Assessing Compliance in Digital Advertising: A Deep Dive into Acceptable Ads Standards

Ahsan Zafar
azafar2@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Anupam Das
anupam.das@ncsu.edu
North Carolina State University
Raleigh, North Carolina, USA

Abstract

Online ads provide essential revenue for millions of websites but often disrupt user experience. To address this, browser extensions emerged to block intrusive ads, prompting the creation of the Acceptable Ads Standards to balance user choice and monetization. The Acceptable Ads Standards, initiated by the Acceptable Ads Committee, seek a balance between user experience and ad effectiveness, allowing certain non-intrusive ads defined by size, placement, and type limitations. This paper analyzes the compliance of digital advertisements with the Acceptable Ads standards by examining 10,000 popular domains intersecting Tranco's top 100K and the Acceptable Ads exception list. Our findings reveal that nearly 10% of these sites display non-compliant ads on landing pages, exposing design flaws in the exception list that allow publishers to bypass size and format restrictions. We propose enhancements to the exception list to better uphold user experience and ad integrity.

CCS Concepts

• **Information systems** → **Display advertising**; • **General and reference** → **Measurement**.

Keywords

Web Measurement; Digital Advertisement; Ad Standard

ACM Reference Format:
Ahsan Zafar and Anupam Das. 2025. Assessing Compliance in Digital Advertising: A Deep Dive into Acceptable Ads Standards. In *Proceedings of the ACM Web Conference 2025 (WWW '25), April 28-May 2, 2025, Sydney, NSW, Australia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3696410.3714725

1 Introduction

In today's digital age, the Internet has become an integral part of our lives, with a significant portion of our time spent navigating through its vast and diverse content. This digital journey often involves encounters with a wide array of online advertisements, which play a crucial role in the economic framework of the web. Advertisements online take various forms, including, but not limited to, display ads such as banners and pop-ups, video ads that often play before or during accessing online content, native ads that blend seamlessly with the content of the webpage, and interstitial ads that appear between page transitions.

While these advertisements are essential for keeping many websites operational and content freely accessible, their disruptive nature has raised concerns. Intrusive ads can lead to negative user experiences, prompting the development of ad standards aimed at improving the web environment. The Better Ads Standards [7] and Acceptable Ads Standard [5] emerged as key guidelines in this respect. The Better Ads Standards, initiated by the Coalition for Better Ads, target eliminating ads deemed excessively intrusive or bothersome to users. On the other hand, the Acceptable Ads Standard, guided by the Acceptable Ads Committee, strives to find a middle ground that allows for ads that are non-intrusive and acceptable to users, thereby ensuring that websites remain profitable without compromising the user experience.

This paper explores the landscape of online advertising, focusing on compliance with the Acceptable Ads Standard—the *default advertising policy* for popular ad blockers like Adblock Plus, impacting around 300 million users worldwide [13]. This standard permits certain non-intrusive ads, making its enforcement crucial. Unlike prior studies on standards such as Better Ads [45], our work is the first to evaluate compliance with the *stricter* Acceptable Ads Standard [1], assessing the effectiveness of its exception rules.

Our work seeks answer to the following research questions to better understand the compliance of acceptable ads.

RQ1: ***Are there non-compliant ads on partner websites exempted under the Acceptable Ads Standard? If so, how prevalent are they?*** We examine online ads on domains that are exempted from ad-blocking under Acceptable Ad's Standard. Additionally, we assessed the role of various ad publishers in contributing to these violations. Our study uncovers patterns of non-compliance and identifies major offenders. From our analysis of Tranco's top 100K websites that include exception rules for acceptable ads, we found that approximately 10% of these sites display at least one ad that fails to meet compliance standards.

RQ2: ***What flaws and limitations exist in the current exception list that contribute to the prevalence of violating ads?*** We utilize our telemetry data of violating ad elements from the web and evaluate the overly permissive rule structures from the exception list to find limitations in the current enforcement of acceptable ads.

RQ3: ***Can the exception list endorsed by the Acceptable Ad committee be enhanced to reduce the non-compliance rate of violating ads on partner websites?*** Based on our findings regarding the limitations and flaws of allowlist rules, we propose ways to enhance the enforcement of acceptable ads by implementing more precise allow rules and avoiding overly permissive ones. We

WWW '25, Sydney, NSW, Australia

ACM ISBN 979-8-4007-1274-6/25/04
https://doi.org/10.1145/3696410.3714725

evaluate our proposed enhancement by demonstrating reduced non-compliant ads when testing on real-world websites.

In summary, we make the following contributions to enhance the understanding and improvement of ad compliance on websites adhering to the Acceptable Ads Standard:

- We developed a web crawling tool that utilizes a proxy-based approach to inject scripts into web pages for measuring properties of web elements. Additionally, we crawl the same page using different configurations of ad filtering rules, enabling us to retrieve the ads of interest. Additionally, the injected script performs in-situ telemetry to identify ad elements that violate the Acceptable Ads Standard. We will also open-source our measurement framework to the public.
- We conducted a comprehensive web measurement study involving 10,000 domains selected from the intersection of Tranco's 100K domains and partner websites listed in the Acceptable Ads exception list. Our findings indicate that one in every ten websites displays at least one violating ad.
- Leveraging the telemetry data collected, we identified overly permissive rules and DOM elements consistently associated with violating ads. We implemented improvements to the exception list and demonstrated a reduction in violating partner websites by 32.4%.

These contributions enhance the existing literature on ad compliance and provide actionable insights for refining ad standards.

2 Background

Online advertising is a vital revenue stream for millions of websites, supporting free content and services. In the U.S. alone, the market exceeded $225 billion in 2023 [16], with ads appearing as text, video, pop-ups, and in-video formats. Despite their economic importance, intrusive ads often disrupt user experience, fueling the rise of ad-blocking software. These tools offer a cleaner, more private browsing experience by blocking or hiding unwanted ads but threaten ad-driven revenue models, prompting the creation of acceptable ad standards to balance user control and monetization.

2.1 Ad-Filtering and EasyList

A cornerstone of ad-filtering is the use of blocklists, which define the specific rules for identifying and blocking ads. One of the most prominent is EasyList [11], an open-source, community-maintained list widely adopted by ad blockers. It covers various ad types, including banners, pop-ups, and tracking elements. Integrated into popular tools like AdBlock Plus [3] and uBlock Origin [20], EasyList enables seamless filtering by specifying rules for URL patterns, CSS selectors, and scripts.

The list is continually updated by a group of volunteers and contributors who review and add new rules based on user submissions or as new advertising techniques emerge. EasyList also contains regional variations, known as supplementary lists, to accommodate language- and region-specific ads. While EasyList is highly effective at reducing intrusive advertisements, it can hurt the economic perspectives of domain owners who may rely on ad monetization. This has led to efforts to develop standards for non-intrusive ads that meet both user and publisher needs.

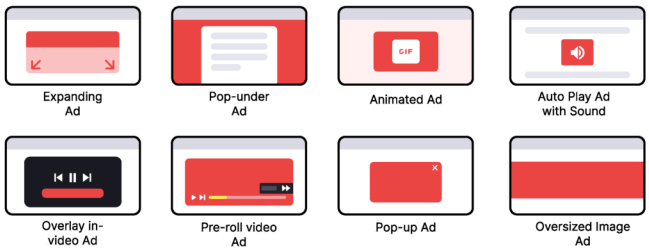

Figure 1: Prohibited ad formats under Acceptable Ads, with ad content highlighted in 'red.'

2.2 Acceptable Ad Standard

The Acceptable Ads Standard aims to balance user experience with website monetization by allowing certain non-intrusive ads that are less disruptive. It creates an exemption for non-intrusive ads by declaring rules in an Exception list [4] that is structured similarly to the Easylist. The standard describes in detail the distinction, size, and placement of the ads in the following manner:

- **Size:** Ads must occupy a reasonable amount of screen space, with specific size and dimension limitations.
- **Placement:** Ads should be clearly distinguishable from the primary content.
- **Labeling:** All ads must be clearly labeled as such.

In contrast, the following types of ads, shown in Figure 1, are deemed unacceptable and are considered violations of the standard:

- **Pop-ups and Pop-under Ads:** Ads that appear in separate frames, windows or tabs, either above or below the current context.
- **Animated Ads:** Advertisements with rapid animations or flashing effects.
- **Audio/Video Ads with Sound:** Advertisements that play audio or video with sound automatically upon loading of the page.
- **Ads Covering Content:** Ads that cover significant portions of the webpage's content.

These criteria formed the foundation for our heuristics to automate ad vetting, ensuring that only ads conforming to the Acceptable Ads standard are allowed. The heuristics, which we discuss in Section 4.3, were tailored specifically to desktop ads, with mobile ads being outside the scope of this analysis.

3 Related Work

Online Ads. The subject of digital advertising has attracted significant attention from various stakeholders within the online ecosystem [24, 30, 44]. Economic incentives have driven research into the effectiveness of different advertising formats and the key factors that capture user attention [28, 38, 40]. Some studies approach this issue from a privacy perspective, highlighting the potential harms posed by targeted advertising to users' privacy [23, 26]. These privacy concerns have fostered the development of tools and privacy controls designed to empower users to defend against tracking by advertising entities [34, 39]. Additionally, security researchers have exposed vulnerabilities in ad systems, demonstrating how they can be exploited to perpetrate fraud against online users [22, 37]. For example, Oentaryo et al. [35] outlines methods for detecting fraudulent ad publishers who generate deceptive ad links aimed at misleading users.

Ad Blocking. Prior research has extensively examined various aspects of online ad experiences, with particular attention to intrusive ads. In response, several tools and extensions have been developed for ad blocking, such as Adblock Plus [3], uBlock Origin [20], and Ghostery [2]. These tools primarily rely on community-maintained blocking lists, like EasyList (for ads) [11] and EasyPrivacy (for trackers) [12], to block specific content URLs. Additionally, some studies have explored automated approaches, such as machine learning classifiers, to adapt to evolving ad and tracker characteristics [29, 33]. These solutions aim to block all types of ads.

Ad Compliance. The issue of ad compliance has become increasingly prominent as it pertains to the quality of online ad experiences. Early on, governmental organizations such as the FTC established guidelines to promote greater transparency among websites and publishers within the digital advertising ecosystem [25]. These governmental frameworks have spurred the creation of self-regulatory bodies by ad publishers to ensure compliance with industry standards [9, 15]. Various studies have evaluated publisher adherence to organizations such as NAI and DAA [32]. More recently, regulations focused on user data, including GDPR and CCPA, have had a profound impact on the digital advertising landscape [8, 27]. Research has examined the effects of these data protection regulations on advertising practices [41, 43], with findings showing that despite such regulations, ad publishers continue to adapt their methods to collect user data for targeted advertising.

In addition to these regulatory frameworks, ad policies like the Acceptable Ads Standard [5] and the Better Ads Standard [7] have provided explicit guidance on ad practices, such as size, placement, and display rules, to minimize disruption to users while allowing site owners to earn through ad monetization. Researchers have studied the impact and privacy implications of these ad policies [42, 47], and Yan et al. have quantitatively assessed the effectiveness of the Better Ads Standards [45]. To our knowledge, however, we are the *first* to conduct a detailed examination of compliance with the Acceptable Ads Standard by partner websites and publishers.

4 Methodology

This section details our methodology for crawling web pages to identify ad types. It covers the technical specifications of our crawler, including configurations for ad discovery and the design of heuristics to detect non-compliant ads and assess compliance rates. To promote transparency in digital advertising research, we have open-sourced our tool. [1]

4.1 Ad Filtering Configurations

We utilize the functionality of AdBlock Plus [3] to block/allow ads on the webpage. The extension allows configuring various blocking and allow lists. For our approach, we develop three configurations that are important to the two-crawl process:

- $C_{\textbf{Ads}}$: The first crawl operates without any list, while the second uses *only* EasyList to block all ads. The difference between these crawls captures all ads, forming the dataset $\mathcal{D}_{\text{Ads}}$, which we use to report the overall prevalence of ads on the web.
- $C_{\textbf{acceptableAds}}$: The first crawl uses EasyList along with the Exception list, while the second uses only EasyList. The difference between the two captures acceptable ads [2]. We denote this dataset as $\mathcal{D}_{\text{acceptableAds}}$ and use it to report the frequency of prohibited ad types observed despite the enforcement of the Acceptable Ads Standard through the exception list.
- $C_{\textbf{modified}}$:In this study, we propose improvements to the Exception list and evaluate their impact by modifying the list and implementing the following setup: the first crawl uses EasyList with the modified Exception list, while the second uses only EasyList. Comparing the results reveals acceptable ads after filtering out non-compliant ones. We refer to this dataset as $\mathcal{D}_{\text{modified}}$ and use it to demonstrate the effectiveness of our proposed changes in enhancing adherence to the Acceptable Ads Standard.

4.2 Two-Crawl Detection Approach

Our configurations enable the tool to capture the required ads based on the configuration we use. Below, we outline the tool's workflow as it crawls webpages and collects telemetry data.

For each domain, the tool performs two consecutive crawls with a 10-second delay between them, minimizing webpage changes during this period [21]. During each crawl, a script is injected into the webpage's head using `mitmproxy` [14], with the `defer` attribute set to execute before the `DOMContentLoaded` event. This script scans the page and lists all elements, including media content like images, videos, and SVG files. After all resources (scripts, images, subdocuments, etc.) load and trigger the `Load` event, the script waits an additional 10 seconds to ensure rendering completion before traversing the `DOM`. If the `Load` event does not trigger, the tool stays on the page for up to 60 seconds before terminating. If loading fails, the domain is retried once before being excluded.

Since frames are isolated by same-origin-policy, we use `mitmproxy` to modify Cross-Origin Resource Sharing (CORS) flags and configure the browser to disable web security. This ensures the script can access all resources loaded in the browser.

The script captures details such as CSS properties, class names, XPaths, optimized XPaths, and other attributes of each element. It is important to note that ad resources are typically placed in well-defined sections of the `DOM`, and repeated visits to the webpage will render ads in the same locations. Ad-blocking tools utilize this deterministic behavior to hide the DIV elements assigned to ad networks. We also leverage this determinism to identify ads wherever they appear on the rendered webpage.

By comparing the lists of web elements generated from the two crawls using XPaths, we can identify the content blocked by AdBlock Plus and determine which elements were flagged as ads (as shown in Figure 2).

4.3 Detection of Non-Compliant Ad Categories

Both the Better Ads Standards and Acceptable Ads Standards block certain common ad formats, such as pop-ups, pop-unders, autoplay media, interstitials, and overlays. Drawing inspiration from Yan et al.'s work on ad types prohibited under the Better Ads Standard [45], we developed heuristics for identifying forbidden ad types within

[1] Our tool is open-sourced at: https://github.com/ahsan238/Ad-Compliance

[2] Acceptable Ads are a subset of generic web ads.

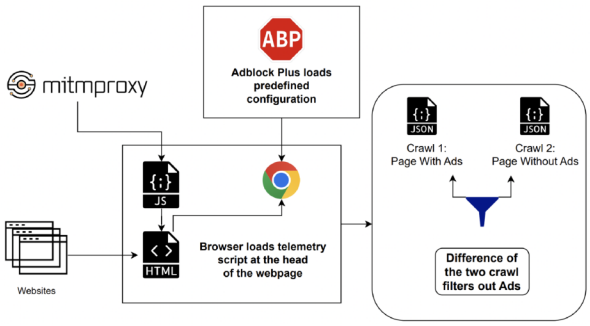

Figure 2: Two-crawl ad detection. `Mitmproxy` injects a scanner script that traverses DOM object. Difference of the two crawl reveals ads.

the Acceptable Ads Standard. We define heuristics for six specific non-compliant ad types based on their CSS properties. These formats were selected for their clear violation criteria and detectability, as highlighted in prior research [45]. These include:

- **Over-Sized Image Ads**: While The Acceptable Ads Standard prohibits 'Generally Oversized Image' ads but does not specify exact size limitations. For our analysis, we consider ads occupying more than 80% of the screen's width or height blatantly oversized.
- **Autoplay Media**: Media ads are identified if they have the `autoplay` attribute enabled or are automatically preloaded. This applies to both videos and images.
- **Overlay Ads**: There are ads with sticky ads that are placed on the top or at the edge of the viewport and have `fixed` or `absolute` positioning.
- **Popup Ads**: Popups are detected by their high `z-index` combined with `fixed` or `absolute` positioning.
- **Popunder Ads**: Popunders are similar to popups but are positioned beneath content, indicated by a negative `z-index`.
- **Interstitial Ads**: These are full-screen ads detected if they cover more than 75% of the viewport and have `fixed` or `absolute` positioning.

These rules enable the systematic identification of different types of non-compliant ads across crawled domains.

5 Data Collection

To collect data for measuring online advertisements and their compliance with Acceptable Ads Standard, we developed a specialized tool, the details for which have been described in Section 4.2. This section focuses on the process and setup used for data collection, including the selection of websites and the technical infrastructure used for crawling.

5.1 Website Selection

We aimed to analyze advertisements on a broad and diverse set of domains. To achieve this, we chose to crawl websites that are in the intersection of two specific lists: the Tranco top 100K websites [19] and the first-party domains found in the Acceptable Ads Standard's Exception list [4]. The Tranco list is a frequently updated ranking of the most popular websites on the Internet, ensuring that our dataset reflects domains with significant user traffic. The exception list, on the other hand, contains rules for allowing ads on certain partner domains, provided they comply with Acceptable Ads Standard.

Table 1: Number of partner domains per rank division analyzed in our measurement.

Rank Division	Domain Count
1-1,000	326
1,001-10,000	1,791
10,001-100,000	7,883

Table 1 shows the number of partner domains and their corresponding rank divisions. By selecting domains present in both the Tranco top 100K and the Acceptable Ads Exception list, we ensured our dataset included high-traffic websites that display ads and are subject to compliance regulations. Although the intersection yielded 11K common domains, we focused on the top 10K for analyzing offending ads. After excluding inactive or publicly inaccessible domains (e.g., CDNs), we were left with a set of 9,463 domains for our analysis.

5.2 Crawling Setup

The data collection process was conducted on a server with 32 cores and 64 GB of RAM, enabling us to run 30 parallel crawling processes to expedite the data collection workflow. Each crawling process was tasked with visiting the selected domains and capturing the ad content displayed on the webpages. For automation, we used Puppeteer [17], a `Node.js` library that provides a high-level API for controlling headless browsers. To avoid detection by bot-detection algorithms, we employed Puppeteer's stealth plugin [18]. Additionally, we incorporated randomization in the scrolling behavior during each crawl to mimic human interaction patterns (e.g., each iteration of scroll-up and scroll-down had a random factor of movement), further evading potential bot detection.

5.3 Crawling Process

Our crawling was limited to each domain's landing page. We detected page load failures by inspecting the `mitmdump` file generated by the proxy, which records network exchanges. If the file indicated a server failure, we reattempted the crawl once.

The browser remained on the page until the `Load` event was triggered, signaling that the initial content had fully loaded. To ensure all dynamically-rendered content, including ads, was captured, we allowed an additional 10 seconds of idle time. Additionally, we took screenshots of the web interaction before closing the browser, which were later used to report the violations.

Crawling the 10K domains took approximately two days. Data collection was conducted in the United States to ensure consistency and avoid regional variations in ad serving practices. By leveraging parallel processes and a curated domain set, our infrastructure enabled efficient and comprehensive data collection, facilitating an effective evaluation of ad compliance on high-traffic websites.

5.4 Ethical Considerations

We were mindful of the ethical concerns associated with web measurement and data collection. Our crawling method injected a telemetry script into web pages, solely for inspecting on-page elements without altering or manipulating the content. No additional

Table 2: Counts of forbidden ad types using $C_{\text{acceptableAds}}$ and C_{Ads}.

Ad Types	$\mathcal{D}_{\text{acceptableAds}}$	$\mathcal{D}_{\text{Ads}}$
Oversized Ads	3,410	22,878
Autoplay Media	20	297
Overlay Ads	3,865	23,891
Interstitial Ads	121	3,503
Popup Ads	465	2,529
Popunder Ads	29	358
Total	**7,910**	**53,456**

requests were made to website servers; the approach mirrors functions available through browser developer tools like Chrome Dev Tools. To minimize our impact, we limited site visits to a maximum of 60 seconds before closing the browser, ensuring no undue server load or risk of disruptions such as DDOS attacks.

6 Prevalence of Violating Ads

This section will discuss our analysis on non-compliant ads found in our datasets, thereby answering **RQ1**: *What is the prevalence of non-compliant advertisements on partner websites exempted under the Acceptable Ads Standard?*

6.1 Effectiveness of Acceptable Ads Standard

As a first step towards understanding the prevalence of non-compliant ads on the web, we selected 10,000 domains shared between Tranco's top 100K and the Acceptable Ads Standard's Exception list. This list provides filter rules to unblock ads on specific domains that are expected to comply with the acceptable ad standard.

Our analysis focused on six categories of non-compliant ads defined by the Acceptable Ads Standard. Table 2 summarizes the detected instances across all domains, highlighting frequencies in the dataset $\mathcal{D}_{\text{acceptableAds}}$. For comparison, we also present counts from $\mathcal{D}_{\text{Ads}}$. The contrast between these configurations reveals the prevalence of ads that fail to meet the Acceptable Ads Standard.

For example, the clean browser profile encountered 4.86 times more overlay ads than the ad-blocking-enabled profile. A Chi-Squared test [36] comparing ad type distributions between configurations with and without Adblock Plus revealed a highly significant difference ($\chi^2 = 1907.24$, $p < 0.0001$), indicating statistically distinct ad type distributions. Specifically, ad types such as oversized images, overlay ads, and interstitial ads show substantial differences, confirming that Adblock Plus effectively blocks or reduces certain types of ads more than others.

While the combination of EasyList and the Exception list effectively reduces intrusive ads, it does not fully eliminate non-compliant ones. Our study found that 9.91% of websites in the dataset displayed at least one violating ad among those considered acceptable. Ideally, the $C_{\text{acceptableAds}}$ configuration should block all ads that breach the Acceptable Ads Standard. However, our findings indicate that the Exception list fails to fully enforce compliance, allowing ads that violate size and type restrictions to be displayed.

Finding 1: *Despite the effectiveness of EasyList in combination with the Exception list at reducing non-compliant ads significantly, we found violating ads to be present in 9.91% of the domains that were partners in Acceptable Ads Standard's exception list.*

Among the 9.91% of domains displaying violating ads, we analyzed the types of violations. Figure 3 illustrates their distribution, with each peak representing the frequency of a specific ad type per domain. The most common violations were Oversized Image Ads and Overlay Ads. Notably, some domains exhibited higher violation frequencies. For example, `naszemiasto.pl` displayed 103 Overlay Ads despite filtering under the Acceptable Ads Standard, driven by continuous scrolling that triggered the heuristics outlined in Section 4.3. Similarly, `express.co.uk` featured oversized ads below the primary content, violating size restrictions. Autoplay Ads were the least frequent violation but appeared on high-traffic sites such as `gsmarena.com`. Appendix B provides example screenshots of these violations.

Finding 2: *Oversized Image Ads and Overlay Ads were observed to be the most common non-compliant ads in our dataset.*

6.2 Non-compliant Ad Publishers

We also assess the contribution of various ad publishers who display violating ads in $\mathcal{D}_{\text{acceptableAds}}$. As discussed in Section 4.2, our crawl gathers CSS properties of ad elements. For more complex cases, we also navigate through parent nodes, collecting class names and other CSS selectors if available. This method provides a rich metadata inventory for each ad, which can be used to trace the ad publishers that display the ad.

Additionally, we match this data with filter rules from the exception list to identify the publisher responsible for showing the ad. Figure 4 shows an example of a metadata report generated after crawling `gsmarena.com`. It highlights metadata from an overlay ad, which violates the Acceptable Ads Standard, published by `brwsrfrm.com`. The `filter_rule` refers to the exception list rule that allows this ad. In this instance, any ad element within a `DIV` with the classname `ad_label` is permitted, even if it violates other properties, such as triggering sticky or overlay behaviors.

```
{
    "ad_src": '/d/img/1108/fd3b2628-68dc-4bab-94a8-
        d6caa14bd2bf/14398?bid=0&w=300&h=600'
    "violation": ['OVERLAY'],
    'tag': 'IMG',
    'parent_tag': 'DIV',
    'parent_class': 'ad-label'
    'parent_src': 'https://brwsrfrm.com/d/if/.../14398?bid
        =0&w=300&h=600/.../'
    'filter_rule': 'gsmarena.com,aternos.org#@#.ad-label'
}
```

Figure 4: Example of metadata of an overlay ad found on `gsmarena.com`

From this example, we can conclude that `brwsrfrm.com` is an ad publisher associated with a non-compliant ad instance. Our inventory of non-compliant ads identified during our crawling revealed multiple other publishers displaying invasive ads in a similar manner. To determine which ad publishers frequently displayed these

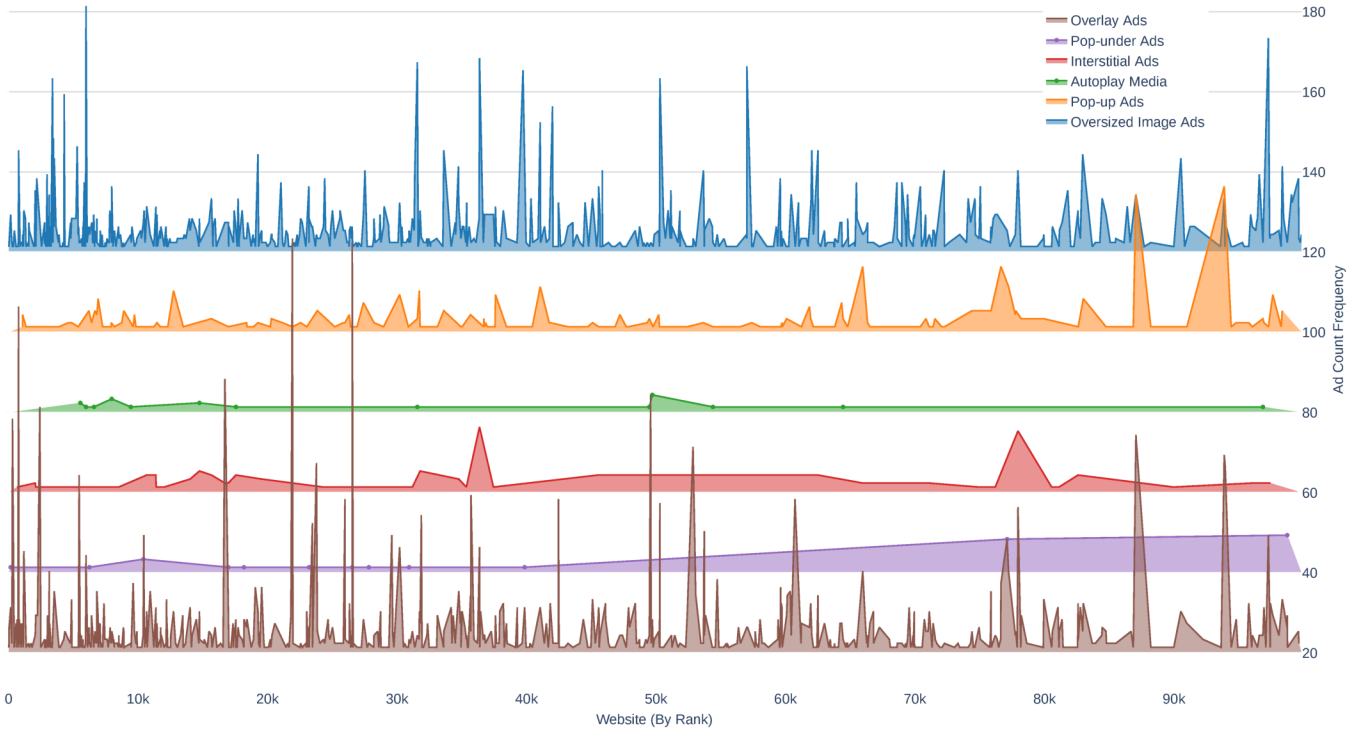

Figure 3: Distribution of violating ads under Acceptable Ads Standard across the different ranked websites. Violations appear across all ranks.

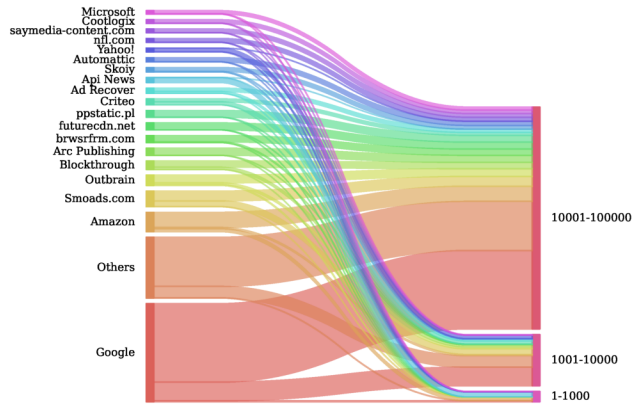

Figure 5: Different ad networks' role in the distribution of non-compliant ads.

violations, we utilized our metadata inventory to identify the parent companies that own these networks. First, we extract the URL of the ad element itself. If the URL is absent, we fetch the container frame's URL. Then, we extract the `tld+1` (Top-level domain + 1) out of the URL. Lastly, we match these domains with entity and services domains found in the list of ad companies, along with the respective services under them, curated by Disconnect [10]. Figure 5 presents a Sankey diagram visualizing the flow of violating ads. On the left, the diagram highlights the top ad publishers as the sources, with the width of each flow representing the volume of violations. On the right, these flows are categorized into three distinct buckets based on domain ranks. The top five ad publishing companies displaying violating ads are `Google`, `Amazon`, `Smoads.com`, `Outbrain`, and `Blockthrough`. This analysis underscores the role of various ad networks in perpetuating non-compliant advertising practices.

> **Finding 3**: *The top five ad publishing companies that had displayed violating ads are Google, Amazon, Smoads.com, Outbrain and Blockthrough.*

6.3 Temporal Consistency of Violations

The presence and nature of ads on the web are subject to significant fluctuations due to various dynamic factors. Live global events, time of day, regional politics, and evolving user interests all influence the ads displayed. Additionally, ad publishers engage in real-time bidding [46], competing for ad space, which can further impact what users see at any given moment. Due to these variables, performing a second round of crawl for all the domains is crucial for assessing the consistency of ad violations. This approach ensures a more accurate evaluation of compliance over time, accounting for the changing landscape of digital advertising.

To assess the consistency of violating ads, we conducted two rounds of crawls: one on September 7, 2024 (Crawl 1) and the other on October 7, 2024 (Crawl 2) using the $C_{\text{acceptableAds}}$ configuration. We compared the proportions of various ad types found to be in violation. The proportions of violating ad types in both crawls were largely consistent. For example, oversized images accounted for

Table 3: Counts of non-compliant ads using $C_{\text{acceptableAds}}$ at two distinct time stamps.

Ad Type	09/7/2024	10/7/2024	Change (%)
Oversized Image	3,410	3,203	6.07% ↓
Autoplay Media	20	19	5.0% ↓
Overlay Ads	3,865	3,593	7.03% ↓
Interstitial Ads	121	105	13.2% ↓
Popup Ads	465	440	5.37% ↓
Popunder Ads	29	28	3.44% ↓

43.11% (3,410) of violations in Crawl 1 and 43.61% (3,203) in Crawl 2. Overlay ads made up 48.86% (3,865) of violations in Crawl 1 and 48.38% (3,593) in Crawl 2. Differences in other ad types were similarly minimal. Further details on these discrepancies are provided in Table 3.

A Chi-squared test was performed to assess the statistical significance of differences in ad distributions, resulting in a statistic of $\chi^2 = 1.9432$ with 5 degrees of freedom with $p = 0.8569$ (since there are 6 ad categories, the degrees of freedom is 5), indicating no significant difference. The expected and observed frequencies for each ad type were closely aligned, suggesting consistency between the two crawls. Furthermore, the Kullback-Leibler divergence [31] of 0.000259 further confirms minimal divergence, reinforcing that the distribution of violating ad types remained consistent across the two time periods.

Finding 4: *The results of the temporal analysis indicate no significant differences in the distribution of violating ad types between the two crawls.*

7 Improving Acceptable Ads Standard

This section aims to identify the underlying causes of non-compliant ads and explore ways to improve the Exception list by removing problematic rules that allow these ads to disrupt the user experience. We achieve this by parsing the list and analyzing the telemetry data collected during our crawl of the violating ads.

7.1 Primary Causes for Non Compliant Ads

We provide the primary causes that lead to the display of non-compliant ads, thereby addressing **RQ2**: *What limitations and flaws currently exist in the Exception list?* To answer this, we inspect the potential sources of violating ads in our dataset.

Over-Permissive Rule.

ccWe analyzed the Exception list and identified rules that matched with the violating domains in $C_{\text{acceptableAds}}$. In some cases, we found that `^$document` unblocking rules were being enforced. These rules effectively create an exception for the entire domain by bypassing any ad-blocking restrictions across the site [6]. Among the 9.91% of domains displaying non-compliant ads, 52 domains (5.34% of the non-compliant sites) were found to have this `^$document` allowlisting rule. This high prevalence highlights a significant flaw in the enforcement mechanism, as such unrestricted allowlisting can result in the display of invasive ads. We argue that such rules undermine the intent of the Acceptable Ads Standard, which seeks to balance user experience with monetization.

Finding 5: *Among the 9.91% of domains exhibiting non-compliant ads, 52 domains utilized the overly permissive* `^$document` *allowlisting rule.*

Offending Element Unblocking Rule. We also identify the container elements where non-compliant ads are rendered. These containers are detected and unblocked by their `classNames` using Exception list rules. Figure 4 shows an example of the metadata of violating ads, compiled during and after the crawl. This metadata includes the class and tag type of the parent node where the ad is embedded. From both crawls (09/07 and 10/07), we observe that these parent nodes consistently display the same violating ad format, regardless of the ad publisher, across multiple site visits.

To mitigate the impact of violating elements, we gather all relevant rules that correspond to these elements in our dataset. 6 highlights some examples of rules within the Exception list that unblock offending content.

```
speedtest.net#@#.ads-right
cnn.com#@#.stack__ads
@@||teva.com^$document
knowyourmeme.com#@#.ad-unit-wrapper
pagesix.com,decider.com,nypost.com#@#.billboard-ad
```

Figure 6: Examples of exception rules that are removed from the Exception list.

7.2 Improving Compliance

In this section, we address **RQ3**: *Can the Exception list be refined to improve ad compliance rates?* Building on insights from Section 7.1, we identify overly permissive rules and those that enable exceptions for parent containers consistently serving non-compliant ads. Importantly, we limit rule modifications to the specific domains where violations were detected, ensuring that compliant domains remain unaffected.

Using these updates, we established a new configuration, Cmodified, incorporating the revised Exception list. We configured the crawler with this setup to assess the occurrence of violating ad types across the 937 offending domains identified in $\mathcal{D}$acceptableAds. Table 4 summarizes the changes, revealing a significant 32.4% reduction in violating domains. To evaluate the impact on non-violating ads incorrectly blocked, we randomly sampled 100 domains from the offending set and manually examined ads from the six categories. Based on a detailed analysis of 982 reported violations, we achieved a precision of 82.9% and a recall of 87%. Our modifications to the Exception list resulted in the unintended blocking of 122 non-offending ads (out of 982) due to the absence of fine-grained exception rules that could replace the overly permissive ones. Developing such nuanced and flexible rules is beyond the scope of this paper and remains an area for future research.

The results indicate a notable reduction in the overall counts of ad types, with oversized image ads and overlay ads decreasing by 22.7% and 25.9%, respectively. In contrast, interstitial ads saw a slight increase of 0.83%, likely due to fluctuations in this ad type. Marginal changes were observed for autoplay media, pop-up, and

Table 4: Counts of violating domains and forbidden ad types found using $C_{\text{acceptableAds}}$ and C_{modified}

Category	$\mathcal{D}_{\text{acceptableAds}}$	$\mathcal{D}_{\text{modified}}$	Change (Δ)
Violating Domains	937	634	32.4% ↓
Oversized Image	3,410	2,636	22.7% ↓
Autoplay Media	20	19	5.0% ↓
Overlay Ads	3,865	2,864	25.9% ↓
Interstitial Ads	121	122	0.83% ↑
Popup Ads	465	458	1.51% ↓
Popunder Ads	29	28	3.44% ↓

pop-under ads. These findings suggest that removing CSS class identifier exemptions and the `^$document` allowlisting rules primarily impacted oversized and overlay ads.

These findings underscore the effectiveness of refining the Exception list in reducing non-compliance and enhancing user experience. Consequently, we recommend removing overly permissive and offending-element unblocking rules to further strengthen compliance.

Finding 6: *Improving Exception list by removing offending element-hiding and overly permissive* `^$document` *rules predominantly impacts oversized image and overlay ads.*

8 Discussion

This study was driven by the heavy reliance of approximately 300 million users on the Acceptable Ads Standard, which promises a browsing experience free from invasive ads. The framework's standards are essential for enforcing accountability among partner domains and ad publishers, striking a balance between protecting the user experience and allowing domain owners to generate revenue through advertising. Despite the existence of these policies and a well-defined Exception list, our findings reveal that non-compliant ads remain prevalent. Specifically, our analysis shows that one in ten partner websites still displays at least one type of violating ad. The violations include oversized images, autoplay media, overlay ads, interstitial ads, and popups—all of which degrade the user experience that the Acceptable Ads Standard aims to safeguard. These results highlight gaps in enforcement and suggest the need for stronger measures to ensure compliance.

To address these issues, we refined the Exception list by identifying and removing rules that were either overly permissive or unblocked DOM elements consistently containing violating ads. This effort marks a crucial step toward ensuring that standards impacting millions of users are continuously improved. Our telemetry data offers valuable insights to guide future modifications, fostering a more compliant advertising ecosystem and enhancing the overall user experience.

Ethical Disclosure. The findings from this study were shared with the Acceptable Ads Committee. Following discussions with the Acceptable Ads Monitoring Manager, who reviewed and confirmed the sample of offending ads, it was revealed that most violations stemmed from programmatic ads. These ads are inherently less predictable in terms of content and format, making compliance enforcement more challenging. Although the committee actively monitors and penalizes non-compliant ad partners, addressing violations within programmatic ad frames remains an ongoing challenge.

Limitations and Future Work. Our work marks an important first step in assessing the enforcement of the Acceptable Ads Standard by developing an online telemetry pipeline to detect non-compliant ads. However, some limitations remain. First, our method relies on the deterministic embedding of third-party resources to detect ads across two consecutive crawls. Although we wait for the `LOAD` event (indicating all resources have loaded), it does not ensure that all ads are in display mode, potentially resulting in an *underestimation* of violating ads.

Second, our work primarily focused on six key types of ads identified as unacceptable. While the Acceptable Ads Standard imposes strict limits on ad size, placement, and type, the complexities of measuring web elements restricted our ability to develop a fully comprehensive set of heuristics. The number of violations reported in our study may represent a lower bound.

Future research should aim to expand the set of heuristics to capture additional violations and improve the thoroughness of compliance checks. Secondly, future research can also focus on the detection of non-compliant advertisements from other vantage points, including but not limited to device-centric ads (e.g. mobile ecosystem) or geolocation-based to assess the impact of local laws (e.g., GDPR vs. CCPA) on the distribution of offending ads. Furthermore, advancements in machine learning, particularly large language models (LLMs), could significantly enhance the detection of placement violations and help identify non-compliant ads across a broader range of contexts on the web.

9 Conclusion

In this paper, we assess the compliance of digital ads allowed under the Acceptable Ads Standard. We develop a crawling framework that identifies online ads and logs CSS properties during page crawls to detect non-compliant ads. Using this framework, we crawl 10K websites from the intersection of Tranco's top 100K and Acceptable Ads' Exception list. Despite these websites being allowed to show non-intrusive ads, our analysis finds that one in ten displays at least one non-compliant ad.

We identify two key issues in the Exception list that lead to these violations: overly permissive `^$document` rules and unblocking rules for offending elements. We demonstrate that improving the Exception list can reduce the number of violating domains by 32.4%. Overall, our findings highlight the need for continuous evaluation and enhancement of ad standards to better align with user expectations and improve the web experience.

Acknowledgments

We thank the anonymous reviewers for their helpful feedback. This material is based upon work supported in parts by the National Science Foundation (NSF) under grant number CNS-2138138. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] 2021. *Better Ads vs. Acceptable Ads: What publishers need to know.* https://www.ad-shield.io/blog/better-ads-vs-acceptable-ads-what-publishers-

need-to-know
[2] 2021. *Ghostery.* https://www.ghostery.com/
[3] Adblock Plus 2024. *Acceptable Ads.* Adblock Plus. https://adblockplus.org
[4] Eyeo 2024. *Acceptable Ads: Exception List.* Eyeo. https://easylist-downloads.adblockplus.org/exceptionrules.txt
[5] 2024. The Acceptable Ads Standard. https://acceptableads.com/standard/.
[6] 2024. Adblock Plus filters explained. https://adblockplus.org/filter-cheatsheet.
[7] 2024. Better Ads Standard. https://betterads.org/standards/.
[8] 2024. California Consumer Privacy Act. https://oag.ca.gov/privacy/ccpa.
[9] 2024. Digital Advertising Alliance. https://digitaladvertisingalliance.org/.
[10] 2024. disconnect-tracking-protection. https://github.com/disconnectme/disconnect-tracking-protection.
[11] 2024. *EasyList.* https://easylist.to/
[12] 2024. *EasyPrivacy.* https://easylist.to
[13] 2024. How Acceptable Ads evolved into an ecosystem. https://resources.eyeo.com/acceptable-ads-evolution-feature-ecosystem.
[14] 2024. Mitmproxy. https://mitmproxy.org/.
[15] 2024. Network Advertising Initiative. https://thenai.org/.
[16] Statista 2024. *Online advertising revenue in the United States from 2000 to 2023.* Statista. https://www.statista.com/statistics/183816/us-online-advertising-revenue-since-2000/
[17] 2024. Puppeteer. https://pptr.dev/.
[18] 2024. Puppeteer Stealh Plugin. https://www.npmjs.com/package/puppeteer-extra-plugin-stealth.
[19] Tranco 2024. *Tranco: A Research-Oriented Top Sites Ranking Hardened Against Manipulation.* Tranco. https://tranco-list.eu/
[20] 2024. *uBlock Origin.* https://ublockorigin.com/
[21] Eytan Adar, Jaime Teevan, Susan T Dumais, and Jonathan L Elsas. 2009. The web changes everything: understanding the dynamics of web content. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*. 282–291.
[22] Suhaib Al-Rousan, Abdullah Abuhussein, Faisal Alsubaei, Lynn Collen, and Sajjan Shiva. 2020. Ads-guard: Detecting scammers in online classified ads. In *Proceedings of the 2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 1492–1498.
[23] Athanasios Andreou, Márcio Silva, Fabrício Benevenuto, Oana Goga, Patrick Loiseau, and Alan Mislove. 2019. Measuring the Facebook advertising ecosystem. In *Proceedings of the Network and Distributed System Security Symposium (NDSS)*. 1–15.
[24] Hana Choi, Carl F Mela, Santiago R Balseiro, and Adam Leary. 2020. Online display advertising markets: A literature review and future directions. *Information Systems Research* 31, 2 (2020), 556–575.
[25] Federal Trade Commission et al. 2000. Advertising and Marketing on the Internet: Rules of the Road. (2000).
[26] José Estrada-Jiménez, Javier Parra-Arnau, Ana Rodríguez-Hoyos, and Jordi Forné. 2017. Online advertising: Analysis of privacy threats and protection approaches. *Computer Communications* 100 (2017), 32–51.
[27] European Commission. 2016. 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) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2016/679/oj
[28] Brett R Gordon, Kinshuk Jerath, Zsolt Katona, Sridhar Narayanan, Jiwoong Shin, and Kenneth C Wilbur. 2021. Inefficiencies in digital advertising markets. *Journal of Marketing* 85, 1 (2021), 7–25.
[29] David Gugelmann, Markus Happe, Bernhard Ager, and Vincent Lenders. 2015. An Automated Approach for Complementing Ad Blockers' Blacklists. *Proceedings on Privacy Enhancing Technologies* 2015 (02 2015), 282–298.
[30] Jiaping Gui, Stuart Mcilroy, Meiyappan Nagappan, and William GJ Halfond. 2015. Truth in advertising: The hidden cost of mobile ads for software developers. In *Proceedings of the 2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, Vol. 1. IEEE, 100–110.
[31] James M. Joyce. 2011. *Kullback-Leibler Divergence.* Springer Berlin Heidelberg, Berlin, Heidelberg, 720–722. doi:10.1007/978-3-642-04898-2_327
[32] Saranga Komanduri, Richard Shay, Greg Norcie, and Blase Ur. 2011. Adchoices-compliance with online behavioral advertising notice and choice requirements. *Journal of Law and Policy for Information Society* 7 (2011), 603–638.
[33] Hieu Le, Jamie Hayes, Salma Elmalaki, Milad Nasr, Athina Markopoulou, Matthew Jagielski, Zubair Shafiq, Vikash Sehwag, Xin Guo, Florian Tramèr, et al. 2023. AutoFR: Automated Filter Rule Generation for Adblocking. In *Proceedings of the 32nd USENIX Security Symposium (USENIX Security)*. 7535–7552.
[34] Kiho Lee, Chaejin Lim, Beomjin Jin, Taeyoung Kim, and Hyoungshick Kim. 2024. AdFlush: A Real-World Deployable Machine Learning Solution for Effective Advertisement and Web Tracker Prevention. In *Proceedings of the ACM on Web Conference 2024*. 1902–1913.
[35] Richard Oentaryo, Ee-Peng Lim, Michael Finegold, David Lo, Feida Zhu, Clifton Phua, Eng-Yeow Cheu, Ghim-Eng Yap, Kelvin Sim, Minh Nhut Nguyen, et al. 2014. Detecting click fraud in online advertising: a data mining approach. *The Journal of Machine Learning Research* 15, 1 (2014), 99–140.
[36] Karl Pearson. 1900. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is to be regarded as significant. *Philos. Mag.* 50, 303 (1900), 157–172.
[37] Shadi Sadeghpour and Natalija Vlajic. 2021. Ads and Fraud: a comprehensive survey of fraud in online advertising. *Journal of Cybersecurity and Privacy* 1, 4 (2021), 804–832.
[38] Navdeep S Sahni, Sridhar Narayanan, and Kirthi Kalyanam. 2019. An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research* 56, 3 (2019), 401–418.
[39] Rohan Singh, Siddhartha Milkuri, Yaling Liu, and Qinghua Li. 2024. TraceMonitor: A Novel AI-Based Chrome Extension to Enhance User Awareness of Privacy Leakage towards Green Web Browsing. In *Proceedings of the 2024 IEEE Green Technologies Conference (GreenTech)*. IEEE, 59–63.
[40] Vinay Singh, Brijesh Nanavati, Arpan Kumar Kar, and Agam Gupta. 2023. How to maximize clicks for display advertisement in digital marketing? A reinforcement learning approach. *Information Systems Frontiers* 25, 4 (2023), 1621–1638.
[41] Tobias Urban, Dennis Tatang, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2020. Measuring the impact of the GDPR on data sharing in ad networks. In *Proceedings of the 15th ACM Asia Conference on Computer and Communications Security*. 222–235.
[42] Robert J Walls, Eric D Kilmer, Nathaniel Lageman, and Patrick D McDaniel. 2015. Measuring the impact and perception of acceptable advertisements. In *Proceedings of the 2015 Internet Measurement Conference*. 107–120.
[43] Pengyuan Wang, Li Jiang, and Jian Yang. 2024. The early impact of GDPR compliance on display advertising: The case of an ad publisher. *Journal of Marketing Research* 61, 1 (2024), 70–91.
[44] Weihang Wang, I Luk Kim, and Yunhui Zheng. 2019. Adjust: runtime mitigation of resource abusing third-party online ads. In *Proceedings of the 2019 IEEE/ACM 41st International Conference on Software Engineering (ICSE)*. IEEE, 1005–1015.
[45] Yutian Yan, Yunhui Zheng, Xinyue Liu, Nenad Medvidovic, and Weihang Wang. 2023. Adhere: Automated detection and repair of intrusive ads. In *Proceedings of the 45th International Conference on Software Engineering (ICSE)*. IEEE, 486–498.
[46] Shuai Yuan, Jun Wang, and Xiaoxue Zhao. 2013. Real-time bidding for online advertising: measurement and analysis. In *Proceedings of the seventh international workshop on data mining for online advertising*. 1–8.
[47] Ahsan Zafar, Aafaq Sabir, Dilawer Ahmed, and Anupam Das. 2021. Understanding the Privacy Implications of Adblock Plus's Acceptable Ads. In *Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security*. 644–657.

A Validation of Heuristics

The heuristics used in this study were partially inspired by prior work [45]. However, detecting ad categories outlined in the Acceptable Ads Standard required adjustments to account for the size and placement restrictions defined by the Standard. Additional heuristics, such as those for large and overlay ads, were developed independently. Care was taken to set height, width, and area thresholds at strictly higher limits to ensure that only ads falling within the definition of unacceptable ads were detected. To further validate these rules, we shared the heuristic designs and samples of offending advertisements with the Acceptable Ads Committee, whose representative confirmed their accuracy and effectiveness.

B Ad Violations Examples

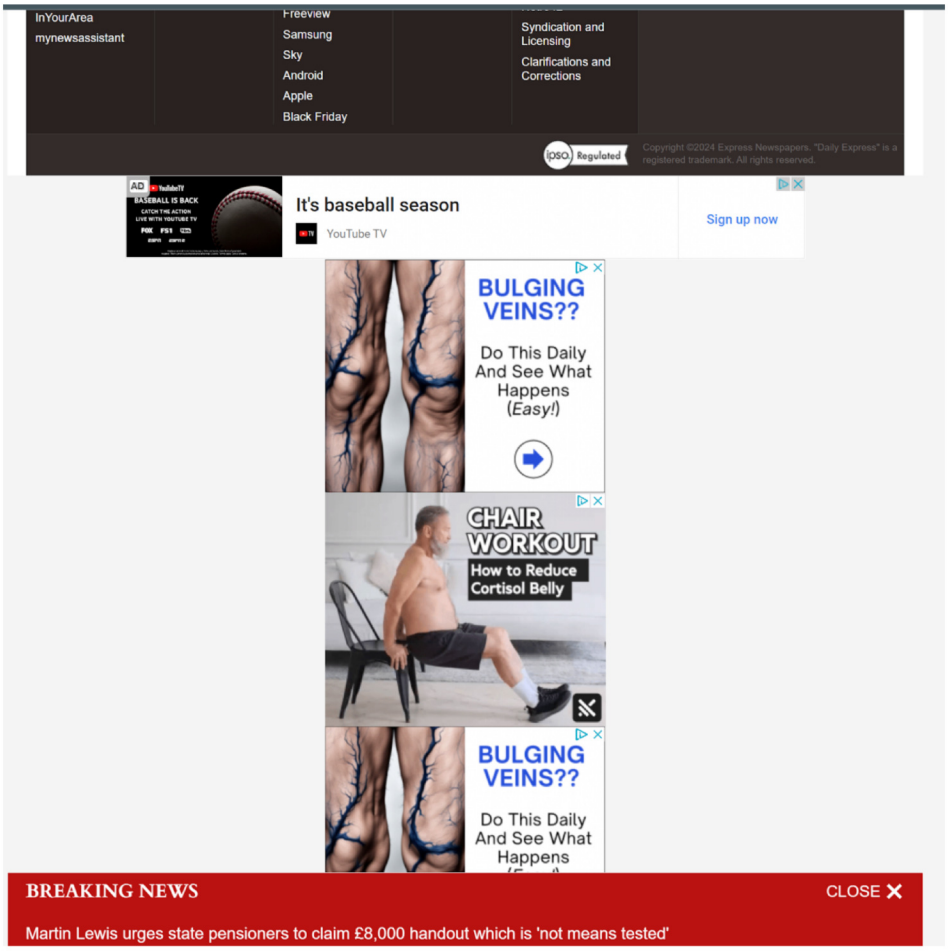

Figure 7: Oversized Image Ad from express.co.uk

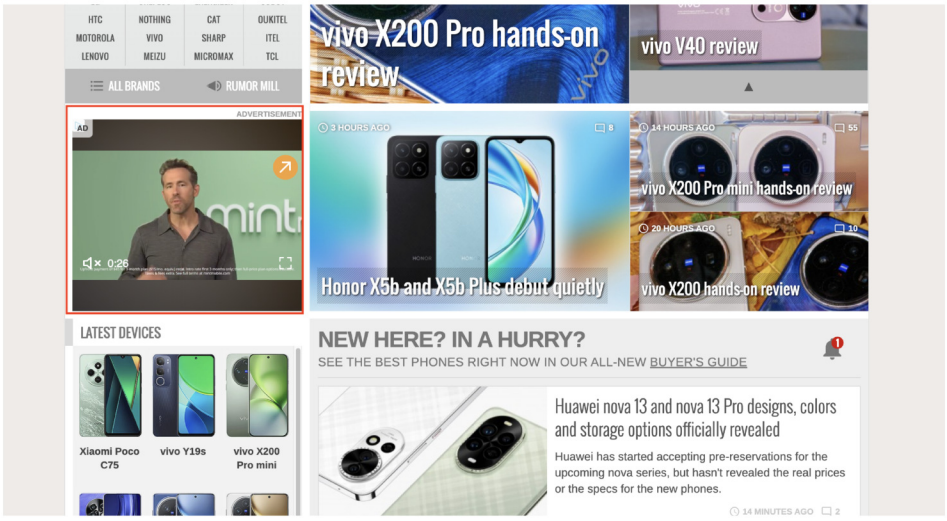

Figure 8: Autoplay video Ad from gsmarena.com

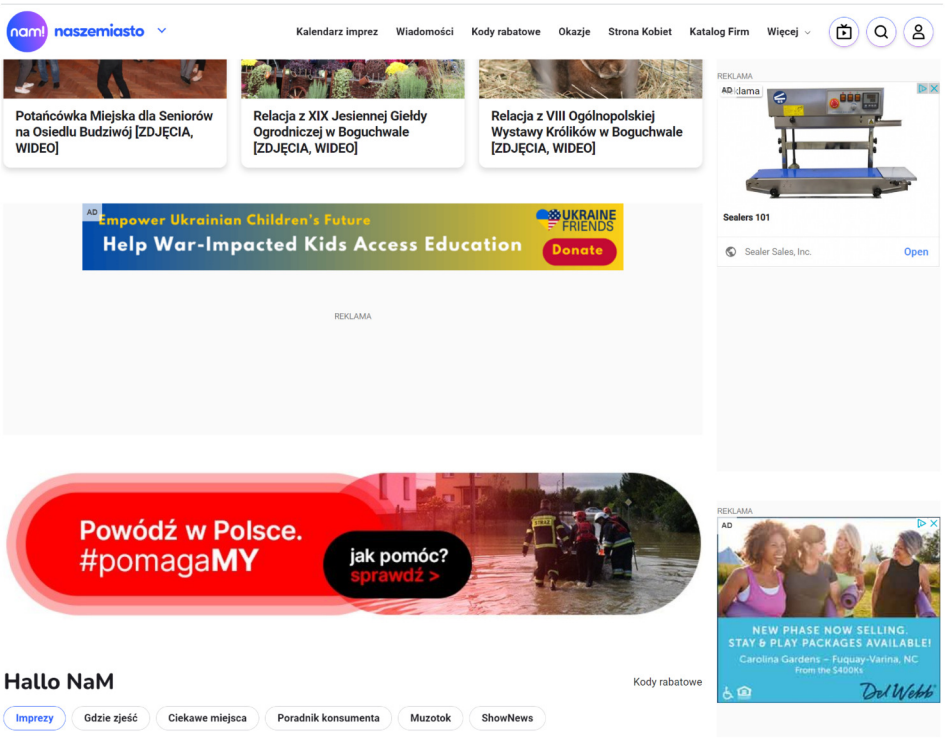

Figure 9: Overlay Image Ad from naszemiasto.pl

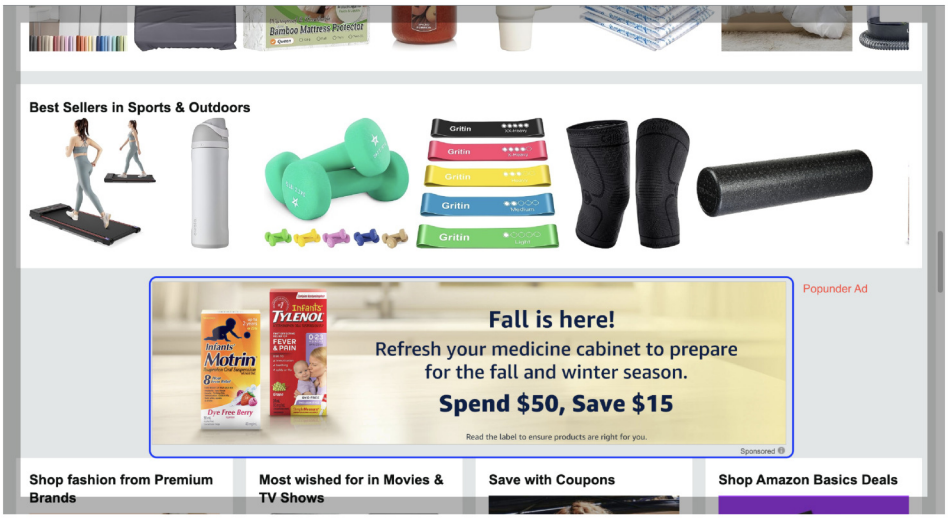

Figure 10: Popunder Ad from Amazon.ca

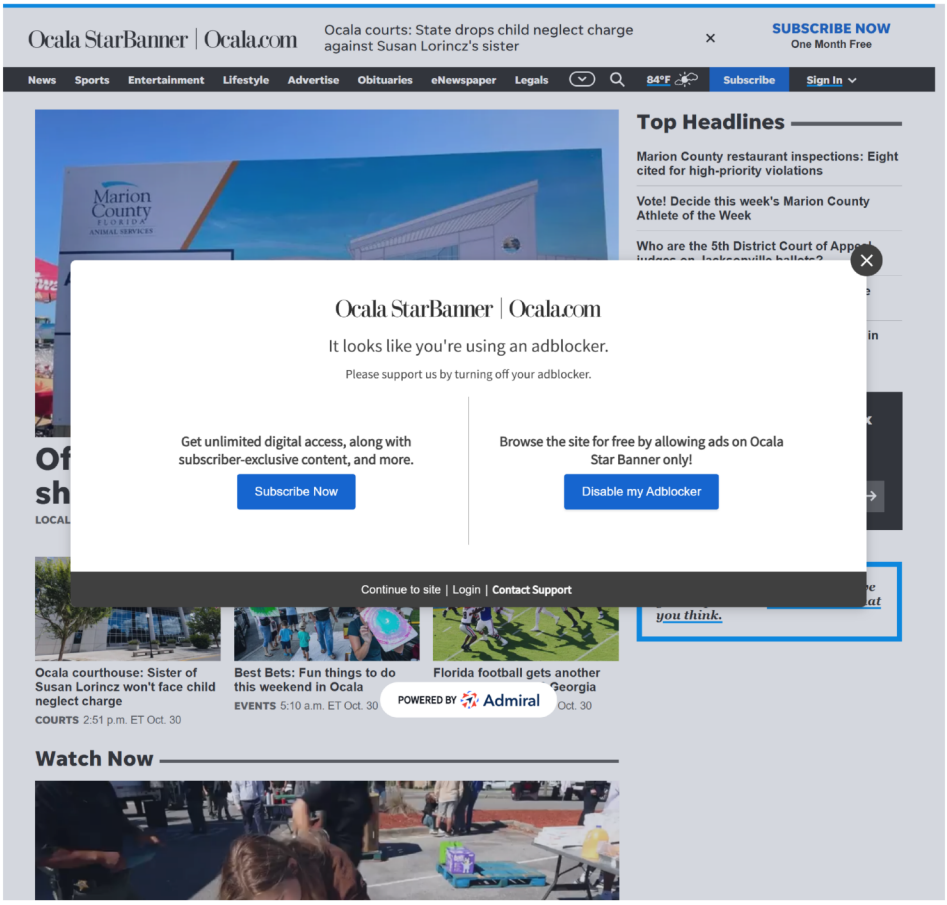

Figure 11: Popup Ad from ocala.com

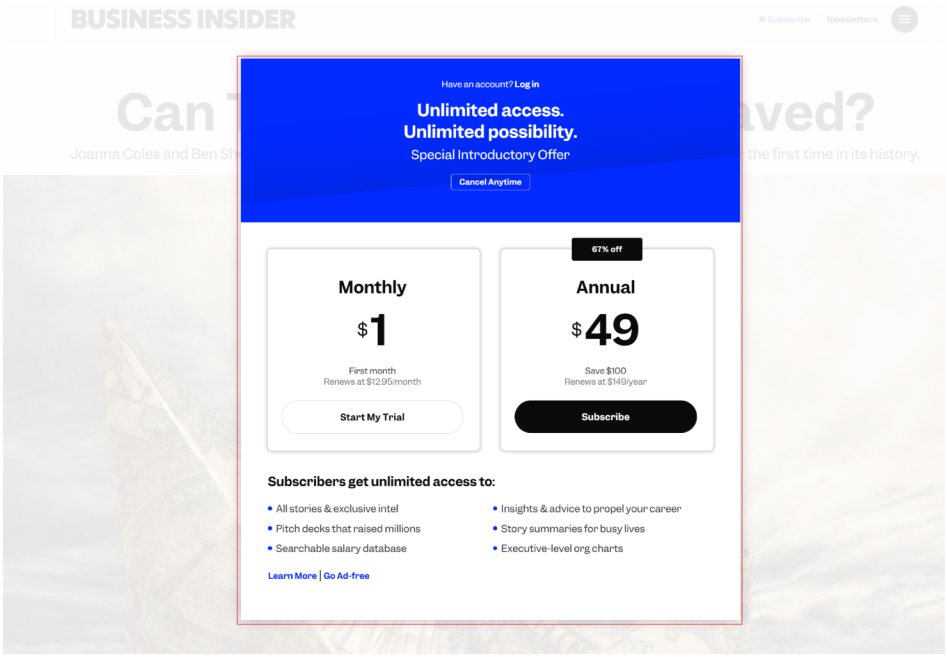

Figure 12: Interstitial from businessinsider.com