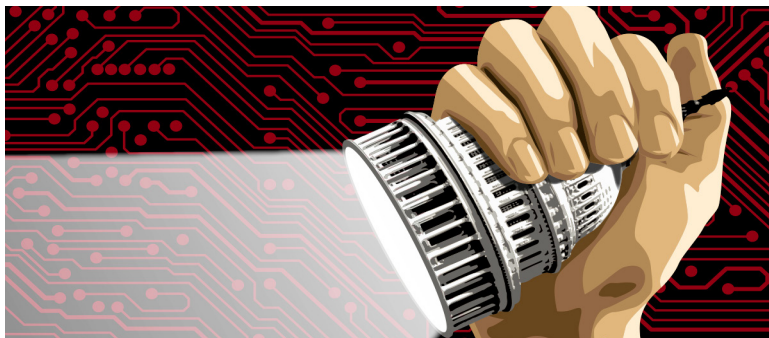

OCCASIONAL PAPERS



A Standard for Universal Digital Ad Transparency

**By Laura Edelson, Jason Chuang,
Erika Franklin Fowler, Michael M. Franz,
and Travis Ridout**



KNIGHT
FIRST AMENDMENT
INSTITUTE at
COLUMBIA UNIVERSITY

The Knight Institute's Occasional Papers series aims to bring to a broad audience thoughtful, provocative work from scholars and experts who usually write for more specialized audiences. The open-ended series features papers that address urgent questions at the intersection of speech, privacy, and technology.

The series is conceptualized by Knight Institute staff, including Jameel Jaffer, executive director; Katy Glenn Bass, research director; Alex Abdo, litigation director; and Larry Siems, chief of staff. The series is edited by Glenn Bass with support from Lorraine Kenny, communications director; Adam Glenn, writer/editor; Kushal Dev, research fellow; and Madeline Wood, research coordinator.

The full series is available at knightcolumbia.org/research/

EXECUTIVE SUMMARY

ELECTION MEDDLING and public health misinformation have fueled calls for greater transparency around online advertising on social media platforms. In response, many platforms have begun to make voluntary disclosures about some of the ads they run, but these efforts remain woefully insufficient. Platforms often neglect to provide data about how ads microtarget users, and only disclose data about a narrow category of political ads, failing to capture the larger ecosystem of digital advertising that is shaping our political, social, and economic environment.

Moreover, current regulatory measures have not kept up as digital advertising has proliferated, allowing misinformation and hate speech to spread rapidly through paid online channels that lack meaningful oversight, driven by algorithmic systems poorly understood by users and policymakers concerned with such threats to the democratic process. And to make matters worse, social media companies have also demonstrated hostility towards researchers who have tried to study advertising on their platforms.

Historically, digital advertising has dealt with issues around ad content that is “illegal, discriminatory, misleading, and manipulative in a wide range of categories beyond political ads.” This status quo demonstrates the urgent

need for interventions that support researcher access to platform-held data to increase transparency around advertising practices on platforms, and to enable regulators to ensure existing law around any disclosures is being followed.

The paper proposes a technical standard for universal digital ad transparency, defining the criteria for determining which ad platforms would need to comply and what content should be made transparent. Additionally, it compares the proposed standard to the platforms' current voluntary transparency practices, as well as to related U.S. legislative efforts.

Key Findings/Insights

THE PROBLEM

- Digital ads can be served to a single user at a time. Unlike broadcast communication, which is seen by many people at a time, highly tailored and manipulative messaging can be aimed at an individual without detection. Such microtargeted delivery can be and has been used to disseminate medical disinformation; promote scams, predatory, discriminatory advertising practices; and promote content that is divisive and violent.
- Many platforms, including Facebook and Google, use limited to no human review of ads that are submitted for promotion on their platforms. Instead, they rely on inherently porous AI-based content filters that outsource content moderation to users.
- We argue this allows political and health misinformation, hate speech, and other abusive content to bypass algorithms and proliferate through paid promotion (from which platforms profit).

PROPOSED SOLUTION

The universal digital ad transparency proposal calls for all major digital ad platforms to disclose an ad's impressions, targeting placement, delivery, and creative data to the public.

- We also push for disclosures around content removal details and decision-making processes.
- The FTC would store this data in public databases available for up to

seven years to users and researchers.

- The standard would apply to any social media company that uses microtargeting in advertising, does not use human review of advertisements, or has a reach exceeding one-third of the United States adult population.

INTERACTION WITH EXISTING LEGISLATION

We explain that our proposal aligns to varying extents with legislative proposals on the table calling for increased transparency but is more exhaustive and inclusive.

Despite these challenges, we emphasize that greater transparency is absolutely necessary for understanding the mechanics of digital advertising, which can help quell the spread of misinformation, hate speech, and abuse that proliferate through paid online channels today.

Note: The release of the universal digital ad transparency proposal follows recent news that Facebook suspended the personal accounts of New York University researchers—including that of Laura Edelson—who were using the proprietary Ad Observatory tool they built to study targeted political advertising trends on the platform.



I. INTRODUCTION

OVER THE PAST FEW YEARS, against the backdrop of ever-louder calls for public transparency of digital advertising, many platforms that deliver ad content have begun taking steps to make the ads they run more visible to the public on a voluntary basis [2, 3, 70]. These efforts towards greater transparency are to be applauded. However, if the benefits of such transparency are to be fully realized, a common standard for public ad disclosure is needed. In this paper, we argue not only for a common standard for digital ad transparency but also for a single common repository for such data. We also argue that all digital ads from certain ad platforms should be made transparent. Current industry ad archiving efforts often cover only narrowly defined electoral ads or a broader but, depending on the platform, differently defined range of political ads [50, 66, 4]. The relatively short history of digital advertising includes content that is illegal [20], discriminatory [22], misleading [49], and manipulative [39] in a wide range of categories beyond political ads. As we detail in Section II, digital ads are served uniquely to a single user at a time, rather than broadcast via a public means, as are print, television, or radio ads. This direct-to-user delivery mechanism means that

researchers' traditional methods of monitoring broadcast communications to identify ad content simply cannot capture online ads. We argue that digital ads are categorically different than their analog counterparts. Therefore, more public transparency may both improve public trust and enable regulators to verify that existing law is being followed.

In this essay, we begin in Section II by reviewing existing ad disclosure and transparency systems in the United States. Currently, these only exist for limited categories of ads. We also describe generally how digital ad networks function and the voluntary transparency measures that digital ad platforms have taken thus far. In Section III, we discuss legal and ethical issues in digital ads and ad targeting. We suggest that all digital ads, not just electoral or political ads, need to be made publicly transparent to allow community watchdogs, including journalists and civil society, to properly fulfill their roles.

We outline several particular characteristics that make some digital ad platforms of greater concern to the public, either because they increase the likelihood that the platform will run ads that violate the law or because they are of such scale that community oversight may improve public trust. The disjunctive criteria we propose for triggering the need for universal ad transparency on the part of ad platforms are:

1. No human review of ads
2. Microtargeting of ads
3. Platform reach exceeding one-third of U.S. adult population

We define these criteria in detail in Section IV and argue that ad platforms that meet any one of these criteria should comply with the standard that we propose.

In Section V, we describe what data about ads should be made transparent and justify why that information is necessary. In Section VI, we describe details of the technical implementation for how ad data can be stored and transmitted. As part of our standard, we propose established solutions for managing large volumes of data, as well as updates and corrections. We also recommend methods for handling illegal content that may appear in ad data. In addition, we consider potential user privacy concerns and suggest methods to protect user privacy.

We compare existing transparency efforts on the part of ad platforms in Section VII. We identify particularly successful implementations that inspired aspects of the standard we propose, but ultimately our experiences working with these various products lead us to believe that a single, combined data repository would be a better solution. In Section VIII, we discuss current legislative proposals in the United States that would require digital ad transparency in some form. We compare our proposed standard to each of these, highlighting key areas of overlap. Lastly, in Section IX, we review potential problems that the adoption of universal ad transparency may cause. We also review possible mitigation strategies for some of these problems, but we acknowledge that our proposal is not without costs, financial or otherwise.

i. Our perspective

As researchers and experts in advertising activity, we see a genuine need for more transparency in digital advertising, and in particular, we believe that a uniform standard for all platforms would benefit all. The amount of effort that is currently required for researchers to standardize database fields in order to compare ad activity across platforms is a substantial barrier to knowledge, one that jeopardizes collective efforts to detect bad actors, and to provide more actionable information to platforms, advertisers, scholars, journalists, advocates, and the interested public alike. Many parties share an interest in greater transparency and clear regulatory guidelines, and we, as researchers, represent only one of those core audiences. Thus, our thinking may not cover all of the necessary perspectives through which policy proposals must pass scrutiny. Still, we believe that our views do represent those of the most active experts currently utilizing digital advertising data, and we offer the following standard as one that we believe would greatly benefit the broader community for the reasons we further outline below.

II. BACKGROUND

SEVERAL LAWS AND REGULATIONS ensure a fairly high degree of transparency in spending on and sponsorship of election activity-related television advertising in the United States [43]. These laws

were largely created when social media and online advertising were in their infancy, and have not kept up with the rapid expansion of and diversity in online activity. We review the existing reporting requirements for election activity as a way of highlighting and explaining why more transparency for online advertising is needed, before describing why digital ad transparency may be needed for a much broader set of ads than simply those that are election-related. Lastly, we provide a basic overview of how modern digital ad platforms function, review the historical context for the voluntary transparency efforts of those digital ad platforms, and discuss the most relevant of those voluntary transparency efforts.

i. FCC reporting requirements

One regulation that ensures a fairly high degree of transparency in spending and sponsorship of television advertising in the United States is the Federal Communication Commission (FCC) requirement that broadcasters maintain a “political file” open to public inspection, a requirement that has existed since 1965 [72]. This requirement applies to both television and radio stations. Included in the file are records of ad purchase requests, the rate paid for each advertisement, the date on which each ad aired, the class of advertising time purchased (e.g., daypart or whether the time is preemptible), and information about the sponsor of the ad, such as the candidate’s authorized committee and its treasurer [63]. Files must be kept for any ad that mentions a candidate for federal office and for any ad that mentions a “political matter of national importance” [59]. Because of this last criterion, information about ads sponsored by interest groups, even if they do not mention a candidate’s name specifically, will generally also be included in the station’s political file.

When the requirement for a political file was first instituted, stations would keep a physical file at their offices, but in August 2012 the FCC required that television broadcasters in the largest media markets upload their political files to an FCC website (publicfiles.fcc.gov). Over time, the requirement for web access has been expanded to all broadcast television and radio stations, satellite radio stations, direct broadcast satellite television, and the largest cable television systems. Thus, access to such information is easier than ever.

Nevertheless, these data are far from perfect [44]. For one, the search tools available on the FCC website resemble more of a library filing system

than a true search engine. One must start with a particular station, drill down to a specific year, and then a specific candidate and week (or however often the station bills). Moreover, the final product is a PDF file containing contracts and invoices, and there is no standard format across stations. Another drawback is that political files do not contain information on the content of the political ads aired. Timing is also an issue. Sometimes there is a substantial lag between when the ad aired and when the station posts the information to the FCC [62], and files are only required to be maintained for two years [72], making it likely that historical data could become unavailable.

Of course, none of these FCC requirements exist for digital platforms, such as video streaming services, websites, or social media companies that sell political ads, and so this imperfect but vital source of information is entirely lacking for digital advertising.

ii. FEC reporting

There is also some disclosure of television and online ad spending through filings that political committees must make periodically with the Federal Election Commission (FEC). According to U.S. Code, Title 52, Section 30104, all registered committees must report expenditures to the FEC on a regular basis, and they must include designations indicating the purpose of each expenditure.¹ As such, one can download a committee's expenditure reports and assign each to broader categories, such as media, voter mobilization, staff, and so on. Still, the FEC does not have a uniform reporting standard, meaning designations of purpose often vary significantly across filers. Some reports may simply list expenditures as "ads" or "media." If ads are purchased for a candidate or outside group by a political consultant or ad firm, the FEC expenditure reports may only reflect the payment to the consultant instead of to the television station or the digital platform. A candidate may also use broad designations, such as "political ads," to apply to both online and televised advertising, making it difficult to segment ad spending by type. Moreover, as with the political files reported to the FCC, users do not have access to the content or creatives associated with political ad expenditures.

¹ All federal candidates and political party committees must register with the FEC, but outside groups do not register if they assert that their major purpose is not federal electioneering. As such, many so-called 501(c) nonprofits do not file regular fundraising and expenditure reports with the FEC.

There are also reporting gaps in the FEC mandates. Outside groups that do not regularly file reports with the FEC, often 501(c) groups, and that spend money on political ads outside certain reporting windows (60 days before a general election and 30 days before a primary) do not have to report expenditures to the commission if the ads do not expressly advocate for the election or defeat of a federal candidate [43].² Whereas the FCC regulations include ads about a “political matter of national importance,” or what we will generally refer to as *issue ads*, the FEC reporting guidelines on political ads are limited to ads that expressly advocate for or against a candidate, and TV ads that mention a federal candidate within the reporting windows. Such issue ads are not generally reported if they air on TV or radio outside the windows, or if they are distributed as digital ads.³

iii. Disclaimers

Also contributing to transparency in television advertising is the requirement that ads have disclaimers. Indeed, the FEC requires that television ads contain a “clearly readable” written statement disclosing the ad’s sponsor at the end of the ad, lasting at least four seconds [84]. And to ensure its readability, regulations state that the disclaimer must be contrasting in color to the background and must occupy at least four percent of the vertical height of the screen on which the ad airs [84].

Further, any ad paid for by a candidate’s campaign committee must feature video of the candidate (or audio combined with the candidate’s image) that identifies the candidate and approves the advertisement. Noncandidate committees that sponsor ads must include audio from a representative of the committee, group, union, corporation, or other sponsor that reveals who is responsible for the content of the advertising. Different states have different (sometimes stricter) disclaimer requirements for television ads aired in state-

² The Bipartisan Campaign Reform Act of 2002 required that all outside groups report any expenditures on radio or television ads that pictured or mentioned a federal candidate within these reporting windows. The legislation did not specifically mention digital or online ads (which were in their infancy at the time of the law’s writing).

³ More specifically, ads from unregistered political groups are not reported to the FEC at all if they air on TV or radio outside the windows or are distributed as digital ads. Registered committees do report these expenditures, but as part of their larger expenditure-level reporting, not as part of a political ad database specifically.

wide and local races [84].

When it comes to requirements for disclaimers in online ads, federal regulations are unclear. There are two major loopholes here. The first is that the FEC allows for an exception to the disclaimer requirement if its inclusion would be impractical or inconvenient. While it is clear that disclaimers are required for webpages and email messages, it has not been settled whether such disclaimers are mandatory for small online banner ads, a sponsored Facebook post or a Google search. Instead of providing clear regulations on the matter, the FEC has considered disclaimer requirements on a case-by-case basis [84].

In 2018, the agency did launch a rule-making process on the issue of disclaimers in online ads, with the goal of improving clarity about it. Hearings on the issue were held in that year [75]. But the agency has yet to provide clear regulations, meaning that groups can still place online ads that lack disclaimers. While social media platforms, such as Facebook and Google, do have their own policies requiring political advertisers to include disclaimers in their ads [51, 37], such disclaimers are not always useful because the same entity may use multiple disclaimers [42].

iv. Overview of digital ad platforms

All major digital ad platforms of which we are aware function broadly in the following manner. First, advertisers upload ad creative data, such as text, image, video, and outbound link to the ad platform's self-service ad platform [34, 41]. The advertiser selects the audience it wants to target, choosing from options that are very general, such as metro area and age, to ones that are very specific, such as an interest in dogs. The advertiser chooses what metric it wants to monetize, usually selecting from options [83] such as paying per impression, per click, or per customer interaction (such as a purchase), and how much it is willing to pay. Lastly, it will choose when it wants the ad to run and usually set a maximum budget for how much it wants to spend for all the deliveries of that ad. Before it runs, the ad will likely be reviewed by automated processes for compliance with the ad platform's policies.

Once the ad has been reviewed and is approved to be delivered, it will enter the real-time auctions [78] that are run by ad platforms to determine which ad will be shown to a user. Real-time ad auctions function as follows:

When a user who the ad platforms believe matches the target audience of one or more of its advertisers is about to be served an ad, all the ads that are competing for that user's view are evaluated. Three factors are taken into consideration: the price an advertiser has agreed to pay for a given user impression or interaction, the platform's estimate of the user's likelihood of taking any advertiser-specified action, and the ad platform's evaluation of ad quality. Ad quality is a somewhat amorphous concept and differs by platform. In general, however, content that is clear and engaging is considered high quality, and content that is misleading or upsetting is considered low quality. Consideration of ad quality is done because platforms believe that low-quality ads not only decrease the likelihood that users will interact with that particular ad but also the likelihood that they will interact with other ads in the future [78]. The exact formulae for weighting these factors are specific to each ad platform, but after consideration of these factors, the auction winner will be determined and the ad will be served.

v. Historical background: Revelations of Russian election meddling via social media political ads

When examining the current state of voluntary transparency efforts, it is useful to understand the impetus behind the launch of the first political ad libraries in the United States—those of Facebook, Google, and Twitter. While there were some basic steps toward ad transparency that came before each of the three major social media platforms launched stand-alone sections of their websites for viewing political ads, these ad libraries were the first products from ad platforms where ads would be archived and viewed separately from other social media content. All three of these ad libraries launched between May and July 2018, and followed the revelations of Russian meddling in the 2016 election that we discuss below, as well as congressional hearings about Russian election disinformation on social media that took place in April 2018.

These libraries have changed and improved in the years since they were first developed. In particular, where the Facebook and Google ad libraries have arrived now reflects a long path of experimentation with new features added on a regular basis. In general, there has been increasing transparency over time, often the result of revelations in the media of serious legal and ethical concerns. For example, Facebook has made housing, employment,

and credit ads transparent through the web portal of its ad library, a significant step forward. However, this step was taken only after reports of discriminatory targeting of these legally regulated categories of ads, as we discuss in Section III.

vi. Facebook ad library and FORT

Currently, Facebook’s political ad library is officially known as the Library for Ads on Social Issues, Elections, and Politics [2]. Prior to April 2019, it was known as the Political Ad Archive. Facebook requires that ads about issues of national importance [12], politicians, or elections and ballot measures on any level of government be disclosed as such by the advertiser, and be made transparent through its political ad library. It is available via a web portal and an API. The web portal is publicly available, and the API is available in the United States to any Facebook user who goes through Facebook’s identity and address verification process. Facebook has said that ads will be archived for seven years after they stop running. Generally, Facebook makes available the ad creative data and the ad impression data called for by our standard transparent, although it does not make targeting data publicly available through either of these methods.

While Facebook does not make ad targeting data available through its ad library generally, it has made a limited ad targeting dataset [54] available through its FORT (Facebook Open Research and Transparency) initiative. FORT is only available to academic researchers who apply and are approved for access. The dataset only includes targeting information for ads on social issues, elections, and politics that were shown in the 90 days leading up to the 2020 U.S. election, and it only includes information for ads with greater than 100 impressions. For reference, ads with fewer than 100 impressions made up a majority of political ads on Facebook during this time period.

See Table 4 in appendix for full details on data accessible through Facebook’s ad library API. Additionally, Facebook makes ads about housing, employment, and credit available for seven years, although these ads are only available through the ad library web portal, not the API. All ads are viewable through the ad library web portal while the ad is running. However, users cannot use a keyword search to find all active ads; they can only search by advertiser.

vii. Google’s Transparency Report and Ads Transparency Spotlight

Similar to Facebook, Google provides a web portal for users to access its political ad library, which it refers to as its Transparency Report [3]. Instead of an API, however, Google provides a publicly available database through its BigQuery platform. Also in contrast to Facebook, Google’s political ad library could better be described as an electoral ad library. Google’s rules for inclusion [67] in its ad library only specify that ads by or about candidates or officeholders for state or federal office, about state or federal political parties, or about ballot measures, be included. Notably, ads about social issues are not disclosed. Google’s political ad library makes ad creative data and ad targeting transparent in a manner similar to what our standard proposes. See Table 5 in appendix for full details.

Google has also published a Chrome browser extension called Ads Transparency Spotlight [14], a user-facing transparency mechanism. This extension allows users, as they browse the web, to see information about the ads that are being shown to them. This only functions for ads that have implemented Google’s new “ad disclosure schema,” [15] but Google is encouraging others to adopt this standard [85].

viii. Other ad platform ad transparency efforts

Twitter launched an Ads Transparency Center in June 2018 [35] to provide information about political ads on its platform. While Twitter originally intended to archive political ads for seven years, the product has since been removed [52]. Political ads that ran during the period when the Ads Transparency Center was active, approximately 1 1/2 years of data, now exist as publicly available files that can be downloaded from Twitter. The files contain advertiser information, links to the ad creatives, and ad impression and delivery data, but no targeting parameters. We note, however, that tweets are often publicly available. Paid content receives promotions but is otherwise still visible to and accessible by nontargeted users.

Snapchat released its Political Ads Library [70] as publicly available downloadable files in CSV format that can be read by standard software. Each file contains a year’s worth of records. The records include ad creative, impression, targeting, and delivery data, as well as advertiser identifier.

Targeting parameters contain both inclusion and exclusion criteria.

Roku released its Political Ad Archive [5] as a folder of files. The master spreadsheets themselves, containing the links to the other files, are publicly available. However, accessing the links and the linked documents requires users to first log in using Roku's Box.com account. Generally speaking, the documents in the folder contain scanned advertising contracts. The documents in different subfolders can be of different formats, and contain codes that are not themselves explained in the folder. While advertiser information is available in these contracts, no delivery, impression, or targeting information is available.

Xfinity released its political ad archive via a searchable web portal [6]. Note that the user interface is broken and, without modifying the script, users cannot advance and scroll through the results returned by the search engine. While the search interface itself is public, accessing information returned by the search interface requires users to first log in using a Microsoft account.

III. MOTIVATIONS FOR UNIVERSAL DIGITAL AD TRANSPARENCY

AS WE DISCUSSED in the prior section, much historical regulation of transparency and disclosure in advertising has been aimed at elections and politics. However, we believe that the necessity of digital ad transparency is not solely, or even primarily, related to the public's interest in transparency of political advertising. As noted earlier, digital advertising is bought, sold, and delivered to audiences in radically different ways compared to traditional broadcast advertising. Therefore, it presents different practical, legal, and ethical concerns to the public and to regulators.

i. Legal and ethical concerns in digital ads

Because of the real-time and distributed nature of the marketplace for digital advertising, enforcement of existing laws regulating truth in advertising is challenging. Deceptive digital ads and clickbait are by now a well-known problem. In 2020, the Federal Trade Commission (FTC) received 2.2 million

complaints from consumers [61], with the second-largest category of those complaints stemming from online shopping.

The problems with deceptive advertising, however, extend beyond click-bait in display ads. Digital advertising has introduced entirely new categories of advertising, including influencer marketing [28, 53]. Influencer marketing refers to ads that run on the websites, blogs, or social media channels of people who have amassed a large audience for their content, commercial or otherwise. It is estimated that more than \$3 billion will be spent on influencer marketing in 2021 [36]. Influencer marketing is subject to all existing laws regulating advertising, but compliance issues on the part of influencers are well documented [38]. This may be because influencer marketing is decentralized both in the sale and delivery of advertising, and therefore lacks a neat analog to any type of advertising in the pre-digital era.

A 2019 study by Mediakix [11] found that of the sponsored posts by the 50 top celebrity influencers on Instagram, only seven percent appeared to be fully FTC-compliant. Influencers often do not disclose that ads they run are paid endorsements rather than uncompensated reviews, or if they do disclose, it may not be easily visible to their audience. An example of an insufficient disclosure would be a #sponsored note at the bottom of a long post, which a user would be less likely to see.

ii. Ad targeting issues

In addition to problems with ads themselves, significant legal issues have arisen regarding how ads are targeted. In 2016, ProPublica reported that Facebook allowed advertisers to systematically exclude racial minorities from housing ads [20]. When ProPublica followed up on this story the following year, it found that the issue had not been fixed, and Facebook was still allowing advertisers to target ads in discriminatory ways [22]. Also in 2017, The New York Times and ProPublica jointly reported that Facebook routinely allowed employers to exclude certain age groups from job ads [21], which could constitute illegal age discrimination. In 2019, the U.S. Department of Housing and Urban Development sued Facebook for violating the Fair Housing Act [26], although this case is still ongoing at the time of this writing [86]. In 2019, Facebook also settled several lawsuits surrounding allegations that the platform allowed housing, employment, and credit to be targeted in discriminatory ways [26].

In September 2017, ProPublica reported that Facebook’s advertising platform allowed advertisers to reach users based on available user categories such as “Jew haters” [23]. Ads that included “Jew haters” as part of their targeting parameters were approved within 15 minutes and ran on the platform. In response, Facebook acknowledged that an ad can “surface on our platform that violates our standards.” The company further stated that “Jew haters” and other similar user categories were generated by an algorithm without human intervention but “didn’t appear to have been widely used.”

However, users’ ability to scrutinize problematic ad targeting by themselves is limited. Ad platforms, including Facebook, Google, and others, do provide users with some mechanisms to see certain information about how an ad that they are being shown was targeted to them [79, 81, 82]. However, ad targeting data disappear as soon as the ad is gone. If a user doesn’t click on the menu item right away, they can no longer retrieve the relevant ad targeting information for the ad in question. The timing and the complexity of retrieving targeting information for *one ad* means that, while users may inspect occasional ads, it is not possible for them to review how they are targeted on a regular basis. Critically, if an ad is targeted in a discriminatory way by *excluding* certain users from seeing an ad, the ad will not be served to those users. Because users cannot see the ad, they cannot access the ad’s targeting information, including any discriminatory or illegal exclusion criteria aimed at them. Finally, researchers have found that the textual description of the ad targeting parameters is incomplete, which makes it hard to assess how useful these disclosures are [19].

iii. Ongoing concerns with misinformation in digital advertising

Misinformation in advertising has a long history in the United States. Even in the early 1800s, campaign pamphlets would feature scurrilous rumormongering [77]. But misinformation may be more of a concern today, due to how much of it is distributed through social media platforms whose technology accelerates its reach and spread.

Misinformation occurs in a number of substantive domains, including health policy. This could include product advertisements (think of misleading claims about a dietary supplement or the effectiveness of a prescription

drug) and recently with regard to COVID-19 [27], vaccinations against the disease [32].

Misinformation, of course, also occurs in political advertising. Scholars have studied lies and misleading claims in televised political ads for a long time (see [18] for an overview), but it was Russian-backed advertising in the 2016 U.S. presidential election that was, to some extent, a game changer. Serious attention to this issue in the United States began in 2017 and 2018, when the public first became aware of the scope of Russian disinformation efforts during the 2016 election that utilized, among other things, Facebook ads [25, 31]. The Russian Internet Research Agency (IRA) used the platform's ad targeting tools to buy ads that often mentioned issues of race and (some suggest) were intended to demobilize voters [87]. Although the IRA spent only about \$100,000 on Facebook ads in the U.S. (most of them "issue" ads) [42], the inability of researchers, politicians, or journalists to track these ads or their sponsors led to calls for more transparency. High-profile attention to this issue is largely credited with prompting platform efforts to create ad libraries.

iv. Voluntary disclosure efforts don't include all ads

While federal regulations leave many disclosure loopholes when it comes to online advertising, there have been voluntary efforts to provide a measure of disclosure for ads placed on major social media platforms, as we discussed in Section II. We will review the efficacy of these various efforts and compare them to our proposed standard in Section VII. However, it is worth stating at the outset that there are significant gaps in all of these existing transparency efforts, and because these efforts are voluntary, not all ad platforms have adopted them or carried them out consistently.

Notably, TikTok has no ad transparency or researcher access tools at all. According to Pew Research [24], TikTok is the ninth-largest platform overall, but fifth among U.S. users ages 18-29, so this lack of tools obscures researchers' understanding of a meaningful portion of the overall ad landscape. Any critiques we have of other transparency tools should be understood in this context: Such tools are not legally required to exist at all, and as we described with Twitter in the previous section, platforms can choose to take them away at any time.

There is no ad platform that currently makes information about all ads on that platform available to researchers. As we will detail in Section VII, Facebook does offer an extremely limited web portal [2] to search for all ads on the platform while they are active. However, given the many limitations and the lack of access to bulk data, we do not believe this tool is functionally useful for research. Other platforms have taken an even more limited approach to which ads they will make transparent. For example, Google’s [45] inclusion policy covers only election content. Of course, differences across platforms, such as in the universe of ads and the variables included, make comparisons and monitoring of activity difficult.

IV. AD PLATFORM INCLUSION

IN THIS SECTION, we define which ad platforms should be obligated to provide universal digital ad transparency, and provide the criteria for inclusion and their justifications. We use the term “advertising platform” or “platform” to refer to the companies that maintain the platforms where ads are sold and displayed, rather than the organizations that pay to run the ads. This term applies both to companies that primarily run ads on websites that they own, such as Facebook, and also to companies like Google that deliver ads on third-party websites that they themselves do not own [7]. There are many websites that do not manage the sale or delivery of ads themselves, but rely on larger ad platforms to do this for them. Such websites are not themselves ad platforms unless they control the sale of the ad space or choose which ads will be delivered. By contrast, we will refer to the ultimate payer(s) of an advertisement as the “advertiser(s).”

We propose three criteria for inclusion of advertising platforms in the universal digital ad transparency standard, and argue that advertising platforms that meet any one of the following criteria should adopt that standard. An ad platform should provide universal digital ad transparency if the ad platform:

-
1. Allows advertisers to purchase and run ads without human review (i.e., using an automated self-service platform)—*No human review of ads.*
 2. Allows advertisers to target, or the platform to optimize, the delivery of ads based on narrow user characteristics—*Microtargeting of ads.*
 3. Reaches greater than one-third of the U.S. adult population annually—*Impact of ad network.*

We discuss each of these criteria in detail below.

i. No human review of ads

One of the central conundrums of online advertising is that, for the most part, ad platforms already have policies prohibiting the types of ads that the public is most concerned about and/or are illegal under existing laws. Given that, why do so many offending ads continue to run?

We believe one possible answer may lie in the rise of self-service ad platforms. Many digital ad platforms, including Facebook [40] and Google [45], allow advertisers to place ads without ever interacting with a human employee of the ad platform and do not subject every ad they run to human review before they are served to users. For example, according to Facebook’s ad policies, “The Facebook ad review system relies primarily on automated tools to check ads against our Advertising Policies” [40].

An automated pipeline for ads is efficient for the platforms and their advertisers. Advertisers can quickly place and run a large number of ads using automated software tools without human intervention. The platforms can quickly scale up their operations, accommodate a greater number of advertisers, and deliver more ads.

Platforms with an automated pipeline for ads rely on a reactive review strategy with several components. Ads are screened by algorithms at the time of submission. Active running ads are then reviewed by a mixed system of people and machines, *if* problematic content is detected.⁴ According to an article Facebook posted to help its advertisers understand why ads that had been initially approved to run might later be removed, “Previously approved

⁴ It is worth noting that research has highlighted how platforms themselves have struggled to articulate and enforce their own policies with regard to political content [56], which is yet another complicating factor in enforcement.

ads can be selected for another review for a number of reasons, including if people hide, block, report or otherwise provide negative feedback about an ad, all of which could help indicate that we missed something during the review process.” [80] If an ad in question is deemed to be illegal or to violate platform policy, it then stops running and is removed. This multitiered enforcement strategy is necessitated by the very nature of a self-service ad delivery system and is inherently reactive.

However, the efficiency of a self-service platform comes with a cost to the users and to the public. Platforms that function in this manner will inevitably run ads that are illegal or violate their policies. Users will inevitably be exposed to such ads. This is not a bug in the system; it is the nature of how reactive review systems operate. Self-service ad platforms effectively outsource to their users the responsibility to ensure that content that runs on their platform conforms with the law. The pipeline does have a mechanism to correct itself, but it can do this only after users are exposed to illegal ads, report them, and trigger a review and removal process. In this case, the cost savings of self-service platforms come at the cost of legal compliance and user experience.

We argue that platforms that allow advertisers to place ads without human review should comply with universal ad transparency for two primary reasons: First, because they have not actually verified that ads they run comply with the law, content on their platforms should be subject to additional scrutiny. Second, there is the potential that by making ad content publicly transparent, they can gather feedback from a wider audience to improve their existing enforcement.

ii. Microtargeting of ads

Next, we argue that greater transparency is also needed for platforms that deploy behavioral or algorithmic targeting, and algorithmic delivery optimization of ads. Used appropriately, microtargeting allows advertisers to “reach the right audience” efficiently, without having to pay to show ads to users they know would not be interested in their products. However, the same tools have in the past been used to prey on low-income users and target them with poor financial products [60], as we discussed in Section III. Public polls [48] have also shown that users are not comfortable with ads being targeted to

them based on characteristics that platforms infer without their direct input, or even ones that are targeted very narrowly.

For these reasons, platforms that enable ad targeting or delivery optimization based on characteristics of users, other than a geographic area no more specific than a zip code, should comply with universal digital ad transparency. We suggest a historical benchmark: Before the rise of digital advertising, most advertising was done via television or newspapers. Advertisers could reach everyone reading the metro section of their local newspaper, or everyone watching the local network affiliate's broadcast of the nightly news. It was difficult (although not impossible) to advertise to very small groups. If we were to compare these historical broadcast advertising mechanisms to audiences that can be targeted, the rough equivalent would be targeting an ad to people viewing a particular piece of content, or to people viewing a piece of content in a fairly large geographic area. Therefore, we will define *microtargeting* as:

The targeting or delivery optimization of an ad based on the personal characteristics of the user (as opposed to characteristics of the content to which the ad is adjacent), other than geographic region the size of a zip code or larger.

Here, we note that it is not our objective to capture ad platforms that themselves are focused on particular small audiences; for example, online news outlets that serve small communities. Only if an ad platform allows advertisers to exclusively specify an audience by personal characteristics (other than geography by zip code or larger) to the exclusion of other users who do not have those personal characteristics, would this criterion be met.

iii. Impact of ad network

Currently there are, to our knowledge, only two advertisers that can reach greater than one-third of the U.S. adult population annually: Google and Facebook. Both of these platforms employ self-service ad platforms and allow microtargeting of ads, so this additional threshold is, at the time of this writing, a theoretical one. However, we argue that any advertising platform that exceeds this threshold should comply with our proposed standard on that basis alone. Any advertiser of this size would be able to deliver billions

of impressions for single ads in very short periods of time. Given the potential scale of ads delivered on such platforms, public transparency to allow community watchdogs to monitor and respond to claims made in ads is vital for ensuring a healthy public discourse. For similar reasons, government regulators are often most concerned as well with ads delivered on these largest platforms—ads that violate existing law have the potential to be shown to millions of people before regulators can act to notify ad sponsors that their ads are violative. We must acknowledge that this cutoff is arbitrary. We do not have enough data to say precisely how big an ad platform can be before the risks of its size alone merit public transparency. It is our hope that increased ad transparency may allow for further research on this point.

V. CONTENT INCLUSION

I**N THIS SECTION**, we review which pieces of content should be considered ads, and what information about ads should be made transparent. We also review potential privacy concerns about ad targeting data and suggest limitations to avoid revealing information about users who may have seen an ad.

i. Which ads should be included?

We propose that all forms of compensated content promotion that the advertising platforms can reasonably be expected to be aware of should be included in universal ad transparency efforts. Specifically, in addition to traditional ads, branded content and native advertising should both be included. These are forms of advertising that are styled and can appear very similar to the nonadvertising content of the platform on which it is placed. Of course, there will always be some small portion of advertising that advertising platforms have no reasonable means of being aware of, such as influencer advertising that is not disclosed, and we do not believe it is reasonable to expect platforms to make such content transparent. However, given the rise of robust branded content disclosure mechanisms, such as Facebook's in order to ensure compliance with existing FTC regulations, we believe it is reasonable to expect that platforms will be able to identify most branded content advertising that they host.

ii. Which fields should be included?

A complete list of fields, descriptions, and types that we propose as the standard for universal ad transparency can be found in Table 3 in the appendix. Here we describe the categories of ad data we include, and discuss the motivations for why these data need to be publicly transparent.

II.1. AD CREATIVE DATA

Ad creative data includes text, images, audio, video, and outbound links that were presented to the ad viewer. The central motivation to require transparency of these fields is to allow regulators to ensure compliance with the law. An additional motivation would be to allow community watchdogs to monitor particular categories of advertising, such as political or financial services ads, for content that might be misleading even if it is not actually false.

II.2. AD IMPRESSION DATA

Ad impression data includes information about when the ad was shown, to whom it was shown, and the context in which it ran. This includes the dates an ad was active, the advertiser who paid for the ad, how much was spent, how many impressions (or distinct views) it received, and nonidentifying data about the geographic and demographic distribution of those impressions. We suggest that the geographic distribution of impressions be specified at the zip code level, where available. We suggest that demographic distribution of impressions be specified by gender and 10-year age brackets. We would also include a field indicating whether the content was removed by the advertiser for a violation of policy after it ran.

II.3. AD TARGETING AND DELIVERY DATA

Ad targeting data includes nonidentifying information about how and to whom an ad has been targeted. Many platforms allow advertisers to list not just criteria for users to target but also criteria for users to exclude. We suggest that both inclusion and exclusion parameters be specified. In general, we require the type of targeting to be disclosed (customer list, content-based, behavioral, demographic, geographic, etc.), as well as the details of that type of targeting, such as the demographic group or related content targeted or excluded.

Of great concern is ensuring that particular users who may have been targeted with an ad cannot be identified by the transparency data provided. So, hypothetically, if an ad platform were to allow an advertiser to specify a single street address or a single email address as the target of an ad, we would not want that data to be unmasked. Less trivially, and more of a real-world problem, we would not want a user to be able to be identified based on the specified combination of ad targeting parameters. We also need to consider the combination of ad targeting parameters and ad impression information, which do reveal some information about users who have seen them.

This privacy concern is not always an issue. Often, ads are targeted on the basis of customer lists or other methods that reveal no information about the audience. Also, we are aware of no ad platform that allows advertisers to target individual users as we suggest in our hypothetical example. However, serving an ad to an audience of one is not illegal and may happen in the future. To prevent the identification of very small audiences of ads, we propose the following limitation on what targeting parameters should be revealed: If the audience publicly described in the targeting parameters is smaller than 100 people, the most restrictive parameters should be removed from what is disclosed until the targeting parameters publicly describe an audience greater than 100 people. In such cases, platforms could still indicate that some categories of targeting parameters were removed. Let us consider several examples to illustrate this limitation:

1. An ad is targeted to a new business's list of 50 customers. In this case, the targeting type would be "customer list" and the targeting details might contain the name of the business. The audience size is smaller than our 100-person threshold, but that isn't knowable based on the publicly disclosed targeting information, so the targeting type of "customer list" should be included.
2. An ad is targeted to all the people who live on a street with only five houses. The total number of people who would see this ad is smaller than our 100-person threshold, and this is knowable based on the disclosed targeting information, so our limitation would apply. In this case, all that should be disclosed is a larger geographic area, such as zip code.
3. An ad is targeted to all the people who live in a particular zip code, are

female, and have an income greater than \$350,000. The ad platform calculates that fewer than 100 people meet that audience description, so the most restrictive parameter, an income greater than \$350,000, is removed. The ad platform indicates that one ad parameter was removed, but doesn't specify which one.

Beyond intentional ad targeting selected by advertisers, we advocate for greater transparency into the optimization or *delivery* process by which a platform assigns users to ads. Ad delivery is a data-intensive process to partition a population, facilitated by algorithms. By its design, the optimization maximizes platform and advertiser interests, which can produce outcomes that are undesirable for users or the public. Researchers and legislators have also expressed concern about ad delivery optimizations, albeit for different reasons. We discuss these concerns in detail in Section III, but in general, research has shown that delivery optimization can be discriminatory [17].

II.4. AD PLACEMENT METADATA

Lastly, our standard specifies that metadata about when and how an ad was shown should be made transparent. This includes the specific dates an ad was active. We note that this is not the same as the date range an ad was active. Because ads can be displayed on noncontiguous days, the date range can be misleading. Ad placement metadata also includes how much was spent on the ad each day, the currency in which it was paid, and the ultimate payer. The last field we specify that falls into this category is “placement_details,” which will differ for different types of platforms. For ad platforms that run ads on third-party websites, the website on which the ad was displayed should populate this field. For sites that run different categories of ads on their own websites, the particular category of the ad can be indicated via this field.

VI. TECHNICAL STANDARD IMPLEMENTATION

IN THIS SECTION, we discuss the technical implementations of the ad standard we propose. We will assume, when contemplating the technical realities of the combined transparency dataset that we propose, that platforms that meet the criteria for inclusion that we specify run a combined average of 5 million new ads per day. Unfortunately, no major ad platform that we are aware of publishes data about how many ads they run on a daily basis; therefore this number is at best an educated guess. We will use it only as a reference point to understand the relative scale of data.

i. Common data warehouse

We propose that the FTC assume responsibility for building and maintaining a single public repository for digital ad data. As we discussed in Section II, there is a well-established history of other government agencies maintaining public datasets of election-related information. In the United States, the FTC is the government body tasked with ensuring that advertising complies with existing law, among its many responsibilities [13]. We believe that creating a single data repository for information about all ads would have several advantages. First, it would reduce the burden of compliance with this standard on ad platforms. Instead of having to develop their own digital ad archives, platforms would simply be responsible for regularly sending their ad data to the FTC. Second, as we will discuss in Section VII, there are significant hurdles to researchers when digital ad archives are fragmented across many platforms. Some of these problems would be ameliorated by the common technical standard we propose, but the logistical difficulty of having to search perhaps dozens of platforms to find ads would remain. Additionally, it may be useful for the FTC to administer a small fund that could be established to offset the costs of compliance for new ad platforms that enter the market. While we believe that the FTC would be a natural home in government for the data warehouse we suggest, this role could also be fulfilled by a different governmental body or an independent third-party institution. We propose that ad platforms that meet our criteria for inclusion described in Section IV transmit the daily ad summary files to the FTC. We will discuss how platforms can do this later in this section. The FTC can combine the files

it receives from all ad platforms and publish unified ad summary files daily (see Table 1 below). We recommend that the HDF5 [73] format be used for these daily summary files. HDF5 is part of a family of file formats that were initially developed by the National Center for Supercomputing Applications [74]. HDF5 is open source and free for general use. This file format is already in wide use by those who need to store large amounts of data of mixed types, including by other U.S. government agencies such as NASA [47]. There is a large ecosystem of libraries and tools for working with HDF5 files in a wide range of languages that should minimize the costs for the FTC to use it, and for researchers to work with the data. We recommend that the FTC retain and publish these files for two years from the date they are received.

	Unified Daily Summary Files	Snapshot Dataset	Advertiser Summary Files
Format	HDF5	HDF5	CSV
Fields Included	All standard fields	All standard fields	Advertiser ID, platform name, total spend
Versioning	All record versions preserved	Most recent versions only	Most recent versions only
Frequency	Daily	Daily	Weekly, Monthly, Yearly
Duration	2 years	7 years	7 years

Table 1: Comparison of proposed FTC file formats.

In addition to the unified daily ad summary files, the FTC should also maintain and publish an overall “snapshot” dataset with current data for all ads less than seven years old. In this snapshot dataset, all corrections, updates, and deletions will have been applied as appropriate to all records. The FTC can also publish weekly, monthly, and yearly summary files listing which advertisers ran ads on which platforms, along with the IDs for those advertisers, and their spending amounts and ad counts. We would recommend that these summary files also be kept for seven years.

These windows of time are significantly shorter than other data retention policies for public datasets published by the U.S. government, but we are sensitive to the storage size and cost requirements that a longer data retention policy would require.

ii. Public transparency tools

In addition to datasets in downloadable file format, we recommend that the FTC provide a publicly available API to allow interested parties to request data from this snapshot dataset by, at minimum, date range, advertiser, and platform. The advertiser summary files we propose should allow researchers to query this API efficiently in cases where they are only interested in a subset of political advertising activity. To supplement this machine-readable data access appropriate for researchers, we recommend that the FTC additionally provide a web-based user interface to allow users to run basic searches over the snapshot dataset (see Table 2 below). Building such a web interface to complement data availability through data files would be similar to the approach taken by the FEC [29]. The FEC’s public website for querying election data allows users to search by time period, geographic area, or contributor. We would recommend that this web interface allow users to search for ad records by advertiser, date, content text, ad impression criteria, or ad targeting criteria.

	Snapshot Dataset API	Snapshot Dataset Web Portal
Fields Included	All standard	All standard
Access	Machine readable	Human readable
Search Parameters	Date range, geo area, advertiser, and platform	Date range, advertiser, and platform geo area, demo group impressions

Table 2: Comparison of proposed FTC dataset search methods.

iii. Data retention and transmission

We estimate, based on our experience housing ad archives, that 5 million ads in a mix of content types (text, image, video, etc.) being disclosed according to our standard would take up approximately 1 terabyte of disk space. We believe that ad platforms that comply with our standard should not have difficulty moving their portion of this volume of data on a daily basis. We recommend that ad platforms complying with our standard publish or transmit a file in the HDF5 format on a daily basis with the information about the most recent day’s ads. We recommend that platforms retain these daily ad summary files for one year. This time window is long enough to allow for errors

in transmission or formatting to be observed and corrected without data loss, and short enough to not impose excessive storage costs on ad platforms.

iv. Record updates, corrections, and deletions

There are several situations in which previously shared records would need to be updated, corrected, or potentially even deleted. Therefore, in addition to the data-related fields specified in Table 3 (see appendix), we recommend two additional metadata fields, “record_type” and “record_version.” “Record_type” should specify one of the following:

1. **NEW:** New records that have not previously been sent.
2. **UPDATE:** Changes to records that have previously been sent. This will occur when ad records that have previously been sent need to be updated; for example, because more spending and impressions have accrued to them. “UPDATE” should be used to indicate fields that were correct in the past, but have now changed.
3. **CORRECT:** Changes to records that have previously been sent that should be backfilled. This can happen for a variety of reasons. For example, if the ad payer had been incorrectly specified, it may be updated with the correct ad payer. Also, if data previously sent was incorrect because of a software bug, any field might need to be corrected. “CORRECT” should be used to indicate fields that were not correct in records that were previously sent.
4. **LEGAL_CORRECT:** Similar to “CORRECT” records but indicating that the ad record sent in the prior version contained illegal content. We discuss this situation in detail in subsection v.
5. **DELETE:** Deletions of records that have been previously sent. We foresee no reason for sending deletions for any actually delivered ads, but it is entirely possible that records that do not correspond to real ads could be sent due to software bugs. These buggy records would need to be deleted when the bugs were detected. For purposes of auditability, these should be considered soft deletions, where the records are marked as deleted but not actually removed.

Any ad, the first time it is transmitted, should be sent as a “NEW” record.

Thereafter, any changes to that record should be sent as an “UPDATE” or “CORRECT” record, as appropriate. The “UPDATE” and “CORRECT” records only need to contain the record ID and the fields that have been changed; unchanged data should not be resent. If a record is determined to have been sent in error, which is to say that the record does not correspond to a real ad that has been shown to users at least once, a “DELETE” record should be sent. “DELETE” records should only send the record ID of the record that is being deleted. “NEW” records should always have a “record_version” of “1.” Each subsequent update sent should increment the “record_version” by 1. When a correction is sent, the record version the correction should be applied to should be specified in the “record_version” field. Deletions will not require a “record_version.” We recommend that the FTC not apply “CORRECT” or “DELETE” records to their compiled daily records of ad data, so that researchers can maintain an auditable trail of changes to ad records. We do recommend, for the ease of researchers, that the FTC additionally maintain a snapshot dataset with all corrections, updates, and deletions applied.

v. Handling illegal content

In planning for universal ad transparency, we must expect that because an extremely wide range of content will be advertised, some of that content will be illegal to store. The two major categories of illegal content are child sexual abuse material and violent terrorist imagery. If a platform discovers that ad content that is illegal to store has been run on their platform, rather than sending the illegal content, they should send a hashed version of the illegal image or video in question, using PhotoDNA [64]. PhotoDNA is currently the industry standard to identify known illegal images and is already in widespread use. This hashed creative data should be sent as the initial creative if the illegal content is discovered before it is first transmitted, or as a “LEGAL_CORRECT” to the relevant fields if the illegal content is detected later.

We would expect that platforms would separately refer such content to law enforcement. The FTC should immediately apply “LEGAL_CORRECT” changes to data in its dataset and periodically conduct audits to ensure that this record type is being used correctly. It should be expected that some of the ad content that is transmitted is objectionable. However, this method should only be used to handle content that is illegal to store, not merely content that

violates the policies or community standards of the platform on which it ran. Both the FTC and outside parties will have visibility into how this record is used, because the FTC itself will continue to publish the “LEGAL_CORRECT” records it receives in its own time series data, both for auditing purposes and to allow the downstream consumers of ad transparency data to remove illegal content from any ad records they may have received in the past.

VII. INDUSTRY TRANSPARENCY EFFORTS

IN THIS SECTION, we compare the most prominent existing ad platform transparency efforts and our proposed standards. We described these efforts in detail in Section II. Here, we review strengths of these voluntary implementations that have informed our standard, as well as common shortcomings to these existing approaches that our standard seeks to address.

i. Modes of access

In this section, we will focus on researcher-oriented modes of access. While web portals for viewing ads are useful tools for a general audience, because of the sheer volume of ads, they are not a useful interface for researchers who want to understand the advertising ecosystem on a particular platform, or even the activity of a single major advertiser. For this reason, ad platforms have built a variety of systems to allow researchers to interact with bulk data.

Facebook offers an API [8] that allows researchers to download data that matches search criteria programmatically, with ad creative data such as images and videos available at additional links that need to be separately downloaded and parsed. Since the initial release, Facebook has gradually added additional parameters that allow users to filter the search results on select data fields. The API returns results that are accurate at the time of the request, but the only way to see what data has changed over time is to constantly reconduct the searches and re-collect the data. In contrast to the web version of the ad library, access to the ad library API is granted only to users who agree to the Facebook Terms of Service (TOS), and who complete Facebook’s verification requirements. We do not consider such a mode of

access as “public.” Researchers cannot access the ad data unless they first sign the TOS, which they may disagree with. Platforms can unilaterally revoke researcher access to their ad library APIs and have done so in the past [46]. Even without platform actions, the threat of the potential loss of access to programmatic data may compel researchers to self-censor and avoid research topics at odds with the platforms’ interests.

Google hosts its ad library as a SQL database on BigQuery, a cloud data warehouse. Similar to Facebook’s design, ad creative data is not included in the library itself, but referenced through outbound links, and needs to be downloaded and parsed separately. Otherwise, however, Google provides users with direct access to ad data (instead of interacting with the data through an API). Storing ad data in a standard SQL database has several advantages. SQL has been standardized for over 35 years [33], and provides efficient built-in capabilities for search, filtering, and sorting the results, and for generating summary statistics. Users have access to these standard tools without having to wait for the platforms to write the functionalities into the API. The ad data can be downloaded as CSV files or exported as a SQL database, and then analyzed using any standard analysis software or database software.

Roku and XFINITY offer downloadable document dumps of PDF and Excel files relating to ad activity.

As researchers, we have found Snapchat’s solution the most comprehensive and easiest-to-use mode of access. Snapchat provides a downloadable zip file, updated daily, that contains ad metadata, creative data, and downloadable links to ad images in a single zipped CSV file. This mode of access provides researchers with a single collection point for all data, and CSV files can easily be compared to versions collected at another time to identify differences over time. While we do not think that this CSV file structure is the right one for the larger volume of ads that our standard calls for, this overall access implementation was the inspiration for the single-file mode of access our standard suggests.

ii. Changes over time

One of the major shortcomings of existing ad platform transparency efforts is the lack of provision for researchers to track changes over time. All existing

transparency efforts of which we are aware give researchers a view of the platforms' ads at the current moment, without information on changes made to the ads over time. This is understandable as a naive approach, but our experience trying to use these transparency systems to understand the digital advertising ecosystem and identify bad actors within it has led us to the conclusion that the ability to track changes over time is absolutely necessary. We view our standard's approach, requiring that separate "UPDATE," "CORRECT," and "DELETE" messages be sent when an ad's data changes, as a significant improvement with respect to existing transparency efforts.

iii. Handling of policy violations

As with modes of access, ad platforms have taken a range of approaches to handling policy violations. Facebook hides ad creatives that have been found to violate policy behind click-through walls that warn the user they are about to see an ad that violated Facebook's policies. This has the effect of making the ads impossible for researchers to collect by automated means. Google simply removes the content and substitutes a message that the content violated its platform rules. We believe it is entirely reasonable for platforms not to want to host content that violates their community standards, or to at minimum warn users that they may see content that is offensive. However, such rule-violating content is often the most important for researchers to collect. This conflict is one that we believe illustrates the inherent shortcomings of these voluntary efforts, and one that our proposed standard addresses by turning over violating content, and the responsibility to host it, to a third party.

iv. Ad and advertiser inclusion

Ad platforms have also taken a wide range of approaches to deciding which ads and advertisers should be included in their transparency efforts. Here, Facebook deserves much credit for making social issue content transparent, in addition to electoral content. Facebook also deserves credit for the limited transparency measures it has taken around nonpolitical ads. While it only provides very limited web portal access to nonpolitical ads during the time the ads are active, and not at all after ads are deactivated, Facebook is the only platform of which we are aware that provides at least some transparency

into all ads. It should serve as an example to others that universal ad transparency systems are possible.

v. Ad creative data inclusion

Ad creative data inclusion is one of the few points of commonality between ad platforms' ad libraries. Facebook, Google, and Snapchat all make ad creative text(s), image(s), link(s), and video(s) available. Xfinity and Roku do not, and this absence makes their ad libraries significantly less useful to researchers.

vi. Ad targeting data inclusion

Ad targeting data is another area where different ad platforms have taken different approaches to transparency. In our experience, this is an additional instance where Snapchat has an excellent implementation that has informed our proposed standard. Notably, Snapchat makes both targeting inclusion and exclusion parameters available. It also makes a complete description of interest and segment targeting available, as well as advanced demographics. At first glance, Google's ad targeting transparency appears to be significantly less robust than Snapchat's because Google only lists geographic and demographic targeting information. However, this is because Google does not allow ads that they consider electoral to be targeted by interests or segments, so these other types of targeting do not apply [88]. Xfinity and Roku also make ad targeting (as it applies on their platforms) available in limited form. The only platform we review that does not make any form of ad targeting data publicly transparent is Facebook.

vii. Diversity of approaches is harmful for functional transparency

As we have shown, voluntary transparency efforts have taken a wide variety of forms that research must approach in entirely different ways. This "marketplace of ideas" approach has certainly been a useful one for testing out a variety of approaches to transparency. However, we believe that the time is right for a standardized approach. The variety of technical and policy approaches that different ad platforms have taken has created a situation where only the best-resourced researchers can afford to build the custom technical solutions required to use the transparency tools that the platforms

have built. Additionally, even researchers who can invest the technical resources required to comprehensively collect ad data from multiple platforms are left with datasets that are not directly comparable because of the different policy decisions made by platforms.

VIII. PENDING LEGISLATION

WE ARE AWARE of several major legislative efforts that require some form of digital ad transparency. We review those efforts here and compare them with our proposed standard in order to show how our proposal exceeds what is already being called for.

i. The Honest Ads Act

The Honest Ads Act

expands source disclosure requirements for political advertisements by (1) establishing that paid internet and paid digital communications may qualify as “public communications” or “electioneering communications” that may be subject to such requirements, and (2) imposing additional requirements relating to the form of such source disclosures and the information contained within. The bill requires certain online platform companies to maintain publicly available records about qualified political advertisements that have been purchased on their platforms. [55]

This bill was first introduced in 2018 by Sens. Amy Klobuchar, Mark Warner, and John McCain, and in 2021 was folded into the For the People Act [68], an omnibus democracy promotion bill. The bill has been endorsed by both Facebook and Twitter [57, 65].

In contrast to our proposal, the Honest Ads Act requires transparency only of political advertisements. Also, the bill requires that digital ad platforms maintain and publish their own archives of political ads, rather than creating a central database of ads from all platforms, as our proposal does. However, where our proposal calls for all ads from ad platforms that meet our criteria for inclusion to be made transparent, the Honest Ads Act calls

for all digital political ads to be made transparent, regardless of size of the platform. Additionally, the Honest Ads Act calls for payer disclosure of all digital political ads when they are displayed to users, something that is not currently required, and which is a form of transparency that our proposal does not speak to. Therefore, we largely view our proposal as complementary to the Honest Ads Act, rather than overlapping.

ii. The DATA Act

The Social Media Disclosure and Transparency of Advertisements Act, or DATA Act [76], was introduced in 2021 by Rep. Lori Trahan. The DATA Act “will ensure academic researchers have access to data related to the targeting of online digital advertisements” [1] by “mandating that large digital advertising platforms maintain an ad library for academic researchers that includes details about the advertisements.” [1]

Like our proposed standard, the DATA Act would require transparency of all digital ads from qualifying ad platforms, rather than only applying to political ads as the Honest Ads Act does. However, the act differs from our proposed standard in two key ways: First, it specifies that rather than being made transparent to the public, platforms would only be required to make their ad libraries available to “academic researchers.” Second, like the Honest Ads Act, it leaves the responsibility to build and maintain databases of ads with the ad platforms themselves.

iii. The Algorithmic Justice and Online Transparency Act

The Algorithmic Justice and Online Transparency Act was introduced in 2021 by Sen. Edward Markey and Rep. Doris Matsui [69]. While the primary focus of the bill is to ensure public transparency of content promotion and to target algorithms, it also contains a provision to require “any online platform (except for a small business) that uses personal information in combination with an algorithmic process to sell or publish an advertisement shall take all reasonable steps to maintain a library of such advertisements.” [58]

Of all the proposed legislation that we review, the Algorithmic Justice and Online Transparency Act is most closely aligned with our proposal. Unlike the DATA Act, this bill requires *public* transparency of all ads from

covered ad platforms. And similar to our proposal’s microtargeting criteria for ad platform inclusion, this bill’s criteria for which ad platforms should comply is tied to the use of personal information in combination with an algorithmic process. The primary difference between this bill and our proposed standard is that we additionally suggest that platforms that do not subject ads to human review before running them also comply with universal ad transparency.

IX. POTENTIAL PROBLEMS AND SOLUTIONS

IN THIS SECTION, we review potential problems with and costs of our proposal. We suggest potential remedies to some of these concerns where possible, but note that this is not always possible.

i. Privacy concerns

In our view, the primary issue that must be addressed by any attempt at greater transparency of digital ad data is privacy. There are real concerns about the privacy of users who may have been shown ads, advertisers who may have inadvertently shared their own private data in their ads, and third parties whose private information or images were inappropriately included by the advertiser in their ads. We will discuss these three concerns separately.

1. User privacy: We see the risk of identifying users who may have seen an ad as the single most serious problem to be avoided. We discuss strategies for doing this in Section V, but we stress that all parties who are involved in the process of making ad data transparent should take steps to ensure user privacy. First, advertisers should always store and transmit any private data that they have about targeted audiences in a secure fashion. Ad platforms that allow ads to be targeted by parameters that are identifying should anonymize those parameters by the methods we suggest in Section V. When determining whether very small audiences could be identified, both ad targeting and ad-impression data must be considered. Further, we recommend that ad platforms continue not to allow advertisers to target very small audiences

for ads. Lastly, we recommend that the FTC or any other organization that publishes an ad transparency dataset have double-check procedures in place to ensure that ad targeting data does not identify very small audiences.

2. Advertiser privacy: It is our view that advertisers' ad content is not private data and therefore the privacy concerns that we have about users being identifiable do not apply to advertisers. However, it is entirely possible that advertisers may inadvertently include their own private data in their ads and thus have a legitimate interest in having that private data be masked, for example, by accidentally listing their personal rather than their business address. In such cases, we believe the FTC should accept petitions for data masking of private data contained in ad content from advertisers.
3. Third-party privacy: Another possibility we must consider is that there may be third parties, aside from the advertiser or the user who sees the ad, whose privacy is compromised in ad content. The simplest such situation to contemplate is that of a person whose image is used by an advertiser without that person's permission. Currently, we know of no ad platform that has a specific process for people who believe that ads are compromising their private data or images to then request that those ads be taken down. We would encourage all ad platforms that adopt our standard to put such a process in place. We recommend that people who believe that an ad contains their private data or images be able to request that the ad in question be taken down and their private data masked from all transparency datasets. They should be able to make this request to the ad platform in question if the ad is still active, or to the data warehousing organization if the ad is no longer running.

ii. Access concerns

We are keenly aware that the technical standard we recommend presents a high barrier to entry for less technically savvy users. The file format HDF5 is not human-readable. The overall record-type structure we suggest of "NEW" and "UPDATE" records allows for time series to be observed but are more complicated to process. Unfortunately, we view this problem as inherent in the sheer volume of ad data that exists, and one that must be managed.

It is our hope that researchers will create tools to allow the public a more user-friendly view of ad data than the data warehouse we propose would allow. We believe this hope is well-founded, as there are several such tools that might serve as a model for such public-facing tools, such as OpenSecrets for FEC data [89], and the Wesleyan Media Project for television and election-related digital ad data [90]. We also wish to sound a note of caution: We have experience building and maintaining public transparency datasets, and this endeavor is both technically complex and expensive. Meaningful financial support of public-facing tools will be required to create meaningful functional transparency of digital ads. We note that these existing third-party public information and tool providers have always struggled to raise enough money to cover the costs of their operations, despite their excellent track record of informing the public, and the widespread use of their tools and data. We expect that the financial picture for providers of public ad transparency tools would likely not be substantially better. We discuss these financial considerations in further detail below.

iii. Cost concerns

We acknowledge the financial costs of what we suggest. Ad platforms that comply with our standard will incur costs, primarily in terms of additional engineering time to compile data into the formats we suggest and to transmit them appropriately. However, we do not recommend that platforms maintain their own ad transparency archives, which is a step that many ad platforms have felt the need to take. We believe the costs to ad platforms of complying with our standard to be significantly lower than those of building their own transparency systems, since we do not require them to store historical ad data or to make that data available to users through an interface that needs to be maintained. Thus, we expect that many platforms that we recommend comply with our standard would actually save money by doing so. Of course, this is possible because we recommend shifting to the FTC the costs of archiving ad data and of making that archive available to the public. Estimating the financial cost of such a data warehouse is outside the scope of this paper, but we suspect it would be similar to the data warehousing costs of other large projects undertaken by different branches of the U.S. government. Lastly, we expect that in addition to the data warehouse built by the FTC, third-party

tools to improve public understanding of ad transparency data would be needed. Proposing a structure for funding of strong functional transparency tools for public use again falls outside the scope of this paper. However, we believe it would be reasonable for ad platforms to, in some way, contribute to the cost to build and maintain tools for public access to ad transparency data because, as we noted earlier, we expect our proposal to be less expensive for them than building their own stand-alone public transparency tools.

iv. Competitive concerns

Lastly, we believe that both ad platforms and some advertisers may have concerns that digital ad transparency could place them at a competitive disadvantage. Digital marketers, in particular, may see information about their ad targeting strategies, and the details of how they A/B test images and messaging, as a vital component of their competitive edge. Ad platforms, particularly ones that control larger portions of the digital ad ecosystem, may perceive that the level of transparency that our standard requires would decrease their market power. We agree that the transparency we recommend will have disparate competitive effects. We do note that several businesses already exist to surveil and provide business intelligence on digital ads [9, 10]. Generally, though, the current situation in ad markets is one with a great deal of information asymmetry, whereas many decades of economic research [16, 71, 30] suggest that more information tends to make markets more efficient and less prone to failure.

X. CONCLUSION

We hope these efforts to survey the existing landscape of ad transparency and measure it against a proposed common (likely evolving) standard will inform continuing and ongoing work by regulators, researchers, and public interest advocates in this quickly evolving space.

REFERENCES

- [1] Congresswoman Lori Trahan. 2021. Fact Sheet: The Social Media DATA Act of 2021. Retrieved from https://trahan.house.gov/uploadedfiles/social_media_data_act_two-pager.pdf.
- [2] Facebook. 9.a. Ads About Social Issues, Elections or Politics. Retrieved from https://www.facebook.com/policies/ads/restricted_content/political.
- [3] Google. 2021. Political Advertising on Google. Retrieved from <https://transparencyreport.google.com/political-ads/home>.
- [4] Snap Inc. 2021. Snap Political and Advocacy Ads Library. Retrieved from https://businesshelp.snapchat.com/s/article/political-ads-library?language=en_US.
- [5] Roku. 2021. Roku's Political Ad Archive. Retrieved from <https://advertising.roku.com/en-gb/Roku-s-Political-Ad-Archive>.
- [6] Xfinity. 2021. Search for political advertisers. Retrieved from <https://poladarchive.comcast.com/>.
- [7] Google. The DoubleClick Ad Exchange. Retrieved from <https://static.googleusercontent.com/media/www.google.com/en//adexchange/AdExchangeOverview.pdf>.
- [8] Facebook. 2021. Facebook Ad Library API. Retrieved from <https://www.facebook.com/ads/library/api>.
- [9] Adalytics Research. 2021. Who is paying to get your attention? Retrieved from <https://adalytics.io/>.
- [10] Oracle. 2020. Moat by Oracle Advertising. Retrieved from <https://moat.com/>.
- [11] Mediakix. 2019. 93% of Top Celebrity Social Endorsements Violate FTC Rules. Retrieved from <https://mediakix.com/blog/celebrity-social-media-endorsements-violate-ftc-instagram/>.
- [12] Facebook. 2021. About Social Issues. Retrieved from <https://www.facebook.com/business/help/214754279118974?id=288762101909005>.
- [13] Federal Trade Commission. 2021. What We Do. Retrieved from <https://www.ftc.gov/about-ftc/what-we-do>.
- [14] Chrome web store. 2020. Ads Transparency Spotlight (Alpha). Retrieved from <https://chrome.google.com/webstore/detail/ads-transparency-spotlight/gkbmnjmlhjnakmfcejhlhpnibcbjdn/>.
- [15] GitHub Ads Transparency Spotlight team. 2020. Implement the Ads Transparency Spotlight (Alpha) Data Disclosure schema. Retrieved from <https://github.com/Ads-Transparency-Spotlight/documentation/blob/main/implement.md>.
- [16] George A. Akerlof. 1978. The Market for “Lemons”: Quality Uncertainty and the Market Mechanism. In *Uncertainty in Economics*, Peter Diamond and Michael Rothschild (Eds.), Academic Press, Inc., New York, NY, 235–251.
- [17] Muhammad Ali et al. 2019. Discrimination through Optimization: How Facebook's Ad Delivery Can Lead to Biased Outcomes. In *Proceedings of the ACM on Human-Computer Interaction 3*, CSCW, November 2019, 1–30. <https://doi.org/10.1145/3359301>
- [18] Barbara Allen and Daniel Stevens. 2018. *Truth in Advertising?: Lies in Political Advertising and How They Affect the Electorate*. Rowman & Littlefield, London.
- [19] Athanasios Andreou et al. 2018. Investigating Ad Transparency Mechanisms in Social Media: A Case Study of Facebook's Explanations. In *Network and Distributed Systems Security (NDSS) Symposium*, February 18-21, 2018, San Diego, CA. <http://dx.doi.org/10.14722/ndss.2018.23191>
- [20] Julia Angwin and Terry Parris, Jr. 2016. Facebook Lets Advertisers Exclude Users by Race. Retrieved from <https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race>.
- [21] Julia Angwin, Noam Scheiber, and Ariana Tobin. 2017. Facebook Job Ads Raise Concerns About Age Discrimination. Retrieved from <https://www.nytimes.com/2017/12/20/business/facebook-job-ads.html>.
- [22] Julia Angwin, Ariana Tobin, and Madeleine Varner. 2017. Facebook (Still) Letting Housing Advertisers Exclude Users by Race. Retrieved from <https://www.propublica.org/article/facebookadvertising-discrimination-housing-race-sex-national-origin>.
- [23] Julia Angwin, Madeleine Varner, and Ariana

-
- Tobin. 2017. Facebook Enabled Advertisers to Reach “Jew Haters.” Retrieved from <https://www.propublica.org/article/facebook-enabled-advertisersto-reach-jew-haters>.
- [24] Brooke Auxier and Monica Anderson. 2021. Social Media Use in 2021. Retrieved from <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/>.
- [25] Brian Barrett. 2018. For Russia, Unraveling US Democracy Was Just Another Day Job. Retrieved from <https://www.wired.com/story/mueller-indictment-internet-research-agency/>.
- [26] Katie Benner, Glenn Thrush, and Mike Isaac. 2019. Facebook Engages in Housing Discrimination With Its Ad Practices, U.S. Says. Retrieved from <https://www.nytimes.com/2019/03/28/us/politics/facebook-housing-discrimination.html>.
- [27] J Scott Brennen et al. 2020. Types, sources, and claims of COVID-19 misinformation. Reuters Institute 7(3), 1. Retrieved from <https://reutersinstitute.politics.ox.ac.uk/types-sources-and-claims-covid-19-misinformation>.
- [28] Duncan Brown and Nick Hayes. 2008. *Influencer Marketing* (1st. ed.). Routledge.
- [29] Federal Election Commission. Campaign finance data. Retrieved from <https://www.fec.gov/data/>.
- [30] John Cawley and Tomas Philipson. 1999. An Empirical Examination of Information Barriers to Trade in Insurance. *American Economic Review* 89, 4 (Sept. 1999), 827–846. DOI: <https://doi.org/10.1257/aer.89.4.827>
- [31] Daniel R. Coats. 2019. Worldwide Threat Assessment of the US Intelligence Community. Retrieved from <https://www.dni.gov/files/ODNI/documents/2019-ATA-SFR---SSCI.pdf>.
- [32] Warren Cornwall. 2020. Officials gird for a war on vaccine misinformation. *Science* 369, 6499 (Jul. 2020), 14-15. DOI: <https://doi.org/10.1126/science.369.6499.14>.
- [33] National Institute of Standards and Technology. Database Language SQL. Retrieved from https://www.itl.nist.gov/div897/ctg/dm/sql_info.html.
- [34] Google Ads. How it works. Retrieved from <https://ads.google.com/home/howit-works/>.
- [35] Laura Edelson et al. 2019. An Analysis of United States Online Political Advertising Transparency. arXiv:1902.04385. Retrieved from <https://arxiv.org/abs/1902.04385#>.
- [36] Insider Intelligence Editors. 2021. US influencer spending to surpass \$3 billion in 2021. Retrieved from <https://www.emarketer.com/content/us-influencer-spending-surpass-3-billion-2021>.
- [37] Google. Election advertising verification. Retrieved from <https://support.google.com/adspolicy/troubleshooter/9973345?hl=en>.
- [38] Mediakix. 2017. Breaking Down the FTC Letters Sent to Over 90 Celebs, Influencers, & Brands. Retrieved from <https://mediakix.com/blog/ftc-letters-notices-celebrities-brands-influencers/>.
- [39] U.S. House of Representatives Permanent Select Committee on Intelligence. Exposing Russia’s Effort to Sow Discord Online: The Internet Research Agency and Advertisements. Retrieved from <https://intelligence.house.gov/social-media-content/>.
- [40] Facebook. Advertising Policies. Retrieved from <https://www.facebook.com/policies/ads/>.
- [41] Facebook for Business. Ads Manager. Retrieved from <https://www.facebook.com/business/tools/ads-manager>.
- [42] Erika Franklin Fowler, Michael M. Franz, and Travis N. Ridout. 2020. Online Political Advertising in the United States. In *Social Media and Democracy: The State of the Field, Prospects for Reform*, Cambridge University Press, Cambridge, UK. 111–38. DOI: <https://doi.org/10.1017/9781108890960>.
- [43] Erika Franklin Fowler, Michael M. Franz, and Travis N. Ridout. 2016. *Political Advertising in the United States*. Westview Press, Boulder, CO.
- [44] Michael M. Franz, Erika Franklin Fowler, and Travis N. Ridout. 2020. Accessing information about interest group advertising content. *Interest Groups & Advocacy* 9 (Sept. 2020), 373–383. DOI: <https://doi.org/10.1057/s41309-020-00083-z>
- [45] Google. 2021. Google Ads policies. Retrieved from https://support.google.com/adspolicy/answer/6008942?visit_id=63762148095732510-1867564195&rd=1.
- [46] Taylor Hatmaker. 2021. Facebook cuts off NYU researcher access, prompting rebuke from

lawmakers. Retrieved from <https://techcrunch.com/2021/08/04/facebook-ad-observatory-nyuresearchers/>.

[47] Earthdata. 2021. HDF5 Data Model, File Format and Library—HDF5 1.6. Retrieved from <https://earthdata.nasa.gov/esdis/eso/standards-and-references/hdf5>.

[48] Paul Hitlin and Lee Rainie. 2019. Facebook Algorithms and Personal Data. Retrieved from <https://www.pewresearch.org/internet/2019/01/16/facebook-algorithms-and-personaldata/>.

[49] Jeff Horwitz. 2020. Political Groups Elude Facebook's Election Controls, Repost False Ads. Retrieved from https://www.wsj.com/articles/political-groups-elude-facebooks-electioncontrols-repost-false-ads-11604268856?st=5jm4uo293bxa50&reflink=desktopwebshare_permalink.

[50] Facebook. 2021. How are ads about social issues, elections and politics identified on Facebook? Retrieved from <https://www.facebook.com/help/180607332665293>.

[51] Facebook. How Disclaimers Work for Ads About Social Issues, Elections or Politics. Retrieved from <https://www.facebook.com/business/help/198009284345835?id=288762101909005>.

[52] Andrew Hutchinson. 2021. Twitter Quietly Removes its Ads Transparency Center. Retrieved from <https://www.socialmediatoday.com/news/twitter-quietly-removes-its-adstransparency-center/594485/>.

[53] Mike Isaac and Taylor Lorenz. 2021. Facebook Wants to Court Creators. It Could Be a Tough Sell. Retrieved from <https://www.nytimes.com/2021/07/12/technology/facebook-creators-tiktok-youtube.html>.

[54] Kiran Jagadeesh et al. 2021. Introducing new election-related ad data sets for researchers. Retrieved from <https://research.fb.com/blog/2021/02/introducing-newelection-related-ad-data-sets-for-researchers/>.

[55] Amy Klobuchar. 2018. S.1989 - Honest Ads Act. Retrieved from <https://www.congress.gov/bill/115th-congress/senate-bill/1989>.

[56] Daniel Kreiss and Shannon C McGregor. 2019. The “Arbiters of What Our Voters See”: Facebook and Google's Struggle with Policy, Process, and Enforce-

ment Around Political Advertising. *Political Communication* 36, 4 (Jun. 2019), 499–522. DOI: <https://doi.org/10.1080/10584609.2019.1619639>

[57] Taylor Hatmaker. 2018. Twitter endorses the Honest Ads Act, a bill promoting political ad transparency. Retrieved from <https://techcrunch.com/2018/04/10/twitter-honest-ads-act/?guccounter=1&guccounter=1>.

[58] Edward J. Markey. 2021. S. 1896 - Algorithmic Justice and Online Platform Transparency Act. Retrieved from <https://www.congress.gov/bill/117th-congress/senate-bill/1896/text>.

[59] Meredith McGehee and Rachel Moran. 2016. Who's Behind That Political Ad? Retrieved July 21, 2021 from <https://campaignlegal.org/document/whos-behind-politicalad>.

[60] John Detrixhe and Jeremy B. Merrill. 2019. The fight against financial advertisers using Facebook for digital redlining. Retrieved from <https://qz.com/1733345/the-fight-against-discriminatory-financial-ads-on-facebook/>.

[61] Federal Trade Commission. 2021. New Data Shows FTC Received 2.2 Million Fraud Reports from Consumers in 2020. Retrieved from <https://www.ftc.gov/news-events/press-releases/2021/02/new-data-shows-ftc-received-2-2-million-fraud-reports-consumers>.

[62] Jay Newell and Jeffrey Layne Blevins. Transparency in Political Advertising: Assessing the Utility and Validity of the FCC's Online Public Inspection File System. *Journal of Information Policy* 8 (2018), 417–441. DOI: <https://doi.org/10.5325/jinfopoli.8.2018.0417>

[63] David Oxenford. 2018. Beware of the Political File Obligations in this Hot Political Advertising Year. Retrieved from <https://www.broadcastlawblog.com/2018/10/articles/beware-of-the-political-file-obligations-in-this-hot-political-advertising-year/>.

[64] Microsoft. 2021. PhotoDNA. Retrieved from <https://www.microsoft.com/en-us/photodna>.

[65] Twitter Public Policy. 2018. Twitter is moving forward on our commitment to providing transparency for online ads. Retrieved from <https://twitter.com/Policy/status/983734920383270912>.

-
- [66] Google. Political advertising on Google FAQs. Retrieved from <https://support.google.com/transparencyreport/answer/9575640>.
- [67] Google. Political content. Retrieved from https://support.google.com/adspolicy/answer/6014595?hl=en&ref_topic=1626336.
- [68] John P. Sarbanes. 2021. H.R.1 - For the People Act of 2021. Retrieved from <https://www.congress.gov/bill/117th-congress/house-bill/1>.
- [69] Ed Markey, U.S. Senator for Massachusetts. 2021. Senator Markey, Rep. Matsui Introduce Legislation to Combat Harmful Algorithms and Create New Online Transparency Regime. Retrieved from <https://www.markey.senate.gov/news/press-releases/senator-markey-rep-matsuiintroduce-legislation-to-combat-harmful-algorithms-and-create-new-onlinetransparency-regime>.
- [70] Snap. 2021. Snap Political Ads Library. Retrieved from <https://www.snap.com/en-US/political-ads>.
- [71] Michael Spence. 1978. Job Market Signaling. In *Uncertainty in Economics*. Elsevier, Amsterdam, Netherlands. 281–306. DOI: <https://doi.org/10.1016/C2013-0-10586-7>.
- [72] Christopher Terry. 2017. Candidate Appearances, Equal Time, and the FCC’s Online Public File Database: Empirical Data on TV Station Compliance During the 2016 Presidential Primary. *Catholic University Journal of Law and Technology* 25, 2 (May 2017), 341–352. Retrieved from <https://scholarship.law.edu/jlt/vol25/iss2/6>.
- [73] The HDF Group. 2021. The HDF5 Library & File Format. Retrieved from <https://www.hdfgroup.org/solutions/hdf5>.
- [74] National Center for Super Computing Applications. 2021. The National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign. Retrieved from <http://www.ncsa.illinois.edu/>.
- [75] Sara Swann. 2019. The struggle is real: FEC stalled on regulations for online political ads. Retrieved from <https://thefulcrum.us/open-government/fec-no-internet-ad-regulation>.
- [76] United States Congresswoman Lori Trahan. 2021. Trahan Leads Introduction of Social Media DATA Transparency Legislation. Retrieved from <https://trahan.house.gov/news/documentsingle.aspx?DocumentID=2112>.
- [77] Gil Troy. 2021. *See How They Ran: The Changing Role of the Presidential Election*. Simon and Schuster, New York, NY.
- [78] Hal R. Varian. Online Ad Auctions. *American Economic Review* 99, 2 (May 2009), 430–34. DOI: <https://doi.org/10.1257/aer.99.2.430>
- [79] Meta. 2019. Why Am I Seeing This? We Have an Answer for You. Retrieved from <https://about.fb.com/news/2019/03/why-am-i-seeing-this/>.
- [80] Facebook for Business. 2021. Why Some Ads are Approved, Then Rejected. Retrieved from <https://www.facebook.com/business/help/133691315402558?helpref=search&sr=14&query=review>.
- [81] Google. Why you’re seeing an ad. Retrieved from <https://support.google.com/ads/answer/1634057?hl=en>.
- [82] Your Ad Choices. 2021. YourAdChoices Gives You Control. Retrieved from <https://youradchoices.com/>.
- [83] Xingquan Zhu et al. 2017. Ad Ecosystems and Key Components. In *Fraud Prevention in Online Digital Advertising*. Springer International Publishing, Cham, Switzerland. 7–18. Retrieved from <https://link.springer.com/book/10.1007/978-3-319-56793-8>.
- [84] Federal Election Commission. Advertising and Disclaimers. Retrieved from <https://www.fec.gov/help-candidates-and-committees/advertising-and-disclaimers/>.
- [85] Google. Overview of the Ads Transparency Spotlight (Alpha) extension. Retrieved from https://services.google.com/fh/files/misc/ads_transparency_spotlight_overview.pdf.
- [86] Kaya Yurief. 2019. Facebook settles lawsuits alleging discriminatory ads. Retrieved from <https://edition.cnn.com/2019/03/19/tech/facebook-discriminatory-ads-settlement/index.html>.
- [87] Nick Penzenstadler et al. 2020. We read every one of the 3,517 Facebook ads bought by Russians. Here’s what we found. Retrieved from <https://www.usatoday.com/story/news/2018/05/11/what-we-found-facebook-ads-russians-accused-election-meddling/602319002/>.

[88] Scott Spencer. 2019. An update on our political ads policy. Retrieved from <https://www.blog.google/technology/ads/update-our-political-ads-policy/>.

[89] Open Secrets. 2021. Open Secrets: Following the Money in Politics. Retrieved from <https://www.opensecrets.org/>.

[90] Wesleyan Media Project. 2021. Wesleyan Media Project. Retrieved from <https://mediaproject.wesleyan.edu/>.

APPENDIX

Proposed Standard Field	Description	Type
platform_archive_id	Id associated with the ad while being presented in a platform archive	bigint
texts	Text(s) of ads displayed	text[]
images	Image(s) of ads displayed	[]
videos	Video(s) of ads displayed	[]
links	Link(s) and displayed text(s)	[]
captions	Caption(s) associated with links displayed	text[]
disclosure_string	Text "paid for by" disclosure displayed to the user	text
advertiser	Advertiser name displayed to the user	text
ultimate_payer	Legal name of the advertiser who paid for the ad	text
platform_advertiser_iden- tifier	Platform specific id associated with the advertiser	bigint
national_advertiser_iden- tifier	Tax or election commission id associated with the advertiser if applicable	text
ad_creation_date_time	Date, time when ad was created	datetime
ad_active_dates	Dates when ad was active	date[]
displayed_creative_com- bination	The combinations of ad creative elements as they were displayed to the user	json
budget	Daily budget amount for ad	float
spend_by_day	Amount paid by the advertiser for the ad	float[]
currency	Currency of spend	text
removed	Whether the ad was removed	boolean
removed_reason	reason the ad was removed if applicable	text
removed_date	date, time when the ad was removed	datetime
total_impressions_by_day	all paid and organic ad impressions broken down by day	json
paid_impressions_by_day	paid ad impressions broken down by day	json
engagement	total ad engagement by type if applicable	json
total_impressions_by_ge- ography	all paid and organic ad impressions broken down by geographic region such as zipcode	json
paid_impressions_by_ge- ography	paid ad impressions broken down by geographic region such as zipcode	json
total_impressions_by_ demo_group	all paid and organic ad impressions broken down by age and gender group	json
paid_impressions_by_ demo_group	paid ad impressions broken down by age and gender group	json
total_impressions_by_plat- form	all paid and organic ad impressions broken down by platform (when applicable)	json
paid_impressions_by_plat- form	paid ad impressions broken down by platform (when applicable)	json
placement_details	other details about where an ad appeared	text
targeting_type	targeting categories used	text
targeting_details	textual description of targeting	json
targeting_inclusive_pa- rameters	parameters of inclusion in ad targeting	json
targeting_exclusive_pa- rameters	parameters of exclusions in ad targeting	json
delivery_platform	platform on which the ad was shown to the user	text
delivery_platform_optimi- zations	targeting optimizations performed by the platform	text
delivery_advertiser_input	input from advertiser which is used by the platform to optimize delivery	text
delivery_proxy_or_exter- nal_data	proxy optimization via influencers or additional external data	text

Table 3: Field description for universal ad transparency standard.

Proposed Standard Field	Facebook API Field	Facebook API - Notes Relative to the Proposed Standard Field
platform_archive_id	id	
texts	ad_creative_body	This field only contains one text in the case of dynamic creatives. Use is not recommended.
images		Can be collected from link in ad_snapshot_url
videos		Can be collected from link in ad_snapshot_url
links	ad_creative_link_description ad_creative_link_title	This field only contains one link in the case of dynamic creatives. Use is not recommended.
captions	ad_creative_link_caption	This field only contains one link in the case of dynamic creatives. Use is not recommended.
disclosure_string	funding_entity	
advertiser	page_name	
platform_advertiser_identifier	page_id	
ultimate_payer		Theoretically, funding_entity should always be the ultimate payer, but this is often not the case.
ad_creation_date_time	ad_creation_time	
ad_active_dates	ad_delivery_start_time ad_delivery_stop_time	Ads can be delivered noncontiguously, so there is no guarantee that an ad was delivered every day between the start and stop day
displayed_creative_combination	ad_snapshot_url	This URL can only be collected when the access token is active. All creative elements should be collected from this URL.
spend	spend	Spend is presented as a range instead of exact value
currency	currency	
impressions_by_geography	region_distribution	
impressions_by_demo_group	demographic_distribution	
impressions_by_platform	publisher_platforms	This field only specifies which platforms have had at least one impression, not a full breakdown by platform.

Table 4: Mapping of Facebook ad library API to universal ad transparency standard; not all proposed standard fields are present.

Proposed Standard Field	Google Field	Google - Notes Relative to the Proposed Standard Field
platform_archive_id	creative_stats.ad_id	
texts		Can be collected from link in creative_stats.ad_url
images		Can be collected from link in creative_stats.ad_url
videos		Can be collected from link in creative_stats.ad_url
links		Can be collected from link in creative_stats.ad_url
captions		Can be collected from link in creative_stats.ad_url
disclosure_string	creative_stats.ad_url	
advertiser	creative_stats.advertiser_name	
platform_advertiser_identifier	creative_stats.advertiser_id	
national_advertiser_identifier	advertiser_declared_stats. advertiser_declared_regulatory_id	
ultimate_payer		Theoretically, advertiser_name should always be the ultimate payer, but this is often not the case
ad_active_dates	creative_stats.first_served_timestamp creative_stats.last_served_timestamp creative_stats.num_of_days	Ads can be delivered noncontiguously so there is no guarantee that an ad was delivered every day between the start and stop day.
displayed_creative_combination	creative_stats.ad_url	
spend	creative_stats.spend_range_min_usd creative_stats.spend_range_max_usd	Spend is presented as a range instead of exact value
impressions_by_day	creative_stats.impressions	Impressions are not broken down by day
placement_details	creative_stats.ad_type	
targeting_inclusive_parameters	creative_stats.age_targeting creative_stats.gender_targeting creative_stats.geo_targeting_included	
targeting_exclusive_parameters	creative_stats.geo_targeting_excluded	

Table 5: Mapping of Google Ad Library to universal ad transparency standard; not all proposed standard fields are present.

Field	XFinity Web Portal	XFinity Notes
disclosure_string	id	Only present on effectv documents
ultimate_payer	name of organization purchasing time; on behalf of:	This field has various names on the various "political record of request" forms.
ad_active_dates	start date end date	This field is always present in "contract revision" documents.
impressions_by_day	quantity/impressions	Column name varies; always present in "contract revision" documents.
impressions_by_geography		All ad buys are presented in a DMA, so this is implied in impressions.
placement_details	seat details/placement name	
targeting_type	audience targets/targeting	
targeting_details	audience targets/targeting	
targeting_inclusive_parameters	audience targets/targeting	

Table 6: Mapping of XFinity to universal ad transparency standard; not all proposed standard fields are present.

Disclosure Product		Scope of Access & Disclosure		Types of Data Disclosed				Notes & Problems
Platform	Ad Archive(s)	Access	Types of Ads Disclosed	Ad Creative	Ad	Ad Targeting	Ad	
PROPOSED PLAN								
Platforms meeting any of the following criteria: 1. No human review of ads	Universal digital ad transparency proposal	Publicly accessible through database maintained by the FTC, in the form of: <ul style="list-style-type: none">• Daily HDF5 summary files (available for two years)• Daily HDF5 snapshot datasets (available for seven years)• Monthly and yearly advertiser summary CSV datasets (available for seven years)• Public-facing API/web interface	All forms of advertising on social media, including paid, branded, native, etc.					
				✓	✓	✓	✓	
EXISTING PLATFORM APPROACHES								
Facebook	Library for Ads on Social Issues, Elections, and Politics	Publicly available web portal (archives after seven years)	Social issues, electoral politics, housing, employment, credit	✓	✓	x	x	Per-ad spending is only disclosed in ranges
Facebook	Facebook Open Research & Transparency (FORT)	Only available via API to academic researchers who apply for access; only available in the U.S.	Political/electoral, only in the 90 days before the 2020 election with > 100 impressions	✓	✓	✓	x	Ads with fewer than 100 impressions make up a majority of political ads on Facebook during this period, so important data is missing
Google	Transparency Report	Publicly available web portal, report, and database through BigQuery platform	Electoral only: ads about candidates and/or officeholders related to state and federal elections, and about ballot measures	✓	✓	✓	x	
Google	Ads Transparency Spotlight	Browser extension for Chrome users	Allows users to browse information about ads targeted to them	✓	✓	✓	x	This information is revealed only if the advertiser has adopted Google's "ad disclosure schema"
Twitter	Ads Transparency Center	Public downloadable database	Political ads	✓	✓	x	x	Twitter discontinued this product in 2019, meaning only 1.5 years worth of data is collected and available
Snapchat	Political Ads Library	Public downloadable database	Political ads	✓	✓	✓	✓	Targeting parameters includes both inclusion and exclusion criteria
Roku	Political Ad Archive	Folder of files, but requires users to log in through Roku's Box.com account	Scanned advertising contracts from political advertisers, likely identifies the advertisers themselves	x	x	x	x	Contracts are in differing formats and contain codes not explained in the documents
Xfinity	Political ad archive	Publicly available searchable web portal, but requires users to log in using a Microsoft account	Political ads	Unknown	Unknown	Unknown	Unknown	Interface is broken; without modifying the script, users cannot access information



About the Authors

LAURA EDELSON is a Ph.D. candidate in computer science at New York University and co-director of NYU Cybersecurity for Democracy.

JASON CHUANG is a researcher in data science and analytics at Mozilla.

ERIKA FRANKLIN FOWLER is professor of government at Wesleyan University and co-director of the Wesleyan Media Project.

MICHAEL M. FRANZ is professor of government and legal studies at Bowdoin College and co-director of the Wesleyan Media Project.

TRAVIS N. RIDOUT is the Thomas S. Foley Distinguished Professor of Government and Public Policy in the School of Politics, Philosophy and Public Affairs at Washington State University and co-director of the Wesleyan Media Project.

Acknowledgments

Cybersecurity for Democracy at NYU's Center for Cybersecurity has been supported by Democracy Fund, Luminate, Media Democracy Fund, the National Science Foundation under grant 1814816, Reset, and Wellspring. The Wesleyan Media Project gratefully acknowledges support from the Democracy Fund, and the John S. and James L. Knight Foundation.

© 2021, Laura Edelson, Jason Chuang, Erika Franklin Fowler, Michael M. Franz, and Travis N. Ridout.

About the Knight First Amendment Institute

The Knight First Amendment Institute at Columbia University defends the freedoms of speech and the press in the digital age through strategic litigation, research, and public education. It promotes a system of free expression that is open and inclusive, that broadens and elevates public discourse, and that fosters creativity, accountability, and effective self-government.

knightcolumbia.org

Design: Point Five

Illustration: © Daniel Hertzberg



**KNIGHT
FIRST AMENDMENT
INSTITUTE** at
COLUMBIA UNIVERSITY