Ojas Kapre
Browser Security Project Report

1. Introduction

1.1 Overview

The "Website Cookie Preference Compliance" project investigates the critical role of cookie banners in aligning with data protection laws like GDPR and CCPA. As digital privacy becomes increasingly important, understanding and ensuring the proper use of cookie banners is essential.

1.2 Importance of Cookie Banners

Cookie banners are more than a legal requirement; they're key to maintaining user privacy on the internet. They inform users about cookie use and seek consent, upholding transparency and trust online.

1.3 Project Goals

This project focuses on several key objectives:

1. Studying the importance of cookie banners in the context of current privacy regulations.
2. Investigating how websites can align with GDPR and CCPA guidelines.
3. Summarizing findings from four research papers on the compliance of cookie banners.
4. Creating a tool using Python, Selenium, and OpenAI's GPT-3.5-turbo API to detect and analyze cookie banners on various websites.
5. Building custom GPTs using the ChatGPT plus to evaluate the cookie consent interfaces based on the principles discussed in the research paper[1]

The project combines theoretical research with practical application, providing valuable insights into the realm of online privacy and the critical role of cookie banners in user data protection.

2. Methodology

2.1. Literature review

The foundation of the project was established through a comprehensive literature review. Four pivotal research papers were analyzed, each offering unique insights into the realm of cookie consent interfaces and compliance challenges:

1. **"Okay, whatever": An Evaluation of Cookie Consent Interfaces:** This paper provided a critical analysis of how cookie consent interfaces are often designed to manipulate user choices.
2. **CookieEnforcer: Automated Cookie Notice Analysis and Enforcement:** This study introduced an architecture for an application aimed at enforcing user cookie preferences, emphasizing the need for automated compliance tools.

3. **A Systematic Study on the Impact of GDPR Compliance on Organizations:** This paper offered a broader view of how GDPR compliance affects organizations, highlighting the widespread implications of these regulations.
4. **Thou Shalt Not Reject: Analyzing Accept-Or-Pay Cookie Banners on the Web:** This research focused on the "accept-or-pay" model in cookie banners, scrutinizing the ethical and legal aspects of this approach.

Each paper is summarized in a detailed 2-3 page document, ensuring a thorough understanding and foundation for the project's practical components.

2.2. Cookie Detection Module Implementation

The concept of cookie enforcement, as detailed in the "CookieEnforcer" research paper[2], provided a foundational understanding of the complexities involved in managing cookie consents. Recognizing the intricacy of the entire enforcement architecture, the project narrowed its focus to the development of a cookie detection module, which is a crucial first step in the process.

Process Outline
The cookie detection process was streamlined into four main steps, as visualized in the provided process outline image:
1. **Identify Candidates Using z-index:** Initially, the module attempted to identify potential cookie banner elements based on their z-index value, a CSS property that determines the stack order of elements on a web page.
2. **Filter and Check Candidates Using String Matching:** The elements were then filtered through string matching, using keywords typically associated with cookie banners, such as "cookies," "preferences," "privacy," and "consent."
3. **GPT 3.5 Analysis to Classify Cookie Banner:** To refine the detection process, the GPT-3.5-turbo OpenAI API was employed. The model analyzed the HTML code of candidate elements to ascertain whether they were indeed cookie banners.
4. **Save Information and Screenshot:** Upon confirmation, the metadata and screenshots of the detected cookie banners were saved for further analysis.

Importance of Saved Metadata
The metadata captured from detected cookie banners—including various CSS properties and HTML attributes—was crucial. It provided insights into the common design and implementation patterns of cookie banners across different websites.

The Necessity of Machine Learning
The initial hypothesis was that cookie banners would possess the highest z-index value since they typically overlay the content on web pages. However, this assumption did not hold true across all websites. For instance, on the Barclays website, the navigation bar had a higher z-index due to its sticky nature, and on Wikipedia, the high z-index banner served a purpose other than cookie consent.
To address these inconsistencies, the module was designed to:
1. Collect elements with a numerical z-index value, excluding those set to 'auto.'
2. Sort these elements in descending order of z-index within a dictionary.

3. Iterate through each element, verifying its role as a cookie banner through keyword matching and the presence of buttons or links related to cookie consent.

This method significantly improved accuracy, yet it was insufficient for certain sites like YouTube, where elements such as the sidebar could be falsely identified as cookie banners due to matching keywords and the presence of links like "Privacy."

Thus, the application of a machine learning model, specifically using the OpenAI GPT-3.5-turbo, became imperative. The model assessed the HTML code of each potential banner and classified it with a simple prompt, enhancing the detection accuracy beyond the basic heuristic checks. The Prompt used was

```
Determine if the following HTML code is a cookie consent banner
and answer with 'True' or 'False' only: \n\n{html_code}
```

2.3 Cookie Consent Interface Evaluation

The evaluation of cookie consent interfaces is a crucial aspect of the project, aimed at assessing the compliance and user-friendliness of cookie banners based on the principles discussed in the research paper "Okay, Whatever." The evaluation module focuses on the following principles:

1. **Clarity and Transparency:** The text should be comprehensible without ambiguous or misleading phrases.
2. **Ease of Choice:** The interface should facilitate straightforward acceptance, rejection, or customization of cookie settings.
3. **Accessibility:** The banner must be navigable for users with disabilities, ensuring proper contrast, text size, and compatibility with screen readers.
4. **No Misleading Visuals:** Visual elements should not deceive or pressure users into certain actions.
5. **Option Balance:** Accept and reject options should be presented neutrally, avoiding undue emphasis on either.
6. **Forced Action:** Users should not be compelled to take action to access standard website functionalities.
7. **Bait and Switch:** The interface must deliver on what it promises, without misleading outcomes.
8. **Hidden Costs:** There should be no concealed costs or selections in the consent process.
9. **Disguised Ads:** Ads or opt-ins should not be camouflaged as other content types.
10. **Confirmshaming:** The design should not use guilt or shame to coerce consent from users.
11. **Obstruction:** Actions such as unsubscribing or rejecting cookies should not be unnecessarily complicated.
12. **Friend Spam:** The interface should not trick users into divulging excessive personal information.

Evaluation Process

Initially, the project used a basic prompt-based approach with screenshots to evaluate cookie banners against these principles. However, this method resulted in inconsistent assessments for different screenshots. To achieve consistency and streamline the process, custom GPT models were created using ChatGPT Plus.

The custom GPT models were designed with embedded instructions tailored to analyze the cookie banners against the 12 evaluation principles. Each screenshot submitted to the GPT was scored on a scale of 1-5 for each principle. The GPT provided not only the scores but also detailed reasoning behind each score, offering a transparent and consistent evaluation method.

Advantages of Custom GPTs

Building custom GPT models presented several benefits:

1. Consistency: Each banner was evaluated with the same set of criteria, ensuring uniformity across the board.
2. Efficiency: With the instructions pre-built into the model, there was no need to enter prompts repeatedly, saving time and reducing the potential for human error.
3. Depth of Analysis: The custom GPTs were able to delve into the nuances of each principle, providing a more profound level of insight than a surface-level evaluation.
4. Reasoning: The models offered explanations for each score, which was invaluable for understanding the context and rationale behind the evaluations.

3. Results

3.1. Cookie Banner Detection Module

The cookie banner detection module demonstrated exceptional performance, successfully identifying cookie banners across all websites included in the study. It effectively discerned the presence of banners and correspondingly saved the results; for websites without cookie banners, no data was captured, illustrating the module's accuracy.

Illustrative Results

```
{
    "position": {
        "x": 0,
        "y": 742
    },
    "size": {
        "height": 338,
        "width": 1920
    },
    "z_index": "1000",
    "class": "",
    "id": "sp-cc",
    "background_color": "rgba(246, 246, 246, 1)",
    "color": "rgba(15, 17, 17, 1)",
    "border": "1px solid rgb(221, 221, 221)",
    "screen_shot_path": "./results/amazon/main_banner.png",
    "links": [ …
    ],
    "number_of_links": 7,
    "buttons": [],
    "number_of_buttons": 0,
    "html_code": "<form id=\"sp-cc\" method=\"post\" action=\"/privacyprefs/retail/v1/acceptall\"><input
    "text": "Continue without accepting\nSelect Your Cookie Preferences\nWe use cookies and similar tools
}
```

Image 1: The above image captures the metadata of a detected cookie banner, detailing its position, size, and other CSS properties, confirming the module's ability to precisely identify and record banner characteristics.
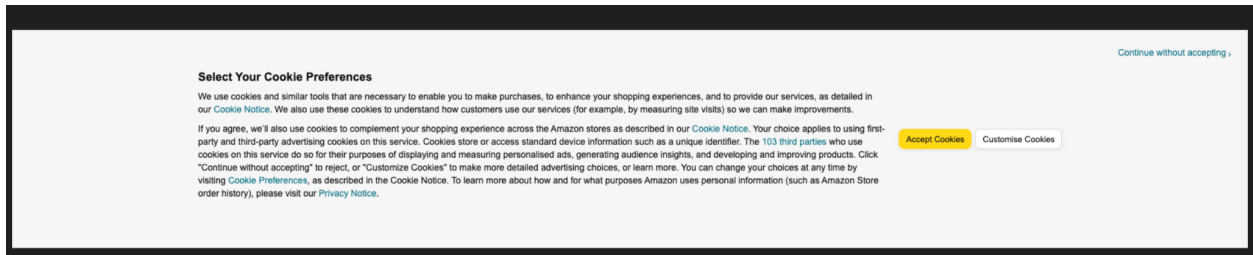
Image 2: Here, we see a screenshot of a website with a detected cookie banner. The visual representation corroborates the module's functionality in real-time application, visually verifying the accuracy of the detection process.
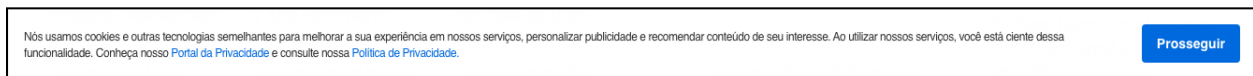
Image 3: The screenshot showcases a cookie banner in Portuguese, highlighting the module's multilingual detection capability. This ability is particularly valuable in ensuring compliance across diverse geographic and linguistic user bases.
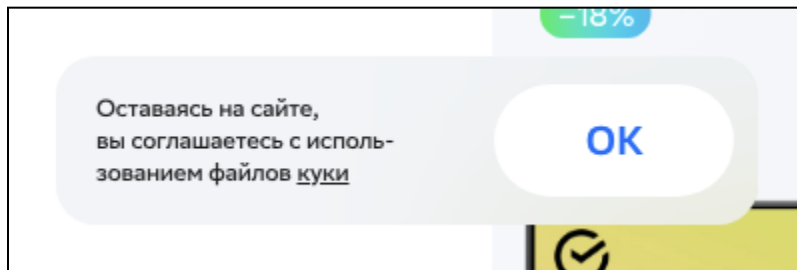

Image 4: Depicting a cookie banner in Russian, this image further affirms the module's linguistic versatility. It underscores the global applicability of the tool, capable of handling various languages with ease.

The results indicate not only the technical efficacy of the detection module but also its practical utility in a real-world context. The incorporation of GPT-3.5-Turbo has been instrumental in extending the module's capabilities, enabling it to navigate the complexities of language-specific banners.

With these findings, the project stands as a testament to the potential of combining machine learning techniques with heuristic analysis to create a robust tool for cookie banner detection.

3.2. Cookie Consent Interface Evaluation

The evaluation of cookie consent interfaces was conducted using a custom GPT model, designed to assess compliance with user-centered design principles. The results are presented in the form of a bar chart, reflecting the distribution of scores across the twelve principles for the analyzed websites.

Number of websites scoring 1-5 for each principles
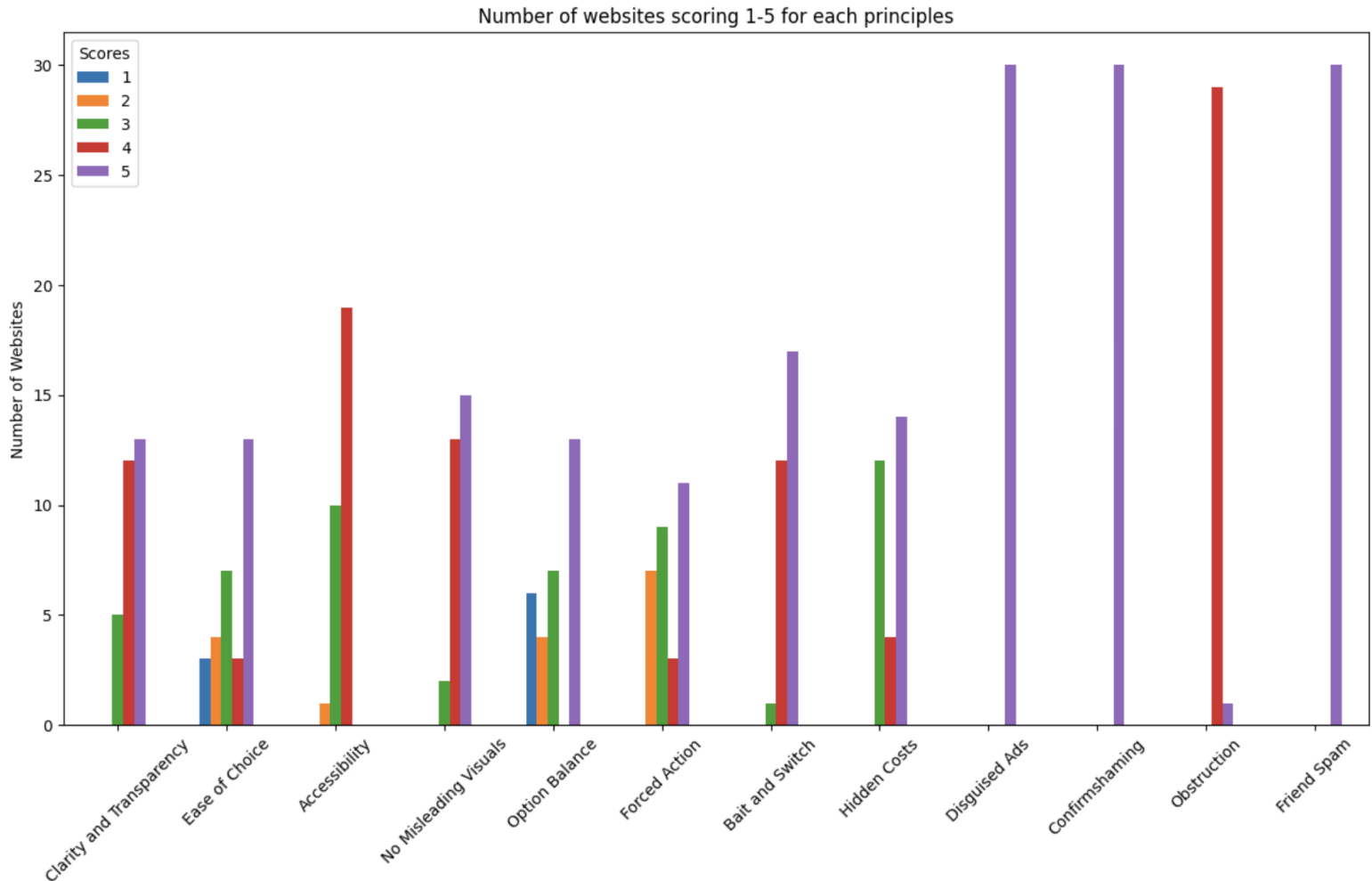
Bar Chart Analysis

The bar chart displays the number of websites that scored from 1 to 5 for each principle, with 1 being the lowest (non-compliance) and 5 being the highest (full compliance). The following observations can be made from the chart:

1. **Clarity and Transparency:** A mixed distribution of scores suggests that while some websites provide clear information, others still employ ambiguous language or provide insufficient details about their cookie usage.
2. **Ease of Choice:** There is a notable lean towards higher scores, indicating that a majority of websites allow users to easily accept or customize their cookie settings, though some still use pre-selected options or have less visible reject buttons.
3. **Accessibility:** The scores are varied, highlighting a disparity in how accessible cookie banners are for users with disabilities.
4. **No Misleading Visuals:** High scores in this principle demonstrate a general avoidance of deceptive designs among the sampled websites.
5. **Option Balance:** The distribution leans towards higher compliance, with many websites presenting accept and reject options without bias.
6. **Forced Action:** The presence of lower scores indicates that some websites still force users into consenting to cookies to use the site fully.

7. **Bait and Switch:** Low scores in this category are concerning, suggesting that some banners are not transparent about the consequences of user choices.
8. **Hidden Costs:** The results are skewed towards higher compliance, showing that most banners do not hide additional costs or selections.
9. **Disguised Ads:** The high scores indicate a general avoidance of disguising ads or opt-ins as other types of content.
10. **Confirmshaming:** The scores reveal that confirm shaming is less prevalent, with most websites avoiding the use of guilt or shame to compel user action.
11. **Obstruction:** Many sites scored well, implying they do not make it difficult for users to reject cookies.
12. **Friend Spam:** The high scores across the board suggest that misleading users into sharing personal information is not a common practice.

Custom GPT Model for Independent Analysis
For those interested in conducting their own analysis, a custom GPT model has been made available at the following link: Cookie Banner Analyst. This tool allows users to evaluate cookie banners against the same principles used in this study, providing an opportunity for broader engagement and verification of the findings.

Interpretation of Results
The results from the custom GPT model provide a comprehensive understanding of the current landscape of cookie consent interfaces. They reveal areas where websites are generally performing well, such as avoiding misleading visuals and disguised ads, and areas where improvement is needed, like clarity and forced actions. These insights can help inform better practices and guide website owners towards more ethical and user-friendly cookie consent strategies.

4. Summary/Lessons Learned

Throughout the "Website Cookie Preference Compliance" project, a number of valuable insights were gained. Here's a summary of the key lessons learned:

The Complexity of Compliance
Compliance with privacy regulations like GDPR and CCPA is not a straightforward task. It requires a nuanced understanding of legal requirements, user rights, and technical capabilities. This project underscored the importance of a comprehensive approach that combines legal theory with practical application.

The Dynamics of User Consent
The project illuminated the delicate balance between legal compliance and user experience. It became clear that obtaining user consent is not just a regulatory requirement but also an exercise in fostering user trust through transparency and empowerment.

Dark Patterns and Ethical Design
Through the analysis of various cookie banners, the concept of 'dark patterns' was explored in depth. The project offered a practical perspective on the ethical considerations designers must navigate to align user interests with business objectives.

Privacy as a User Experience
Privacy was conceptualized not just as a policy or a legal requirement but as a facet of the user experience. This underscored the need for privacy considerations to be woven into the fabric of UX design from the outset.

Adaptive Compliance Strategies
The project revealed the importance of adaptive compliance strategies that can evolve with changing regulations and societal expectations. This concept is pivotal in a digital landscape where static solutions quickly become obsolete.

Transparency Operations
The implementation of transparency as an operational principle stood out as a novel concept. It involves actively ensuring that every aspect of user interaction promotes an understanding of data practices and bolsters user confidence.

Importance of Heuristic and AI Synergy
The initial reliance on heuristic methods, such as the use of z-index values for detecting cookie banners, demonstrated that while useful, such approaches have limitations. The integration of AI, particularly the use of GPT-3.5-Turbo, highlighted the synergy between heuristic methods and machine learning, leading to more accurate and efficient outcomes.

5. References

[1] - "Okay, whatever": An Evaluation of Cookie Consent Interfaces
[2] - CookieEnforcer: Automated Cookie Notice Analysis and Enforcement
[3] - A Systematic Study on the Impact of GDPR Compliance on Organizations
[4] - Thou Shalt Not Reject: Analyzing Accept-Or-Pay Cookie Banners on the Web

6. Links and Deliverables

- Project's Github Link
- Cookie Banner Analyst - Custom GPT for evaluating cookie consent interfaces
- Summary of Research Paper [1]
- Summary of Research Paper [2]
- Summary of Research Paper [3]
- Summary of Research Paper [4]
- Proposal Presentation
- Mid-Term Presentation
- Final Presentation