This paper appears at the Workshop on Technology and Consumer Protection (ConPro '20).

Bad News: Clickbait and Deceptive Ads on News and Misinformation Websites

Eric Zeng, Tadayoshi Kohno, Franziska Roesner
Paul G. Allen School of Computer Science & Engineering
University of Washington

***Abstract*—A key aspect of online ads that has not been systematically studied by the computer security community is their visible, user-facing content. Motivated by anecdotal evidence of problematic content such as clickbait, misinformation, scams, and malware, particularly in native advertising, we conducted a systematic measurement study of ad content on mainstream news sites and known misinformation sites. We provide evidence for significant numbers of problematic ads on popular news and misinformation sites, primarily served through native ad platforms. This work begins a rich, systematic line of inquiry into problematic ad content, ultimately to inform technical and/or regulatory solutions.**

I. Introduction

Online advertisements are an unavoidable fact of the modern web — they are embedded in and financially support the majority of content websites. Significant prior work in the computer security and privacy community has studied the ecosystem of online advertising, particularly in terms of its privacy implications (e.g., [6, 16, 19, 37–39, 48, 56, 65, 66]) or the use of ads to spread malware (e.g., [40, 55, 69, 70]). What has not been substantively considered in the security community, however, is the *visible, user-facing content* of these advertisements (except to the extent it relates to privacy, e.g., people finding highly personalized ad content or ad targeting explanations "creepy" [17, 64]).

Meanwhile, there is significant anecdotal evidence that the content of online advertisements can be deeply problematic [33, 41, 46, 47, 57, 62, 63] — consider the examples in Figure 1, a row of low-quality ads colloquially called a "chumbox". These concerns have been voiced particularly about *native advertising*, that is, ads that appear to be first-party content on the hosting website (such as inline search results or recommended articles) but are actually paid for by an advertiser. Concerns about native ads include the fact that they are deceptive: users are not reliably able to identify them as ads [3, 20, 30, 67, 68] and may click on them thinking that they are reading another story on a news site. Anecdotally, native ads also commonly use "clickbait" techniques or other "dark patterns", e.g., curiosity-provoking headlines or shocking imagery to attract attention and entice users to click. Further, these ads seem to often lead to low-quality content, misinformation, or even outright scams (e.g., cure-all supplements) and malware.

Despite these issues — or perhaps because of them — native ads are appealing to ad networks and hosting websites, as they have the potential to generate a significant amount of revenue. Prior work has shown, and native ad platforms themselves claim that they generate significantly higher clickthrough rates (0.2% vs. 0.05%) than traditional "display ads" [5, 36, 59]. As a result, online news and media publications, which have recently struggled to raise revenue [23, 28], frequently host native advertising on their properties.

Fig. 1. Portion of a "chumbox" native ad banner, showing four ads that use "clickbait" techniques to entice clicks (such as distasteful imagery, sensationalism, provoking curiosity, and urgency). Such ads often lead to low-quality sites, misinformation, or outright scams.

Despite these many growing concerns about problematic content and dark patterns in online advertising, there has been limited systematic, scientific study of this phenomenon. We argue that these issues should be a concern of the computer security and privacy community, alongside the now well-understood privacy concerns regarding how those ads are targeted. First, these ads use misleading, deceptive, and in some cases illegal practices — impacting users financially, wasting their time and attention, and spreading scams, misinformation, and malware. At the same time, not all problematic ads are equally harmful: we must understand the spectrum of problematic ad content practices, their prevalence, and their impacts. Second, the locations where these ads appear can compound their harms: for example, on mainstream news and media websites, deceptive native ads may benefit from the trust that users have in the hosting website. Moreover, there is growing evidence that online ads are used to financially support news and media websites that spread *disinformation* (e.g., [2, 12, 14, 21, 22, 33, 50]). To fully understand the potentially harmful impacts, we must understand where these ads appear on the web and how they are targeted at individual consumers. The security and privacy community has the right tools (e.g., web crawlers, ad and tracker detectors), experience, and mindset to conduct a systematic study of this ecosystem.

In this work, we lay the foundation for such a systematic study of problematic ad content. We present the results from an

initial measurement study of ad content on news, media, and known misinformation websites, and we surface hypotheses and directions for future work in the security and privacy community. Specifically, we focus on the following research questions:

1) How prevalent are different types of problematic ad content on the modern web?
2) How does the prevalence of problematic ad content *differ* across different types of ads (native vs. display), different ad platforms, and different types of websites (news/media vs. known misinformation)?

We performed a mixed-methods measurement study, using quantitative and qualitative techniques to explore ad content on popular news/media and known misinformation[1] sites in January 2020. Among other findings, we present empirical evidence that native ads use problematic techniques significantly more often than traditional display ads. We also find that both popular news sites and misinformation sites both run a significant amount of problematic ads, but that this phenomenon is not evenly distributed — that is, some sites choose to run problematic ads while others do not. Comparing ad platforms, we find that Taboola is responsible for serving the majority of problematic ads in our dataset, that Google also serves a significant number of problematic ads (though these represent a small percentage of their ads overall), and that there are certain (smaller) native ad platforms that appear more frequently on misinformation sites.

Our results and systematic measurement methodology lay a foundation for future work to further understand this ecosystem — e.g., studying the concrete impacts of problematic ads on users, or the ways that these ads may be targeted at more susceptible populations — in order to ultimately inform technical and/or regulatory solutions.

II. Related Work

Our work falls within a tradition of studying the security and privacy implications of the online advertising ecosystem. These prior works focused primarily on the privacy-invasive mechanisms of targeted advertising (e.g., [6, 16, 17, 19, 37–39, 48, 56, 64–66]) and on malicious advertisements that spread malware or perpetrate clickfraud or phishing attacks (e.g., [40, 55, 69, 70]). We argue that the *visible content* of online ads — particularly deceptive or manipulative content — must also be systematically studied via scientific web measurement methodologies.

We note that one piece of this picture *has* been rigorously studied: network traces leading to software engineering download attacks revealed that a large fraction were reached via deceptive online ads [46]. Our view here is broader, considering a spectrum of problematic content ranging from time-wasting clickbait to outright scams and download attacks.

[1]Information that is deliberately false is often called "disinformation", while unintentionally incorrect information is called "misinformation" [31]. For simplicity, we default to the term *misinformation*, as we do not always know — and do not aim to clarify — the underlying intent of the creator.

In addition to studies of deception in native advertising and anecdotal evidence of problematic ad content discussed in Section I, our work is also thematically related to broader discussions of "dark patterns" [8] on the web and in mobile apps (e.g., [7, 27, 49]). Most closely related is recent work systematically studying affiliate marketing on YouTube and Pinterest [43, 60] and dark patterns on shopping websites [42], though neither considered web ads.

Finally, our work adds to a growing, multi-disciplinary literature studying online mis/disinformation. Most related to our investigation, evidence is emerging (mostly anecdotally) that online advertising plays a role in financially supporting mis/disinformation (e.g., [14, 21, 22, 33, 50]) or directly spreading it (e.g., [34]). In this work, we begin systematically exploring one aspect of this relationship, considering the content of the ads that appear on known misinformation sites.

III. Methodology

We designed a rigorous methodology to allow us to study the prevalence of different types of problematic ad content. At a high level, our methodology involved crawling websites of interest, scraping the ads from these sites, and performing a systematic manual qualitative analysis of ad and landing page content for selected samples of ads.

A. Input Datasets

Mainstream News and Media Sites. We collected a dataset of 6714 news and media sites from the Alexa Web Information Service API [4], which categorizes websites in the Alexa top 1 million by topic. We scraped all domains in the "News" category, and all domains in subcategories in other top-level categories that ended in "News and Media" or "Magazines and E-Zines". We excluded known misinformation sites.

Misinformation Sites. We compiled a dataset of 1158 known misinformation websites (spreading political disinformation, hoaxes, conspiracies, and other misleading and false content) based on a combination of existing sources [1, 18, 29, 35, 44, 51, 52, 54]. These lists are surely incomplete, but allow us to study ads on *known* misinformation sites.

B. Crawling Infrastructure

We built a web crawler using Puppeteer [26], a browser automation and instrumentation library for the Chromium browser. Our crawler takes a URL as input, visits the URL, identifies each ad on the page using the EasyList filter list for Adblock Plus [15], a popular list of CSS selectors and domains used by many ad blockers to detect ads and trackers. The crawler screenshots each ad, stores its HTML content, and then clicks on each ad, and screenshots and scrapes the landing page.

Because ads that appear on a site's homepage may differ from those on article pages (e.g., some sites show native ads only at the bottom of articles), we crawled both the homepage and one article page for each site in our dataset. We found the URLs for articles using three heuristics: extracting the RSS feed from the site's HTML metadata, guessing the RSS feed

by appending "/feed" or "/rss" to the domain, and randomly clicking links on the homepage and using Firefox's Readability library [45] (which transforms web articles into a simpler format) as an article-detection heuristic.

Clicking on ads raises ethical questions, since advertisers pay per click. We note that prior works have used similar methodologies [55, 69] and that even loading ads can lead to (smaller) costs (per impression). We believe that our measurements were small-scale compared to the overall business of the companies potentially impacted, and that fully studying this ad ecosystem, including landing pages, is crucial to understanding and reducing problematic content online.

Identifying Ad Platforms. In addition to studying the content of the ads, we are also interested in the ad platforms responsible for delivering ads. The process for serving an individual ad is complex: often many companies are involved in taking an ad from an advertiser to a publisher, via supply side providers, ad exchanges, demand side providers, and ad servers. For the purposes of this study, we attempt to identify the third-party platform used directly by publishers to allow ads to run on their websites, such as Google Ad Manager. These platforms usually appear as a Javascript file or iframe embedded in the publisher's website (i.e., the host website).

To identify these publisher-side ad platforms, we use two complementary approaches. First, we detected well-known platforms like Google Ad Manager, Taboola, and Outbrain using CSS selectors that match HTML classes that we determined to be associated with the platform, based on manual inspection. For native ads that contain multiple ads in a single area, we also built selectors to split each individual ad into a separate record in our database. Second, for each DOM subtree we detected as an ad, we recorded each third-party resource in the subtree (iframes, anchors, images, and scripts), as well as any modifications made to the subtree via third-party Javascript elsewhere in the document. Post-crawl, we manually identified the publisher-side ad platform or other entity (e.g., ad exchange or third-party image host) behind the 100 most popular third-party resources — we did this by examining the resources and reading promotional materials or documentation at the domain of the resources. Lastly, we labeled the ad platforms we identified in both approaches as either native ad platforms or display ad platforms, based on how they describe their own product on their websites.

Studying Site-Based, Not Profile-Based, Targeting. To enable comparisons between ads that appear on different types of sites, we wanted to maximize the chance that if we see a problematic ad, it was served based on the site we were visiting, not on the fact that our crawler has visited many misinformation sites in the past.

We thus visit each site using a separate browser instance in a new Docker container (i.e., containing no tracking cookies or other persistent browser state), to approximate a new user without a tracking profile. However, we must assume that the ad ecosystem may nevertheless successfully track our crawler, even across Docker instances, using fingerprinting, IP targeting, and other techniques [16, 48]. Embracing this reality, we thus "warm" the profile of our crawler by visiting all the sites in our input datasets twice, in random order, collecting data only on the second run (still using new containers for each site in each run). In other words, we ensure that the crawler's browsing profile looks consistent throughout the measurement to any ad networks able to fingerprint our crawler.

Crawls. We created our dataset during January 15-19, 2020, successully crawling 6498 mainstream news sites (plus 5831 articles) and 1055 misinformation sites (plus 863 articles). Across these pages, we detected 81,870 ads, 55,045 of which were visible HTML elements.

C. Qualitative Analysis

Finally, we qualitatively analyzed and labeled the ads we observed on a subset of the websites we crawled. We generated a codebook to describe different types of problematic ads we observed in a preliminary analysis of the dataset, with each code describing a set of ads with similar advertisers, products, and advertising tactics. Our codes ranged from ads for things that could cause material harm, such as potentially misleading ads for supplements and investment pitches, to ads that people find irritating, such as ads for celebrity news content farms. The codebook was informed by prior academic research, regulations, and journalism on deceptive advertising, clickbait, malvertising, and advertising industry practices [20, 33, 41, 46, 47, 57, 62, 63]. The full codebook with definitions is included in the Appendix, and these categories are also listed in each table in our results.

Because we crawled 55,045 ads in total, we could only manually analyze a subsample of our dataset. For this preliminary work, we coded three different samples of websites, focusing on sites that users visit most: (1) 100 of the most popular news sites and their articles, (2) 100 of the most popular misinformation sites and their articles, and (3) 100 news sites (and articles) that have a similar popularity to the misinformation set. For the first and second samples, we discarded sites that our crawler could not reach and supplemented them with additional sites from the ranked lists until our sample size was 100 for each. The third sample allows us to control for the effect of site popularity on the types of ads that appear.

For each site in the samples, we coded each ad that appeared on the home page and article page, using a single code per ad. In total, we coded 2058, 1308, and 2048 ads from the top 100 news sites, top 100 misinformation sites, and 100 similar popularity news sites respectively, for a total of 5414 ads.

D. Limitations

Our dataset contains a significant number ads of that could not be analyzed, because they were *not initialized* and were not rendered, or because of being *occluded* by other site content. Our crawler was unable to take screenshots of approximately one-third of ads detected using Easylist, because they were uninitialized and had zero height and/or width. Of the 5413 ads in our manually labeled sample, 1813 had no screenshot

TABLE I

		Display Ad Platforms					Native Ad Platforms						
Code	Total	Amazon	Concert	Google	TownNews	Subtotal	Outbrain	PowerInbox	RevContent	Taboola	Zergnet	Subtotal	Unknown
Content Farms	283	0	0	13	0	13	0	0	1	178	87	266	4
Insurance Advertorials	96	0	0	21	0	21	0	0	15	59	0	74	1
Investment Pitches	43	0	0	10	0	10	0	2	6	24	0	32	1
Misleading Political Poll	14	0	0	10	0	10	0	0	0		0	0	4
Mortgage Advertorials	29	0	0	4	0	4	0	0	4	21	0	25	0
Potentially Unwanted Software	8	0	0	7	0	7	0	0	0	1	0	1	0
Product Advertorials	103	0	0	8	0	8	0	0	2	92	0	94	1
Sponsored Editorials	50	0	0	29	0	29	0	0	0	9	0	9	12
Sponsored Search	196	0	0	17	0	17	0	0	0	177	0	177	2
Supplements	256	0	0	106	0	106	0	2	38	98	0	138	12
Problematic Ads Subtotal	1078	0	0	225	0	225	0	4	66	659	87	816	37
Charities and PSAs	17	0	0	17	0	17	0	0	0	0	0	0	0
Political Campaign	28	0	0	12	0	12	0	0	0	0	0	0	16
Products and Services	1214	0	8	1050	1	1059	1	2	0	93	0	96	59
Self-Link	82	0	8	28	1	37	5	0	4	17	0	26	19
Benign Ads Subtotal	1341	0	16	1107	2	1125	6	2	4	110	0	122	94
Total Coded	2419	0	16	1332	2	1350	6	6	70	769	87	938	137
Occluded/Uninitialized Ads	2995	1	16	1666	7	1690	292	0	16	72	45	425	876
Grand Total	5414	1	32	2998	9	3040	298	6	86	841	132	1363	1013

Raw counts of coded ads across ad platforms, from the samples of misinformation, top news, and similar popularity news sites. Subtotals for native ad platforms and display ad platforms are listed inline. The percentage of a particular ad code contributed by a platform can be calculated by dividing the cell by the row-wise total (e.g., 55% of ads for investment pitches were served by Taboola, or 24/43). The percentage of ads within a platform of a specific code can be calculated by dividing the cell by the column-wise total (e.g., 8% of Google Ads were labeled as "Supplements", or 106/1332).

(33.5%), and 1182 (21.8%) were occluded or otherwise did not contain meaningful content. While the percentage of occluded and uninitialized ads were similar across our three samples of coded ads (40-47%), we observed that a substantially larger number of display ads, primarily from Google, were not rendered compared to native ads (56% vs. 31%).

To sanity check the quality of the data we collected via our crawler, we ran the ad detection algorithm described above in a standard desktop browser on 10 randomly sampled news and misinformation sites, and found 55.2% of ads were uninitialized, occluded, or otherwise false positives, compared to 58.3% on the same sites in our crawled dataset, suggesting that our crawled data is similar to what users actually see.

We suggest several reasons why some ads were not loaded or visible: (1) the elements were false positives in the ad blocker's detection algorithm, (2) the Docker environment and virtual frame buffer interfered with the browser's rendering, (3) content on the website, such as sign-up interstitials or cookie banners, occluded the ad content, and (4) the ad platform chose not to fill the ad space, e.g., due to detecting our visits as anomalous, low demand for ads, or high latency during real-time bidding. In drawing our conclusions, we assume that the distribution of problematic content among the ads that did not load because of the crawling environment is similar to that among the ads that did. Future work must validate this assumption and address this measurement challenge.

Additionally, our method for identifying ad platforms was not comprehensive (we did not identify ad platforms for 20.7% of the ads crawled), nor does it perfectly describe the entity "responsible" for working with problematic advertisers. For example, sites might configure Google Ad Manager to allow ads from a third-party ad exchange, where many third-party supply-side providers may bid on the site's ad inventory. Nevertheless, we chose to investigate the ad platforms used directly by publishers, as these platforms often have content policies in place against malicious and harmful content [24, 25, 61].

IV. Results

A. Which ad platforms show problematic ads?

We first investigate whether native ad platforms are the primary culprit for problematic content in ads on news and misinformation sites. Table I shows the count of each ad content code across *all* of our samples, comparing their prevalence across native and traditional display ad platforms.

Based on the subtotals for all native ad platforms and display ad platforms, we highlight several high-level conclusions. First, a significant fraction of all coded ads contain some kind of problematic content: of the 2419 ads we coded, 1078 of (44.6%) them were labeled as problematic. Second, native ads are indeed primarily responsible for these issues: 87% of native ads (that loaded during the crawl) were labeled as problematic, compared to 20% of display ads. Third, however, display ads do also include non-trivial numbers of problematic ads (particularly for supplements) — thus, conversations about ad content should not focus exclusively on native ads.

Next, we analyze the prevalence of problematic ads on specific ad platforms, from two perspectives. First, which ad platforms serve the largest absolute number of problematic ads, contributing most to what users see? Second, which ad providers serve disproportionately many problematic ads, as a fraction of all ads they serve?

We observe that Taboola served the largest number of problematic ads in our samples (61.1% of all problematic ads), and that proportionally, most of the ads served by Taboola were problematic (85.7%). Taboola also served a large diversity of problematic ads: we saw examples for all categories in our codebook except for misleading political polls. By contrast, other native ad platforms with significant numbers

TABLE II

	Top 100 News				Top 100 Misinfo				100 Popularity Adjusted News			
	Homepage		Article		Homepage		Article		Homepage		Article	
Code	n	%	n	%	n	%	n	%	n	%	n	%
Content Farms	46	12.5%	66	12.1%	3	1.5%	62	15.4%	37	9.4%	69	13.7%
Insurance Advertorials	18	4.9%	21	3.9%	6	2.9%	20	5.0%	5	1.3%	26	5.2%
Investment Pitches	9	2.4%	10	1.8%	5	2.4%	10	2.5%	2	0.5%	7	1.4%
Mortgage Advertorials	0	0.0%	13	2.4%	0	0.0%	5	1.2%	1	0.3%	10	2.0%
Misleading Political Polls	1	0.3%	1	0.2%	5	2.4%	7	1.7%	0	0.0%	0	0.0%
Potentially Unwanted Software	0	0.0%	1	0.2%	3	1.5%	2	0.5%	0	0.0%	2	0.4%
Product Advertorials	12	3.3%	33	6.1%	0	0.0%	16	4.0%	8	2.0%	34	6.7%
Sponsored Editorials	14	3.8%	11	2.0%	1	0.5%	0	0.0%	14	3.6%	10	2.0%
Sponsored Search	39	10.6%	56	10.3%	6	2.9%	41	10.2%	20	5.1%	34	6.7%
Supplements	22	6.0%	72	13.2%	35	17.1%	73	18.2%	11	2.8%	43	8.5%
Problematic Ads Subtotal	161	43.6%	284	52.1%	64	31.2%	236	58.7%	98	24.9%	235	46.6%
Charities and PSAs	7	2.0%	0	0.0%	3	1.5%	1	0.3%	4	1.1%	2	0.4%
Political Campaigns	0	0.0%	3	0.6%	10	5.0%	13	3.3%	1	0.3%	1	0.2%
Products and Services	179	51.6%	240	45.5%	124	61.7%	143	36.4%	273	69.1%	255	51.7%
Self Links	22	6.0%	18	3.3%	4	2.0%	9	2.2%	18	4.6%	11	2.2%
Benign Ads Subtotal	208	56.4%	261	47.9%	141	68.8%	166	41.3%	296	75.1%	269	53.4%
Total # of Ads Coded	369		545		205		402		394		504	
Occluded/Uninitialized Ads	466		678		255		446		594		556	

Counts of ads we labeled across our samples of news and misinformation sites. Percentages are computed columnwise (with the total number of coded ads as the denominator). We do not see evidence for substantial differences in the prevalence of problematic ad content across these samples.

of ads in our samples served a more concentrated selection of problematic ad types: 100% of the ads on the Zergnet network were for content farm-style articles and slideshows, and 57.6% of all RevContent ads advertised some sort of supplement.

Google was the most popular ad platform in our sample, making up nearly the entirety of the display ads that we coded. While most ads served through Google were benign ads for various products and services, 16.9% of Google-served ads were problematic, accounting for 20.8% of problematic ads in our samples. While Google's platform does not serve as many problematic ads proportionally, due to its large volume of ads in general, we note that the number of problematic ads it serves is substantial, second only to Taboola.

These results suggest that while the advertising ecosystem is large and complex, a large proportion of problematic content flows through large platforms popular with publishers, like Google Ads and Taboola. Efforts to eliminate problematic ads could start by focusing on regulating or improving ad content moderation on these platforms.

B. Are problematic ads more frequent on misinfo. sites?

We next consider whether problematic ads appear disproportionately more often on misinformation sites, compared to legitimate news/media sites. We initially hypothesized that we would see such a difference, because news sites might choose to include higher quality ads, and/or because the ad targeting ecosystem might be more likely to serve problematic ads to misinformation sites. Table II investigates this relationship, breaking down labeled ads between the three samples of websites, considering both homepages and article pages.

We draw several conclusions. First, although we see some differences, the numbers are small — overall, we do *not* see evidence for significant differences between the types of sites.

TABLE III

	Mainstream News		Misinformation	
Avg. # of Ads/Page	Homepage	Article	Homepage	Article
All Ads	5.80	6.37	2.37	5.16
Display Ads	5.57	5.22	1.78	2.79
Native Ads	0.23	1.15	0.59	2.37
All Coded Ads	8.27	12.35	4.90	8.88
Coded Display Ads	6.89	9.05	4.53	6.58
Codes Native Ads	1.38	3.30	0.37	1.38

Average number of ads per page. Top: Ads on *all* crawled pages. Bottom: Manually labeled ads. While mainstream news sites tend to have more ads on the homepage, misinformation sites run more native ads. (Note that the native ad fraction is an underestimate, since uncommon, unknown ad providers are considered display ads here.)

In other words, it does not appear that visitors to popular misinformation sites are significantly more likely to encounter problematic ads. Second, in all samples, problematic ads appear more on articles than homepages. This may be because some sites "hide" problematic ads beyond the homepage.

Without automated classification of problematic ads, we cannot consider the prevalence of these issues below the top-ranked websites that we studied manually. However, recall that our measurement infrastructure automatically identifies a set of popular ad providers associated with ads. Based on this metadata, which is available even for ads that did not load properly, we can estimate the proportion of native ads in our *whole* crawled dataset, i.e., thousands of sites.

Table III shows the average number of native and display ads per page, for sites in our full dataset. On average, we see that news sites run more ads than misinformation sites on their homepages, but both run similar numbers of ads on their articles. However, news sites appear to use a significantly greater fraction of display ads compared to misinformation

TABLE IV

	Top 100 News		Top 100 Misinfo		100 Popularity Adjusted News	
	Homepage	Article	Homepage	Article	Homepage	Article
Some Problematic Ads	43	44	24	42	29	40
Ads, None Problematic	54	47	47	34	61	50
No Ads	3	9	29	24	10	10

Counts (or percents) of sites in our three samples that include no ads, only "clean" ads, and at least one problematic ad. Problematic ads are clustered: a large fraction of sites in each sample include only "clean" ads. We caution that these are underestimates, due to ads that were not loaded.

TABLE V

		Misinformation				Mainstream News				
		Home Page		Article		Home Page		Article		
Platform	Ad Format	n	%	n	%	n	%	n	%	Total
Ad Butler	Display	1	0.0%	1	0.0%	109	0.3%	102	0.3%	213
Amazon	Display	5	0.2%	10	0.2%	23	0.1%	50	0.1%	88
AuctionNudge	Display	0	0.0%	0	0.0%	2	0.0%	1	0.0%	3
Concert	Display	0	0.0%	0	0.0%	77	0.2%	72	0.2%	149
Google	Display	1322	52.8%	1572	35.3%	25753	68.3%	20947	56.3%	49594
TownNews	Display	0	0.0%	0	0.0%	452	1.2%	321	0.9%	773
Connatix	Interactive	1	0.0%	0	0.0%	20	0.1%	35	0.1%	56
Insticator	Interactive	3	0.1%	4	0.1%	168	0.4%	169	0.5%	344
AdBlade	Native	13	0.5%	47	1.1%	0	0.0%	18	0.0%	78
content.ad	Native	163	6.5%	495	11.1%	3	0.0%	165	0.4%	826
FeedNetwork	Native	8	0.3%	0	0.0%	0	0.0%	0	0.0%	8
MGID	Native	58	2.3%	234	5.3%	0	0.0%	75	0.2%	367
Outbrain	Native	15	0.6%	148	3.3%	694	1.8%	1470	4.0%	2327
PowerInbox	Native	20	0.8%	43	1.0%	0	0.0%	1	0.0%	64
RevContent	Native	197	7.9%	567	12.7%	55	0.1%	418	1.1%	1237
Taboola	Native	111	4.4%	452	10.2%	1336	3.5%	5866	15.8%	7765
Zergnet	Native	69	2.8%	277	6.2%	89	0.2%	558	1.5%	993
Unknown		517	20.7%	601	13.5%	8928	23.7%	6939	18.6%	16985
Total		2503		4451		37709		37207		81870

Counts of ads from each ad platform, across *all* crawled pages. Percentages are computed columnwise (with the total number of ads as the denominator). Google Ads and Taboola are similarly popular across both populations, but many smaller native ad platforms are present on misinformation sites but rare on news sites, such as content.ad and RevContent.

sites, *when considering the full dataset*. This result suggests that as we consider lower-ranked news and misinformation sites, the gap between the quality of ads on those sites might be larger than what we observe for the popular subset.

More broadly, Table III also provides large-scale evidence that misinformation sites heavily leverage the targeted ad ecosystem for monetization — supporting recent reports [12, 22] and underscoring the need for advertisers and ad platforms to consider their role in supporting (or combating) these actors.

C. Are problematic ads evenly distributed across sites?

The previous section showed that problematic ad content appears roughly equally often, on average, on different samples of sites. However, this result does not imply that all sites include equal numbers or fractions of problematic ads.

Table IV divides sites into three categories: those that contain problematic ads, those that do not, and those that do not have any ads at all. What we find is that sites do indeed differ on this point: the problematic ads we see are clustered in 32%–57% of the ad-supported sites in each sample, though we do not see evidence for large differences between the samples. In other words, certain sites use ad platforms or preferences that allow problematic ads to run, but others run only or primarily "clean" ads.

Due to the challenges with many ads not loading discussed above, and because ads that appear are not consistent across page loads, the number of sites that run problematic ads may be an *underestimate*. Due to this concern, we manually investigated a sample of "clean" sites, which indeed appeared to only include display ads for benign products and services.

Also anecdotally, we observed that sites *with* problematic ads are also not created equal: some sites include a mix of "clean" display ads and one native ad, while others contain 10+ problematic native ads.

D. Do misinformation sites use a different set of ad providers?

Lastly, we investigate whether misinformation sites use different ad platforms than news sites. Are there specific ad platforms that are more popular among misinformation site operators? We might expect to see such difference because, for example, these site operators tolerate lower quality advertisements, or because certain ad platforms are willing to work with misinformation sites but not others.

Table V shows the distribution of ad platforms used by misinformation sites compared to news sites across our entire dataset. We see that Google Ads are common in both populations, comprising 52.8% of ads on misinformation homepages, and 68.3% of ads on news site homepages. Taboola is the

second most common ad platform and most common native ad platform, especially on article pages, making up 10.2% and 15.8% of ads on misinformation and news article pages. However, we see that certain native ad providers, such as content.ad, RevContent, and Zergnet, are much more popular among misinformation sites. We note that these ad platforms also run high proportions of problematic ads (see Table I).

Our results suggest that misinformation sites appear largely to be able to work with the same types of ad platforms as mainstream news sites — i.e., we do not see much evidence that they have been systematically "deplatformed" by any major providers. These results are consistent with prior work from GDI showing that misinformation sites generate revenue from roughly the same ad exchanges as mainstream news sites [22]. Our data also suggests that with the exception of Taboola, mainstream news sites tend to avoid using many native ad platforms that misinformation sites use, perhaps due to the low quality ad content served by those platforms.

V. Discussion and Conclusion

We argue that problematic content of online — particularly native — ads should be a subject of systematic study by the computer security and privacy community. In this paper, we provide initial results, which raise many additional questions and lay a foundation for future work. For example:

Larger-Scale Systematic Measurement. Our work considers a small set of popular sites. While these are (by definition) the sites users are most likely to visit, our results raise the question of how things look in the longer tail. For example, perhaps lower-ranked sites tolerate more problematic ads, or perhaps (as suggested by Table III) lower-ranked misinformation sites are worse than similarly-ranked news sites. One key challenge to a larger-scale analysis is the need for *automated classification* of problematic ad content; future work might build on our labels in the Appendix and prior work on clickbait or adversarial ad detection (e.g., [10, 53, 58]). The methodology we present can also lay a foundation for future measurements, but we highlight several additional *measurement challenges* that must be addressed: (1) classifying problematic ads often requires considering both the ad itself and the landing page, but automatically clicking on ads should be thought through carefully given that it impacts the ad ecosystem, and (2) many ads were not loaded by our crawler (perhaps due to anomaly detection by ad networks due to our clicking). Prior work on tracking detection either did not have to contend with the challenge of ads not loading due to anomaly detection (because it did not require clicking on ads) or did not notice the limitation (because it did not inspect ads visually).

Role of Ad Targeting. The types of ads that appear on a website result from a combination of the ad platform's policies and partners, options chosen by site's owner, and the ad platform's targeting of the end user. We described and used a methodology that isolated ad targeting based on hosting site, not the user. While we studied news and misinformation sites, other types of sites warrant investigation (e.g., sites targeted at children). Additionally, we hypothesize that there is an interplay between problematic ad content and the fine-grained (and privacy-invasive) *user targeting* enabled by today's online ad ecosystem. Who is being targeted with different types of problematic ads? Are there some potentially vulnerable populations (e.g., seniors, or people who frequently visit known misinformation sites) being disproportionately exploited?

Understanding and Differentiating Impacts on Users. Beyond studying the ad ecosystem technically via web measurement, it is crucial to also study the actual *human impacts* of these problematic ad practices. Not all of the practices we discuss are equally harmful, and to combat them, particularly through policy and regulation, we must understand their relative harms. For example, false advertising and scams are not only problematic but illegal under existing regulations. But is "clickbait" merely annoying, or actively harmful? Future work should conduct user studies to help clarify these harms. For example, how do people actually perceive and interact with these ads? How much time do people spend on low-quality sites reached via ads, and how do they value that time compared to the time they spend elsewhere on the web? How well do the various "dark patterns" we see work in practice, and on which types of users — are some manipulative techniques disproportionately successful, and are some users particularly vulnerable? While prior work in the marketing literature has considered related issues (e.g., [9, 11, 13, 32]), these works typically focus on deception in legitimate product ads and/or do not include large-scale measurement studies.

Defenses: Policies, Regulations, Tools. Many ad platforms, including native ad platforms, have explicit policies against problematic ad content (e.g., [25, 61]). In our analysis, however, we saw many examples of ads that either violate these policies or only technically meet them. Understanding the root cause of this discrepancy requires further investigation: perhaps some violating ads are difficult to detect, some policies are inconsistently enforced, or the policies as written are insufficient to prevent the types of ads we identified as problematic. At the same time, some types of problematic ads may be annoying, but are not sufficiently problematic to ban outright (especially by U.S. regulatory agencies, which are constrained by the First Amendment). Combining systematic web measurements with user studies (proposed above) to understand the concrete impacts on end users may provide clarity on where to draw the line. Beyond policy, technical defenses may play an immediate role in helping end users. For example, future work might explore designing and evaluating a browser extension that detects and warns users of problematic content in ads, or that blocks only problematic ads.

Last Word. The potential harms of online ads have become a core interest of the computer security and privacy community in the last decade. In this work, we expand that focus to consider the visible content of advertisements. We aim for our work to lay the foundation to rich future investigations into this aspect of the online ad ecosystem, ultimately reducing the spread of misinformation and other low-quality content online.

Acknowledgements

We thank Ryan Calo for helpful discussions and feedback on earlier versions of this work. We also thank Christine Chen and Ivan Evtimov for help developing and refining the qualitative codebook. This work is supported in part by the National Science Foundation under Awards CNS-1565252 and CNS-1651230.

References

[1] Alexa Web Information Service, "Alexa - Top Sites by Category - Top/News/Alternative," https://www.alexa.com/topsites/category/Top/News/Alternative, accessed on 2019-07-24.

[2] Alexander Smith and Vladimir Banic, "Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies," NBC News, December 2016, https://www.nbcnews.com/news/world/fake-news-how-partying-macedonian-teen-earns-thousands-publishing-lies-n692451.

[3] M. A. Amazeen and B. W. Wojdynski, "Reducing native advertising deception: Revisiting the antecedents and consequences of persuasion knowledge in digital news contexts," *Mass Communication and Society*, vol. 22, no. 2, pp. 222–247, 2019.

[4] Amazon, "Alexa Web Information Service API," https://awis.alexa.com/.

[5] A. Aribarg and E. M. Schwartz, "Native advertising in online news: Trade-offs among clicks, brand recognition, and website trustworthiness," *Journal of Marketing Research*, vol. 57, no. 1, pp. 20–34, 2020. [Online]. Available: https://doi.org/10.1177/0022243719879711

[6] M. A. Bashir, S. Arshad, W. Robertson, and C. Wilson, "Tracing information flows between ad exchanges using retargeted ads," in *25th USENIX Security Symposium*, 2016.

[7] C. Bösch, B. Erb, F. Kargl, H. Kopp, and S. Pfattheicher, "Tales from the dark side: Privacy dark strategies and privacy dark patterns," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 4, p. 237–254, jul 2016.

[8] H. Brignull, "Dark patterns," 2019, https://www.darkpatterns.org/.

[9] C. Campbell and P. E. Grimm, "The challenges native advertising poses: Exploring potential federal trade commission responses and identifying research needs," *Journal of Public Policy & Marketing*, vol. 38, no. 1, pp. 110–123, 2019.

[10] A. Chakraborty, B. Paranjape, S. Kakarla, and N. Ganguly, "Stop clickbait: Detecting and preventing clickbaits in online news media," in *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016.

[11] A. W. Craig, Y. K. Loureiro, S. Wood, and J. M. Vendemia, "Suspicious minds: Exploring neural processes during exposure to deceptive advertising," *Journal of Marketing Research*, vol. 49, no. 3, pp. 361–372, 2012.

[12] L. G. Crovitz, "How Amazon, Geico and Walmart fund propaganda," The New York Times, Jan. 2020, https://www.nytimes.com/2020/01/21/opinion/fake-news-russia-ads.html.

[13] P. R. Darke and R. J. B. Ritchie, "The defensive consumer: Advertising deception, defensive processing, and distrust," *Journal of Marketing Research*, vol. 44, no. 1, pp. 114–127, 2007.

[14] P. Dave and C. Bing, "Russian disinformation on YouTube draws ads, lacks warning labels: researchers," Jun. 2019, https://www.reuters.com/article/us-alphabet-google-youtube-russia/russian-disinformation-on-youtube-draws-ads-lacks-warning-labels-researchers-idUSKCN1T80JP.

[15] Easylist Filter List Project, "Easylist," https://easylist.to.

[16] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," in *ACM Conference on Computer and Communications Security (CCS)*, 2016.

[17] M. Eslami, S. R. K. Kumaran, C. Sandvig, and K. Karahalios, "Communicating algorithmic process in online behavioral advertising," in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2018.

[18] FactCheck.org, "Misinformation directory," https://www.factcheck.org/2017/07/websites-post-fake-satirical-stories/.

[19] I. Faizullabhoy and A. Korolova, "Facebook's advertising platform: New attack vectors and the need for interventions," in *Workshop on Consumer Protection (ConPro)*, 2018.

[20] Federal Trade Commission, "An Exploration of Consumers' Advertising Recognition in the Contexts of Search Engines and Native Advertising," https://www.ftc.gov/reports/blurred-lines-exploration-consumers-advertising-recognition-contexts-search-engines-native, December 2017.

[21] Global Disinformation Index, "Cutting the Funding of Disinformation: The Ad-Tech Solution," May 2019, https://disinformationindex.org/wp-content/uploads/2019/05/GDI_Report_Screen_AW2.pdf.

[22] ——, "The Quarter Billion Dollar Question: How is Disinformation Gaming Ad Tech?" Sep. 2019, https://disinformationindex.org/wp-content/uploads/2019/09/GDI_Ad-tech_Report_Screen_AW16.pdf.

[23] B. Goggin, "7,800 people have lost their jobs so far this year in a media landslide," https://www.businessinsider.com/2019-media-layoffs-job-cuts-at-buzzfeed-huffpost-vice-details-2019-2, December 2019.

[24] Google, "Ad Manager and Ad Exchange program policies - Prevent malware in ad content," https://support.google.com/admanager/answer/181490?hl=en&ref_topic=28145.

[25] ——, "Google Ads policies," https://support.google.com/adspolicy/answer/6008942.

[26] ——, "Puppeteer," https://developers.google.com/web/

tools/puppeteer/.

[27] C. M. Gray, Y. Kou, B. Battles, J. Hoggatt, and A. L. Toombs, "The dark (patterns) side of UX design," in *CHI Conference on Human Factors in Computing Systems*, 2018.

[28] E. Grieco, N. Sumida, and S. Fedeli, "About a third of large U.S. newspapers have suffered layoffs since 2017," https://www.pewresearch.org/fact-tank/2018/07/23/about-a-third-of-large-u-s-newspapers-have-suffered-layoffs-since-2017/, July 2018.

[29] C. Herbert, "The fake news codex," http://www.fakenewscodex.com, December 2018.

[30] D. A. Hyman, D. J. Franklyn, C. Yee, and M. Rahmati, "Going Native: Can Consumers Recognize Native Advertising? Does it Matter?" 19 Yale J.L. & Tech. 77, 2017.

[31] C. Jack, "Lexicon of lies: Terms for problematic information," Data & Society, Aug. 2017.

[32] G. V. Johar, "Consumer involvement and deception from implied advertising claims," *Journal of Marketing Research*, vol. 32, no. 3, pp. 267–279, 1995.

[33] Joshua Gillin, "The more outrageous, the better: How clickbait ads make money for fake news sites," https://www.politifact.com/punditfact/article/2017/oct/04/more-outrageous-better-how-clickbait-ads-make-mone/, October 2017.

[34] D. Keating, K. Schaul, and L. Shapiro, "The Facebook ads Russians targeted at different groups," The Washington Post, Nov. 2017, https://www.washingtonpost.com/graphics/2017/business/russian-ads-facebook-targeting/.

[35] Kim LaCapria, "Snopes' field guide to fake news sites and hoax purveyors," Snopes, January 2016, https://www.snopes.com/news/2016/01/14/fake-news-sites/.

[36] Laura Kloot, "Native Ads vs. Display Ads: What are the differences?" https://www.outbrain.com/blog/native-ads-vs-display-ads/, July 2018.

[37] M. Lécuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu, "Xray: Enhancing the web's transparency with differential correlation," in *23rd USENIX Security Symposium*, 2014.

[38] M. Lecuyer, R. Spahn, Y. Spiliopolous, A. Chaintreau, R. Geambasu, and D. Hsu, "Sunlight: Fine-grained targeting detection at scale with statistical confidence," in *ACM Conference on Computer and Communications Security (CCS)*, 2015.

[39] A. Lerner, A. K. Simpson, T. Kohno, and F. Roesner, "Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016," in *25th USENIX Security Symposium*, 2016.

[40] Z. Li, K. Zhang, Y. Xie, F. Yu, and X. Wang, "Knowing your enemy: Understanding and detecting malicious web advertising," in *ACM Conference on Computer and Communications Security (CCS)*, 2012.

[41] J. Mahoney, "A Complete Taxonomy of Internet Chum," The Awl, June 2015, https://www.theawl.com/2015/06/a-complete-taxonomy-of-internet-chum/.

[42] A. Mathur, G. Acar, M. Friedman, E. Lucherini, J. Mayer, M. Chetty, and A. Narayanan, "Dark patterns at scale: Findings from a crawl of 11k shopping websites," *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, 2019.

[43] A. Mathur, A. Narayanan, and M. Chetty, "Endorsements on Social Media: An Empirical Study of Affiliate Marketing Disclosures on YouTube and Pinterest," *Proceedings of the ACM on Human-Computer Interaction (CSCW)*, vol. 2, Nov. 2018.

[44] Media Bias/Fact Check Team, "Media Bias/Fact Check: The Most Comprehensive Media Bias Resource," https://mediabiasfactcheck.com/fake-news/.

[45] Mozilla, "mozilla/readability," https://github.com/mozilla/readability.

[46] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad, "Towards measuring and mitigating social engineering software download attacks," in *25th USENIX Security Symposium*, 2016.

[47] C. Newton, "You might also like this story about weaponized clickbait," The Verge, apr 2014, https://www.theverge.com/2014/4/22/5639892/how-weaponized-clickbait-took-over-the-web.

[48] N. Nikiforakis, A. Kapravelos, W. Joosen, C. Krügel, F. Piessens, and G. Vigna, "Cookieless monster: Exploring the ecosystem of web-based device fingerprinting," *IEEE Symposium on Security and Privacy*, pp. 541–555, 2013.

[49] M. Nouwens, I. Liccardi, M. Veale, D. Karger, and L. Kagal, "Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence," in *CHI Conference on Human Factors in Computing Systems*, 2020.

[50] A. Ohlheiser, "This is how Facebook's fake-news writers make money," Washington Post, Nov. 2016, https://www.washingtonpost.com/news/the-intersect/wp/2016/11/18/this-is-how-the-internets-fake-news-writers-make-money/.

[51] OpenSources Contributors, "Opensources," https://github.com/OpenSourcesGroup/opensources, April 2017.

[52] Politifact, "Fact-checking U.S. politics," https://www.politifact.com/.

[53] M. Potthast, T. Gollub, M. Hagen, and B. Stein, "The clickbait challenge 2017: Towards a regression model for clickbait strength," 2018.

[54] PropOrNot Team, "Is it propaganda or not?: Your friendly neighborhood propaganda identification service, since 2016!" http://www.propornot.com/p/home.html.

[55] V. Rastogi, R. Shao, Y. Chen, X. Pan, S. Zou, and R. Riley, "Are these ads safe: Detecting hidden attacks through the mobile app-web interfaces," in *NDSS*, 2016.

[56] F. Roesner, T. Kohno, and D. Wetherall, "Detecting and defending against third-party tracking on the web," in *USENIX Symposium on Networked Systems Design and Implementation (NSDI)*, 2012.

[57] Sapna Maheshwari and John Herrman, "Publishers Are Rethinking Those 'Around the Web' Ads," https://www.nytimes.com/2016/10/31/business/media/publishers-rethink-outbrain-taboola-ads.html, October 2016.

[58] D. Sculley, M. E. Otey, M. Pohl, B. Spitznagel, J. Hainsworth, and Y. Zhou, "Detecting adversarial advertisements in the wild," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2011.

[59] Sharon Hurley Hall, "Native Ads vs. Display Ads," https://blog.taboola.com/native-ads-vs-display-ads/, November 2019.

[60] M. Swart, A. Mathur, and M. Chetty, "Is This An Ad? Help Us Identify Misleading Content On YouTube," Freedom to Tinker, Jul. 2019, https://freedom-to-tinker.com/2019/07/09/is-this-an-ad-help-us-identify-misleading-content-on-youtube/.

[61] Taboola, "Advertising content policies overview," https://help.taboola.com/hc/en-us/articles/115007287467-Advertising-Content-Policies-Overview.

[62] J. Temperton, "We need to talk about the internet's fake ads problem," Wired, March 2017, https://www.wired.co.uk/article/fake-news-outbrain-taboola-hillary-clinton.

[63] K. Tiffany, "A mysterious gut doctor is begging americans to throw out "this vegetable" now. but, like, which?" Vox, May 2019, https://www.vox.com/the-goods/2019/5/8/18537279/chum-box-weird-sponsored-links-gut-doctor.

[64] B. Ur, P. G. Leon, L. F. Cranor, R. Shay, and Y. Wang, "Smart, useful, scary, creepy: Perceptions of online behavioral advertising," in *Symposium on Usable Privacy and Security (SOUPS)*, 2012.

[65] G. Venkatadri, A. Andreou, Y. Liu, A. Mislove, K. P. Gummadi, P. Loiseau, and O. Goga, "Privacy Risks with Facebook's PII-Based Targeting: Auditing a Data Broker's Advertising Interface," in *IEEE Symposium on Security and Privacy*, 2018.

[66] P. Vines, F. Roesner, and T. Kohno, "Exploring ADINT: Using Ad Targeting for Surveillance on a Budget - or - How Alice Can Buy Ads to Track Bob," in *Workshop on Privacy in the Electronic Society (WPES)*, 2017.

[67] B. W. Wojdynski, "The deceptiveness of sponsored news articles: How readers recognize and perceive native advertising," *American Behavioral Scientist*, vol. 60, no. 12, pp. 1475–1491, 2016.

[68] B. W. Wojdynski and N. J. Evans, "Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising," *Journal of Advertising*, vol. 45, no. 2, pp. 157–168, 2016.

[69] X. Xing, W. Meng, B. Lee, U. Weinsberg, A. Sheth, R. Perdisci, and W. Lee, "Understanding malvertising through ad-injecting browser extensions," in *24th International Conference on World Wide Web (WWW)*, 2015.

[70] A. Zarras, A. Kapravelos, G. Stringhini, T. Holz, C. Kruegel, and G. Vigna, "The dark alleys of Madison Avenue: Understanding malicious advertisements," in *ACM Internet Measurement Conference*, 2014.

APPENDIX

The table below provides detailed explanations of the problematic ads labels we used.

TABLE VI
PROBLEMATIC AD CODEBOOK

Category	Definition
Content Farms	News sites and blogs that contain a high density of ads, often broken up into slideshows to artificially increase ads loaded. The content of the articles are typically about human interest news, celebrity news, or political news.
Insurance Advertorials	Ads appearing to be news articles about people saving money on car or health insurance, to persuade consumers to give personal information to insurance companies for quotes. The landing page does not clearly disclose that it is an ad.
Mortgage Advertorials	Ads for mortgage refinancing, promising large savings, sometimes citing changes to government policies. The goal is to collect consumers' personal information and send it to lenders for quotes. Unclear advertising disclosure.
Investment Pitches	Ads for investment opportunities that make sensationalist claims about their returns, "secret stock picks", or predictions of imminent economic turmoil. The advertisers are not affiliated with established brokerages or financial institutions.
Misleading Political Polls	Ads that appear to be political opinion polls, about politically polarizing candidates or issues, but require users to submit names and email addresses — likely for fundraising or advertising purposes.
Potentially Unwanted Software	Ads for software downloads that primarily consist of misleading UI elements, like large buttons labeled "Download" or "Watch Now", rather than advertising the name of the product or its functionality.
Product Advertorials	Ads for consumer products written in the style of a blog post or news article that do not obviously disclose that they were written by the advertiser, other than in the fine print in the header or footer of the page.
Sponsored Editorial	Articles hosted on news sites paid for and/or authored by an advertiser, to sell products or promote their views.
Sponsored Search	Ads for products or travel packages, but rather than linking to a specific business, links to search results for the product.
Supplements	Ads for supplements which claim about solve various chronic medical conditions, such as tinnitus, dark spots, weight loss, and toe nail fungus, but are not FDA approved.
Charities / PSAs	Charitable causes, public service announcements, class action lawsuit settlements, and other ads in the public interest.
Political Campaigns	Ads for political candidates or advocacy organizations, intended to spur people into taking action, including voting, signing petitions, donating, or other forms of political participation.
Products and Services	Straightforwards ads for various consumer products. No deception about the intent or identity of the ad is used.
Self Links	Ads that link to a page on the parent domain. Some native ad platforms will recommend both sponsored content and 1st party articles from the publisher.

Labels used to describe ads in our qualitative analysis. The top section includes ad content we consider problematic, based on prior work, while the bottom section includes more neutral ad content.