracy and flexibility. The system could modify recommendations in real-time by including real-time feedback mechanisms depending on user interactions and cumulative behaviour patterns [24]. This would improve user experience and ensure compliance with evolving privacy standards. These enhancements pave the way for the real-world implementation of user-centred privacy technologies. They would be a great addition to the knowledge of consent management systems.

References

[1] Daniel Anderson and Richard von Seck. The gdpr and its impact on the web. *Network*, 1, 2020.

[2] Martin Degeling, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz. We value your privacy... now take some cookies: Measuring the gdpr's impact on web privacy. *arXiv preprint arXiv:1808.05096*, 2018.

[3] Soha Jiwani, Rachna Sasheendran, Adhishree Abhyankar, Elijah Bouma-Sims, and Lorrie Cranor. Crumbling cookie categories: Deconstructing common cookie categories to create categories that people understand. *Proceedings on Privacy Enhancing Technologies*, 2024.

[4] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. (un) informed consent: Studying gdpr consent notices in the field. In *Proceedings of the 2019 acm sigsac conference on computer and communications security*, pages 973–990, 2019.

[5] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. Dark patterns after the gdpr: Scraping consent pop-ups and demonstrating their influence. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pages 1–13, 2020.

[6] Wacharawan Intayoad, Chayapol Kamyod, and Punnarumol Temdee. Reinforcement learning for online learning recommendation system. In *2018 Global Wireless Summit (GWS)*, pages 167–170. IEEE, 2018.

[7] Célestin Matte, Nataliia Bielova, and Cristiana Santos. Do cookie banners respect my choice?: Measuring legal compliance of banners from iab europe's transparency and consent framework. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 791–809. IEEE, 2020.

[8] Vitor Jesus. Towards an accountable web of personal information: The web-of-receipts. *IEEE Access*, 8:25383–25394, 2020.

[9] Harshvardhan J Pandit, Axel Polleres, Bert Bos, Rob Brennan, Bud Bruegger, Fajar J Ekaputra, Javier D Fernández, Roghaiyeh Gachpaz Hamed, Elmar Kiesling, Mark Lizar, et al. Creating a vocabulary for data privacy: The first-year report of data privacy vocabularies and controls community group (dpvcg). In *On the Move to Meaningful Internet Systems: OTM 2019 Conferences: Confederated International Conferences: CoopIS, ODBASE, C&TC 2019, Rhodes, Greece, October 21–25, 2019, Proceedings*, pages 714–730. Springer, 2019.

[10] Vitor Jesus and Harshvardhan J Pandit. Consent receipts for a usable and auditable web of personal data. *IEEE Access*, 10:28545–28563, 2022.

[11] OneTrust. Onetrust api reference, 2024. Accessed: 2024-07-25.

[12] Sven Bock, Ashraf Ferdouse Chowdhury, and Nurul Momen. Partial consent: a study on user preference for informed consent. In *HCI International 2021-Late Breaking Papers: Design and User Experience: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings 23*, pages 198–216. Springer, 2021.

[13] Gary Burkhardt, Frederic Boy, Daniele Doneddu, and Nick Hajli. Privacy behaviour: A model for online informed consent. *Journal of business ethics*, 186(1):237–255, 2023.

[14] Sebastian Zimmeck, Eliza Kuller, Chunyue Ma, Bella Tassone, and Joe Champeau. Generalizable active privacy choice: Designing a graphical user interface for global privacy control. *Proceedings on Privacy Enhancing Technologies*, 2024.

[15] Sebastian Zimmeck, Oliver Wang, Kuba Alicki, Jocelyn Wang, and Sophie Eng. Usability and enforceability of global privacy control. In *23rd Privacy Enhancing Technologies Symposium (PETS 2023)*, pages 265–281, 2023.

[16] Dino Bollinger, Karel Kubicek, Carlos Cotrini, and David Basin. Automating cookie consent and {GDPR} violation detection. In *31st USENIX Security Symposium (USENIX Security 22)*, pages 2893–2910, 2022.

[17] Ralf Gundelach and Dominik Herrmann. Cookiescanner: An automated tool for detecting and evaluating gdpr consent notices on websites. In *Proceedings of the 18th International Conference on Availability, Reliability and Security*, pages 1–8, 2023.

[18] Midas Nouwens, Rolf Bagge, Janus Bager Kristensen, and Clemens Nylandsted Klokmose. Consent-o-matic: Automatically answering consent pop-ups using adversarial interoperability. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pages 1–7, 2022.

[19] Sebastian Zimmeck, Eliza Kuller, Chunyue Ma, Bella Tassone, and Joe Champeau. Generalizable active privacy choice: Designing a graphical user interface for global privacy control. *Proceedings on Privacy Enhancing Technologies*, 2024.

[20] Yuanguo Lin, Yong Liu, Fan Lin, Lixin Zou, Pengcheng Wu, Wenhua Zeng, Huanhuan Chen, and Chunyan Miao. A survey on reinforcement learning for recommender systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[21] Chia-Yu Lin, Li-Chun Wang, and Kun-Hung Tsai. Hybrid real-time matrix factorization for implicit feedback recommendation systems. *Ieee Access*, 6:21369–21380, 2018.

[22] Kira Radinsky, Krysta Svore, Susan Dumais, Jaime Teevan, Alex Bocharov, and Eric Horvitz. Modeling and predicting behavioral dynamics on the web. In *Proceedings of the 21st international conference on World Wide Web*, pages 599–608, 2012.

[23] Jiangpeng He, Runyu Mao, Zeman Shao, and Fengqing Zhu. Incremental learning in online scenario. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13926–13935, 2020.

[24] Dimitris Kalimeris, Gal Kaplun, Preetum Nakkiran, Benjamin Edelman, Tristan Yang, Boaz Barak, and Haofeng Zhang. Sgd on neural networks learns functions of increasing complexity. *Advances in neural information processing systems*, 32, 2019.

[25] Dominique Machuletz and Rainer Böhme. Multiple purposes, multiple problems: A user study of consent dialogs after gdpr. *arXiv preprint arXiv:1908.10048*, 2019.