Understanding Topics API in the Wild: Dubious Usage and Stale Adoption

Received 11 April 2025; revised 26 August 2025; accepted 17 September 2025. Date of publication 1 October 2025; date of current version 15 October 2025. The review of this article was arranged by M. Humbert.

Digital Object Identifier 10.1109/TP.2025.3615120

Understanding Topics API in the Wild: Dubious Usage and Stale Adoption

ALBERTO VERNA [1], NIKHIL JHA [1], MARTINO TREVISAN [2], AND MARCO MELLIA [1] (Fellow, IEEE)

[1]Politecnico di Torino, 10129 Torino, Italy
[2]Università degli Studi di Trieste, 34127 Trieste, Italy

CORRESPONDING AUTHOR: ALBERTO VERNA (email: alberto.verna@polito.it).

This work was supported in part by the Spoke 1 "FutureHPC & BigData" of ICSC - Centro Nazionale di Ricerca in High-Performance-Computing, Big Data and Quantum Computing, funded by European Union - NextGenerationEU and in part by the AI4CTI FISA under Project #FISA-2023-00168, funded by the Italian Ministry of University and Research (MUR).

ABSTRACT Among several proposals for a privacy-preserving replacement for third-party cookies, Google's new Topics API is widely discussed as a possible solution to balancing privacy and utility for online targeted advertising. Despite being promoted as a key solution in a post-cookie world, the new paradigm still faces skepticism from researchers and regulators. This paper provides a first complete independent study of current adoption of the Topics API technology in the wild, to observe whether popular advertising platforms are already deploying it and how. We study the practices they adopt and their interplay with privacy policies and consent acquisition mechanisms. For this, we deploy a crawler to record the usage of Topics API across tens of thousands of popular websites worldwide, enriching our results in a twofold dimension: first, we run a seven-months-long campaign to understand how the adoption of Topics API is evolving over time — showing that typical problems of early deployments are still to be solved. Second, we observe how the use of the Topics API changes when observed from different regions, each adopting different privacy regulations. We find that users are likely to encounter the Topics API while surfing the Web, regardless of the geographic position. Our results show that this technology has not yet replaced third-party cookie technology for targeted advertising.

INDEX TERMS Topics API, web measurement, web privacy.

I. INTRODUCTION

Since the birth of online advertising, advertisers and trackers have followed users' browsing habits by the means of cookies, small chunks of text installed in the client's browser, which allow the server to identify the same user on subsequent visits. Third-party cookies — i.e., cookies set by a domain other than the one a user is currently visiting — allow trackers to follow the user on different websites, reconstruct their browsing patterns [1], [2], [3], and build their profiles to ultimately provide targeted ads and personalised content. This approach impacts users' privacy and still sparks debates and countermeasures — i.e., tracker blockers [4], [5], [6], privacy-friendly browsers and search engines [7], [8]. Recently, some browsers started blocking third-party cookies [9], [10], with Google Chrome being among the few ones to still consent to their usage. Legislators faced this unlimited data collection by mandating users to consent before the use of any personal information. European Union's GDPR [11], California's CCPA [12], or Brazil's LGPD [13] are the most prominent and well-known laws that regulate the online personal data collection.

These regulations challenge the online ads ecosystem, and industries are looking for alternative paradigms. Google, as one of the largest players in this arena, has proposed new solutions, including the Topics API [14], a key component of its larger Privacy Sandbox framework [15]. The Topics API moves all the tracking activity inside the browser: the latter observes the sites the user visits, maps them into topics, and shares some of them when asked by an enabled third party.

The Topics API lets the advertisers access valuable information about topics the user is interested in, without giving access to private information such as the specific website or page the user visits. After some initial setbacks,[1] Google

[1]https://blog.google/products/chrome/update-testing-privacy-sandbox-web/, accessed on September 25, 2025.

initially considered enabling the Topics API as part of the Privacy Sandbox for 1% of the Chrome users in the first quarter of 2024 and aimed to complete the third-party cookies phase-out by the end of the year,[2] a deadline that was later postponed not earlier than 2025.[3] However, in July 2024, the company has changed direction, announcing the cancellation of its third-party cookie phase-out in favour of a more informed user choice model.[4] As of March 2025, the Privacy Sandbox has reached general availability, rolling out to 99% of Chrome installations.

Websites and third parties have started to experiment with the Topics API. But what does the actual picture look like? We instrument a headless browser to collect Topics API statistics and run a carefully engineered measurement campaign visiting the top-ranked 50,000 websites for months. We observe which services enable the Topics API, identifying interesting patterns and witnessing unexpected facts as well. We observe requests for topics issued before the user gives consent, or the usage of Topics API by first and third parties not entitled to do so. We then expand our experiment by testing the usage of Topics API over seven months, showcasing the ambivalent penetration of the technology in the market. To complement our results, we finally run measuring campaigns in five different regions over three continents (Italy, Utah, California, Brazil, and Japan), to explore possible different usage patterns based on the user's location and regulations.

Similar to us, other works have explored the spread and penetration of the Privacy Sandbox toolset. For instance, [16] — backed by Sincera's Privacy Sandbox monitoring platform[5] — explores the penetration of the Topics API from an economic perspective, revealing a stagnant and skewed market. Other works are also starting to explore the solutions inside the Privacy Sandbox, other than the Topics API, focusing especially on the Protected Audience API [16], [17], [18]. In perspective, our work testifies how a new technique for behavioural advertising suddenly gained momentum before a subsequent stall. Our work complements the related measurements on the classical cookie-based approaches [2], [19], [20] and controversial techniques such as device fingerprinting [21], [22].

Some interesting facts emerge from our results:

- Popular ads platforms already adopt the Topics API and appear to be running live A/B tests to compare their effectiveness with the current cookie technology;
- A user encounters a legitimate party calling the Topics API in one website every four;
- During our period of observation, we witness the technology growing mature, reducing inconsistent deployment, questionable integration with privacy regulations, and even erroneous support in the Chrome browser. However, some problems still persist, that could potentially have consequences from a legal point of view, as some behaviour we observed around the Topics API could go against privacy regulations.
- The experimentation runs on a global scale, with no major difference across countries with different privacy regulations.

We believe our work sheds some light on this new technology. We hope this will stimulate other researchers to explore it and monitor its deployment. For this, we offer our tools and dataset to the community. These include the source code to perform a crawling campaign[6] and a simple tool to detect the usage of Topics API on a target website.[7] This paper is the extended version of our preliminary work [23]. We have broadened our analyses to include the temporal and spatial dimensions, monitoring the adoption of the Topics API over its first year and measuring how it affects users in different regions of the world.

The remainder of the paper is organized as follows: in Section III, we briefly describe the Topics API functioning, together with details on our data collection procedure, and a high-level view of the resulting dataset. In Section IV, we present our findings on legitimate usage. In Section V and Section VI, we describe two different families of non-compliant use of this technology. In Section VII we describe our temporal measurement campaign, while in Section VIII we focus on measurements collected in different regions. Finally, we draw conclusions in Section IX.

II. TOPICS API PRIMER

The Topics API is a component of Google's Privacy Sandbox, the new online advertisement ecosystem promoted and designed by Google to look for a new balance between offering valuable information to advertisers and respecting users' privacy.

The functioning of the Topics API is articulated as follows. First, the browser internally monitors the browsing activity of the user. During each *epoch* (lasting one week), the browser collects the visited websites and assigns to each of them one or more labels, called *topics*, using a predefined language model. At the end of the epoch, the browser computes the top 5 most relevant topics according to their frequency and importance to an advertiser, and stores them in a list. These processes happen inside the browser so that no external entity has access to potentially private information. When a user visits a website, any third-party service (e.g., advertisers) on the page — henceforth Calling Party (CP) — can invoke the API to ask the browser for some topics the user is interested in. The browser returns three topics, one for each of the last three epochs, choosing each randomly from the epoch's top 5

[2]https://developers.google.com/privacy-sandbox/blog/cookie-countdown-2023oct, accessed on September 25, 2025.

[3]https://privacysandbox.com/news/update-on-the-plan-for-phase-out-of-third-party-cookies-on-chrome/, accessed on September 25, 2025.

[4]https://privacysandbox.com/news/privacy-sandbox-update/, accessed on September 25, 2025.

[5]https://app.sincera.io/privacysandbox

[6]https://github.com/Novant8/priv-accept-topics, accessed on September 25, 2025.

[7]https://smartdata.polito.it/is-this-site-calling-the-topics-api/, accessed on September 25, 2025.

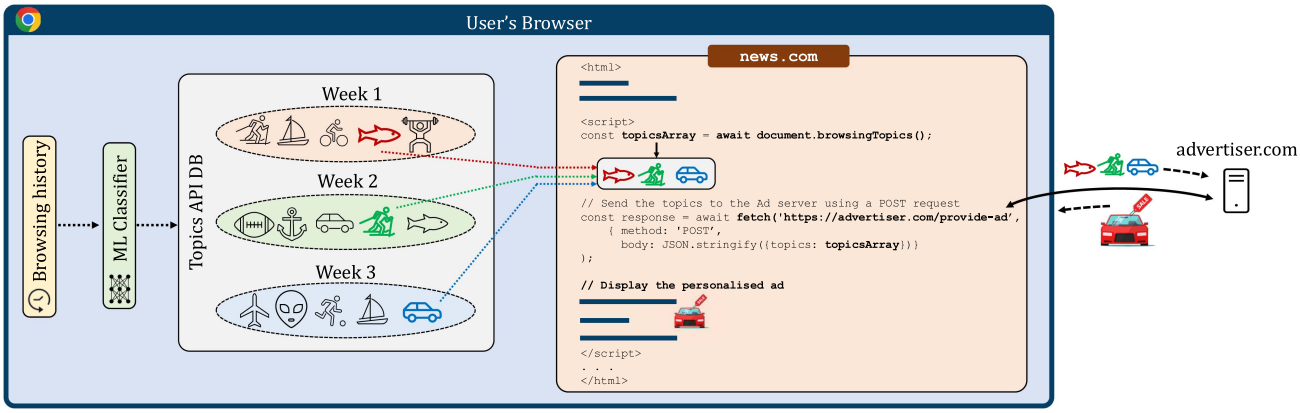


FIGURE 1. **Topics API operation and use in a JavaScript.**

topics.[8] We exemplify this mechanism in Fig. 1. The Topics API implements specific mechanisms to protect the users' privacy: for instance, to add some *plausible deniability*, 5% of the offered topics are replaced by a random topic. This should make it more difficult to build the user's profile and gives all topics a minimum exposure probability. To access the Topics API, services must complete an enrollment and attestation process. This provides a mechanism to verify which entities can call the API, adds transparency to who is accessing data, and mitigates attempts to misuse the API to gather more data than intended (see below for technical details).

The Topics API is implemented in Chromium and Chrome on both their mobile and desktop versions since Chrome version 101 of March 2022. Researchers showed that this tool is an improvement over the old "all-allowed" cookie-tracking jungle, while some theoretical [24], [25] and practical [26], [27], [28] results show to various extent that some privacy leak may still occur. Similarly, some public control bodies and other organizations, such as the United Kingdom's Competition and Markets Authority (CMA)[9] and the W3C[10], have expressed concerns about users' privacy guarantees under the new paradigm.

The privacy issues, the setbacks from rival firms' browsers such as Firefox and Safari (which, at the moment of writing, are not implementing the Topics API[11]), and uncertainty about third-party cookies disruption are slowing down the large-scale introduction of the Topics API on the market, as we will show in the following.

Parties interested in the usage of Topics API must complete an onboarding process. To enforce this, the browser checks whether the CP is included in the local *allow-list* file stored in the browser's configuration folder. If present, the browser allows the call; otherwise, it blocks it. Every time the browser is opened, it updates an *allow-list* file. We call the parties that are present in this list as *Allowed*. We initially use the file obtained in June 6th, 2024, and update it henceforth for the temporal measurement campaign to observe evolutions.

In addition, the Privacy Sandbox policy mandates all the CPs using the Topics API to offer an attestation JSON file in a predetermined URL path, namely <domain>/.well-known/privacy-sandbox-attestations.json. This attestation serves as a declaration by the CP that it will not use the Topics API for re-identification purposes and is considered by Google as part of the verification process for new enrollments.[12] For every first and third party we encounter (i.e., for every domain), we verify whether a valid attestation file is present. If so, we label the party as *Attested*.

During our experiments, we purposefully corrupted the local allow-list of our Chromium browser. Interestingly, we observe the browser allows any first and third parties to call the Topics API in this case. Investigating this, we found that the current implementation permits any Topics API calls as the *default* case when the internal database is corrupted or missing. By removing the *allow-list*, one can thus observe whether non-allowed callers are trying to request topics to the API. This would allow any caller to access the Topics API independently whether *Allowed* / *Attested* or not, permitting them to collect users' topics and possibly abuse this information.[13]

[8] A caller can only obtain the topics that have been observed by it in the previous three epochs—that is, topics of websites where the caller invoked the Topics API previously. Callers can also choose not to record the website of a specific invocation by setting a `skipObservation` flag.

[9] https://assets.publishing.service.gov.uk/media/65ba2a504ec51d000dc9f1f5/A._CMA_Q1_2024_update_report_on_Google_Privacy_Sandbox_commitments_24.4.24.pdf, accessed on September 25, 2025.

[10] https://github.com/w3ctag/design-reviews/issues/726, accessed on September 25, 2025.

[11] https://developer.mozilla.org/en-US/docs/Web/API/Topics_API#browser_compatibility, accessed on September 25, 2025.

[12] https://github.com/privacysandbox/attestation, accessed on September 25, 2025.

[13] To actually corrupt the internal database, an attacker should gain access to the file system of the user's device, deleting or modifying the database in a way that is no longer readable by the browser itself. This would require a high-privileged access inside the user's system and appears impractical, although the actual feasibility of an attack goes beyond the scope of this paper.

We notified Google and Chromium developers about the error in June 2024: they recognized the problem and declared to fix it in a future release. As of writing, no fix has been released.

III. DATA COLLECTION AND DATASETS

In this section, we present how we engineered our custom Web crawler and the dataset we gathered.

A. PER-WEBSITE DATA COLLECTION

To collect Topics API usage information, we rely on a Selenium-based crawler running on Chromium browser version 122.0.6261.128, where we manually enable the Privacy Sandbox's components. When the crawler visits a website, we collect the URL of each first- and third-party object downloaded to render the page and record every call to the Topics API by modifying Chromium's source code. Such information includes the domain calling the Topics API, the domain of the website on which the call happened, and the timestamp of the call. We modify the handler to additionally log the API call type [29] (JavaScript, Fetch or IFrame) and record possible multiple calls from the same CP on the same webpage. We show an example of a JavaScript call in Fig. 1.

As shown in [30], reliable crawling campaigns in the wild must account for the presence of consent banners: if ignored, crawls provide a biased and incomplete view of website behaviour. Building on this insight, we adopt the working assumption that the Topics API must follow the same regulatory framework that protects users' privacy—i.e., users have to agree to the privacy policy of a website and explicitly authorise the usage of any personal data. For instance, the GDPR clearly mandates any website to collect the user's explicit consent before using any personal information. While we do not claim to provide a definitive legal interpretation, we rely on this assumption to structure our methodology and tell legitimate usage from illegitimate one. In our crawling, we mimic the user who grants the usage of personal data by interacting with the Consent Banner shown during the user's first visit. To this end, we build on the *Priv-Accept* tool presented in [30] that automatically provides consent by interacting with Privacy Banners, if present. In a nutshell, for each website, we first visit it and record statistics *before* accepting the privacy policy; we then grant consent to personal data usage and, if successful, visit the site *after* acceptance. We delete the browser cache to load again all objects. We call these two visits *Before-Accept* and *After-Accept* as in [30]. Note that if we are not able to find a banner and allow the usage of personal data, we do not proceed with the *After-Accept* visit. This may happen because i) the banner is not actually present, or ii) *Priv-Accept* fails to recognise the "Accept" button. *Priv-Accept* looks for keywords and supports five languages — i.e., English, French, Spanish, German and Italian. The authors of [30] show that it is 92—95% accurate with banners in such languages.

As with any automated crawling effort, there is the possibility that websites or embedded third parties detect that the visit is performed by a crawler and change their behaviour accordingly (cloaking mechanism). In such cases, our measurements may not fully reflect the behaviour experienced by real users. We acknowledge this as a potential source of bias in our datasets.

To facilitate further exploration of Topics API usage, we built a web application prototype that allows users to test the described methodology on any URL of their choice. We provide its source code to the community.[7]

B. MEASUREMENT CAMPAIGNS

To paint a complete picture of Topics API usage, we ran three different measurement campaigns. In the first one, we crawl the top-50,000 websites according to the Tranco list [31], as of March 26th, 2024, using the methodology described in the previous section. We start the crawling on March 30th, 2024, ending after about one day, from a server located in Italy.

Second, we repeat this process weekly (starting every Monday at midnight) from the same vantage point. We begin on August 18, 2024 and continue until March 10, 2025, covering seven months in total.

Third, we run measurement campaigns from five different geographic locations, each subject to different privacy regulations. To achieve this, we use ExpressVPN[14], a commercial VPN service to tunnel the machine's traffic and make it appear to the visited websites as if it were originating from each location. Section VIII describes in depth the location choice and technical details. We perform the crawling sequentially for each location, from January 10 until January 24, 2025. Differently from the previous crawls, we only visited 10,000 websites per location. We compensate for the longer loading times introduced by the VPN service by relaxing the crawler's timeouts.

C. INITIAL FINDINGS

During the first crawl, we successfully visit 43,405 websites for which we obtain a *Before-Accept* visit — the remaining websites failed due to domain name resolution or connection-related errors. We refer to this dataset as D_{BA}. It includes 19,534 unique third parties in addition to 43,405 first parties.

For 14,719 websites (about 30% of the active sites) *Priv-Accept* accepts the privacy policy and consents to the usage of personal information. For these, we execute an *After-Accept* visit and save the data in a dataset we call D_{AA}. This percentage is in line with [30]. In the remaining cases, the website does not implement any banner, or *Priv-Accept* misses language or keyword. Table 1 summarises our results:

- We find 193 *Allowed* domains. These are the only ones allowed to use the Topics API. They include popular advertisers.
- We check all these 193 services to see if they correctly expose the attestation file. 181 do, but 12 do not.

[14]https://www.expressvpn.com/, accessed on September 25, 2025.

TABLE 1. Overall Status of Topics API Usage. In Red, the Anomalous Usage. In Blue, the Questionable Usage.

	Allowed	*Attested*		Notes
Enrolled	✓	✓	181	Regular
	✓	✗	12	No attestation file
			193	Total
D_{BA}	✓	✓	**28**	Questionable
	✗	✓	**1**	Not allowed
	✗	✗	**1,312**	Anomalous
D_{AA}	✓	✓	47	Legitimate
	✗	✓	**1**	Not allowed
	✗	✗	**2,633**	Anomalous

- In our crawls, we encounter 47 CPs that call the Topics API during the *After-Accept* visit (D_{AA} dataset). The 193-47=146 potential CPs may not have activated it, or we did not encounter them during our crawling.
- We find that one CP — namely dstillery.com — has the attestation file timestamped on November 2023. Yet, it is not included in the *allow-list*. This possibly reflects the attestation process is still ongoing, or that Dstillery has no interest in completing it: in fact, we observe it using the Topics API on the dstillery.com website only, hinting at initial testing.
- Considering the D_{BA}, we would expect no usage of the API because the user has not yet consented to the privacy policy. However, we find 28 *Allowed* and *Attested* CPs that call the API even if the user did not consent. We will investigate this *questionable* usage in Section VI.
- Surprisingly, we observe thousands of websites and CPs that call the API even if they are not among the *Allowed* ones. We investigate this *anomalous* usage in Section V.

In a nutshell, we observe a very confused deployment, with apparent violations and questionable implementations. In the following, we dig into regular and unexpected cases.

D. TEMPORAL AND GEOGRAPHICAL DATASETS

Concerning the temporal measurements, we collect 29 snapshots in total out of 32 weeks. Three weeks are missing due to technical issues. For each week, we count a number of successful visits ranging from 38,324 to 40,223. Similarly, the number of websites where the crawler found and clicked a Privacy Banner ranges between 13,534 and 14,310.

As for the geographical analysis, we obtain five datasets, one for each chosen location. Each contains usage information for the first 10,000 successful visits. The crawler's accuracy in detecting a Privacy Banner varies across locations, ranging from 21.6% to 40.1% of visited websites. We attribute this variation to different Privacy Banner implementations across different locations.

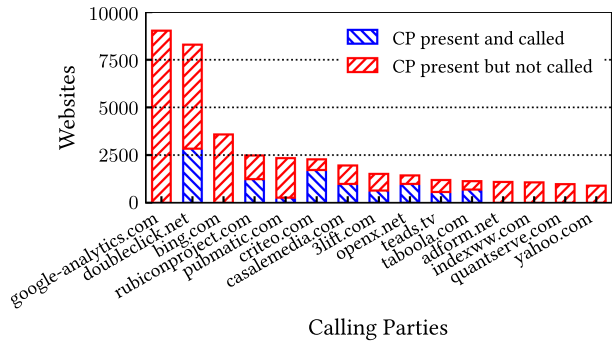


FIGURE 2. Number of websites where a CP is present and subset where it calls the Topics API. D_{AA} and all *Allowed* and *Attested* parties.

IV. LEGITIMATE USAGE

In this section, we characterise the Topics API's penetration inside the Web ecosystem. Here, we take into consideration only legitimate uses of the Topics API: hence, we only include interactions from the 47 CPs that are both present in the *Attested* and the *Allowed* sets, and we encounter after successfully accepting the privacy policy (D_{AA} dataset).

Processing each CP attestation file, we observe the onboarding process for the use the Topics API by extracting the attestation certificate issue date. Enrolments kicked off in June 2023, the first attestation being on the 16th. Until May 2024 the enrolment process continues at a low pace: each month, approximately a dozen new services obtain the attestation for the Topics API. On October 17th, 2024, many of the enrolled CPs had to update their attestations to include the new `enrollment_site` field and other minor changes. We do not report the complete timeline for the sake of brevity.

We now focus on the extent to which popular ad-related platforms adopt the Topics API. In D_{AA}, we observe at least one API call from a legitimate CP in 27.3% of visited websites. That is, every fourth website one visits already hosts a CP that invokes the Topics API — not surprising given the pervasiveness of ad services.

Fig. 2 details the number of websites on which a given CP is present (red pattern). We show the top-15 most pervasive CPs. Unsurprisingly, this list is dominated by established and recognisable players in the online advertising ecosystem. In blue we highlight the fraction of websites in which a CP invokes the Topics API. google-analytics.com, although both *Allowed* and *Attested*, never calls the Topics API (not being an ad-related service) while Google's doubleclick.net employs the Topics API on about one third of the websites we found it. Conversely, bing.com (also *Allowed* and *Attested*) never calls the Topics API. criteo.com, rubiconproject.com, and casalemedia.com are leveraging the Topics API the most. Moreover, none of them enable it on all websites and visits.

In general, results show that all the major ad-related players are adopting the Topics API. Yet, we seldom observe consistent usage, hinting they are still in a testing phase. We next investigate this aspect.

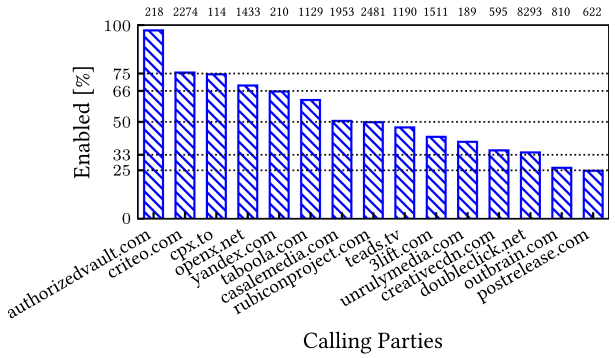


FIGURE 3. Fraction of times a CP calls the Topics API over the total times it is present on (in the top row). D_{AA} and *Allowed* and *Attested* services.

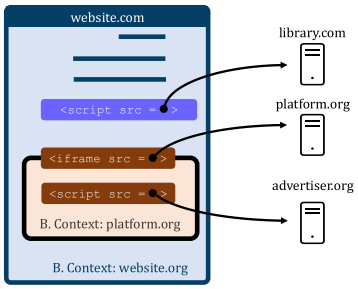


FIGURE 4. Graphical explanation of the context mechanism with scripts and iframes.

Given a CP that uses the Topics API, we count the fraction of times it uses them over the total number of times we observe it. We show the CPs with the highest enabled percentage in Fig. 3. We highlight some notable fractions on the y-axis to simplify reading the results. The top of the figure details the number of websites in which we observe such a party — we only filter CPs that are at least present in 100 websites. We notice a clustering of behaviours: for instance, authorizedvault.com, present on 218 websites, calls the Topics API almost every time. criteo.com and cpx.to call it 75% of the time, yandex.com 66% of the time, etc. We impute this to some form of A/B testing taking process, with percentages that look predetermined. They test how well the Topics API paradigm behaves compared with the standard third-party cookie solutions for their business metric: even if the Topics API is still below the surface, the most prominent advertising companies in the market are deeply studying its effectiveness.

We additionally run repeated tests to observe the policy some CPs use to enable/disable Topics API. We notice consistent alternating periods: for some time, CP, and website, the usage of the API is ON for all visits, followed by some time when it is OFF. This supports the hypothesis that there are some ongoing A/B tests which consider the same population and website but at different times.

Beyond adoption frequency, we also examine *how* CPs invoke the Topics API. Among the calls in D_{AA}, we find that JavaScript is the most prominent call type, accounting for 56% of total legitimate uses. Fetch comes at a close second at 39.4%, while IFrame is comparatively rare, found in only 4.6% of cases. Notably, doubleclick.com is the only *Allowed* CP that makes use of the IFrame call type.

V. ANOMALOUS USAGE

We now concentrate on the thousands of CPs in D_{AA} ($\approx 5\%$ of all the contacted domains) that call the Topics API despite not being *Allowed*. Recall that we observe them because we removed the *allow-list* from our crawler. With the correct configuration, the browser would not allow such a call. This behavior allows us to highlight incorrect implementations of the Topics API. First, we investigate in which context the call is executed — i.e., the environment in which the browser displays a Web page.[15] The non-*Allowed* CP often appears to coincide with the website we are visiting, i.e., the first party directly calls the Topics API, despite not being *Allowed*. Out of the 4,835 Topics API anomalous calls, 72.8% of them come from the same second-level domain of the website we are visiting, e.g., www.foo.com and ad.foo.net. A manual check on the remaining 27.2% reveals similar situations: i) the same company owns the two domains (e.g. windows.com and microsoft.com); ii) the visited website redirects to a second website which then calls the API — both websites being owned by the same company.

Second, all these calls use the JavaScript `browsingTopics()` function. This suggests that some popular JavaScript libraries may be erroneously accessing the Topics API. If loaded by a website, such a library would execute some calls from the website context (which is not *Allowed*).

To find a possible explanation, we observe the presence of Google Tag Manager's (GTM)[16] JavaScript scripts on 76.3% of the websites where anomalous calls occur, as opposed to 22% of penetration in websites where we do not observe anomalous calls. GTM does indeed contain a call to the `browsingTopics()` function, although it is neither *Allowed* nor *Attested*.

We then investigate why the call is executed as coming from the websites and not the GTM context. The browser downloads the script from a Google server with a link similar to https://www.googletagmanager.com/gtm.js?id=<ID>. The script is then executed within the *root* browsing context, resulting in having its browsing context origin[17] set to the website instead of the GTM context. As sketched in Fig. 4, this happens because the relevant `<script>` tag is placed directly inside the HTML content of the website's page and not included inside an `<iframe>` with an external source. As such, the crawler executes the Topics API call that appears as generated by the website itself.

[15]https://developer.mozilla.org/en-US/docs/Glossary/Browsing_context, accessed on September 25, 2025.

[16]https://tagmanager.google.com/, accessed on September 25, 2025.

[17]https://developer.mozilla.org/en-US/docs/Glossary/Origin, accessed on September 25, 2025.

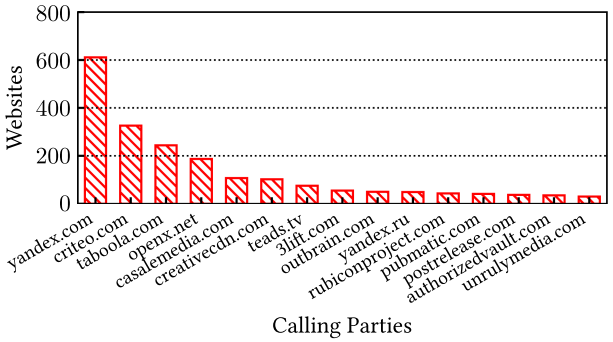


FIGURE 5. Number of questionable API calls by *Allowed* and *Attested* services. D_{BA} dataset.

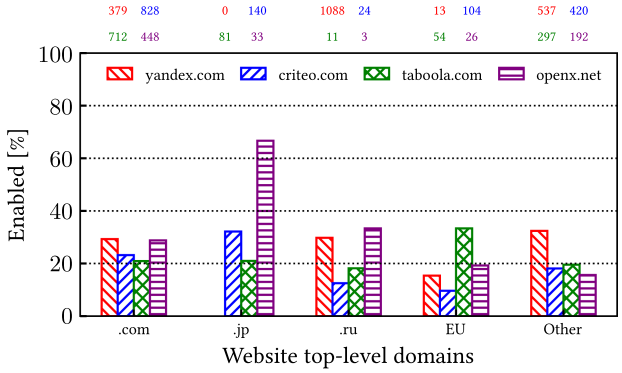


FIGURE 6. Share of websites where a CP calls the Topics API among all the websites under a given questionable TLD where it appears (D_{BA}).

The "wrong-context" problem is general and could complicate the deployment of Topics API solutions. This sort of behavior suggests that websites implementing the Topics API will have to be very careful about the inclusion of third-party scripts (like GTM), as they may cause unexpected and unwanted privacy issues. We contacted Google about this issue as well, but at the moment of writing we did not receive any response.

VI. QUESTIONABLE USAGE

We focus now on those Topics API calls that are performed in the *Before-Accept* visit by an *Attested* and *Allowed* CP. Ideally, we would expect no API usage since the user has not consented to the use of any personal data. However, as reported in Table 1, we observe 28 CPs that perform a call during the first visit. Given that we run our crawling campaign from Europe, we appear as a European citizen protected by the GDPR [11]. The above cases are all questionable and can be seen as a violation of said regulations, as one could consider the Topics API usage equivalent to using third-party cookies. Whether this could be considered an actual violation of the current legislation is outside of the scope of this paper. The fact that some services respect this interpretation reinforces our position.

Fig. 5 shows the number of websites where we observe a violation for a given CP. yandex.com comes first (611 calls in *Before-Accept*), even if it is not among the top legitimate callers (138 calls in *After-Accept*). In general, we observe little correlation with the service popularity. For instance, doubleclick.net, the top caller, does not perform any call in *Before-Accept* (and more than 2,500 in *After-Accept*). This corroborates the assumption that no call shall be issued in the *Before-Accept* visits.

At least two cases can justify this behaviour: (1) the website does not include any Privacy Banner — and privacy-invasive technologies can be used in every visit. This could be the case with a website outside EU.[18] (2) The website does not correctly implement a privacy banner, e.g., in a shallow-but-in-good-faith behaviour.

We investigate these cases by checking the top-level domain (TLD) of the websites where we observe a violation. We use the TLD as a coarse indication of website country. Here, we focus on the top 4 questionable CPs and break down their API calls by geographic region: `.com`, Japan (`.jp`), Russia (`.ru`), European Union (30 TLDs for EU countries where the GDPR is in force) and all the remaining TLDs. Fig. 6 reports the share of websites in which the given CP invokes the Topics API over the number of websites the CP is embedded in. The top x-axis indicates the latter number. We first observe that the presence of CPs strongly varies in different regions. Yandex, a Russian company, is not present in Japan and almost absent in the EU. Conversely, Criteo, based in France, has a worldwide marketplace. Looking at the different bars, we do not identify any clear trend. While the sizeable differences among CPs can be caused by different deployment strategies, we do not identify radical diversity across the geographical regions. We even observe questionable API calls also for websites in the EU, where the GDPR is expected to apply.

Next, we check if this questionable behaviour can be due to missing or incomplete configuration of the Privacy Banners by the website administrator. For this, we look for the Consent Management Platform (CMP) a website uses, if any. CMPs are commercial products which simplify the implementation of Privacy Banners. They offer standard libraries that control all the third parties embedded in the websites (such as advertisers or trackers), enabling them only after the user consents to the personal-data collection policy. They require minimal configuration by the website administrator. In case the CMP does not block all third party calls, third parties can exhibit non-GDPR-compliant behaviour, i.e., being active in *Before-Accept* [30]. As such, a website that adopts a CMP but allows CPs to call the Topics API on the *Before-Accept* visit (i.e., without user consent) is due to a CMP misconfiguration or bad CMP implementation.

We check which CMP is in use when we visit websites by relying on the list of the most widespread CMPs (identified

[18]This would still be a GDPR violation which protects Europeans even when accessing international services.

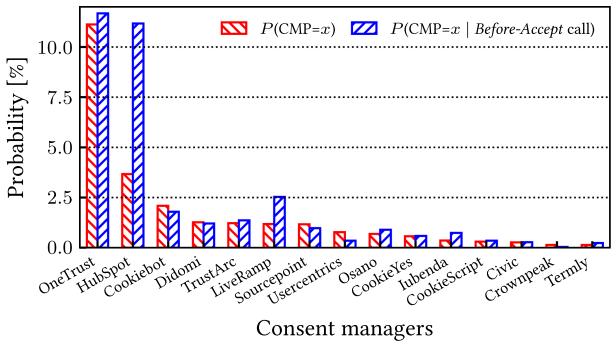


FIGURE 7. Probability of observing a CMP given (or not) a questionable API call (D_{BA}).

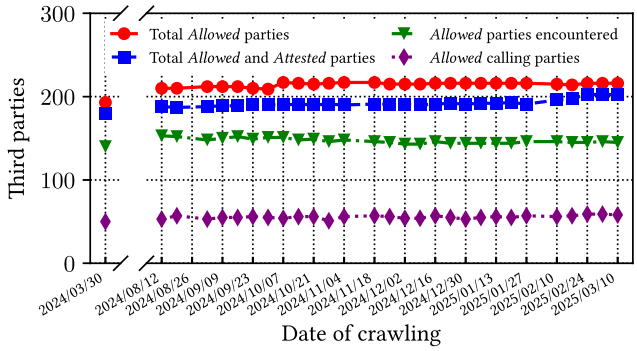


FIGURE 8. Evolution of *Allowed* and *Attested* parties using the Topics API from August 19, 2024, to March 10, 2025.

by their domain name) offered by Wappalyzer.[19] In Fig. 7, we show side-by-side the probability of observing a CMP over all websites ($P(\text{CMP} = x)$, red bars) and over websites where we observe a call to Topics API call in *Before-Accept* ($P(\text{CMP} = x \mid \text{Before-Accept call})$, blue bars). We conclude that the popularity of CMPs is generally independent of the presence of calls in *Before-Accept*, being the two probabilities equal: all CMPs suffer from the same problem. Some notable exceptions emerge: Hubspot has a probability of being the CMP in use given a call in *Before-Accept* which is $\approx 3\times$ the probability of observing it. As such, Hubspot shows potential shortcomings in its handling of the Topics API: the probability of a call in *Before-Accept* given the CMP is Hubspot is 12%, twice as big as the average probability. Liveramp shows a similar behaviour.

In a nutshell, the complexity of configuring and managing the privacy options has yet to properly integrate the support for the Topics API, allowing possible violations of privacy regulations. This leaves space for more in-depth analysis that we leave for future work.

VII. TEMPORAL ANALYSIS

We now extend our analysis to examine how the Topics API usage has evolved from our initial assessment in March 2024 until March 2025. To achieve this, we conduct a weekly measurement campaign over approximately six months, from August 19, 2024, to March 10, 2025. On a weekly basis, we visit the top-50,000 ranked websites according to a freshly-updated Tranco list. We access each website from our location in Italy, with a *Before-Accept* and (potentially) *After-Accept* visit.

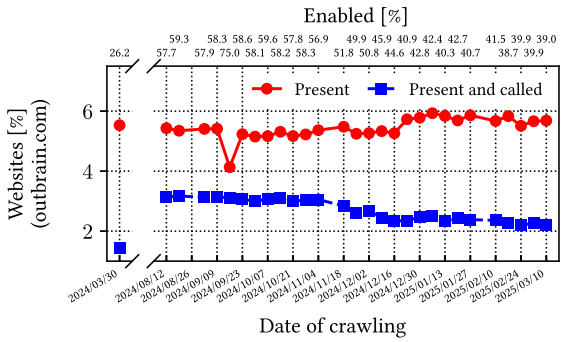
A. LEGITIMATE USAGE OVER TIME

We first start by observing the evolution of the number of legitimate (and active) parties in the system. Fig. 8 shows the evolution of four different counters: i) the number of *Allowed* domains (red line, round markers), ii) the number of domains that are both *Allowed* and *Attested* (blue line, square markers), iii) the number of domains in *Allowed* that we found at least once during our crawling on the given date (green line, triangle markers), iv) the number of domains in *Allowed* that called the Topics API at least once on the given date (purple line, diamond markers). Starting from August, Fig. 8 shows very little changes. Between March 2024 and August 2024, 18 new domains were onboarded — thus became *Allowed* — while only 3 left. Since then, the process slowed down, as in more than six months only 21 domains were included in the *Allowed*, and only 2 more were found to be *Attested*. Some domains left as well (15 from the *Allowed* and 9 from the *Attested*), for a net increase of 6 and 14, respectively. Curiously, we identify 5 domains that never expose any attestation file, and yet remain *Allowed* throughout the entire period of our analysis. Moreover, there are 3 domains that removed their attestation file at some point during this period but continue to remain *Allowed*. These parties, however, only call the Topics API on a negligible amount of websites.

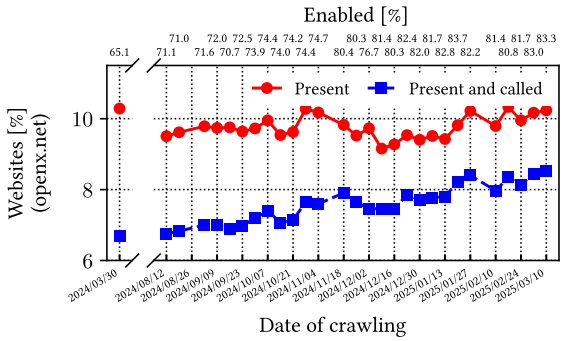
As observed in Table 1, only 40-50 out of the more than 200 *Allowed* domains actively use the Topics API during our measurement campaigns, with no increase over time. We argue that the uncertainty that still lingers around the Topics API and the Privacy Sandbox as a whole, the back and forth with W3C and the U.K. CMA, and the decision by Google to indefinitely postpone the phase-out of the third-party cookies makes potential players waiting for next events, without causing a growth for the Topics API usage, nor pushing to dismiss its usage.

We now focus on the individual behaviour of CPs. We take as an example Outbrain and OpenX, which exhibit illustrative trends. In Fig. 9, we show how their behavior changes over time. Fig. 9(a) reports the percentage of websites Outbrain is present in (red solid line) and the percentage of websites where it also invoked Topics API (blue dashed line). The upper x-axis reports the ratio between the two quantities — i.e., how frequently Outbrain invokes the Topics API. Outbrain is present in $\approx$5.5% of the websites, with a slight increase (never above 6%) in the end of 2024. We observe that, comparing March and August 2024, the enabled percentage increased

[19] https://www.wappalyzer.com/, accessed on September 25, 2025.

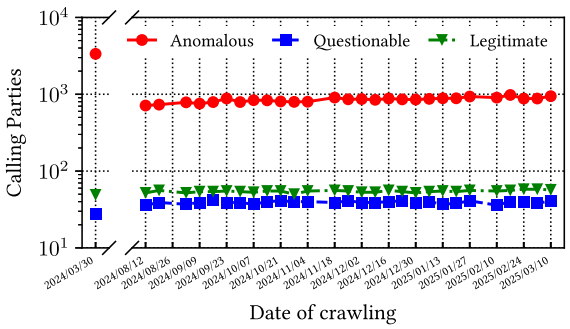


(a) Timeline for Outbrain.

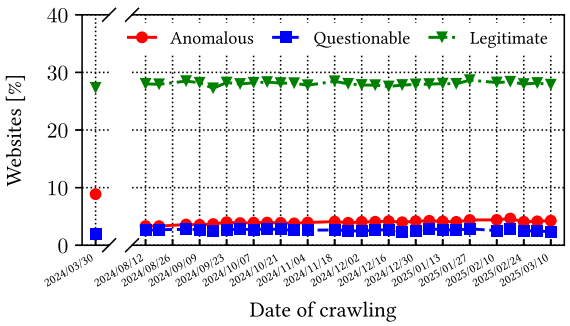


(b) Timeline for OpenX.

FIGURE 9. **Percentage of websites where the CP is present and where it enables the Topics API, across time. In the top row, the enabled percentage. (a) Timeline for Outbrain. (b) Timeline for OpenX.**



(a) Breakdown of the CPs calling the Topics API.



(b) Breakdown of the percentage of websites where calls to the Topics API happen.

FIGURE 10. **Legitimate, questionable, and anomalous calls to the Topics API, across time. (a) Breakdown of the CPs calling the Topics API. (b) Breakdown of the percentage of websites where calls to the Topics API happen.**

from 26.7% to 57.7%, lately starting to decrease again to 39% in March 2025. Conversely, OpenX (Fig. 9(b)) follows a different trend, constantly increasing the Topics API enabled percentage (from 65.1% in March 2024 to 83.3% in March 2025). In the uncertain scenario we were discussing above, every CP is pursuing its own Topics API deployment strategy.

B. QUESTIONABLE AND ANOMALOUS USAGE OVER TIME

We conclude the temporal analysis by breaking down calls to the Topics API in legitimate, questionable, and anomalous usage. Our goal is to understand whether the problems that affected the implementation of the Topics API in the early phase of its deployment are solved, at least partially. Fig. 10(a) shows the temporal evolution of the number of CPs issuing legitimate, questionable, and anomalous calls. Notice that a single CP can issue both Questionable and Legitimate calls (depending on the context), thus being counted in both cases. We first observe that the number of CPs being responsible of anomalous calls decreased from ≈3,300 to ≈700 in the period between March and August 2024, again increasing during the seven months of observations to ≈1,000 (note the log y-scale). We attribute the first drop to several JavaScript libraries having either removed the Topics API calls in their scripts, or fixed their implementation to correctly invoke the API from their own browsing context instead of the websites', as explained in Section VI. Over time, some new libraries may have been introducing the Topics API in their own scripts, which could explain the subsequent growth in anomalous CPs. A manual check on some websites confirms this, as we discover new libraries that added an invocation to their script, such as Google's Protocol Buffer library.

The number of CPs issuing legitimate and questionable calls is much lower, ≈60 and ≈40 respectively. Contrary to the anomalous calls, the trend for legitimate and questionable ones seems to be stable over time, reinforcing the conclusions drawn from Fig. 8.

A different picture emerges if we study the share of websites where a CP invokes the Topics API, shown in Fig. 10(b). In almost 30% of websites, we observe at least a legitimate call, while questionable and anomalous calls appear on less than 5% of websites. The observation that ≈60 CPs are responsible for the calls to the Topics API on almost 30% of the websites confirms the well-known notion that few third parties have tentacular reach in the Web [2], [20], and that such third parties have the scale to run collusion-based attacks as those presented in [26], [28]. The red curve shows that, coherently with what we have observed in Fig. 10(a), the number of websites where we observe anomalous calls is increasing between August 2024 and February 2025, although barely visible due

to the log y-scale. They represent a non-negligible share of the overall websites. This is expected, as anomalous calls are often issued by first parties themselves, due to configuration errors — as we discussed in Section VI. Most of those are first parties, thus impacting a relatively small share of websites.

VIII. GEOGRAPHICAL ANALYSIS

Besides the longitudinal dimension, we dig into the geographic aspect to explore whether the use of the Topics API varies based on the users' location. To this end, we look at measurements from five different countries in three continents. We use a VPN to set the egress point of our traffic to the desired location. We choose them to include different privacy regulation scenarios:

- *USA (Utah):* people in Utah are subject to the UCPA, which regulates tracking-related technologies similarly to the European GDPR, requiring clear and explicit consent by users before collecting or processing their personal data;
- *USA (California):* differently from Utah, the California privacy regulation (CCPA) only imposes websites to inform users of the presence of tracking tools like cookies and Topics API and mandates the presence of an opt-out button;
- *Brazil:* the Brazilian privacy regulation (LGPD) is similar in requirements to the European GDPR;
- *Japan:* the Japanese regulation (APPI) does not require consent to collect users' data.
- *Italy:* to cover the area subject to the GDPR, we repeat a measurement campaign from Italy in the same days as the others and using the same VPN software to have comparable results.

To reduce the impact of extra latency induced by the VPN, we increase the crawler's page load timeout to 30 seconds. This offers a reasonable trade-off between the probability of completing the page's download and the number of websites we can crawl in one day. In fact, the average webpage load time is ≈7.5 seconds, with only a few hundred exceeding the 30-second limit. We additionally configure the browser to request — when available — the English version of the website, so that there is a higher chance to obtain an English version of the Privacy Banner, which is easily identifiable by our crawler.

For each region, we visit the top websites according to the Tranco list on January 10, 2025, the day of the first crawling. We start by visiting the top-ranked websites and stop after completing a successful visit to 10 000 websites.

Fig. 11 shows the number of legitimate, questionable and anomalous calls for every region. We first compare the number of CPs calling the Topics API with Fig. 11(a). Overall, we observe no major difference: in all countries, ≈400–500 parties cause anomalous calls, while questionable and legitimate calls rest around the 50-CPs mark, meaning that the ad platforms which are experimenting with the Topics API are making worldwide testing. Among the 50 legitimate CPs encountered across all five crawls, we found that 15 do not

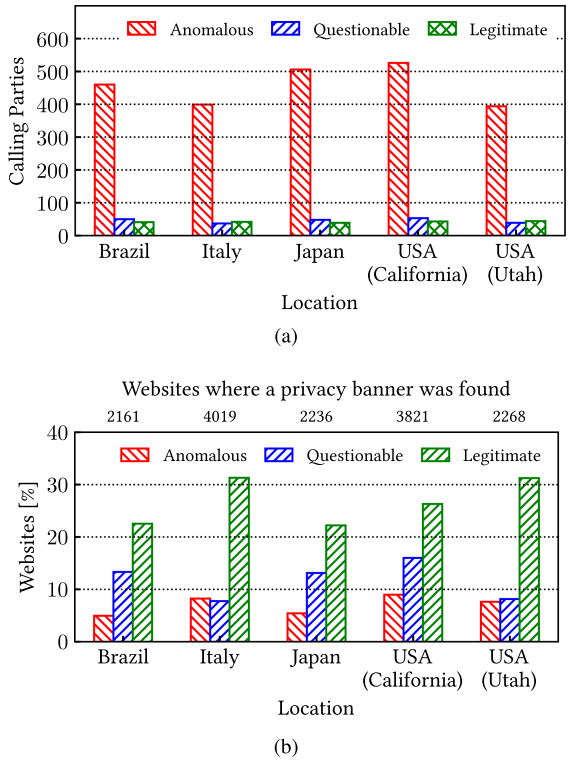


FIGURE 11. Legitimate, questionable, and anomalous calls to the Topics API, measured across different regions.

invoke the Topics API in every region. We notice some of them calling from a single location: namely, adstir.com from Japan, primecaster.com from Utah and relevant-digital.com from Italy. Additionally, we identify a few instances of unexpected behaviour: yahoo.co.jp, a Japanese CP, appears to perform no calls from Japan, while uol.com.br, a Brazilian CP, generates calls from every country except Brazil.

Fig. 11(b) shows the share of websites where each type of call is found. Again, we observe similar behaviour for all the regions, although with some caveats: first of all, we observe that Japan, California, and Brazil have a larger share of websites where we observe questionable calls. This is understandable for the first two, as the local regulations on privacy only mandate an opt-out mechanism, thus data collection is allowed without the need to accept on the privacy banner. Conversely, it is surprising for Brazil, which should follow the same principle as Utah and Italy. Brazil is also closer to California and Japan than Italy and Utah regarding legitimate calls. When filtering only those websites where a legitimate call was recorded in Italy and Utah but not in Brazil, we find that the crawler did not detect any privacy banner in the vast majority of them when accessed from Brazil. Manual visits confirm that, in most cases, the banner indeed does not appear only when the website is accessed from Brazil. In a smaller number of websites, the banner is shown in both Brazil and the other countries, but with a significant delay — potentially taking longer than 30 seconds in the original crawl. Only very few websites display the banner in Portuguese, despite setting

the browser language to English, with keywords that are not covered by *Priv-Accept*. Finally, notice that Italy also shows a larger share of websites with questionable and anomalous calls if compared to the non-VPN scenario — see for instance Fig. 10(b) around January 10, 2025. This is due to the different sets of websites considered in the two analyses: taking only the top-10,000 websites, the share of them observing a questionable or anomalous call increases just below 10%. Similar figures hold for Utah.

IX. CONCLUSION

This paper presented the first in-depth study on Topics API deployment. Using a carefully engineered custom-made crawler, we found that the most popular advertising platforms are already deploying and experimenting with the Topics API. We have measured activity compatible with forms of A/B tests on controlled subsets of websites and users. In this transition period, advertisers and third parties on the Web need to evaluate the potential impact of switching from a cookie-based approach to a Privacy-Sandbox one. It would be interesting to observe how third parties would act in browsers that no longer support third-party cookies. Unfortunately, these are also the browsers which, so far, do not to implement the Privacy Sandbox tools — most notably, Mozilla Firefox and Apple's Safari.[20]

We also testify how the deployment is in an early stage, and the introduction of such a new technology entails shortcomings, bugs, and unexpected behaviors of which all the stakeholders in the system — advertisers, public opinion, privacy advocates, and Google itself — should be aware: for instance, a non-negligible portion of websites and third parties fail in properly handling this new technology, invoking the Topics API even when the user has not explicitly opted in on personal data usage. In turn, we were able to discover such phenomena thanks to an issue in Chromium's Topics API implementation. This potential vulnerability is not isolated inside the Privacy Sandbox ecosystem [18], indicating the urge for a third-party risk analysis of the whole initiative. Continuous measurements across 2024 and early 2025 show that time mitigated initial issues with the implementation of the Topics API, reducing the amount of anomalous calls — that still persist nonetheless. A final, worldwide campaign confirms that the experimentation with the Topics API is running worldwide, irrespective of the users' location and the privacy regulations they are subject to.

Being proposed by Google, the Topics API may become the *de facto* standard for behavioral advertising and one of the pillars of the future Web ecosystem. According to our measurement (comforted by results obtained in [16]), the commercial acceptance of this technology is however still uncertain. Advertisers base their business model on fine-grained user profiling, which allows them to track a user's interest in a particular topic, brand, or even product. The Topics API, which is explicitly designed to create boundaries, may not be well-received, making the long-term implications of this technology difficult to predict. In this sense, Google's announcement that the Privacy Sandbox will indeed not replace third-party cookies, which will continue to be available in Chrome until further notice, casts shadows on the privacy-oriented goal of the initiative — the back-and-forth decision about the project recalling the concept of "cynical resignation" [32]. Google, as part of the Privacy Sandbox initiative, is developing other technologies to support behavioural advertising from different angles. They are designed to complement and partially relax the rigidity of the Topics API, allowing remarketing and real-time bidding directly from the browser. As a result, they may be more favorably received by advertisers, so their adoption and longevity will be worth watching closely.

[20]https://developer.mozilla.org/en-US/docs/Web/API/Topics_API, accessed on September 25, 2025.

REFERENCES

[1] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," in *Proc. 9th USENIX Symp. Networked Syst. Des. Implementation (NSDI 12)*, San Jose, CA, USENIX Association, Apr. 2012, pp. 155–168. [Online]. Available: https://www.usenix.org/conference/nsdi12/technical-sessions/presentation/roesner

[2] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, New York, NY, USA, Association for Computing Machinery, 2016, pp. 1388–1401, [Online]. Available: https://doi.org/10.1145/2976749.2978313

[3] B. Krishnamurthy and C. Wills, "Privacy diffusion on the web: A longitudinal perspective," in *Proc. 18th Inter. Conf. World Wide Web*, New York, NY, USA, 2009, pp. 541–550, [Online]. Available: https://doi.org/10.1145/1526709.1526782

[4] "AdBlock," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://adblockplus.org

[5] "Ghostery," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://www.ghostery.com

[6] "Disconnect," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://disconnect.me

[7] "Brave," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://brave.com/

[8] "DuckDuckGo," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://duckduckgo.com/

[9] "Safari," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://www.apple.com/safari/

[10] "Firefox," 2024. Accessed: Aug. 25, 2025. [Online]. Available: https://www.mozilla.org/en-US/firefox/

[11] European Parliament and Council of European Union, "Directive 95/46/EC. general data protection regulation," 2016. Accessed: Aug. 26, 2025. [Online]. Available: http://data.consilium.europa.eu/doc/document/ST-5419-2016-INIT/en/pdf

[12] California State Legislature, "California consumer privacy act of 2018," Accessed: Aug. 26, 2025. [Online]. Available: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[13] Brazilian President of the Republic, "Lei geral de proteção de dados pessoais," 2018. Accessed: Aug. 26, 2025. [Online]. Available: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709compilado.htm

[14] API Topics, 2024. Accessed: Aug. 26, 2025. [Online]. Available: https://developers.google.com/privacy-sandbox/relevance/topics

[15] "Sandbox Privacy," 2024. Accessed: Aug. 26, 2025. [Online]. Available: https://privacysandbox.com/

[16] G. Johnson, "Unearthing privacy-enhancing ad technologies (PEAT): The adoption of google's privacy sandbox," *SSRN*, 2024, Art. no. 58. [Online]. Available: https://dx.doi.org/10.2139/ssrn.4983927
[17] M. Philipse, G. Acar, and C. Utz, "Post-Third-Party Cookies: Analyzing Google's Protected Audience API," Masters Thesis, Radboud Univ., Nijmegen, NL, 2024. [Online]. Available: https://www.cs.ru.nl/masterstheses/2024/M_Philipse___Post-Third-Party_Cookies_Analyzing_Google's_Protected_Audience_API..pdf
[18] G. Calderonio, M. M. Ali, and J. Polakis, "Fledging will continue until privacy improves: Empirical analysis of google's privacy-preserving targeted advertising," in *Proc. 33 rd USENIX Secur. Symp. (USENIX Secur. 24)*, Philadelphia, PA, USA Aug. 2024, pp. 4121–4138. [Online]. Available: https://www.usenix.org/conference/usenixsecurity24/presentation/calderonio
[19] J. R. Mayer and J. C. Mitchell, "Third-party web tracking: Policy and technology," in *Proc. Symp. Secur. Privacy*, 2012, pp. 413–427.
[20] H. Metwalley, S. Traverso, M. Mellia, S. Miskovic, and M. Baldi, "The online tracking horde: A view from passive measurements," in *Proc. Int. Workshop Traffic Monit. Anal.*, Berlin, Germany, 2015, pp. 111–125.
[21] V. Rizzo, S. Traverso, and M. Mellia, "Unveiling web fingerprinting in the wild via code mining and machine learning," in *Proc. Privacy Enhancing Technol.*, 2021, vol. 2021, no. 1, pp. 43–63.
[22] E. Papadogiannakis, P. Papadopoulos, N. Kourtellis, and E. P. Markatos, "User tracking in the post-cookie era: How websites bypass GDPR consent to track users," in *Proc. Web Conf.*, New York, NY, USA, 2021, pp. 2130–2141. [Online]. Available: https://doi.org/10.1145/3442381.3450056
[23] A. Verna, N. Jha, M. Trevisan, and M. Mellia, "A first view of topics API usage in the wild," in *Proc. 20th Int. Conf. Emerg. Netw. EXperiments Technol.*, 2024, pp. 48–54.
[24] M. S. Alvim, N. Fernandes, A. McIver, and G. H. Nunes, "A quantitative information flow analysis of the topics API," in *WPES '23: Proc. 22nd Workshop Privacy Electron. Soc.*, New York, NY, USA, 2023, pp. 123–127. [Online]. Available: https://doi.org/10.1145/3603216.3624959
[25] C. Carey et al., "Measuring re-identification risk," in *Proc. ACM Manag. Data*, Jun. 2023, vol. 1, no. 2, pp. 1–26. [Online]. Available: https://doi.org/10.1145/3589294
[26] N. Jha, M. Trevisan, E. Leonardi, and M. Mellia, "On the robustness of topics api to a re-identification attack," in *Proc. Privacy Enhancing Technol. Privacy Enhancing Technol. Symp.*, 2023, pp. 66–78.
[27] Y. Beugin and P. McDaniel, "Interest-disclosing mechanisms for advertising are privacy-exposing (not preserving)," in *Proc. Privacy Enhancing Technol. Privacy Enhancing Technol. Symp.*, 2024, pp. 41–57.
[28] N. Jha, M. Trevisan, E. Leonardi, and M. Mellia, "Re-identification attacks against the topics API," *ACM Trans. Web*, vol. 18, no. 3, pp. 1–24, Aug. 2024. [Online]. Available: https://doi.org/10.1145/3675400
[29] Topics, "API integration guide| Privacy Sandbox | Google for Developers, 2024. Accessed: Aug. 26, 2025. [Online]. Available: https://developers.google.com/privacy-sandbox/relevance/topics/integration-guidecall_the_topics_api
[30] N. Jha, M. Trevisan, L. Vassio, and M. Mellia, "The internet with privacy policies: Measuring the web upon consent," *ACM Trans. Web*, vol. 16, no. 3, pp. 1–24, Aug. 2022. [Online]. Available: https://doi.org/10.1145/3555352
[31] V. Le Pochat, T. Van Goethem, S. Tajalizadehkhoob, M. Korczyński, and W. Joosen, "Tranco: A research-oriented top sites ranking hardened against manipulation," in *Proc. 26th Annu. Netw. Distrib. System Secur. Symp.*, Feb. 2019, pp. 1–15.
[32] S. Munir, K. Kollnig, A. Shuba, and Z. Shafiq, "Google's chrome antitrust paradox," *Vanderbilt J. Entertainment Tech. Law*, vol. 27, 2024, Art. no. 104.

ALBERTO VERNA received the BSc and MSc degrees in computer engineering, from Politecnico di Torino, Italy, in 2022 and 2024, respectively. He is currently working toward the PhD degree in computer and control engineering with Politecnico di Torino, Italy. His research interests primarily focus on privacy and security within the Web ecosystem, concentrating on the technical development and application of privacy-enhancing technologies.

NIKHIL JHA received the double-degree MSc in ICT for Smart Societies with Politecnico di Torino and in data science and engineering with EURECOM, France, in 2020. He is a Research Fellow at Politecnico di Torino, Italy. He defended the PhD degree in 2024, the thesis being titled "Privacy on the Web: Algorithms, Tools and Measurements". His research interests include 'pivot around users' privacy in the Web ecosystem, with ramifications in the Social Media and financial fields.

MARTINO TREVISAN received the MSc and Phd degrees in computer engineering from Politecnico di Torino, in 2015 and 2019, respectively. He is currently an assistant professor with the Department of Engineering and Architecture of the University of Trieste. During his career, he visited Télécom ParisTech (Paris), Cisco Systems labs (San Josè, US), AT&T (Bedminster, US) and the Universidade Federal de Minas Gerais (Brazil). He has authored or coauthored more than 40 papers in prestigious journals and conferences in the field of networking and Big Data. His research interests are mainly focused on Big Data methodologies for Web and Internet analysis. He also studies the operation of Online Social Networks and their implications on user behaviour.

MARCO MELLIA (Fellow, IEEE) is currently a full professor with Politecnico di Torino, Italy. He has coauthored more than 300 papers published in international journals and presented with leading conferences. His research interests are in the areas of Internet monitoring, users' characterisation, cyber security, and machine learning applied to different sectors. He is also the Editor in Chief of the Proceedings of the ACM on Networking. He won the IRTF ANR Prize with IETF-88, and the best paper awards with IEEE P2P'12, ACM CoNEXT'13, IEEE ICDCS'15, ACM CCR'16, ITC'18.