Radboud University Nijmegen

Faculty of Science

Inside the Cookie Jar: Tracking Consent in Digital Advertising

Assessing the TCF's Impact on Cookie Consent and Adtech Compliance

Thesis BSc Computing Science

Author:
Camile Lendering

Supervisor:
Dr. Ir. Hugo (H.L.) Jonker

Second reader:
Dr. Güneş (M.G.C.) Acar

June 2023

Contents

1 Introduction

Nowadays, European citizens are presented with cookie consent banners on nearly every website they visit. These consent banners, typically facilitated by Consent Management Platforms (CMPs), provide visitors with an often overwhelming variety of choices, including selecting which AdTech vendors are permitted to collect and process the user's personal data and for what purpose. The European Interactive Advertising Bureau (IAB) has developed the Transparency and Consent Framework (TCF) for storing, encoding, and communicating a user's consent choices. The TCF's objective is to ensure that all actors involved in the advertising process adhere to both the General Data Protection Regulation (GDPR) and ePrivacy Directive (ePD) while processing, accessing, and storing user data [G-Bur22]. It is the technology behind many of the consent banners present on European websites today.

The importance of this topic stems from the potential major breach of trust and legality in the digital advertising industry. If CMPs and Adtech vendors fail to comply with user consent preferences, it would suggest that existing systems for safeguarding user data and privacy may not be working properly. This non-compliance could potentially constitute a violation of the GDPR and ePD in Europe, with legal and financial consequences for the companies found to be in breach. Additionally, this would mean that users' personal data is being collected, processed, and possibly shared without their full and informed consent. With the rapid growth of digital advertising and the resulting increase in data collection, it is crucial to understand the effectiveness of consent mechanisms such as the TCF.

Previous studies have examined if Consent Management Platforms (CMPs) registered with the TCF properly communicated the user's consent registered by the CMP's consent banners [MBS20] or web tracking that occurs without a user's explicit consent [PPKM21]. To date, no study has yet been done to investigate the compliance of AdTech vendors registered with IAB Europe's TCF with the user cookie consent provided by the CMP.

This study aims to address this research gap by investigating the compliance of both AdTech vendors and CMPs within the TCF. Our research is driven by the following main research question:

> *To what extent does the TCF ensure compliance of AdTech vendors and Consent Management Platforms registered with IAB Europe in adhering to users' cookie consent preferences in practice?*

To address this main question, we will explore the following sub-questions:

- *To what extent do CMPs registered with IAB Europe comply with the TCF's requirements in terms of obtaining, storing, and communicating user consent?*
- *To what extent do AdTech vendors comply with the communicated cookie consent preferences from CMPs within the TCF framework?*

Contributions. Because of the open-source nature of the TCF, we have the opportunity to conduct the first comprehensive analysis of the compliance of both AdTech vendors and CMPs with user cookie consent preferences. Our main contributions are as follows:

- We devise an automated approach for detecting the implementation of the TCFv2.0 on websites, allowing for efficient identification of websites utilizing the framework (Section 4.1).
- We create an automated method for communicating custom user cookie consent preferences without requiring interaction with the cookie banner (Section 4.2).
- We design an evaluation method to assess the compliance of CMPs by verifying that the consent preferences communicated by the user match the consent preferences communicated by the CMP (Section 4.3).
- We evaluate the compliance of CMPs by verifying that the consent preferences communicated by the user match the consent preferences communicated by the CMP (Section 5).
- We design an evaluation method to assess the compliance of AdTech vendors with user consent preferences communicated by the CMP (Section 6).
- We assess the compliance of AdTech vendors by confirming that the third-party cookies they place align with the consent preferences communicated by the CMP (Section 7).

2 Background

This section first discusses the legal frameworks that govern online privacy, namely the General Data Protection Regulation (GDPR) and the ePrivacy Directive (ePD), and their implications for digital advertising. We then introduce the Interactive Advertising Bureau (IAB) Europe's Transparency and Consent Framework (TCF) as a response to these regulations, explaining its core components and objectives. Finally, we explore the roles and responsibilities of Consent Management Platforms (CMPs) and AdTech vendors in implementing the TCF and ensuring compliance with user consent preferences.

2.1 Legal Basis for Cookie Usage

In the European Union (EU), the legal bases for the use of cookies on websites are primarily outlined in two pieces of legislation: the General Data Protection Regulation (GDPR) and the ePrivacy Directive (ePD).

General Data Protection Regulation (GDPR) The GDPR is the most comprehensive piece of data protection legislation passed by any government agency to date. The GDPR establishes regulations that aim to protect the rights of individuals concerning the processing of their personal data. [W-EC16]. It applies to all organizations, regardless of their location, that process the personal data of individuals within the EU. The regulation also promotes the principles of 'Privacy by Design and Default', stressing that privacy considerations are fundamental in all data processing activities.

GDPR Recital 30 prescribes that, when used to identify a user, cookies are considered **personal data** and are therefore subject to the GDPR. Consequently, the processing of such cookies requires a valid legal basis as specified under the GDPR, namely:

1. **Consent** refers to the data subject[1] explicitly agreeing to the processing of their personal data. Consent must be opt-in, informed, specific, and freely given.
2. **Legitimate interests** form the most flexible lawful basis for processing personal data. Companies can rely on legitimate interests if their interests outweigh the negligible impact on the privacy of the data subjects.
3. **Contractual necessity** applies as a legal basis when processing personal data is required to fulfill a contract with the data subject.
4. **Legal obligation** applies as a legal basis when an organization is required to process personal data to comply with a specific legal obligation under EU law or the law of an EU member state.
5. **Vital Interests** applies as a legal basis when the processing of personal data is necessary to protect the vital interests of a data subject.
6. **Public task** applies as a legal basis when the processing of personal data is required to complete a task that serves the public interest or is part of the official authority granted to the organization.

Veale and Zuiderveen Borgesius concluded [VZ22] that, 'in almost all cases, the data subject's consent is the only available legal basis for personal data processing for RTB[2] and behavioral advertising under data protection law.' This means that privacy by default, which includes obtaining explicit and informed user consent before collecting personal data, is a crucial requirement for CMPs and AdTech vendors.

ePrivacy Directive (ePD) The ePD, commonly referred to as the 'Cookie Law' is a directive issued by the EU that addresses the use of cookies and other tracking technologies [W-02]. It complements the GDPR by providing more specific rules for the use of cookies. The ePD requires website publishers to:

1. Inform users about the use of cookies on their websites, including the purpose of each cookie.
2. Obtain user consent for the use of non-essential cookies, such as those used for advertising or analytics purposes. Essential cookies, like those required for the basic functionality or security of a website, do not require user consent. Essential cookies are required to be anonymous and are not permitted to track browsing activity across other websites.
3. Provide users with the option to withdraw their consent at any time.

In summary, the legal bases for the use of cookies in the EU are outlined in the GDPR and ePD. Website publishers and advertisers must comply with both sets of regulations by obtaining user consent for non-essential cookies, providing clear information about the cookies used, and enabling users to withdraw their consent at any time.

[1] A data subject refers to a person whose personal data is collected, processed, or stored by an organization or entity.

[2] RTB (Real-Time Bidding) is an automated process in online advertising that enables advertisers to bid for the opportunity to display their ads to specific audiences.

2.2 IAB Europe's Transparency and Consent Framework (TCF)

In the web advertising and tracking industry, we distinguish several different actors. The *publishers*, as owners of digital advertising space on their websites, present their websites to *users* (data subjects) and incorporate third-party content provided by *advertisers*, who collect user data and display ads. The GDPR's arrival revealed that the various actors in this ecosystem lacked the necessary tools to adequately gather and share user consent [MBS20].

To address this issue, the IAB Europe created a new category of actors referred to as *Consent Management Providers* (CMPs), responsible for collecting the consent of end users, documenting and storing the obtained consent, and implementing methods to communicate this consent to bidding AdTech vendors. [MBS20].

In an attempt to standardize and regulate the process of obtaining and exchanging user consent regarding data collection and the use of cookies, the TCF was developed. The TCF is promoted as a solution to ensure compliance with GDPR and ePD regulations by all stakeholders involved in the advertising process, including CMPs, vendors, and publishers, in terms of processing, accessing, and storing user data [G-Bur22].

However, it is important to note that the TCF was declared illegal by the Belgian Data Protection Authority in 2020.[3]

In order to participate in the TCF, both CMPs and advertisers are required to register with IAB Europe. The IAB maintains a publicly accessible CMP list, which includes 76 registered CMPs, and a Global Vendor List (GVL), serving as a public record of registered advertisers or "vendors". As of the latest update on March 2nd, 2023, the GVL contains 1185 advertisers.

When an advertiser registers in the GVL, they must indicate one or more of the ten predefined purposes for which data is collected and consent will be utilized (See Table 3 for an overview of cookie purposes). Notice that the listed purposes are distinct from the legal bases for processing personal data under the GDPR. These purposes are typically displayed to the user in the cookie banner interface [W-IABb].

An overview of actors under IAB Europe's TCF ecosystem, as presented in [SNTBR21], can be found in Figure 1 below:

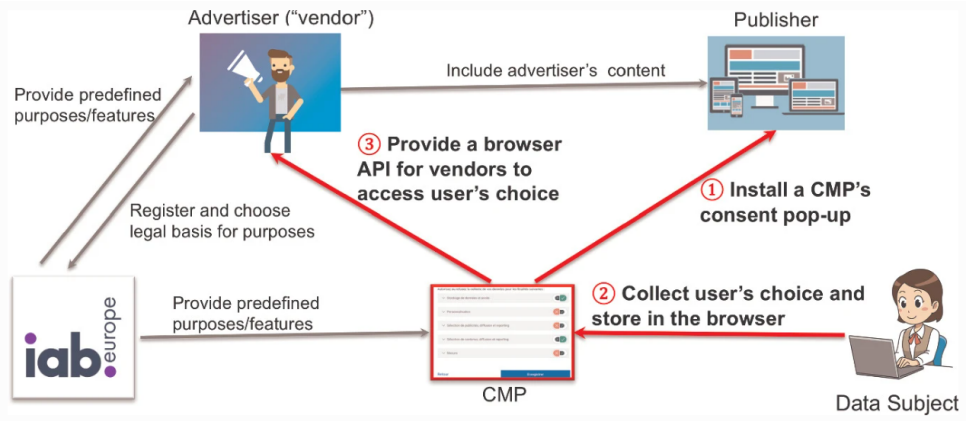

Figure 1: Actors under IAB Europe TCF ecosystem: IAB Europe, Advertisers (called "vendors"), Consent Management Providers (CMPs), Publishers, Data Subjects. Image and caption taken directly from [SNTBR21]

[3]Refer to section 3 for more information regarding the ADP's decision.

2.2.1 Global Vendor List (GVL)

The Global Vendor List (GVL) is a crucial component of the TCF. It serves as a centralized and publicly available registry containing the necessary information about vendors participating in the TCF. The GVL ensures transparency in the digital advertising ecosystem by providing a comprehensive overview of vendors and their data processing activities.

Each vendor listed in the GVL has a unique ID, name, and domain. Moreover, each vendor is required to provide a comprehensive list of purpose IDs for which they collect user data. The GVL also includes details about third-party cookies set by each vendor and their intended purpose(s). This information can be found by visiting the domain specified in the `deviceStorageDisclosureUrl` field of each vendor in the GVL.

To put it simply, any cookie set by a vendor must correspond to one or more purposes listed under that particular vendor's purposes.

Each `deviceStorageDisclosureUrl` contains a JSON file with disclosures related to device storage access and duration, as well as the domains the vendor uses for data processing. It also specifies the storage type -'cookie' for HTTP cookies and 'web' for localStorage and IndexedDB. Furthermore, the file provides the name and purpose(s) of each piece of data stored on the vendor's domain. See the example Device Disclosure JSON object below.

```
{
"disclosures": [
{
"identifier": "i",
"type": "cookie",
"maxAgeSeconds": 31536000,
"cookieRefresh": true,
"domain": ".openx.net",
"purposes": [1,2,7,10]
}
],
"domains": [
{
"domain": ".openx.net",
"use": "Ad serving and cookie syncing with partners."
}
]
}
```

Listing 1: Example device disclosures found on `https://www.openx.com/device-storage.json`

2.2.2 The Consent String

The Transparency and Consent Framework (TCF) specifies a standardized format for CMPs to communicate user consent and legitimate interest preferences to Adtech vendors, known as the *consent string*. The consent string is made up of several components, including:

1. A list of advertisers (vendors) to whom the user has consented to transmit their data.
2. A list of purposes for which the user has given consent for data processing.
3. A list of legitimate interest vendors and the purposes for which they have a legitimate interest to process user data.
4. The CMP identifier and other metadata.

The consent string is encoded as a modified version `base64`. To decode this format, we use a script provided by the IAB [G-Int23]. For example, the consent string `CPq-6wAPq-6wAAHABBENDBCgAAFAAANAAAAAIrwAgDCARXAAAAAA.YCgAAGgAAAAA` obtained on `euronews.com` decodes to:[4]

```
{
  "core": {
    "created": 1682726400000,
    "cmpId": 7,
    "purposeConsents": {
                        "8": true,
                        "10": true
                },
    "vendorConsents": {
      "388": true,
      "1111": true
    },
    "purposeLegitimateInterests": {
                        "7": true,
                        "8": true,
                        "10": true
                },
  "vendorLegitimateInterests": {},
  }
}
```

The `cmpID` identifies the CMP registered with IAB that is responsible for storing and communicating the consent string. The `purposeConsents` correspond to the purposes for data processing that the user has explicitly consented to, and `vendorConsents` identify the vendors on the Global Vendor List (GVL) that the user has explicitly consented to. The `vendorLegitimateInterests` lists vendors that rely on legitimate interest as their legal basis for processing user data, and `purposeLegitimateInterests` indicates the purposes for which they claim legitimate interest [G-Bur22]. Interestingly, this indicates that the consent string does not enable users to assign distinct purposes to different vendors. This highlights a limitation in the granularity of the consent string to capture user consent preferences.

2.2.3 Storage Mechanisms

The TCF$v2.0$ specifications allow a CMP to determine the storage mechanism used for consent strings, including non-cookie storage. For persistent storage, CMPs often use the following methods:

- **Cookies or localStorage:** Typically, CMPs use a cookie or localStorage to store consent on the publisher's domain [MSLH22]. This approach originates from previous versions of the TCF, that required the consent string to be stored in a cookie named `euconsent-v2`.
- **Server-side storage:** CMPs also have the option to store the consent string on their servers rather than client-side.
- **Mobile storage:** For mobile applications, CMPs can use internal data storage, or shared preferences to store consent strings.

The strategy recommended by the IAB is to use server-side storage to retain consent for an extended period and share it across websites while using client-side storage such as cookies or shared preferences to create a local cache that can be quickly accessed [G-Bur22].

However, we believe the recommendation is problematic as it can lead to inconsistencies. Namely, if the consent stored on the server differs from the consent stored client-side,

[4] Only relevant fields are shown.

it can lead to conflicting consent preferences. Additionally, using server-side storage to share consent across different websites raises questions about compliance with the GDPR, which requires explicit, informed, and specific consent.

After a user's consent choices have been saved, whether on the client-side or the server-side, any advertiser on a webpage can access these preferences through the CMP. To accomodate this, CMP's are required to implement standardized API's to facilitate the sharing of user consent data.

2.2.4 Standard APIs for Consent Sharing

The TCFv2.0 specifies an API that each CMP must implement, allowing third-party advertisers to query the CMP for a user's consent preferences on a particular website. Every CMP needs to incorporate a JavaScript function named "`__tcfapi()`" that can be invoked directly by first-party scripts. Additionally, it should include an `iframe` named "`__tcfapiLocator`" that facilitates communication with third-party scripts using the postMessage API. [G-Bur22].

Both API endpoints return a '`TCData`' object containing the encoded and decoded consent string, as well as CMP status information indicating whether a user has finished their interaction with the cookie banner. This is relevant for evaluating Adtech vendor compliance, as a vendor will only respond to an updated consent string if the event status of the CMP updates from '`cmpuishown`' to '`tcloaded`' [G-Bur22].

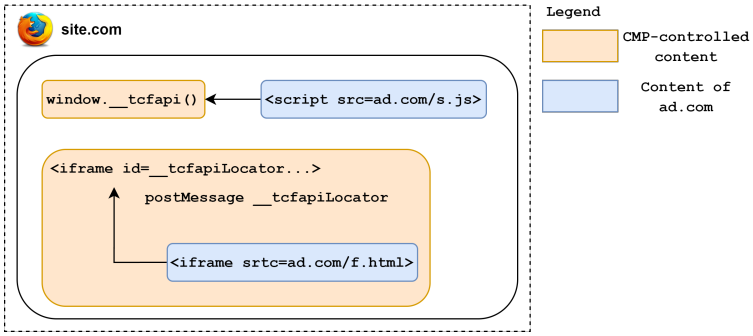

Figure 2: Mechanisms to share consent in IAB Europe's TCFv2.0. (Adapted from [MBS20])

Note that the `__tcfapi()` function cannot be used to set the consent string. Instead, it functions as an interface for Adtech vendors to access a user's consent preferences. The responsibility for generating and setting the consent string lies solely with the CMP. This highlights the fact that manipulating the consent string directly via the `__tcfapi()` is not possible.

3 Related work

The landscape of digital privacy and user consent has become increasingly complex. Evolving regulations and technological advancements within the digital advertising industry have complicated the process of obtaining, processing, and respecting user consent preferences. Many studies have been conducted to understand the various aspects of this issue, forming the basis for further exploration into the extent that the TCF serves

its purpose in ensuring standardization and compliance within the digital advertising ecosystem.

A key area of concern in digital privacy is the use of manipulative design patterns in cookie banners, often referred to as 'dark patterns'. Nouwens *et al.* [NLVKK20] analyzed CMP banner designs of numerous top-ranked websites in the UK and found that dark patterns and implied consent are widespread. The vast majority (88%) of websites they analyzed did not meet the minimum design requirements required by EU law. Compounding this issue, another study by Toth *et al.* [TBR22] revealed that many default consent banners, provided by CMPs, often fail to comply with the law. These findings raise questions about the extent to which users are truly able to exercise informed consent, a concern the IAB aims to address with the TCF.

However, the effectiveness of the TCF in addressing this issue is brought into question by the study conducted by Matte *et al.* [MBS20] revealing inconsistencies in how consent is managed. It found that a significant amount of websites implementing the TCF registered positive consent without explicit user interaction or even against explicit opt-outs.

Another area of concern highlighted by Fouad *et al.* [FSABC20] is the lack of transparency in the use of third-party cookies on websites. They examined over twenty thousand third-party cookies and found that only 12.85% of them had a cookie policy that mentioned the respective cookie. Furthermore, their automatized audit revealed that in 95% of cases, the purposes stated in the cookie policies did not adhere to the purpose 'specification principle'-a key component of the GDPR that mandates organizations to specify their data collection and processing purposes at the time of collection.

The IAB is attempting to tackle these issues through the TCF. However, the effectiveness of these efforts remains under scrutiny. For instance, a study conducted by Matte *et al.* [MSB20] evaluated the data processing purposes declared by all advertisers registered with the TCF. Their findings indicated that numerous advertisers did not adhere to a legal basis compatible with the requirements of the GDPR.

Additionally, a study conducted by Kyi *et al.* [KASRZB23] found that many websites disclosed their data processing for advertising purposes under the justification of 'legitimate interests', a practice that likely violates the GDPR. Notably, in these cases, the cookie banners implemented the TCF, which categorizes various advertising purposes as 'legitimate interests'. This finding aligns with the concerns raised by Hils *et al.* [HWB20], who found that at least 20% of vendors claimed not to require consent for processing personal data for each purpose outlined in the TCF.

These concerns paved the way for the decision by the Belgian Data Protection Authority (APD). The APD's decision states that the TCF failed to establish a legal basis for processing consent strings, and the legal bases proposed by the TCF were inadequate.[5] Multiple studies [RS22; VNS22] argue that the TCF in its current form, is fundamentally incompatible with GDPR requirements, and further state that solving these issues to make the TCF compliant with the GDPR may be impossible.

Despite this, however, the TCF remains the primary mechanism for ensuring regulatory compliance in the digital advertising industry, which makes gaining insights into its actual operation in practice crucial.

In conclusion, the complexity of privacy and user consent in the digital advertising industry, and specifically the role and efficacy of the TCF, has been investigated in numerous

[5]Belgian DPA, Decision on the merits 21/2022 of 2 February 2022, Unofficial Translation from Dutch, Case number DOS-2019-01377, `https://www.gegevensbeschermingsautoriteit.be/publications/beslissingten-gronde-nr.-21-2022-english.pdf` (last accessed: 2023-04-27).

studies. These studies have revealed significant concerns regarding dark patterns, transparency issues, and inconsistencies in consent management. These concerns were further amplified by the decision of the APD, questioning the TCF's fundamental compatibility with GDPR requirements.

Our research aims to add to this discussion by investigating the compliance of both AdTech vendors and CMPs within the TCF. To the best of our knowledge, no previous study has yet assessed the complaince of AdTech vendors registered with IAB Europe's TCF with user cookie consent provided by the CMPs.

4 CMP Compliance Methodology

In this section, we introduce our methodology to evaluate the compliance of Consent Management Platforms (CMPs) with the Transparency and Consent Framework (TCF) guidelines. Our approach consists of three primary objectives:

1. Develop an automated technique to detect the implementation of the TCFv2.0 across a large dataset of websites.
2. Develop a method to automate the process of communicating user consent to the CMP.
3. Develop a semi-automated method for verifying that the consent preferences communicated by the user match the consent preferences communicated by the CMP, thereby streamlining the evaluation of CMP compliance with user preferences.

4.1 Automated Detection of TCFv2.0 Implementation

To detect the presence of the TCFv2.0 API on a large set of domains, we developed an automated methodology that leverages web crawling and JavaScript. Our approach consists of the following steps:

1. **Domain Collection**: We first compile a list of target domains to analyze. This can be obtained from various sources, such as web ranking services (e.g., Tranco Top 1 Million).
2. **Web Crawling**: Using a headless browser or web crawler that supports JavaScript execution (e.g., Selenium), we visit each domain in the collected list. The crawler is configured to load all resources and execute JavaScript code.
3. **JavaScript Function Detection**: During the page load process, we inject a custom script into the browsing context to monitor the global JavaScript environment for the presence of the `__tcfapi` function. If the function is detected, we record the domain as implementing the TCFv2.0 API.

The domains that implement the TCFv2.0 are then used in the next phase of the methodology for CMP compliance evaluation.

4.1.1 Ethical Considerations in Automated Detection

To ensure minimal impact on the domains analyzed during the execution of our automated detection methodology, we implement the following measures:

- **Minimal intrusion:** The web crawler only visits the landing page of each domain and does not navigate further into the website or interfere with its normal operation. This limits the potential load placed on the servers of the domains.
- **Limited interaction:** If a domain requires login or presented a CAPTCHA, we exclude it from our analysis.

This experiment is conducted solely for research purposes, with the goal of understanding the extent of TCFv2.0 implementation across a large set of domains.

4.2 Automating Communication of User Consent Preferences

Manually filling out cookie dialogues, for each website that implements the TCF limits the scalability of our method and introduces the risk of user error. In this section, we evaluate methods for automatically detecting and filling out cookie dialog banners according to predefined consent preferences. This approach is not only relevant for assessing CMP compliance with the TCF guidelines but also plays a crucial role in assessing vendor compliance with the TCF guidelines. By automating consent communication, we can effectively gauge how well vendors adhere to user preferences and comply with the TCF policy.

4.2.1 Browser extensions

There are several browser extensions available that attempt to automatically handle cookie consent banners. The vast majority of existing extensions [W-Goo; W-Dan; W-Cen; W-Ben; W-Rep] detect and block the cookie banner from interrupting the user's browsing experience. This is done by either simply deleting the `iframe` that contains the consent pop-up, or automatically accepting the cookie policy. These extensions are not useful in the context of this study, as they do not provide insight into the compliance of CMPs with the TCF guidelines. Rather than simply removing or accepting the consent banners, our focus is on evaluating the implementation of TCF and verifying the accurate communication of user consent preferences through the CMPs. The Consent-O-Matic browser plugin [W-rol], developed by Nouwens *et al.* [NBKK22] does automatically detect and fill out cookie dialog banners according to predefined consent preferences. Data processing purposes are combined into five categories that can be toggled on or off.

While automating cookie consent with Consent-O-Matic may seem convenient, it would significantly narrow the scope of our research. Currently, there are 76 CMPs registered with the TCF as of February 19, 2023, but Consent-O-Matic only supports 37 CMP banners, and only 11 of them are actively registered with the TCF. This limitation means that we would miss out on data from the publishers implementing CMPs that are not supported by Consent-O-Matic. Additionally, Consent-O-Matic lacks the fine-grained preferences required to automatically deny consent to some or all AdTech vendors. This eliminates our ability to study the behavior of vendors when consent is only given to a select set of vendors.

Finally, we consider the Autoconsent tool [G-Sam], developed for the (discontinued) Cliqz browser. Autoconsent is a library containing rules to automatically fill out consent banners on the web. Using these rules, opt-in and opt-out preferences are selected automatically, without requiring user interaction. This tool offers more flexibility than Consent-O-Matic when defining custom rulesets. As an example, we created a custom

rule for the CMP present on `https://euronews.com` that automatically rejects tracking from all vendors in their vendor list:

```
{
    "name": "Didomi",
    "detectCmp": [{ "exists": "div[class^=didomi-popup-view]" }],
    "detectPopup": [{ "exists": "div[class^=didomi-popup-view]" }],
    "optOut": [
        {"click": "button[didomi-notice-view-partners-link]"},
        {"wait": "1000"},
        {"click": "button[didomi-components-radio__option didomi-components-radio__option--
            unselected]"},
        {"click": "button[didomi-components-button didomi-button didomi-components-button--
            color didomi-button-highlight highlight-button]"},
        {"wait": "1000"},
        {"click": "button[didomi-components-button didomi-button didomi-button-standard
            standard-button]"}
    ],
}
```

Unfortunately, there is no guarantee that the above rule is applicable to all publishers implementing the Didomi CMP because CMPs allow publishers to customize the consent notice. Hence, one CMP may need many rules to cover all possible configurations. Furthermore, separate partial consent rules would need to be made for each experimental condition.

In conclusion, browser extensions are not the ideal solution for assessing CMP and vendor compliance with TCF policy. Existing extensions primarily block or automatically accept consent banners, rather than allowing for the fine-grained control required for assessing both CMP and AdTech vendor compliance witn the TCF policy.

The Consent-O-Matic plugin and Autoconsent tool offer automation of consent preferences but come with limitations in terms of supported CMP banners and applicability to various publisher configurations. These constraints make relying solely on browser extensions insufficient for studying compliance with TCF policy. A more comprehensive methodology is necessary for a thorough assessment.

4.2.2 Leveraging the TCF

As mentioned in Section 2.2, the TCF establishes a standardized format for communicating user consent, known as the consent string. This implies that automating cookie consent on domains implementing the TCF requires us to be able to encode valid consent strings with custom consent preferences. To achieve this, We will make use of the `iab-tcf-v2` client library to read and encode IAB TCF V2.0 consent strings [G-Rem23]. This library enables us to define a custom `TCData` object containing all relevant properties related to user consent preferences, and then encode the object into a valid consent string using the `Encode()` function provided by the library.

In this way, we can automate the communication of user consent preferences to the CMP by creating and encoding a valid TCData object outlining all of the user's consent preferences. We can then inject this valid consent string into the storage mechanism utilized by the CMP (either a cookie or localStorage). Note that injecting consent server-side is not possible. We are only able to manipulate the contents of client-side storage mechanisms.

This approach allows for a more comprehensive evaluation of CMP and vendor compliance with the TCF policy, overcoming the limitations posed by browser extensions.

4.3 Evaluating CMP Compliance

To assess the compliance of Consent Management Platforms (CMPs) with the TCFv2.0 standard, we designed an automated method that examines how domains handle user

consent. Our approach involves the generation and injection of a custom consent string, observing the CMP's behavior in response to the injected consent, and categorizing domains based on their compatibility with this injection method. The categorization helps us identify potential non-compliant CMPs that may require further investigation.

The key steps in this methodology are:

1. **Consent String Injection**: We generate a valid custom consent string and inject it into the domain's storage, such as cookies or localStorage. This simulates a user's interaction with the consent banner.
2. **CMP Response Analysis**: We analyze the CMP's behavior following the injection of the consent string. This analysis involves two steps:
   (a) **Consent String Verification**: We verify whether the consent string retrieved through the CMP API call[6] matches the one stored in the domain's storage. This check ensures that the CMP accurately relayed the injected consent string.
   (b) **Display Status Updates**: We check if the CMP's display status updates to reflect user interaction. This typically manifests as the disappearance of the cookie banner, suggesting the CMP has registered and acknowledged the user's consent choices. Importantly, this status can be queried through the `__tcfapi()` function, providing us with an automated way to validate the CMP's response.

Based on these observations, we classify domains into four categories:

Category	Stores User Consent Accurately	Updates CMP's Display Status
0	No	No
1	Yes	Yes
2	Yes	No
3	No	Yes

Table 1: Domain categorization based on compatibility with the consent string injection method

Domains that fall into category 0, likely use a different storage mechanism, such as server-side storage, that is not susceptible to the consent string injection method. Conversely, domains in Category 1 are fully compatible with the consent string injection method and demonstrate proper handling of user consent. However, domains in Category 2 are only partially susceptible to our method, as the cookie banner persists even after injection. Lastly, Domains in Category 3 are likely non-compliant with the TCF policy and warrant further manual analysis.

By categorizing domains in this way, we can effectively evaluate each domain's compatibility with the custom consent string injection method. Furthermore, this classification system aids in the identification and analysis of non-compliant CMPs.

[6]The `tcfapi()` call, to be exact.

5 Experiment: Evaluating CMP Compliance

In this section, we discuss the outcomes of our evaluations for TCF availability and CMP compliance using the automated methodology described in Section 4. These findings provide insights into the extent of compliance with TCF guidelines across various CMPs, shedding light on possible instances of TCF policy violations.

5.1 Sub-Experiment 1: TCFv2.0 Implementation Detection

We developed a script [G-CLe23] that checks TCFv2.0 availability on the top one million domains using the Tranco list of top domains, accessed on March 1st, 2023[7]. For each domain on the list, we first use the requests library to check if the domain is reachable. If the domain is reachable, we use a headless Selenium web driver to navigate to the domain and to check if the `__tcfapi()` function is available.

5.1.1 Results:

We checked TCF availability on the Tranco top 99, 569 domains, of which 16, 539 (16.6%) timed out after one minute. Of the remaining 83, 030 domains, 4, 460 (5.4%) implement the TCFv2.0. The prevalence of the TCF API was most notable on the `.com`, `.de`, and `.uk` top-level domains, as shown in Figure 3 below.

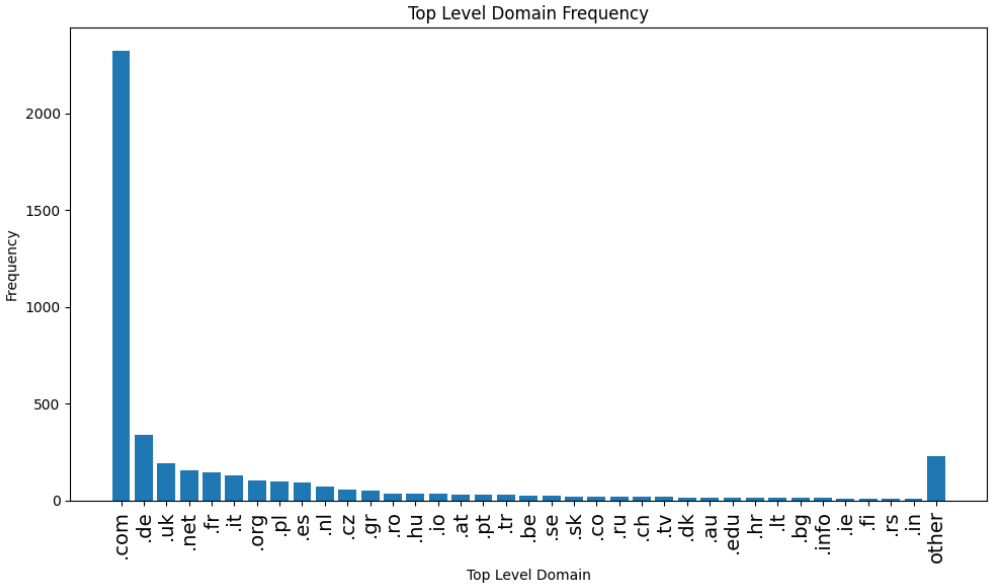

Figure 3: TCF availability vs. Top Level Domain

5.2 Sub-Experiment 2: CMP Compliance Evaluation

In this section, we provide the implementation details of our CMP compliance assessment method, which evaluates the compliance of domains with the TCFv2.0 policy.

We developed a script [G-Len23a] consisting of the following components:

[7]Tranco list id: 'X57KN'.

1. **Domain List and Browser Automation**: We utilize the list of domains implementing TCFv2.0, obtained from the output of the automatic detection method described in section 5.1. Using Selenium to navigate to each domain in a Chrome browser.

2. **Custom Consent String Generation**: We generate and encode a custom, valid consent string using a TCFv2.0 consent string encoding library [G-Rem23].

3. **Consent String Injection**: We inject the generated custom consent string into the domain's storage by setting two key-value pairs as cookies and in localStorage:

   - `euconsent-v2: [generated consent string]`
   - `eupubconsent-v2: [generated consent string]`

   Note that while the above cookie names were previously mandated by the TCF specifications, they are no longer required by any formal standard. However, they have been widely adopted by CMPs as part of their TCF implementation, ensuring compatibility and consistency across different actors within the TCF.

4. **Consent String and CMP Display Status Retrieval**: We use JavaScript code to retrieve the consent string through the `__tcfapi()` call and the CMP's display status. We store this information for comparison after the page reloads.

5. **Page Reload**: We reload the domain's page, simulating a user's revisiting the website with the injected consent string.

6. **Post-reload Analysis**: After the page reloads, we again retrieve the consent string and CMP display status using JavaScript code. We then compare this information with the previously stored data to evaluate the domain's compatibility with the consent string injection method.

7. **Domain Categorization**: Based on the comparison results, we categorize the domains into one of the four categories described in section 4.3. This classification helps us assess the level of CMP compliance and identify potential non-compliant domains.

By following these implementation steps, we can effectively evaluate the compliance of domains with the TCFv2.0 standard using the custom consent string injection method. This allows us to identify and analyze non-compliant CMPs.

5.2.1 Results

We ran the script described in section 5.2 on the $4,460$ domains we found implementing the TCFv2.0, of which, 426 domains timed out after 30 seconds. The remaining $4,034$ domains were categorized as follows: $2,562$ (64%) fall into category 0; 372 (9%) fall into category 1; $1,017$ (25%) fall into category 2; and 83 (2%) fall into category 3. These results are summarised in Figure 4 below.

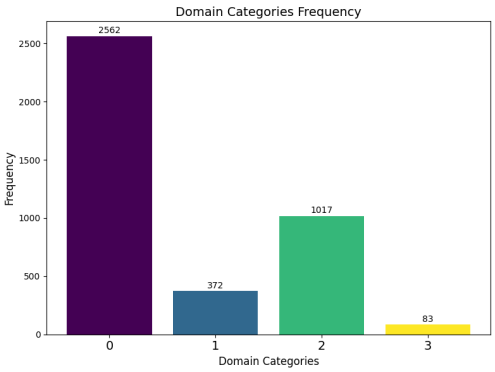

Figure 4: Domain categories and their frequencies

In our manual analysis of domains in Category 3, where we observed that injecting the consent string into the domain storage causes the cookie banner to disappear. However, the consent string communicated by the CMP did not match the injected consent string. We discovered that 10 domains, all using the Seznam.cz, a.s. CMP (cmpId: 247), send an 'accept all'[8] consent signal before users can interact with the cookie banner.

Additionally, we discovered 3 domains (cmpId's 28, 279) that transmitted a modified consent string, consenting to various legitimate interest vendors and purposes. These behaviors potentially violate both the Transparency and Consent Framework (TCF) policy and the General Data Protection Regulation (GDPR).

Finally, we discovered that on 58 domains, the `__tcfapi()` call returned an empty string or nil. Moreover, 8 domains had misconfigured CMPs where the mandatory `ping` API call was not properly defined, making it impossible to query the corresponding cmpId. Lastly, on the remaining 4 domains, the CMP version returned by the `PingReturn` object did not match the actual CMP's own internal version. These findings are summarised in Table 2 below.

Findings	**Frequency**	**cmpId's**
'accept all' consent string	10	[247]
Legitimate interest consent string	3	[28,279]
Empty or nil consent string	58	[5,68,269,300,46,21,31,305]
Misconfigured PingReturn objects	4	[28,47,236]
Undefined ping API call	8	-

Table 2: Summary of findings in the analysis of domains in category 3 and their CMPs

In our manual analysis of domains in category 0, where the consent string retrieved through the `__tcfapi()` call does not match the one stored in the domain's storage and the displayStatus of the CMP remains unchanged, we made the following observations:

1. We identified one domain using the Usercentrics GmbH CMP (CMPId: 5) that sends an 'accept all' consent signal before users can interact with the cookie banner.

2. We observed 121 domains utilizing the Osano, Inc. CMP (cmpId: 279) that sends a consent string that grants consent to specific vendors prior to user interaction with the CMP.

[8]Refer to section A.4 for a description of the 'accept all' consent string.

5.3 Validity

In this section, we discuss the validity of the CMP compliance methodology presented in Section 4. We assess the reliability and accuracy of our method in detecting TCFv2.0 implementation, automating communication of user consent preferences, and evaluating CMP compliance. Furthermore, we identify potential limitations and areas for future improvement.

5.3.1 Validity of Automated Detection of TCFv2.0 Implementation

Our automated method for detecting the implementation of TCFv2.0 relies on web crawling and JavaScript injection to detect the presence of the `__tcfapi` function. This approach provides a highly scalable and accurate means of determining TCFv2.0 implementation on a large number of websites. However, 16.6% of domains analyzed were not accessible due to factors such as domain expiration, or connection timeouts. This limitation may result in a (slight) underestimation of the number of domains implementing the TCFv2.0, as well as the extent of CMP compliance across the sampled domains.

Our method relies on the assumption that the `__tacfapi()` function is defined if and only if a website has implemented the TCF$v2.0$. This assumption allows us to use the presence of the `__tacfapi()` function as an indicator of TCF v2.0 implementation on a website.

However, it is important to acknowledge that this approach may lead to false negatives. While it is uncommon, there is a possibility that certain CMPs have their own methods of implementing TCF$v2.0$ that do not involve defining the `__tacfapi()` function. Although the TCF specifications do require the implementation of `__tacfapi()`, there may be rare cases where alternative approaches are used.

5.3.2 Validity of Automating Communication of User Consent Preferences

The methodology presented for automating the communication of user consent preferences relies on the CMP storing consent in the browser under a specific name. This limitation arises due to the following factors:

- Storage mechanism: The consent string injection method assumes that the CMP uses browser storage (such as cookies or localStorage) to save user consent information. However, not all CMPs use the same storage mechanism. Some CMPs could employ server-side storage, rendering the injection method ineffective for these cases.
- Consent string naming: The method also assumes that the CMP stores the consent string under a specific, known name. If the CMP uses a different or obfuscated name, the injection method is rendered ineffective.

  As mentioned, previous versions of the TCF mandated specific naming requirements for storing consent strings. However, these naming requirements have been lifted in more recent versions of the TCF. As a result, CMPs now have the flexibility to choose their own naming conventions for storing consent strings, which may differ from the previously mandated naming conventions.

  Therefore, it is possible that some CMPs switched to alternative naming conventions for storing consent strings. This variability in naming conventions adds a layer of complexity to the injection method, as the specific name under which the

consent string is stored needs to be known in order for the injection method to function effectively.

- User interaction detection: Some CMPs only update the consent string in response to user interactions with the consent banner. In such cases, the CMP typically relies on user actions, such as clicking on the accept button within the CMP banner, to trigger updating the consent string.

5.3.3 Validity of Automating the Evaluation of CMP Compliance

The automated method for evaluating CMP compliance with the TCFv2.0 standard involves generating and injecting a custom consent string, observing the CMP's behavior in response to the injected consent, and categorizing domains based on their compatibility with the injection method. This categorization helps identify potential non-compliant CMPs that may require further investigation.

Our method is effective in assessing CMP compliance and identifying non-compliant instances. However, it has some limitations:

- The classification system may not cover all possible scenarios of non-compliance. There might be other subtle ways in which CMPs fail to comply with TCFv2.0 policy that our method does not capture.
- While the automated method does effectively filter out compliant domains, classifying the instances involving potential non-compliance still requires manual intervention.

Despite these limitations, our methodology provides a solid foundation for evaluating CMP compliance with TCF guidelines. Future work could focus on refining the categorization system and exploring alternative methods to improve the coverage of CMP compliance evaluation.

5.4 Analysis

The methodology presented in section 5.1 succeeded in detecting TCFv2.0 implementations across a large dataset of websites and automating the communication of user consent preferences. Our evaluation of CMP compliance revealed a varying degree of adherence to the TCFv2.0 policy. The categorization of domains based on their compatibility with the custom consent string injection method was shown to help identify potentially non-compliant CMPs.

The results from the experiments indicate that a significant number of domains (64%) fall into category 0, where the consent string retrieved through the CMP API call does not match the one stored in the domain's storage, and the CMPs display status remains unchanged. Although this category does not necessarily imply non-compliance, it highlights potential limitations of the consent string injection method and the CMPs' handling of user consent.

A small percentage of domains (2%) were classified as potentially non-compliant (category 3) and had discrepancies between the injected consent string and the one communicated by the CMP. Manual analysis of these domains revealed TCF policy violations such as sending an "accept all" consent signal or pre-selecting certain purpose and vendor consents before any user interaction. These results are significant as they indicate non-compliance with TCFv2.0 policies and potentially the GDPR.

Our findings provide valuable insights into the current state of CMP compliance across various websites, offering a starting point for further investigation and potential regulatory action. This methodology also serves as a basis for future research on automating CMP compliance assessments.

6 AdTech Vendor Compliance Methodology

We previously established a method to programmatically create custom valid consent strings with predefined advertiser and purpose consents, and inject them into websites. A portion of these websites would then correctly transmit the consent string to advertising vendors. In this section, we will build on this approach to check if vendors comply with the consent string. ‘

6.1 Defining vendor compliance

For the purposes of this study, we define a vendor to be *compliant* with a consent string if and only if:

1. The vendor only places third-party cookies on a domain if they have obtained explicit consent through the `VendorsConsents` field in the consent string or they have a legitimate interest as specified in the `vendorLegitimateInterests` field. For explicit consent, there must be an entry with the vendor's ID and a value of "true" in `VendorsConsents` in order for them to set third-party cookies on the domain. For legitimate interest, the vendor's ID and a value of "true" should be listed in the `vendorLegitimateInterests` field.

   AND

2. The vendor only sets third-party cookies on a domain if they have obtained explicit consent for the cookie's intended purpose through the `PurposesConsent` field of the consent string or the purpose aligns with a legitimate interest as specified in the `purposeLegitimateInterests` field. If consent is required, there must be an entry with the cookie's purpose ID('s) and a value of "true" in `PurposesConsent` for the vendor to set the cookie on the domain. If the cookie's purpose aligns with a legitimate interest, this purpose ID and a value of "true" should be listed in the `purposeLegitimateInterests` field.

Note that vendors relying solely on legitimate interests are not permitted to set cookies that require explicit consent as defined in the `PurposesConsent` field. Explicit consent must be obtained in these scenarios.

An overview of cookie purpose ID's and their descriptions can be found in Table 3 below.

Purpose ID	Description
1	Store and/or access information on a device.
2	Select basic ads.
3	Create a personalized ads profile.
4	Select personalized ads.
5	Create a personalized content profile.
6	Select personalized content.
7	Measure ad performance.
8	Measure content performance.
9	Apply market research to generate audience insights.
10	Develop and improve products.

Table 3: Cookie purpose IDs and their descriptions [W-IABb]

6.2 Assessing AdTech Vendor Compliance

In this section, we outline our method to assess vendors' compliance with user cookie consent preferences.

In our previous analysis, we identified domains that accurately transmit injected consent strings from their localStorage or cookies to the CMP. Among these domains, we specifically focused on those where the CMP successfully registers user interaction with the consent banner. This interaction triggers the hiding of the banner user interface (UI) after injecting the consent string and effectively signaling the updated consent string to bidding vendors. Thus, these domains are suitable for assessing vendor compliance with the injected consent string.

The key steps in this methodology are:

1. Using a set of eligible domains, injecting varying levels of consent via the consent string:
    - Reject all Vendors, reject all purposes.
    - Accept basic ads.
    - Accept all vendors, reject all purposes.
    - Accept all purposes, reject all vendors
    - Accept all purposes and vendors.
2. Intercepting and storing all third-party cookies set by vendors on each domain.
3. Classifying each third-party cookie by cross-referencing the GVL.[9]
4. Assessing vendor compliance by:
    - Checking if each vendor had obtained explicit consent to set its third-party cookie(s) through the `VendorsConsents` and `vendorLegitimateInterests` fields in the consent string.
    - Checking each vendor had obtained explicit consent for its third-party cookies(s) intended purpose(s) through the `PurposesConsent` and `purposeLegitimateInterests` fields of the consent string.

[9] As was mentioned in section 2.2.1, the GVL contains the names, domains, and purposes of all third-party cookies set by each vendor.

7 Experiment: Evaluating Vendor Compliance

In this section, we outline the experimental setup for assessing vendor compliance with the consent string, the process of extracting and classifying third-party cookies, and the results of our experiment.

7.1 Experiment Setup

To test vendor compliance with the consent string, we used the 372 category 1 domains identified in section 5.2.1[10], that accurately transmit injected consent strings from their localStorage to bidding vendors via the `__tcfapi()` call.

7.1.1 Extracting Third-Party Cookies

To inject a consent string into a domain and subsequently intercept and store all third-party cookies set by vendors on the domain, we wrote a Go script [G-Len23c] that implements a web crawler to extract third-party cookies from websites. The script performs the following tasks:

1. Reads the domain names of the 372 websites identified in Section 5.2.1 from a CSV file.
2. Visits each domain using a web browser controlled by the `Chromedp` [G-Ken23] package. This is an API for controlling and interacting with the Chrome browser using the Chrome DevTools Protocol.
3. Sets a valid consent string with predefined consent preferences for each domain by simulating user interaction with the website, generating a valid consent string for the CMP, and saving it in a cookie and localStorage on the domain. The consent string is generated using the `IAB TCFv2` library.
4. Uses a proxy server (implemented using the `GoProxy` package [G-Lei23]) to intercept and modify HTTP requests and responses. This is needed to extract third-party cookies set by vendors, as the `Chromedp` package can only access first-party cookies.
5. Adds cookies to subsequent requests and extracts cookies from the domain using the modified responses. By adding the captured cookies to the subsequent requests, the script can maintain the continuity of the browsing session before and after injecting the consent string and capture any additional cookies that may be set.
6. Writes extracted cookies to a CSV file.

7.1.2 Classifying Third-Party Cookies Using the GVL

To classify the purpose of the intercepted third-party cookies set for a particular consent string, we wrote Go scripts [G-Len23b] to:

1. Obtain relevant vendor information for every vendor listed in the Global Vendor List (GVL), including vendor name, vendor ID, vendor purposes, and device disclosures URL.

[10]We observed that the domains that updated their `displayStatus` also correctly updated their `eventStatus` signaling the updated consent string to bidding vendors.

2. Extract and interpret data from the JSON file found at the device disclosures URL, including cookie domains, cookie names, cookie purposes, vendor domains, and vendor uses.

3. Compile the gathered information and save it in a CSV file.

4. Cross-reference the third-party cookies found on each domain with the information extracted from the GVL by comparing the cookie's domain and name. The following steps were taken:

   (a) If an entry in the GVL data matches both the cookie's domain and name, record the website, vendor name, vendor purposes, cookie name, cookie domain, and cookie purposes in a CSV file.

   (b) If an entry in the GVL matches only the cookie's domain, save the same information as in the previous case but exclude the specific cookie purposes, as this information is not available in the GVL.

   (c) If no match is found in the GVL, simply document the unmatched cookie in the CSV file.

From this point forward, we will refer to cookies from 4a as **matched** cookies, cookies from 4b as **partial-match** cookies, and cookies from 4c as **unmatched** cookies.

Note: Not all matched cookies set by third parties necessarily contribute to user identification. There are certain cookies, like test cookies, which are not used for identifying individual users but rather for ensuring the correct functionality of the website. Hence, while these cookies may fall under the category of "matched" due to their presence in the GVL, their existence does not necessarily indicate a user privacy concern.

7.2 Validity

The validity of our experiment to assess vendor compliance with the consent string is based on several factors:

- Injecting varying levels of consent enables us to test how vendors respond to different consent signals. This provides a fine-grained understanding of vendor compliance.

- Intercepting and storing third-party cookies allows for detailed examination of the cookies each vendor sets under different consent scenarios. This is key to understanding the extent that vendors respect user consent preferences when setting cookies.

- Cross-referencing the Global Vendor List (GVL) to classify the intercepted third-party cookies not only allows us to identify the purposes of each cookie and the vendor associated with them but also enables us to assess the transparency of vendors by checking if the GVL comprehensively lists all the cookies they use.

- The set of domains used in our experiment accurately communicate the injected consent string and trigger the correct behavior in the CMP's callback functions. This ensures that an event is dispatched to all bidding vendors whenever the consent string is updated.

However, our method does have a few limitations:

- The method relies on the accuracy and completeness of the device disclosures provided by the different vendors, which may not always contain the most up-to-date or comprehensive information about the cookies they use and their purpose. Furthermore, we encountered issues with some vendors' device disclosures containing malformed JSON data, which posed challenges to automated parsing.
- The method can only be used on domains that are susceptible to the consent string injection method outlined in section 5.2, limiting the scope of our assessment.
- Our method, while efficient at capturing cookies set through HTTP headers, might overlook those set via JavaScript. Given that we use a proxy to intercept and analyze cookies, First-party cookies, which are created client-side, may go undetected in some instances. This limitation implies that our results might underestimate the number of cookies set under different consent scenarios.

Despite these limitations, our method provides a systematic and comprehensive approach to assessing vendor compliance with user cookie consent preferences.

7.3 Results

In this section, we present the results of our experiment assessing vendor compliance with injected consent strings.

7.3.1 Results Given the 'Reject All' Consent String

The *reject all* consent string corresponds to a consent string where a user chooses not to allow any data collection or sharing with third parties. This means:

1. The user does not give consent to any vendors.
2. The user does not allow any purposes or legitimate interests for data processing.
3. The user does not give consent to any publisher-specific purposes or legitimate interests.

In short, the user is not permitting any data processing or sharing in this scenario.

We observed **1,562 third-party cookies across 372 websites** using the method described in 6. These cookies were categorized into matched (51.1%), unmatched (30.8%), and partial-match (18.2%) based on cross-referencing with the GVL. The results are summarized in Figure 5.

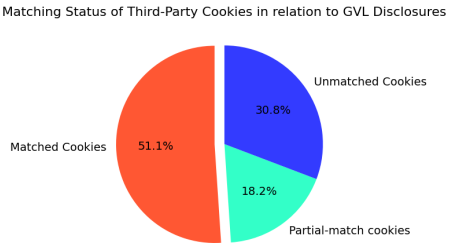

Figure 5: Percentages of the 1,562 third-party cookies that exactly matched, partially matched, and did not match the cookies disclosed in the GVL.

Analysis of matched cookies Figure 6[11] presents the distribution of cookie purposes for each vendor setting third-party cookies on the analyzed domains. Zoom Ltd. was the most prominent vendor, accounting for 26.4% of all matched third-party cookies. Revcontent, LLC represented another significant portion, contributing 18.3% of all third-party cookies. The remaining half of the 798 matched cookies were set by various vendors for a range of purposes.

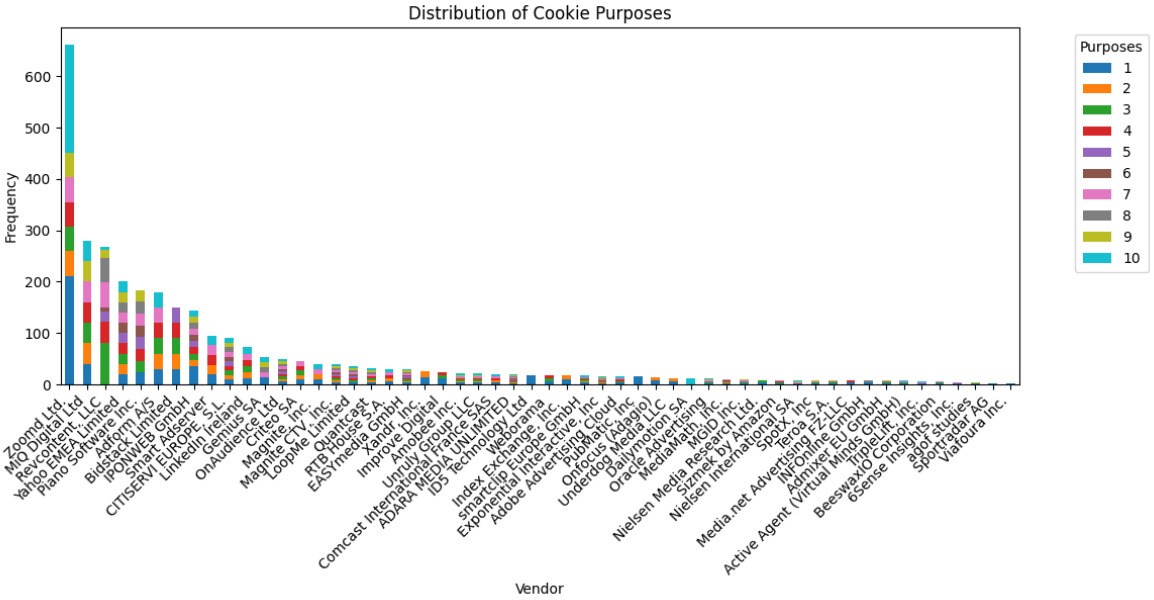

Figure 6: This stacked bar chart illustrates the Frequency of third-party cookies and their purposes for each vendor.

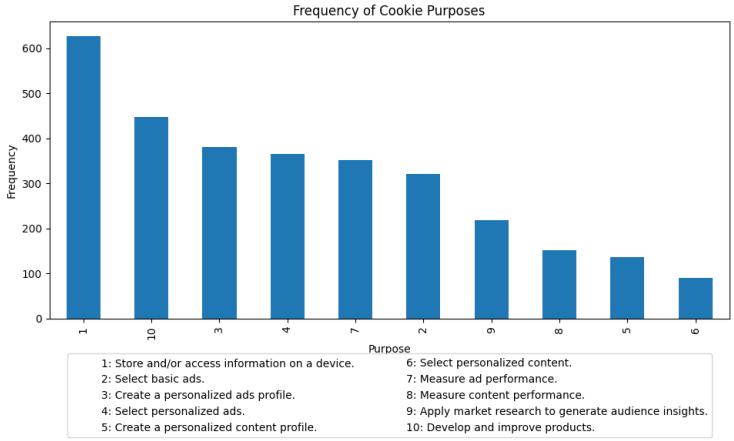

Figure 7: This bar plot shows the Frequency of each cookie purpose category in the matched third-party cookies.

[11]To improve the readability of the plot, we devised '**pseudo-cookies**'. Each of these represents a single purpose derived from each observed multi-purpose cookie.

In Figure 7, the frequency of each purpose category is displayed for the matched third-party cookies. Purpose 1, storage and/or access of information on a device, was the most common purpose, accounting for 79% of all matched cookies. Purpose 10, related to the development and improvement of products, is the second most frequent purpose, with 56.1% of the cookies set for this purpose. Purposes 3 and 4, which are associated with creating and selecting personalized ads profiles, respectively, are the third and fourth most common purposes, accounting for 47.6% and 45.7% of the matched cookies.

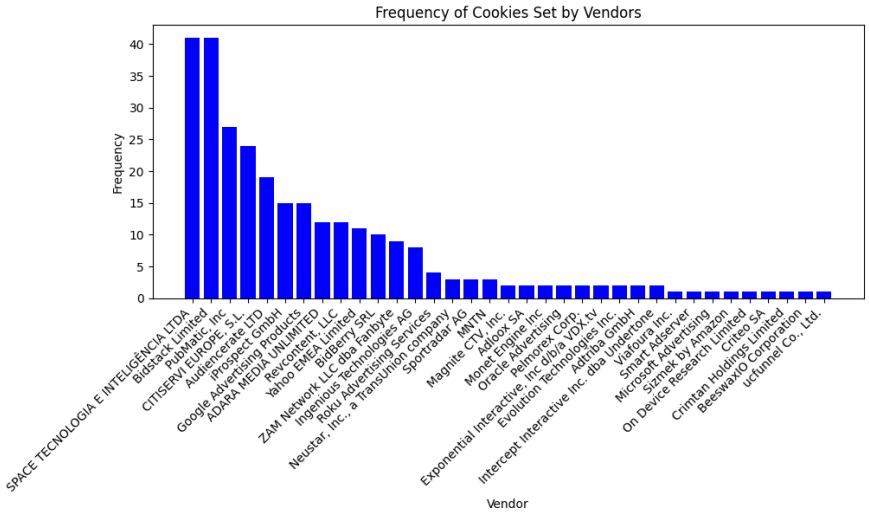

Figure 8: This bar plot shows the Frequency of each vendor that set a partial-match third-party cookie.

Analysis of partial-match cookies In Figure 8 above, we can see that 36 different vendors set partial-match cookies across the analyzed domains. The top four most prominent vendors account for 46.8% of all the partial-match cookies, while the remaining 32 vendors each contribute to at most 6.7% of the partial-match cookies observed.

We have visualized the disclosed vendor purposes[12] in Figure 9 below.

[12]The vendor purposes specified in the Global Vendor List (GVL) specify *all* the intended uses of collected data by a vendor, while cookie purposes define the *specific* purpose(s) for which a cookie is set on a website. Hence, cookie purposes $\subseteq$ vendor purposes.

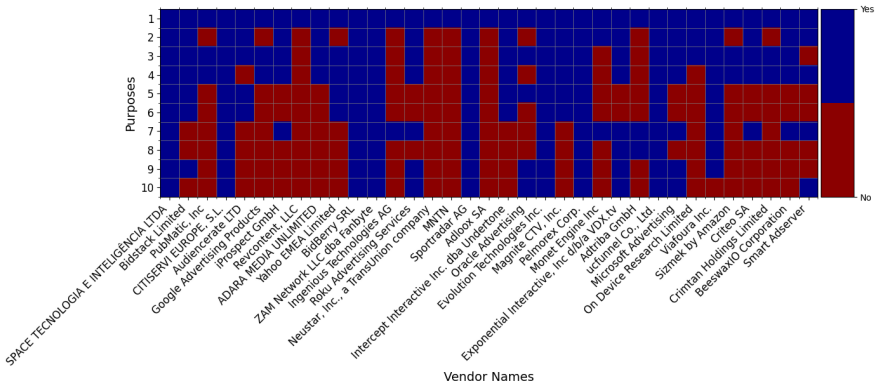

Figure 9: A matrix-plot of partial-match cookie vendors and their disclosed vendor purposes.

Combining the results from Figure 8 and 9 above, we observe that all cookies can be set for accessing and storing information on a device (purpose 1). Approximately one-third (32.4%) of the partial-match cookies can be set for any purpose(s) and the vast majority (80%) of partial-match cookies can be set for creating and selecting personalized ads profiles (purposes 3 and 4 respectively).

No-match cookies were excluded from the discussion, as they are not listed in the GVL and cannot be classified, thus falling outside the scope of this study.

7.3.2 Results Given the 'Accept Basic Ads' Consent String

In our observations of third-party cookies using the 'Accept Basic Ads' consent string, we found a total of 3,590 cookies across 372 websites, with an average of 9 cookies per site. When cross-referencing these cookies with the GVL, we found that 56.8% were matched, 18.5% were unmatched, and 24.7% were partial matches.

For matched cookies, ID5 Technology Ltd. was the most prominent vendor, accounting for 20.5% of all matched third-party cookies. The vendor exclusively sets functional cookies on the `id5-sync.com domain`. Zoomd Ltd. was another prominent vendor, contributing 11.8% of all third-party cookies, while Revcontent, LLC. accounted for 8.9% of all matched third-party cookies. Notably, only 40.4% of the matched cookies had purpose(s) 1 (device storage) and/or 2 (select basic ads), and 80% of vendors set matched cookies that were not compliant with the 'allow basic ads' consent string.

Regarding partial-match cookies, we discovered that 45 different vendors set cookies across the analyzed domains. More than half (56%) of these cookies could potentially be set for purpose 2. Based on the disclosed purposes of each vendor, we deduced that the partial-match cookies' purposes were consistent with the 'accept basic ads' consent string, and no compliance concerns were observed.

For a comprehensive report (similar to section 7.3.1), including figures and charts illustrating the distribution of cookie purposes, the frequency of each vendor setting partial-match cookies, and the matrix plot of partial-match cookie vendors and their disclosed purposes, please refer to Appendix A, section A.1.

7.3.3 Results Given the 'Accept All Vendors, Reject All Purposes' Consent String

In the 'Accept All Vendors, Reject All Purposes' consent string, users grant consent to all vendors but not for any specific purposes, legitimate interests, or special features. Vendors are expected to prioritize the most restrictive preferences and not set any third-party cookies or engage in processing activities that require these consents.

We observed a total of 1,732 third-party cookies across 372 websites, averaging 4 cookies per site. When cross-referencing these cookies with the GVL, 54.6% were matched, 28.2% were unmatched, and 17.2% were partial matches.

Zoomd Ltd. was the most prominent vendor, accounting for 22.9% of all matched third-party cookies, followed by Revcontent, LLC, contributing 16.3% of all matched cookies. Purpose 1 (storage and/or access of information on a device) was the most common purpose, accounting for 81.3% of all matched cookies. No vendors were compliant with the TCF policy as the consent string prohibited the setting of third-party cookies for any purpose or legitimate interest.

We identified 36 different vendors setting partial-match cookies across the analyzed domains. The top five most prominent vendors accounted for 51.1% of all partial-match cookies. From the disclosed vendor purposes, we deduced that 31.5% of partial-match cookies could be set for any purpose(s), 79.2% for creating and selecting personalized ads profiles (purposes 3 and 4), and all cookies could be set for storing and/or accessing information on a device (purpose 1).

Please refer to Appendix A, section A.2 for a more comprehensive report.

7.3.4 Results Given the 'Accept All Purposes, Reject All Vendors' Consent String

The 'Accept All Purposes, Reject All Vendors' consent string represents a contradictory situation, as users give consent to all purposes but not to any specific vendor or legitimate interest vendor. In this case, vendors must, once again, prioritize the most restrictive preferences and refrain from all data collection and processing activities that rely on these permissions. In an analysis of 1,617 third-party cookies across 372 websites, 52.6% of the cookies matched information stored in the GVL, 28.9% were unmatched, and 18.6% were partial matches.

For the matched cookies, Zoomd Ltd. emerged as the most prominent vendor, accounting for 26.2% of all matched third-party cookies set across domains. The majority of their cookies were for purposes 1 and 10. Revcontent, LLC contributed to 18.7% of all matched third-party cookies, primarily setting cookies for purposes 3, 4, 5, 6, 7, 8, and 10. These purposes are associated with creating and selecting personalized ad profiles, as well as measuring ad and content performance, and generating audience insights.

In terms of purpose categories, storage and/or access of information on a device (purpose 1) was the most common, accounting for 78.7% of all matched cookies. Purpose 10, related to the development and improvement of products, was the second most frequent purpose, with 55.6% of the cookies set for this purpose. Purposes 3 and 4, which are associated with creating personalized ads profiles and selecting personalized ad profiles, were the third and fourth most common purposes, accounting for about 47.4% and 45.4% of the matched cookies, respectively.

For partial-match cookies, 38 different vendors set cookies across the analyzed domains, with the top five vendors accounting for 49.8% of all partial-match cookies. The disclosed

purposes of these vendors show that 30.7% of the partial-match cookies can be set for any purpose(s), 81.6% for creating and selecting personalized ads profiles (purposes 3 and 4), and 98.7% for storing and/or accessing information on a device (purpose 1).

Please refer to Appendix A, section A.3 for a more comprehensive report.

7.3.5 Results Given the 'Accept All' Consent String

The 'Accept All' consent string allows all vendors to place third-party cookies for any purpose combination. A total of 11,543 third-party cookies were observed across 372 websites, averaging 31 cookies per website. Of these, 54.6% matched, 14.2% were unmatched, and 31.2% were partial matches with information stored in the GVL. The matched cookies had a wide range of purposes, with Smart Adserver and ID5 Technology Ltd. being the most prominent vendors.

The most common purpose was storing/accessing information on a device (90.9%). Other common purposes included measuring ad performance (58.4%), selecting personalized ad profiles (56.2%), selecting basic ads (53.7%), and creating a personal ads profile (50.6%).

Partial-match cookies were set by 61 different vendors, with the top five accounting for 53.5% of all partial-match cookies. Just under one-sixth (15.1%) of partial-match cookies could be set for any purpose, 71% for creating/selecting personalized ads profiles, and all cookies could be set for storing and/or accessing information on a device (purpose 1).

Please refer to Appendix A, section A.4 for a more comprehensive report.

7.4 Analysis

The results of our evaluation of vendor compliance, as outlined in section 7.3, uncover significant discrepancies in vendor compliance across five distinct user consent scenarios. In this section, we will discuss the implications of these findings.

7.4.1 Analysis Given the 'Reject All' Consent String

Despite communicating the consent string to reject all forms of data collection and processing, many third-party cookies were still set across the analyzed websites. Over half of the observed cookies (51.1%) perfectly matched cookies disclosed in the GVL. This suggests that many vendors continue collecting and processing user data for a whole range of purposes despite explicit negative consent signals disallowing these practices.

The high prevalence of cookies set for creating a personalized ads profile (purpose 3) and selecting personalized ads (purpose 4) among matched and partially-matched cookies is especially concerning, as these purposes often involve privacy-invasive data collection and processing [Var19].

Zoom Ltd. and Revcontent, LLC were the most prevalent vendors setting matched cookies. The large number of cookies placed by these vendors, along with their broad range of purposes, likely indicates that user consent preferences were disregarded.

7.4.2 Analysis Given the 'Accept Basic Ads' Consent String

The results suggest that vendors are not fully compliant with the 'Accept Basic Ads' consent string. Despite consenting exclusively to cookies for basic ad selection, we observed a large number of third-party cookies set for different purposes. Alarmingly, only 40.4% of the matched cookies were set for purposes 1 (device storage) and/or 2 (selecting basic ads).

Furthermore, the majority (80%) of cookies were set by vendors that did not comply with the 'Accept Basic Ads' consent string. This suggests widespread vendor non-compliance with user consent in the digital advertising industry.

7.4.3 Analysis Given the 'Accept All Vendors, Reject All Purposes' Consent String

The results show a large number of third-party cookies set that contradict the 'Accept All Vendors, Reject All Purposes' consent string. Once again, vendors such as Zoomd Ltd. and Revcontent, LLC kept setting cookies for purposes that were explicitly rejected in the consent string.

The substantial number of observed partial-match cookies and their wide range of potential purposes suggests that vendors do not follow the more restrictive consent preference when contradicting signals are present in the consent string.

7.4.4 Analysis Given the 'Accept All Purposes, Reject All Vendors' Consent String

Once again, we observed significant vendor non-compliance with user consent. Vendors such as Zoomd Ltd. and Revcontent, LLC, continued to set cookies for a wide range of purposes despite the consent string explicitly rejecting all vendors.

The large number of partial-match and matched cookies set for personalized advertising purposes is concerning. This suggests that user privacy preferences are not being respected when vendors are presented with contradicting consent signals.

7.4.5 Analysis Given the 'Accept All' Consent String

As expected, we observe a large number of third-party cookies being set in the 'Accept All' scenario. However, the large number of observed partial-match and no-match cookies indicates that many vendors are not disclosing their data processing practices in the GVL properly.

8 Discussion

Our findings reveal several key insights into the extent to which the TCFv2.0 framework ensures compliance of AdTech vendors and CMPs in adhering to users' cookie consent preferences in practice. Despite the limitations of our proposed consent string injection method constraining our experimental scope, we were nevertheless able to detect significant instances of non-compliant CMPs and an alarming amount of vendors not adhering to user consent preferences. Our results emphasize the need for mechanisms to audit and enforce compliance within the TCF. In this section, we examine the implications of our

results, discuss the limitations of our experiments, and propose potential improvements in ensuring compliance with the TCF.

The implications of our findings are significant. They raise important concerns about the TCF's ability to effectively safeguard users' privacy choices. The existence of non-compliant CMPs and blatant disregard for user consent by vendors highlight flaws in the implementation and enforcement of the framework. If left unaddressed, the TCF risks becoming an ineffective tool, undermining its main purpose of serving as a standardized framework for accountability and compliance with provisions of the ePD and the GDPR.

In addition to instances of CMP and vendor non-compliance, our experiments also indicate a relatively low prevalence of the TCF on top-ranked websites. These results align with a previous study conducted by Matte *et al.* in 2019 [MBS20], indicating that the adoption of the TCF has remained stagnant over time. This lack of adoption is concerning as it limits the framework's ability to safeguard user privacy across the web, highlighting the lack of willingness within the digital advertising industry to prioritize user consent and transparency.

It is also important to acknowledge the limitations of our study. While the limited efficacy of the consent string injection method narrowed the scope of our experiments, our findings still uncovered valuable insights into the current state of compliance within the TCFv2.0. The lack of standardization in the storage mechanism of the consent string is the main obstacle, making it difficult to assess compliance across CMPs that use different storage mechanisms. Therefore, we suggest that the TCF establish a requirement specifying a standard storage mechanism for the consent string. Specifically, we argue for utilizing the browser's localStorage for the following reasons:

- **Privacy and Transparency:** localStorage keeps the consent string on the user's device, giving users more control over their data. It also eliminates the need for CMPs to store user identifiers (UIDs) to track user consent.
- **Domain Isolation:** localStorage is domain-specific, reducing the risk of cross-site tracking.
- **Storage capacity:** localStorage has a higher storage capacity compared to cookies, enabling it to store lengthy consent strings.
- **Compliance:** localStorage is easily auditable, aiding in compliance evaluation and enforcement.

Furthermore, when utilizing the consent injection method, we assume that injecting a consent string into the storage mechanism of the CMP results in the same behavior as manually filling out a cookie banner. To validate this assumption, we perform checks to ensure that the CMP's banner is no longer visible and that the CMP accurately indicates an updated consent string. However, there may be differences in unobservable underlying processes within the CMP's API that invalidate this assumption.

Another limitation of our study is that each suspected CMP violation needs to be inspected manually. This constrains the scope of our CMP evaluation method. To address this, we propose the integration of our automated consent string injection method with established existing CMP violation detectors like `Cookinspect` [G-Mat] developed by Matte *et al.* [MBS20]. By incorporating our automated consent string injection method into Cookinspect, we can eliminate the need for manual inspection and the time-consuming task of manually filling in cookie banners. This integration would fully automate the evaluation process and significantly expand the scope of our CMP evaluation method.

At the time of writing, the IAB had just announced the development of their `CMP Validator` [W-IABa] browser extension and web crawler initiative. These new tools are meant to streamline CMP and vendor compliance evaluation. This is a step in the right direction to ensure compliance with the TCF. However, our analysis, conducted prior to these developments, revealed an alarming level of non-compliance. Hence, it remains to be seen how effective these new tools will be in addressing the compliance issues we identified. Additionally, there is a potential conflict of interest present in the TCF auditing itself. To maintain impartiality, we suggest appointing an independent auditor to monitor TCF compliance. Future studies should be conducted to assess the impact of these new tools on TCF compliance once they are fully developed.

Despite the mentioned potential improvements, it may be necessary to reevaluate the cookie consent process as a whole. The current reliance on individual website publishers, CMPs, and Adtech vendors to correctly implement and respect user consent mechanisms may be inherently flawed, leading to the non-compliance we identified in our study.

Echoing the recommendation of Bollinger *et al.* [BKCB22], it might be beneficial for future privacy regulations to mandate the incorporation of a *purpose* flag as a new attribute in cookie headers by browser vendors and the World Wide Web Consortium. This would integrate the management of user cookie consent into the browser itself, potentially rendering both the traditional cookie banner and CMPs obsolete. Until such measures are adopted, the TCF, as it stands, must be constantly monitored, audited, and improved to ensure its efficacy in safeguarding user privacy.

9 Future Work

Given the findings from our study and the potential improvements suggested, we envision several areas for future research:

- **Impact of new tools:** With the recent development of the IAB's `CMP Validator` and web-scraper, future research should investigate its effectiveness in increasing TCF compliance.
- **Automated CMP Violation Detection:** Our study suggests integrating our automated consent string injection method with existing CMP violation detectors like `Cookinspect`. Future research should aim to implement and evaluate this integration. This could substantially improve the scope and speed of compliance evaluations, reducing the number of manual checks.
- **Alternative Consent Frameworks:** Our findings also highlight the potential need for alternative frameworks to the TCF. Future work could include the development and testing of new systems that ensure user privacy and consent in the digital advertising industry. Any new frameworks should prioritize transparency, user control, and strict adherence to the GDPR and ePD.

Our study is a stepping stone in understanding and improving TCF compliance. However, given the ever-changing nature of digital advertising and privacy regulations, ongoing research is crucial to continually safeguard user privacy.

10 Conclusion

Our research reveals significant non-compliance with the TCF among CMPs and AdTech vendors. Our results have demonstrated that numerous CMPs, registered with the IAB Europe's TCF, do not properly handle user consent, showing discrepancies between the user's input and the consent communicated by the CMPs. Moreover, we discovered considerable non-compliance amongst AdTech vendors, who continue to set cookies that blatantly disregard the user's consent choices. These findings call into question the effectiveness of the TCF in ensuring compliance with the General Data Protection Regulation (GDPR) and the ePrivacy Directive (ePD).

This research should serve as a wake-up call for all stakeholders in the digital advertising industry. Comprehensive mechanisms to audit and enforce compliance within the TCF are needed. While the introduction of the `CMP Validator` by the IAB is a step in the right direction, our findings emphasize the need for further measures to address these issues. Moving forward, we recommend continuously monitoring and evaluating CMPs and Adtech vendor compliance. Regular assessments and stringent penalties for non-compliance should be implemented to incentivize adherence to the TCF guidelines.

Our study's main contributions include an automated method to detect the implementation of the TCFv2.0 on websites, an automated method for communicating custom user cookie consent preferences, and a novel evaluation method to assess the compliance of AdTech vendors.

If these issues are left unaddressed, the consequences could be severe, potentially leading to significant legal and financial repercussions for non-compliant Adtech vendors and CMPs. To prevent this, it is crucial for the advertising industry to not only acknowledge these findings but also act upon them. By doing so, we can move towards a digital advertising ecosystem that respects user consent and values privacy.

Academic References

[BKCB22] Dino Bollinger, Karel Kubicek, Carlos Cotrini, and David Basin. "Automating Cookie Consent and {GDPR} Violation Detection". In: *31st USENIX Security Symposium (USENIX Security 22)*. 2022, pp. 2893–2910.

[FSABC20] Imane Fouad, Cristiana Santos, Feras Al Kassar, Nataliia Bielova, and Stefano Calzavara. "On compliance of cookie purposes with the purpose specification principle". In: *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE. 2020, pp. 326–333.

[HWB20] Maximilian Hils, Daniel W. Woods, and Rainer Böhme. "Measuring the Emergence of Consent Management on the Web". In: *Proceedings of the ACM Internet Measurement Conference*. IMC '20. Virtual Event, USA: Association for Computing Machinery, 2020, pp. 317–332. ISBN: 9781450381383. DOI: 10.1145/3419394.3423647. URL: https://doi.org/10.1145/3419394.3423647.

[KASRZB23] Lin Kyi, Sushil Ammanaghatta Shivakumar, Cristiana Teixeira Santos, Franziska Roesner, Frederike Zufall, and Asia J Biega. "Investigating deceptive design in GDPR's legitimate interest". In: *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 2023, pp. 1–16.

[MBS20] 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)*. IEEE. 2020, pp. 791–809.

[MSB20] Célestin Matte, Cristiana Santos, and Nataliia Bielova. "Purposes in IAB Europe's TCF: which legal basis and how are they used by advertisers?" In: *Privacy Technologies and Policy: 8th Annual Privacy Forum, APF 2020, Lisbon, Portugal, October 22–23, 2020, Proceedings 8*. Springer. 2020, pp. 163–185.

[MSLH22] Victor Morel, Cristiana Santos, Yvonne Lintao, and Soheil Human. "Your Consent Is Worth 75 Euros A Year-Measurement and Lawfulness of Cookie Paywalls". In: *Proceedings of the 21st Workshop on Privacy in the Electronic Society*. 2022, pp. 213–218.

[NBKK22] 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*. 2022, pp. 1–7.

[NLVKK20] 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*. 2020, pp. 1–13.

[RS22] Johnny Ryan and Cristiana Santos. "An unending data breach immune to audit? Can the TCF and RTB be reconciled with the GDPR?" In: *Can the TCF and RTB be reconciled with the GDPR* (2022).

[SNTBR21] Cristiana Santos, Midas Nouwens, Michael Toth, Nataliia Bielova, and Vincent Roca. "Consent Management Platforms Under the GDPR: Processors and/or Controllers?" In: *Privacy Technologies and Policy*. Ed. by Nils Gruschka, Luís Filipe Coelho Antunes, Kai Rannenberg, and Prokopios Drogkaris. Cham: Springer International Publishing, 2021, pp. 47–69. ISBN: 978-3-030-76663-4.

[TBR22] Michael Toth, Nataliia Bielova, and Vincent Roca. "On Dark Patterns and Manipulation of Website Publishers by CMPs". In: *PETS 2022 - 22nd Privacy Enhancing Technologies Symposium*. 2022.

[Var19] Kaan Varnali. "Online behavioral advertising: An integrative review". In: *Journal of Marketing Communications* 27.1 (June 2019), pp. 93–114. DOI: 10.1080/13527266.2019.1630664. URL: https://doi.org/10.1080/13527266.2019.1630664.

[VNS22] Michael Veale, Midas Nouwens, and Cristiana Santos. "Impossible Asks: Can the Transparency and Consent Framework Ever Authorise Real-Time Bidding After the Belgian DPA Decision?" In: *Michael Veale, Midas Nouwens and Cristiana Teixeiras Santos, Impossible asks: Can the Transparency and Consent Framework Ever Authorise Real-Time Bidding After the Belgian DPA Decision* (2022), pp. 12–22.

[VZ22] Michael Veale and Frederik Zuiderveen Borgesius. "Adtech and Real-Time Bidding under European Data Protection Law". In: *German Law Journal* 23.2 (2022), pp. 226–256. DOI: 10.1017/glj.2022.18.

Web Links

[W-Ben] Ben Cheshire. *Cookie Notice Blocker*. (Feb. 2023). URL: https://chrome.google.com/webstore/detail/cookie-notice-blocker/odhmfmnoejhihkmfebnolljiibpnednn.

[W-Cen] Ceni Apps. *Remove Cookie Banners*. (Feb. 2023). URL: https://chrome.google.com/webstore/detail/remove-cookie-banners/pacehjmodmfilemfbcahnpdcdmlocjnm.

[W-Dan] Daniel Kladnik. *I Don't Care About Cookies Version 3.4.6*. (Feb. 2023). URL: https://chrome.google.com/webstore/detail/i-dont-care-about-cookies/fihnjjcciajhdojfnbdddfaoknhalnja.

[W-02] *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)*. July 2002. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%5C%3A32002L0058.

[W-EC16] European Parliament and Council of the European Union. *Regulation (EU) 2016/679 of the European Parliament and of the Council*. of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). May 4, 2016. URL: https://data.europa.eu/eli/reg/2016/679/oj (visited on 04/13/2023).

[W-Goo] Goodbye Cookies. *Accept All Cookies Version 1.0.1*. (Feb. 2023). URL: https://chrome.google.com/webstore/detail/accept-all-cookies/ofpnikijgfhlmmjlpkfaifhhdonchhoi.

[W-IABa] IAB Europe. *CMP Validator*. (Jun. 2023). URL: https://chrome.google.com/webstore/detail/cmp-validator/ffhhjklgcfabkpholngojpkijlafjooc.

[W-IABb] IAB Europe. *IAB Europe Transparency & Consent Framework Policies*. (Feb. 2023). URL: https://iabeurope.eu/iab-europe-transparency-consent-framework-policies/.

[W-Rep] RepoGamesStudio. *CookiesBlock - cookie pop-ups*. (Feb. 2023). URL: https://chrome.google.com/webstore/detail/cookiesblock-cookie-pop-u/ajkknbgennjgacpfbhdobipfhhikbldg.

[W-rol] rolfba. *Consent-O-Matic Version 1.0.11*. (Feb. 2023). URL: https://chrome.google.com/webstore/detail/consent-o-matic/mdjildafknihdffpkfmmpnpoiajfjnjd.

GitHub Repositories

[G-Bur22] Interactive Advertising Bureau. *GDPR Transparency and Consent Framework/TCF Implementation Guidelines.md at Master*. Nov. 2022. URL: https://github.com/InteractiveAdvertisingBureau/GDPR-Transparency-and-Consent-Framework/blob/master/TCFv2/TCF-Implementation-Guidelines.md.

[G-CLe23] CLendering. *IAB-vendor-compliance/tcf-availability-crawler/tcf-crawler.py at main · CLendering/IAB-vendor-compliance*. June 2023. URL: https://github.com/CLendering/IAB-vendor-compliance/blob/main/tcf-availability-crawler/tcf-crawler.py.

[G-Int23] InteractiveAdvertisingBureau. *GitHub - InteractiveAdvertisingBureau/iabtcf*. Mar. 2023. URL: https://github.com/InteractiveAdvertisingBureau/iabtcf.

[G-Ken23] Kenny Grant. *GitHub - chromedp/chromedp: A faster, simpler way to drive browsers supporting the Chrome DevTools Protocol.* May 2023. URL: https://github.com/chromedp/chromedp.

[G-Lei23] Elazar Leibovich. *Goproxy*. Feb. 2023. URL: https://github.com/elazarl/goproxy.

[G-Len23a] Camile Lendering. *IAB-vendor-compliance/cmp-compliance-check/inject-custom-consent.go at main · CLendering/IAB-vendor-compliance*. June 2023. URL: https://github.com/CLendering/IAB-vendor-compliance/blob/main/cmp-compliance-check/inject-custom-consent.go.

[G-Len23b] Camile Lendering. *IAB-vendor-compliance/vendor-compliance-check/cross-reference-gvl at main · CLendering/IAB-vendor-compliance*. June 2023. URL: https://github.com/CLendering/IAB-vendor-compliance/tree/main/vendor-compliance-check/cross-reference-gvl.

[G-Len23c] Camile Lendering. *IAB-vendor-compliance/vendor-compliance-check/extract-third-party-cookies.go at main · CLendering/IAB-vendor-compliance*. June 2023. URL: https://github.com/CLendering/IAB-vendor-compliance/blob/main/vendor-compliance-check/extract-third-party-cookies.go.

[G-Mat] Célestin Matte. *Cookinspect*. URL: https://github.com/Perdu/Cookinspect.

[G-Rem23] Thomas Lay Remi Demol. *GitHub - SirDataFR/iabtcfv2: Go client to read TCF v2 consent string*. June 2023. URL: https://github.com/SirDataFR/iabtcfv2.

[G-Sam] Sam Macbeth. *Duckduckgo/autoconsent Version 4.13*. URL: https://github.com/duckduckgo/autoconsent.

A Appendix A

In this section, we give the detailed results of our evaluation of vendor compliance with the following consent strings:

- Accept basic ads (A.1).
- Accept all vendors, reject all purposes (A.2).
- Accept all purposes, reject all vendors (A.3).
- Accept all purposes and vendors (A.4).

A.1 Detailed Report for the 'Accept Basic Ads' Consent String

The *accept basic ads* consent string corresponds to a consent string, where:

1. Only `purposeConsents` 1: Store and/or access information on a device; and 2: Select basic ads; where set to `true` in the Core consent string.
2. All `vendorConsents` in the Core consent string where set to `true`.
3. None of the `purposeLegitimateInterests`, `vendorLegitimateInterests`, or `specialFeatureOptins` in the Core consent string where set to `true`.
4. None of the `purposeConsents` or `purposeLegitimateInterests` in the `publisherTC` section of the consent string where set to `true`.

Essentially, this configuration only allows third-party cookies to be set by vendors to select basic ads for the user. It explicitly disallows all other purposes and legitimate interests.

Once again using the method described in 4.9.1, we observed a total of **3,590 third-party cookies across 372 websites analyzed**. Averaging 9 third-party cookies per website. Cross-referencing these cookies with information stored in the GVL, over half (56.8%) of the cookies were matched, about one-fifth of the cookies (18.5%) were unmatched, and the remaining (24.7%) of the cookies were partial-matches. These results are summarised in Figure 10 below.

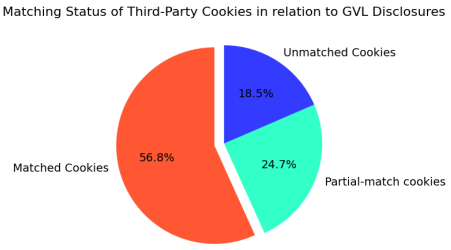

Figure 10: Percentages of the 3,590 third-party cookies that exactly matched, partially matched, and did not match the cookies disclosed in the GVL.

Analyzing the **matched** cookies and their disclosed purposes for each vendor, we, once again, obtained the distribution of cookie purpose in Figure 11 below. The plot is

a stacked bar chart that visualizes the distribution of cookie purposes for each vendor that set third-party cookies on the analyzed domains. Each bar represents a vendor, and the segments within each bar correspond to the different cookie purposes. The height of the segments indicates the frequency of each purpose for the respective vendor. The legend on the right side of the plot provides a mapping of the purpose categories to the colors used in the chart. The height of each bar represents the total number of cookies set by the respective vendor across all analyzed domains.

In Figure 11, we observe a wide range of cookie purposes distributed among vendors. ID5 Technology Ltd. emerges as the most prominent vendor, accounting for 20.5% of all matched third-party cookies set across domains. The vendor exclusively sets functional cookies (purpose 1) on the `id5-sync.com` domain. Note that the setting of these cookies is compliant with the provided consent string, as we consented to purposes 1 and 2. Zoomd Ltd. once again represents another significant portion, contributing to 11.8% of all third-party cookies. They primarily set cookies for purposes $[1, 10]$ and $[1, 2, 3, 4, 7, 9, 10]$. Which is not compliant with provided consent string.

Another noteworthy vendor is Revcontent, LLC. accounting for 8.9% of all matched third-party cookies. They set a number of cookies for purposes in the range $[1, 3, 4, 5, 6, 7, 8, 10]$. Which is not compliant with provided consent string. Meanwhile, the remaining half of the 2,038 matched cookies are being set by various different vendors for a whole range of purposes. In total, only 824 (40.4%) of the matched cookies had purpose(s) 1 and/or 2. The vast majority (80%) of vendors set matched cookies that were not compliant with the *allow basic ads* consent string.

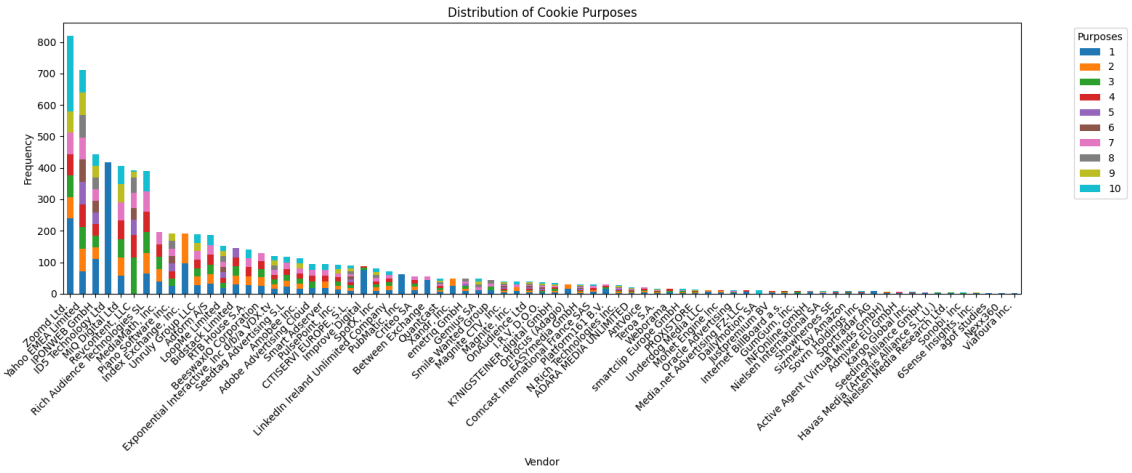

Figure 11: This stacked bar chart illustrates the Frequency of third-party cookies and their purposes for each vendor.

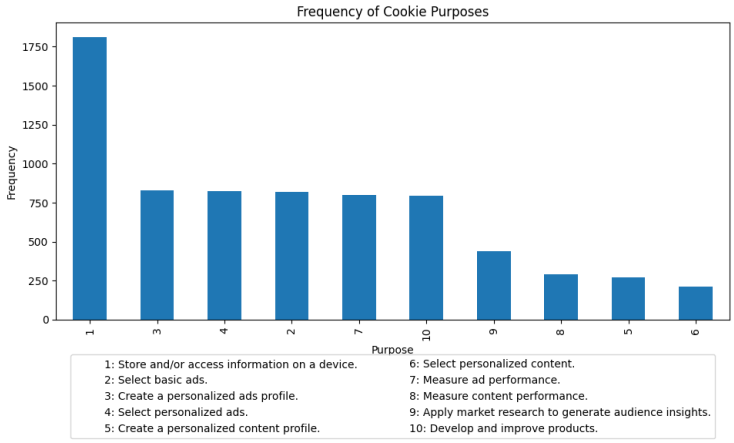

Figure 12: This bar plot shows the Frequency of each individual cookie purpose category in the matched third-party cookies.

In Figure 12 above, the frequency of each purpose category is once again displayed for the matched third-party cookies. The results show that purpose 1, which refers to the storage and/or access of information on a device, is by far the most common purpose, accounting for 89% of all matched cookies. Purposes 3 and 4, which are associated with creating and selecting personalized ads profiles, respectively, are the third and fourth most common purposes, both accounting for 40.6% of the matched cookies, respectively.

Purpose 2, selecting basic ads, is only the fifth most common purpose, with 40.1% of the matched cookies set for this purpose.

Moving on to our analysis of the 887 **partial-match** third-party cookies. (cookies for which the entry in the GVL matches only the cookie's domain)

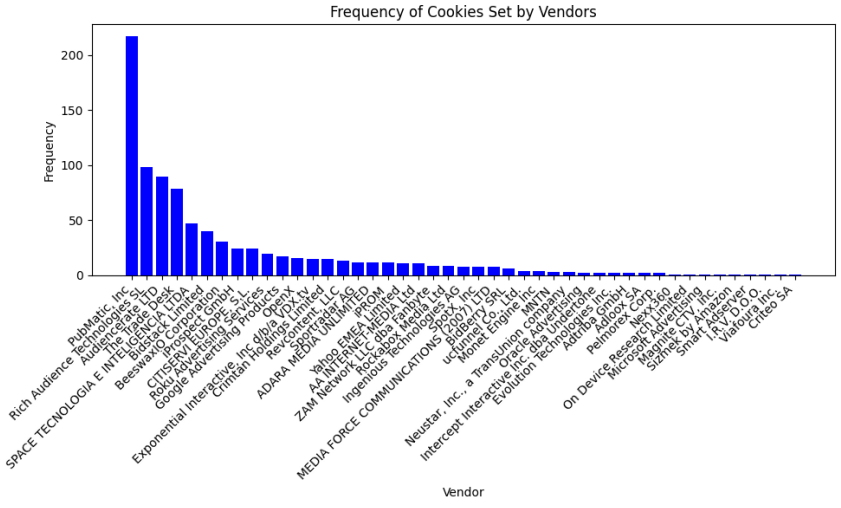

Figure 13: This bar plot shows the Frequency of each vendor that set a partial-match third-party cookie.

In Figure 13 above, we can see that 45 different vendors set partial-match cookies across

the analyzed domains. The top four most prominent vendors account for 54.6% of all the partial-match cookies, while the remaining 41 vendors each contribute to at most 5.3% of the partial-match cookies observed.

Although the purposes associated with each individual partial-match cookie are not disclosed on the GVL, we do once again have access to the disclosed purposes of each individual vendor. We have visualized these disclosed purposes in Figure 9 below.

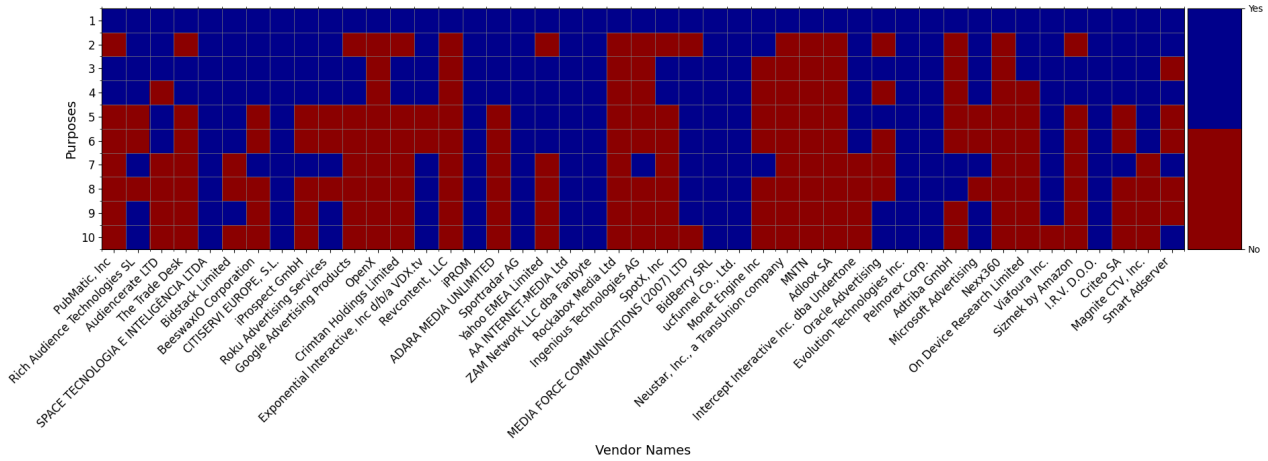

Figure 14: A matrix-plot of partial-match cookie vendors and their disclosed purposes.

Integrating the findings from Figures 13 and 14, we note that every partial-match cookie can be set for purpose 1, while the majority (56%) of these cookies can be set for purpose 2. Considering that all vendors have either purpose 1 or purpose 2, or both, disclosed within their vendor purposes on the GVL (see Figure 14), we can deduce that it is possible that the purposes of the partial-match cookies are consistent with the *accept basic ads* consent string. As a result, we observe no compliance concerns associated with the usage of these partial-match cookies.

A.2 Detailed Report for the 'Accept All Vendors, Reject All Purposes' Consent String'

The *accept all vendors, reject all purposes* consent string corresponds to a consent string, where:

1. All `vendorConsents` in the Core consent string where set to `true`.
2. None of the `purposeConsents`, `purposeLegitimateInterests`, `vendorLegitimateInterests`, or `specialFeatureOptins` in the Core consent string where set to `true`.
3. None of the `purposeConsents` or `purposeLegitimateInterests` in the `publisherTC` section of the consent string where set to `true`.

Essentially, this consent string represents a contradictory situation. While the user has granted consent to all vendors, they have not given consent for any specific purposes or legitimate interests. In this case, vendors are required to prioritize the most restrictive preferences: since none of the specific purposes, legitimate interests, or special features are granted, advertisers should refrain from data collection and processing activities that

rely on these permissions. This means that compliant vendors are expected not to set any third-party cookies for the user or engage in any processing activities that would require these consents.

Once again using the method described in 4.9.1, we observed a total of **1,732 third-party cookies across 372 websites analyzed**. Averaging 4 third-party cookies per website. Cross-referencing these cookies with information stored in the GVL, over half (54.6%) of the cookies were matched, just under one-third of the cookies (28.2%) were unmatched, and the remaining (17.2%) of the cookies were partial-matches. These results are summarised in Figure 15 below.

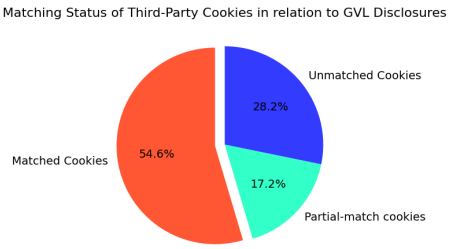

Figure 15: Percentages of the 1,732 third-party cookies that exactly matched, partially matched, and did not match the cookies disclosed in the GVL.

Analyzing the **matched** cookies and their disclosed purposes for each vendor, we, once again, obtained the distribution of cookie purpose in Figure 16 below. The plot is a stacked bar chart that visualizes the distribution of cookie purposes for each vendor that set third-party cookies on the analyzed domains. Each bar represents a vendor, and the segments within each bar correspond to the different cookie purposes. The height of the segments indicates the frequency of each purpose for the respective vendor. The legend on the right side of the plot provides a mapping of the purpose categories to the colors used in the chart. The height of each bar represents the total number of cookies set by the respective vendor across all analyzed domains.

In Figure 16, we observe a wide range of cookie purposes distributed among vendors. Similar to the results given the *reject-all* consent string, Zoomd Ltd. emerges as the most prominent vendor, accounting for 22.9% of all matched third-party cookies set across domains. The vendor, once again, predominantly sets cookies for purposes $[1, 10]$. Revcontent, LLC represents another significant portion, contributing to 16.3% of all matched third-party cookies. They primarily set cookies for purposes $[3, 4, 5, 6, 7, 8, 9, 10]$. Meanwhile, the remaining matched cookies are being set by various different vendors for a whole range of purposes. Since we provided a consent string that does not consent to the setting of third-party cookies for any purpose or legitimate interest, none of the 59 vendors in Figure 16 are compliant with the TCF policy.

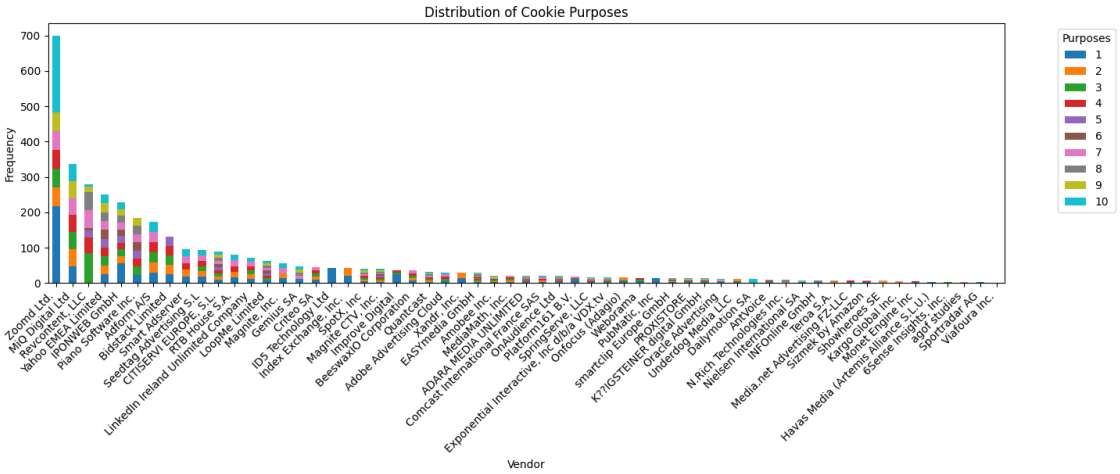

Figure 16: This stacked bar chart illustrates the Frequency of third-party cookies and their purposes for each vendor.

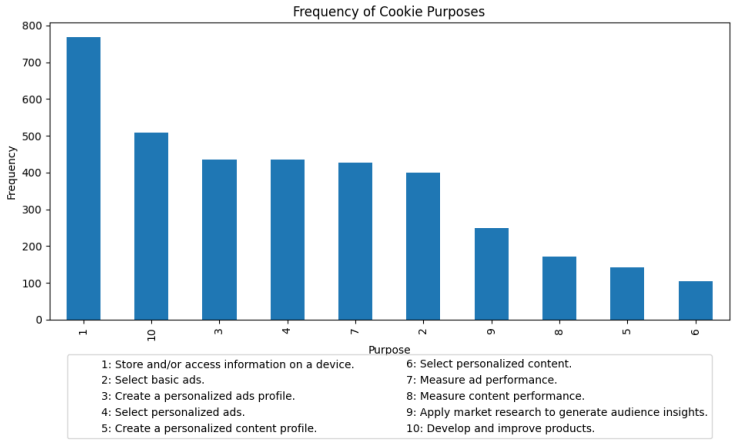

Figure 17: This bar plot shows the Frequency of each individual cookie purpose category in the matched third-party cookies.

In Figure 17 above, the frequency of each purpose category is once again displayed for the matched third-party cookies. The results show that purpose 1, which refers to the storage and/or access of information on a device, is by far the most common purpose, accounting for 81.3% of all matched cookies. Purpose 10, related to the development and improvement of products, is the second most frequent purpose, with 53.8% of the cookies set for this purpose. Purposes 3 and 4, which are associated with creating and selecting personalized ads profiles, respectively, are the third and fourth most common purposes, both accounting for about 46.1% of the matched cookies, respectively.

Purpose 7, measuring ad performance, is the fifth most common purpose, with 45.1% of the matched cookies set for this purpose. Purpose 2, selecting basic ads, is the sixth most common purpose, accounting for 42.8% of the cookies. Purposes 9, related to market research for generating audience insights, and 8, measuring content performance, account for 26.4% and 18.2% of the matched cookies, respectively. Purpose 5, creating

a personalized content profile, is the ninth most common purpose, accounting for 15.1% of all matched cookies. Purpose 6, related to selecting personalized content, has the lowest frequency, accounting for 11% of the matched cookies.

Moving on to our analysis of the 298 **partial-match** third-party cookies. (cookies for which the entry in the GVL matches only the cookie's domain)

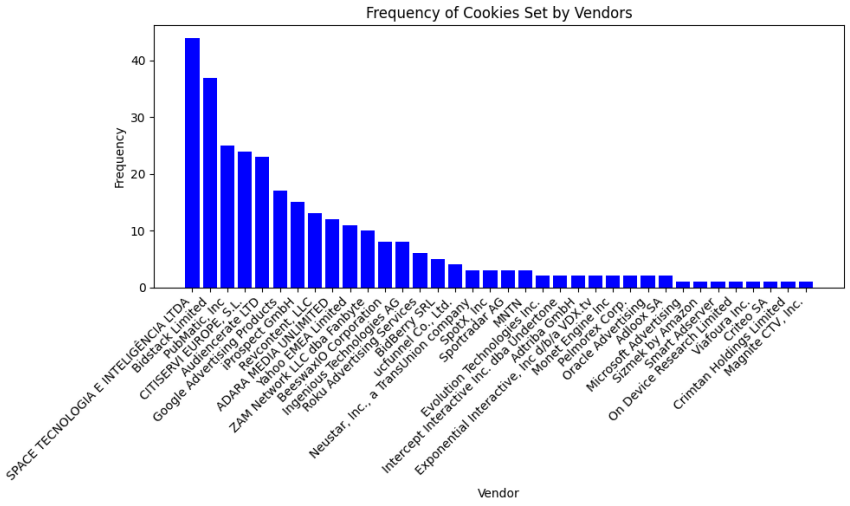

Figure 18: This bar plot shows the Frequency of each vendor that set a partial-match third-party cookie.

In Figure 18 above, we can see that 36 different vendors set partial-match cookies across the analyzed domains. The top five most prominent vendors account for 51.1% of all the partial-match cookies, while the remaining 29 vendors each contribute to at most 5.7% of the partial-match cookies observed.

Although the purposes associated with each individual partial-match cookie are not disclosed on the GVL, we do once again have access to the disclosed purposes of each individual vendor. We have visualized these disclosed purposes in Figure 19 below.

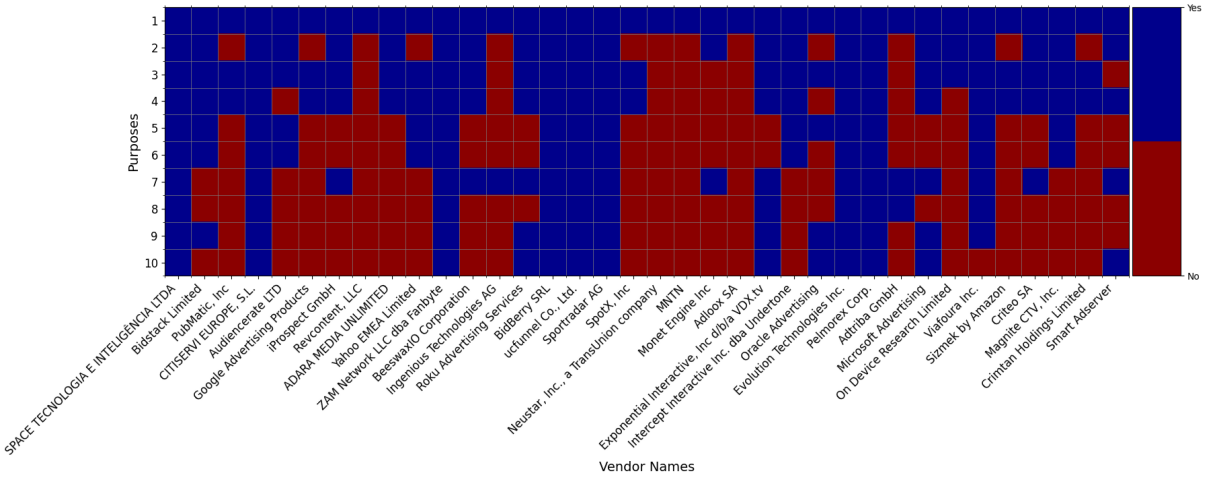

Figure 19: A matrix-plot of partial-match cookie vendors and their disclosed purposes.

Combining the results from Figure 18 and 19 above, we observe that almost one-third (31.5%) of the partial-match cookies can be set for any purpose(s) and the vast majority (79.2%) of partial-match cookies can be set for creating and selecting personalized ads profiles (purposes 3 and 4 respectively). Finally, we observe that all of the partial-match cookies can be set to store and/or access information on a device (purpose 1).

A.3 Detailed Report for the 'Accept All Purposes, Reject All Vendors' Consent String

The *reject all vendors, accept all purposes* consent string corresponds to consent string, where:

1. None of the `vendorConsents` in the Core consent string were set to `true`.
2. None of the `purposeLegitimateInterests`, `vendorLegitimateInterests`, or `specialFeatureOptins` in the Core consent string were set to `true`.
3. None of the `purposeConsents` or `purposeLegitimateInterests` in the `publisherTC` section of the consent string were set to `true`.
4. All of the `purposeConsents` in the Core consent string were set to `true`.

Essentially, this consent string represents another contradictory situation. While the user has granted consent to all purposes, they have not given consent for any specific vendor or legitimate interests vendor. In this case, vendors are required to prioritize the most restrictive preferences: since none of the specific vendors, legitimate interests, or special features are granted, advertisers should refrain from all data collection and processing activities that rely on these permissions. This means that compliant vendors are expected not to set any third-party cookies for the user or engage in any processing activities that would require these consents.

Once again using the method described in 4.9.1, we observed a total of **1,617 third-party cookies across 372 websites analyzed**. Averaging 4 third-party cookies per website. Cross-referencing these cookies with information stored in the GVL, over half (52.6%) of the cookies were matched, just under one-third of the cookies (28.9%) were unmatched, and the remaining (18.6%) of the cookies were partial-matches. These results are summarised in Figure 20 below.

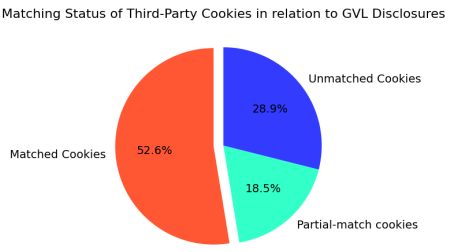

Figure 20: Percentages of the 1,617 third-party cookies that exactly matched, partially matched, and did not match the cookies disclosed in the GVL.

Analyzing the **matched** cookies and their disclosed purposes for each vendor, we, once again, obtained the distribution of cookie purpose in Figure 21 below. The plot is

a stacked bar chart that visualizes the distribution of cookie purposes for each vendor that set third-party cookies on the analyzed domains. Each bar represents a vendor, and the segments within each bar correspond to the different cookie purposes. The height of the segments indicates the frequency of each purpose for the respective vendor. The legend on the right side of the plot provides a mapping of the purpose categories to the colors used in the chart. The height of each bar represents the total number of cookies set by the respective vendor across all analyzed domains.

In Figure 21, we observe a wide range of cookie purposes distributed among vendors. Similar to the results given the *'accept all vendors, reject all purposes'* consent string, Zoomd Ltd. emerges as the most prominent vendor, accounting for 26.2% of all matched third-party cookies set across domains. The vendor, once again, predominantly sets a cookie named `test_cookie` on the `.doubleclick.net` domain for purposes [1, 10], which is observed in 178 out of the 372 websites analyzed, making up 47.8% of the sample. Revcontent, LLC represents another significant portion, contributing to 18.7% of all matched third-party cookies. They primarily set cookies for purposes [3, 4, 5, 6, 7, 8, 10]. Meanwhile, the 469 remaining matched cookies are being set by various different vendors for a whole range of purposes. Since we provided a consent string that does not consent to any specific vendor, none of the 59 vendors in Figure 21 are compliant with the TCF policy.

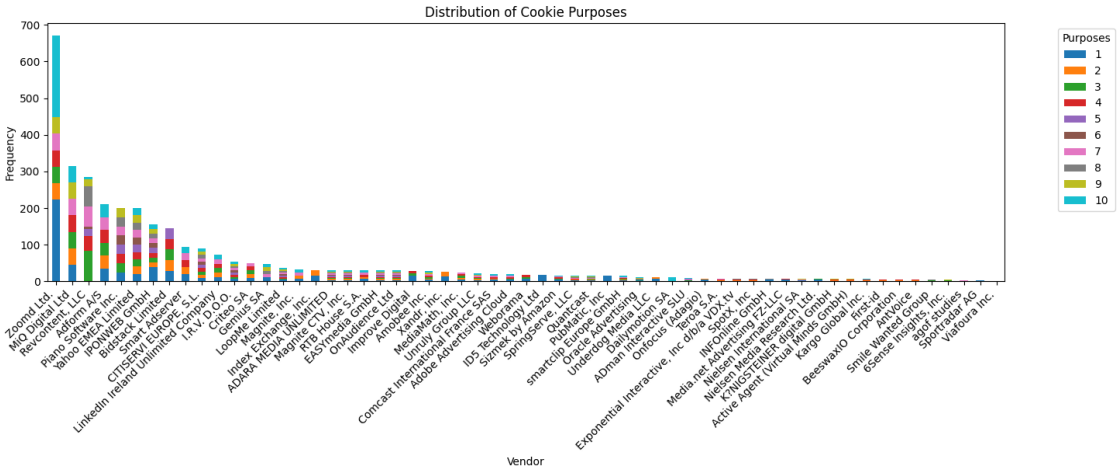

Figure 21: This stacked bar chart illustrates the Frequency of third-party cookies and their purposes for each vendor.

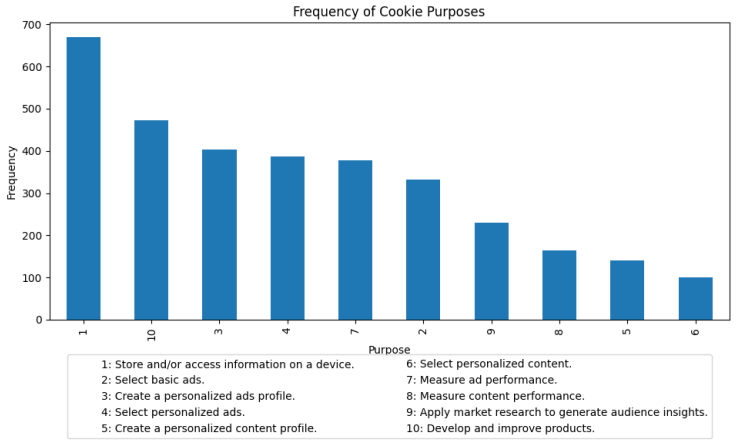

Figure 22: This bar plot shows the Frequency of each individual cookie purpose category in the matched third-party cookies.

In Figure 22 above, the frequency of each purpose category is once again displayed for the matched third-party cookies. The results show that purpose 1, which refers to the storage and/or access of information on a device, is by far the most common purpose, accounting for 78.7% of all matched cookies. Purpose 10, related to the development and improvement of products, is the second most frequent purpose, with 55.6% of the cookies set for this purpose. Purpose 3, creating personalized ads profiles, and purpose 4, selecting personalized ad profiles, are the third and fourth most common purposes, accounting for 47.4% and 45.4% of the matched cookies, respectively.

Purpose 7, measuring ad performance is the fifth most common purpose, accounting for about 44.4% of the matched cookies set. Purpose 2, selecting basic ads, is the sixth most common purpose, accounting for 39% of the cookies. Purposes 9, related to market research for generating audience insights, and 8, measuring content performance, account for 27% and 19.4% of the matched cookies, respectively. Purpose 5, creating a personalized content profile, is the ninth most common purpose, accounting for 16.6% of all matched cookies. Purpose 6, related to selecting personalized content, has the lowest frequency, accounting for 11.8% of the matched cookies.

Moving on to our analysis of the 299 **partial-match** third-party cookies. (cookies for which the entry in the GVL matches only the cookie's domain)

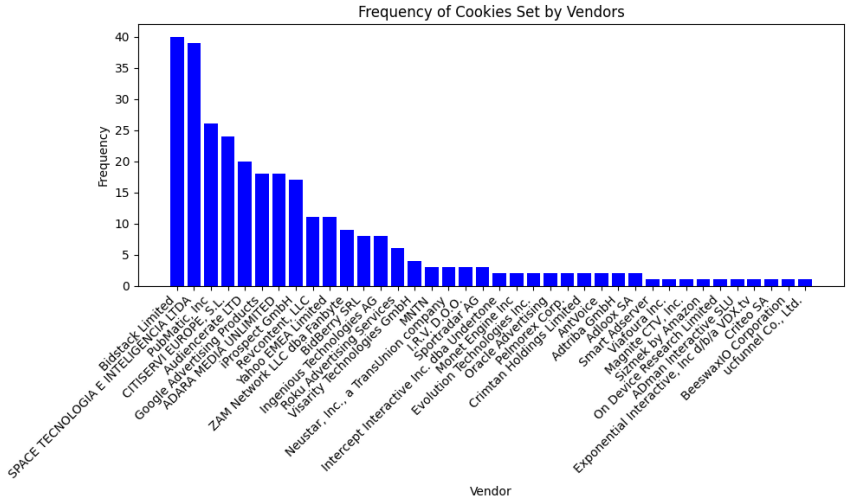

Figure 23: This bar plot shows the Frequency of each vendor that set a partial-match third-party cookie.

In Figure 23 above, we can see that 38 different vendors set partial-match cookies across the analyzed domains. The top five most prominent vendors account for 49.8% of all the partial-match cookies, while the remaining 33 vendors each contribute to at most 6% of the partial-match cookies observed.

Although the purposes associated with each individual partial-match cookie are not disclosed on the GVL, we do once again have access to the disclosed purposes of each individual vendor. We have visualized these disclosed purposes in Figure 24 below.

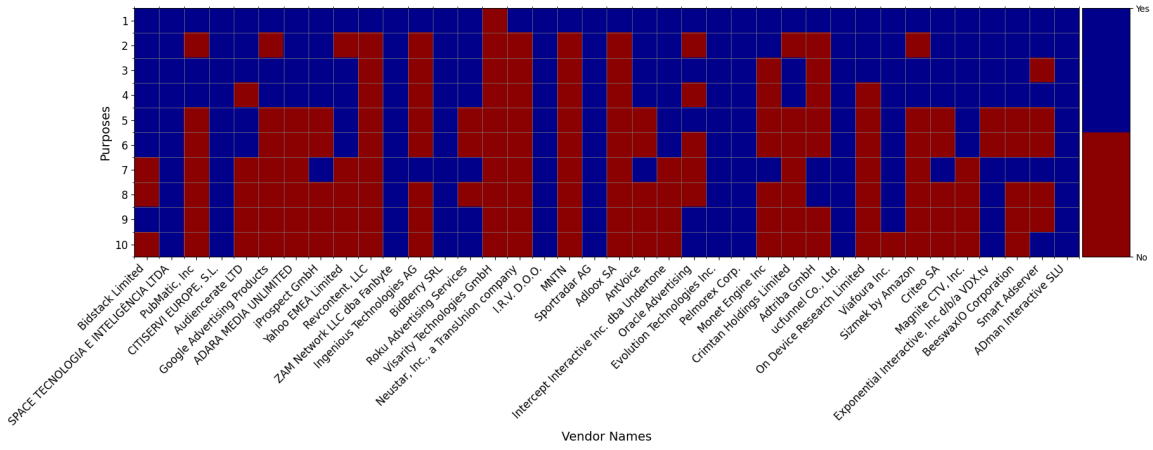

Figure 24: A matrix-plot of partial-match cookie vendors and their disclosed purposes.

Combining the results from Figure 23 and 24 above, we observe that just under one-third (30.7%) of the partial-match cookies can be set for any purpose(s) and the vast majority (81.6%) of partial-match cookies can be set for creating and selecting personalized ads profiles (purposes 3 and 4 respectively). Finally, we observe that almost all (98.7%) of partial-match cookies can be set to store and/or access information on a device (purpose 1).

A.4 Detailed Report for the 'Accept All' Consent String

The *accept all* consent string corresponds to consent string, where:

1. All `vendorConsents` and `purposeConsents` in the Core consent string were set to `true`.
2. None of the `purposeLegitimateInterests`, `vendorLegitimateInterests`, or `specialFeatureOptins` in the Core consent string were set to `true`.
3. None of the `purposeConsents` or `purposeLegitimateInterests` in the `publisherTC` section of the consent string were set to `true`.

Essentially, this consent string allows all vendors to place third-party cookies for any combination of purposes, setting the legitimate interest consents to `true` was not necessary, as all `vendorConsents` and `purposeConsents` were already set to `true`.

Once again using the method described in 4.9.1, we observed a total of **11,543 third-party cookies across 372 websites analyzed**. Averaging 31 third-party cookies per website. Cross-referencing these cookies with information stored in the GVL, over half (54.6%) of the cookies were matched, just under one-sixth of the cookies (14.2%) were unmatched, and the remaining (31.2%) of the cookies were partial-matches. These results are summarised in Figure 25 below.

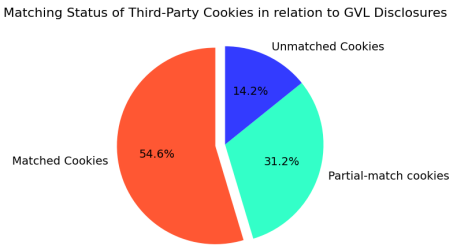

Figure 25: Percentages of the 11,543 third-party cookies that exactly matched, partially matched, and did not match the cookies disclosed in the GVL.

Analyzing the **matched** cookies and their disclosed purposes for each vendor, we, once again, obtained the distribution of cookie purpose in Figure 26 below. The plot is a stacked bar chart that visualizes the distribution of cookie purposes for each vendor that set third-party cookies on the analyzed domains. Each bar represents a vendor, and the segments within each bar correspond to the different cookie purposes. The height of the segments indicates the frequency of each purpose for the respective vendor. The legend on the right side of the plot provides a mapping of the purpose categories to the colors used in the chart. The height of each bar represents the total number of cookies set by the respective vendor across all analyzed domains.

In Figure 26, we observe a wide range of cookie purposes distributed among vendors. Smart Adserver and ID5 Technology Ltd. emerge as the most prominent vendors, accounting for 10.5% and 8.3% of all matched third-party cookies set across domains. Smart Adserver primarily sets cookies for purposes $[1, 2, 4, 7, 10]$ while ID5 Technology exclusively set cookies to store and/or access information on a user's device (purpose 1). IPONWEB GmbH represents another significant portion, contributing to 5.7% of all matched third-party cookies. They also exclusively set cookies to store and/or access

information on a user's device (purpose 1). Meanwhile, the remaining 84.5% of the 6,306 matched cookies are being set by various different vendors for a whole range of purposes. Since we provided a consent string that allows any vendor to set any third-party cookie for any purpose. All of the 67 vendors in Figure 21 are compliant with the TCF policy.

To improve readability, we excluded 29 vendors with a frequency lower than 5 from Figure 26. Including these vendors in the chart would not provide additional meaningful insights into the distribution of cookie purposes across different vendors.

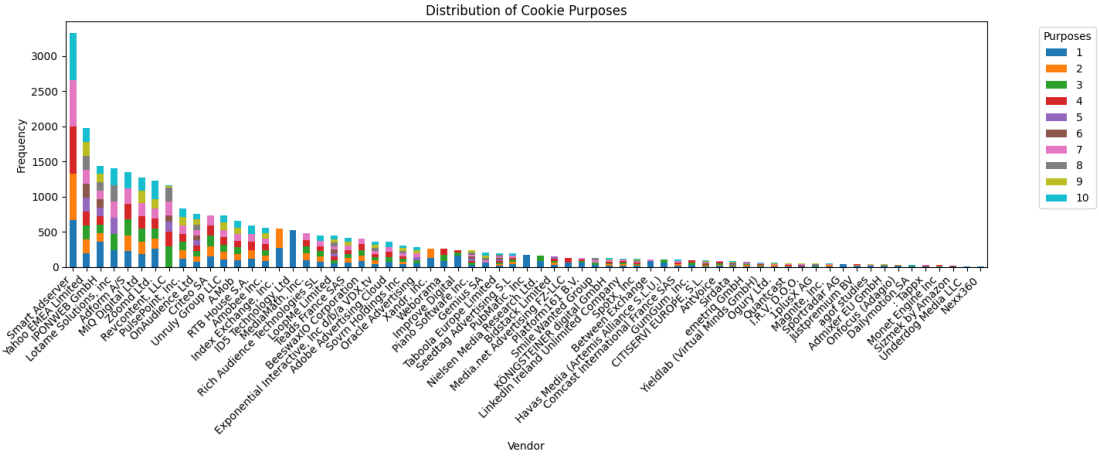

Figure 26: This stacked bar chart illustrates the Frequency of third-party cookies and their purposes for each vendor.

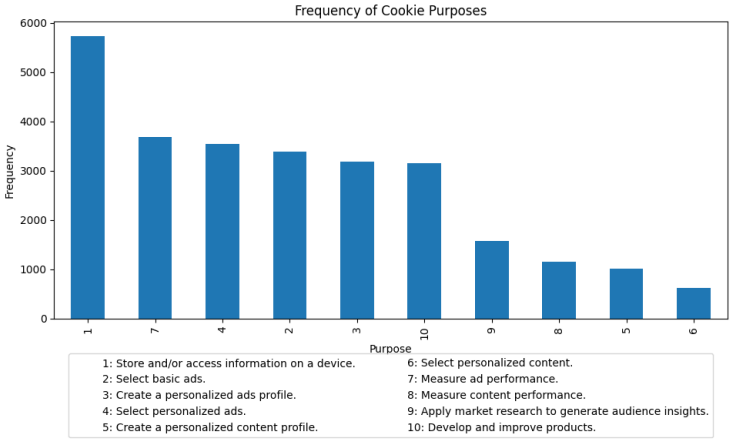

Figure 27: This bar plot shows the Frequency of each individual cookie purpose category in the matched third-party cookies.

In Figure 27 above, the frequency of each purpose category is once again displayed for the matched third-party cookies. The results show that purpose 1, which refers to the storage and/or access of information on a device, is by far the most common purpose, accounting for 90.9% of all matched cookies. Purpose 7, related to the measurement of ad performance, is the second most frequent purpose, with 58.4% of the cookies set for

this purpose. Purpose 4 which is associated with selecting personalized ad profiles is the third most common purpose, accounting for about 56.2% of the matched cookies.

Purpose 2, selecting basic ads, is the fourth most common purpose, accounting for 53.7% of the matched cookies set. Purpose 3, creating a personal ads profile, is the fifth most common purpose, accounting for 50.6% of the cookies. Purpose 10, develop and improve products, is the sixth most common purpose, accounting for 49.9% of the cookies. Purposes 9, related to market research for generating audience insights, and 8, measuring content performance, account for 24.9% and 18.4% of the matched cookies, respectively. Purpose 5, creating a personalized content profile, is the ninth most common purpose, accounting for 16.1% of all matched cookies. Purpose 6, related to selecting personalized content, has the lowest frequency, accounting for 9.9% of the matched cookies.

Moving on to our analysis of the 3,597 **partial-match** third-party cookies. (cookies for which the entry in the GVL matches only the cookie's domain)

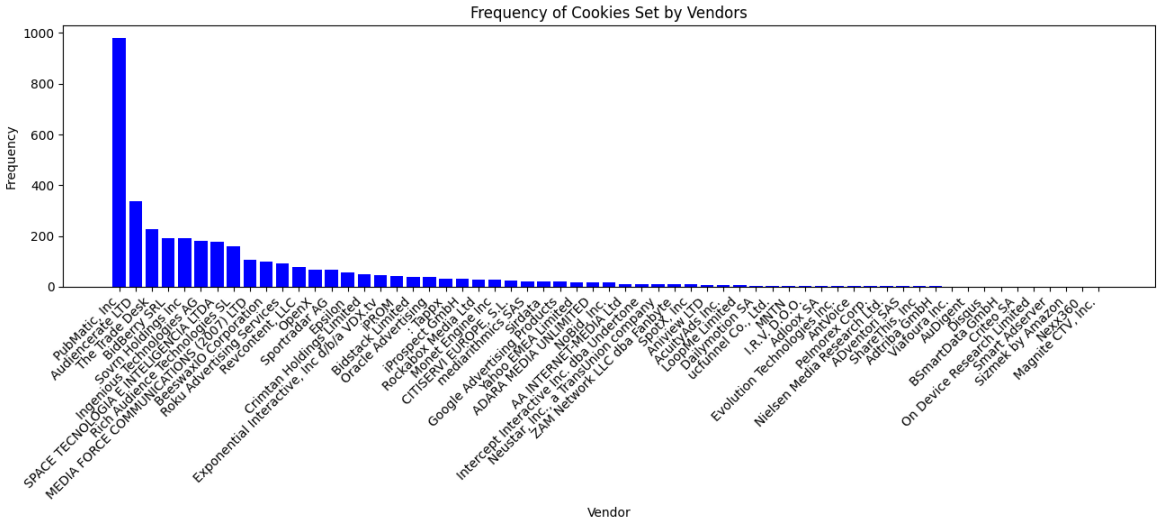

Figure 28: This bar plot shows the Frequency of each vendor that set a partial-match third-party cookie.

In Figure 28 above, we can see that 61 different vendors set partial-match cookies across the analyzed domains. The top five most prominent vendors account for 53.5% of all the partial-match cookies, while the remaining 56 vendors each contribute to at most 5.0% of the partial-match cookies observed.

Although the purposes associated with each individual partial-match cookie are not disclosed on the GVL, we do once again have access to the disclosed purposes of each individual vendor. We have visualized these disclosed purposes in Figure 29 below.

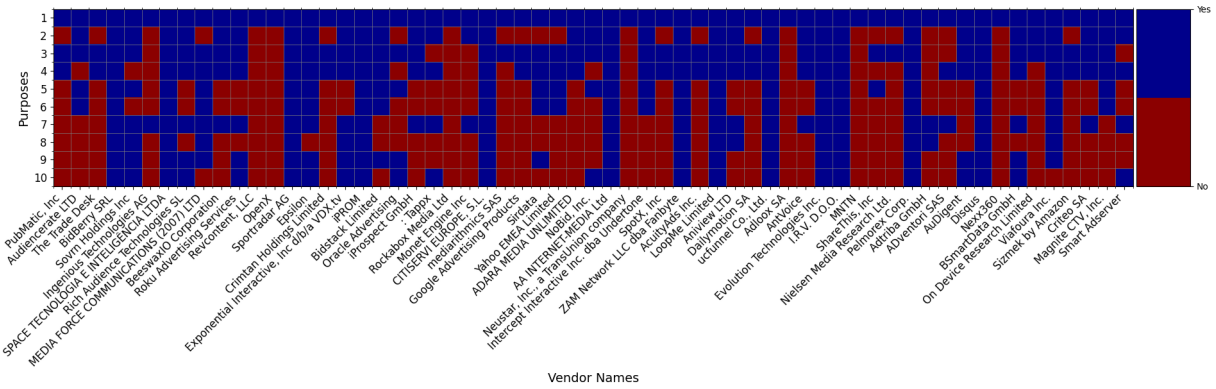

Figure 29: A matrix-plot of partial-match cookie vendors and their disclosed purposes.

Combining the results from Figure 28 and 29 above, we observe that just under one-sixth (15.1%) of the partial-match cookies can be set for any purpose(s) and just below three-fourth (71%) of partial-match cookies can be set for creating and selecting personalized ads profiles (purposes 3 and 4 respectively). Finally, we observe that all partial-match cookies can be set to store and/or access information on a device (purpose 1).