Commissioned by PolicyLink
May 2023

Bias and Discrimination in Digital Advertising

Danaë Metaxa, PhD
University of Pennsylvania

Abstract

This white paper examines the digital advertising ecosystem, with a particular focus on social inequality. I begin in Section 1 by introducing the concept of a sociotechnical system and the challenges of studying such systems empirically: these systems are dynamic, their data are ephemeral, they are embedded deeply in people's everyday lives, and they are highly personalized. Next, in Section 2, I introduce the reader to digital advertising, beginning with a brief history of advertising before the digital era and continuing to explain the technical infrastructure that underlies this ecosystem. In Section 3, I delve into bias and discrimination in digital ads on both a theoretical level (legal context and types of bias) and a summarization of work that has empirically studied discrimination in digital ads. Finally, in Section 4, I conclude by discussing why these issues persist and how we might change them—from better research to law and policy.

Table of Contents

1.0
Introduction: Studying Sociotechnical Systems

Before delving into digital advertising, as an introduction, I first provide an overview of sociotechnical systems—a category that encompasses systems including digital advertising, social media, search engines, and others—in order to establish a baseline level of understanding of these technologies, the academic disciplines that study them, and the challenges faced in doing so.

What Are Sociotechnical Systems?

Simply put, sociotechnical systems are those systems in which *people*, their users, play an integral role. The term was coined in the mid-20th century to describe workforce-related systems that functioned through the complex interconnection of people and technologies (broadly construed, not only digital ones) [Trist, 1981]. This term is used by scholars in a wide range of disciplines, including management science (e.g. [Bentley et al., 2016]), biological science (e.g. [Hyer et al., 1999]), and computer science. My focus in this article is on the use of the term in computer and information sciences, in particular scholars studying computer-supported cooperative work (CSCW). The CSCW community is closely tied to a subfield of computer and information science known as human-computer interaction (HCI). Together, the CSCW and HCI communities (which have significant overlap) explore questions about human users' interactions with digital technologies, communication with and through machines, and use of technology toward collaborative ends (sometimes explicitly work-related but also more broadly). For the interested reader, some key pieces of scholarship in this domain are Lucy Suchman's 1987 book about interface design that contributed to the founding of CSCW as a topic of study [Suchman, 1987], and Mark Ackerman's 2000 article identifying a key challenge over the previous decade of CSCW work, termed the "sociotechnical gap" [Ackerman, 2000].

So what does "sociotechnical system" mean in computer and information science? The term describes digital systems that are tightly tied to their human users—they shape and are in turn shaped by users. This includes, for example, search engines and social media sites. Algorithmic content is a closely related concept: the content to which sociotechnical systems expose users today—on a Facebook (now owned by Meta) user's Feed or a Google Search user's page of search results—is selected and organized algorithmically, rather than being curated by hand. Related research in the field of communication would include such content under the umbrella of "AI-mediated communication" [Hancock et al., 2020].

An aside about algorithmic content: it is partly a result of the pressures of scale; these systems are deployed at a massive scale, with thousands or even billions of users. As more traditional forms of media like newspapers and television reflect, however, scale itself does not necessitate algorithmic tools. Rather, the decision of technologists to focus on *personalization* combines with massive scale to implicate algorithmic content (an idea that will become important later on, when I discuss digital advertising).

In computing, recent work has begun identifying normative and ethical issues in sociotechnical systems and the algorithmic content they rely on and produce. These issues include bias along the lines of race [Buolamwini and Gebru, 2018], gender [Chen et al., 2018], political affiliation [Robertson et al., 2018], and more—biases that are all the more concerning for the far reach and large scope of these sociotechnical systems.

DEEP Challenges

In my research, I have identified four main attributes of today's algorithmically mediated sociotechnical systems that make them challenging to study. Two of these challenges are located within the technology and two on the side of the user, as both sides interact to produce a single system. I term these "DEEP," an acronym for the four attributes (dynamicity, ephemerality, embeddedness, and personalization) and a fitting label for capturing the importance and complexity of the challenges these attributes represent. Consider as a running example through this section a prominent sociotechnical system, the popular search engine Google.

Our first DEEP challenge is located on the side of the technology: sociotechnical systems are *dynamic*; they change constantly, which can pose significant technical challenges in trying to study them. This is especially true when they are powered by algorithms built with machine learning technologies that cause them to change constantly and in response to users' inputs. This makes the systems a moving target. Today's results may not hold next year, or even tomorrow—what's more, the speed at which these changes happen, or the degree of the changes themselves, are generally not known by those not privy to their inner workings. Our example, Google's search engine, changes frequently over time, both in response to external changes in the web and internal changes to its algorithm. If I do a Google search today for "coffee shops near me," those results will look different today than they will next month if a new coffee shop has opened in my neighborhood in the interim. And, those results can also change if Google's algorithm changes how it displays location-based queries like coffee shops or changes something else about its algorithm that affects the prioritization of those results.

Second, the media produced by these technologies are *ephemeral*; a user's experience of their news feed or search engine results disappears without a trace after that interaction, and it is impossible to choose to examine it retrospectively without having planned for that examination ahead of time. Using the same example, let's say I remember finding a great coffee shop in my neighborhood using Google last week, but (due to the dynamicity of those results) today I do the same search and cannot find it. What I would like to do is pull up that search results page I looked at last week, but this is impossible to do. Neither I nor even Google can tell me what that page looked like last week. The data simply does not persist; it is ephemeral.

Our third challenge occurs on the side of the user. These systems are *embedded*; by definition, they are in a loop with users, whose experiences of those systems are a fundamentally important variable of interest. In every step of the process, from system design to evaluation, the (often subjective, highly variable) experiences of users are a critical piece of the system to understand. This makes such systems more complex, requiring broader and interdisciplinary expertise to study than other types of software. Moreover, users are interacting with dozens of these systems simultaneously in a single browser or on a single device throughout their daily lives. This embeddedness also increases the difficulty of building and understanding sociotechnical systems. Using the example of a search, a user looking for coffee shops near them might really be interested in getting directions to it, placing an online order, or sharing information about it with a friend. Deciding how to order results, what information to prioritize displaying to the user, what other services (like, say, Google Maps) to integrate, and so on are all complex questions that require understanding the system users' myriad goals, values, and other attributes. The seemingly straightforward ask—"Does this system work?"—is deeply complicated.

And finally, our last DEEP challenge: these systems are *personalized*. Each user's view of a sociotechnical system can be unique, which poses technical and theoretical challenges in understanding the system's content and impacts. And indeed, the impact these systems have on their users can be very different from one person to another. Going back to our running example, imagine a Google engineer wants to improve the search results shown for location-based queries, like "coffee shops near me." One metric of success might be whether all coffee shops in a reasonable radius, say, 2 miles, are displayed to the user conducting that search. But for a user in the middle of San Francisco, this might be an overwhelming number of results, especially if they are sorted randomly. Meanwhile, for a user not too far away, in rural Northern California, this might frustratingly not produce any search results at all! Personal differences in users' locations, tastes, values, and so on all complicate the task of understanding whether this sociotechnical system works effectively.

These DEEP complexities make sociotechnical systems hard to study, especially in the domain of computer science, a discipline that has not traditionally been very strong in asking such interdisciplinary user-centered questions. But as we will see in the following sections, these challenges are the backdrop for studying such systems and understanding their impacts on the world.

2.0
The Case of Digital Advertising

This section overviews the landscape of the digital advertising industry, after first briefly discussing advertising in the predigital era.

A Brief History of Predigital Advertising

For an excellent and relatively succinct history of advertising with a focus on the United States, see [O'Barr, 2010]; this section briefly summarizes that history.

Advertising's history dates back to at least the 1600s, with the invention of the newspaper; advertisements were run alongside news articles, with advertisers paying the paper to print their ads. In the mid-1800s, newspapers began raising the costs of running advertisements in order to subsidize the cost of the newspaper, making the medium more widely accessible by lowering the cost to the public—another practice reminiscent of today's media ecosystem. Toward the end of that century, advertising agencies came into existence (companies that managed the writing and distributing of ads), as did other advertising practices, like billboards and public signage (for example on trains and by roadsides). By the early 1900s, advertisements appeared in newspapers and magazines, and on the newly invented radio. Gendered advertising also became common, with many ads taking a paternalistic tone to sell home goods to women. Commercial television, becoming widespread in the mid-1950s, also ran commercials between and during its programs. Notably, by this account, virtually all mass media in the United States following the newspaper was founded with advertising as an integral part. These advertising messages were very broad, needing to appeal to a wide base of readers, listeners, or viewers, in contrast with other sales techniques like door-to-door salesmen who made personalized pitches face-to-face in the 1800s and 1900s.

The Internet was invented in 1983, when a standardized set of communication protocols began allowing computers to communicate with each other around the world. In 1989, the World Wide Web was invented, allowing people to navigate to specific content on a web page using a web browser—practices most people today would consider synonymous with "the Internet" [Pew Research, 2014]. Unlike the newspaper, radio, and television, the Internet did not begin as a form of mass media; it was an interconnected computer network on which public-facing content did not appear until the advent of the Web. As mass media began to spread through this medium and commercial activities began to take place online, advertisements began to appear—the first banner ad appeared in a web browser in 1994 [Pew Research, 2014].

By 1999, when the founders of Google wrote the foundational paper leading to the creation of their search engine, online advertising existed, but it was not yet clear that advertising revenue would underpin the Web to a major degree. In that paper, the authors included a subsection entitled, “Manipulation by Commercial Interests,” describing the “worst” risk to their algorithm: “manipulation in the form of buying advertisements (links) on important sites” [Page et al., 1999]. Ads were seen as risks to information quality, not the bedrock of the dominant search engine those authors subsequently founded.

Today's Digital Advertising Landscape

This section overviews the digital ad landscape—its terminology, key players, and technical infrastructure. Disclaimer: it is not necessary to understand the full details of this very complex system in order to grasp Section 3 on inequality in digital advertising, but this level of detail does become a bit more important for Section 4 on technical fixes, as good measurement and policy relies on a solid understanding of the system itself.

Imagine you open your browser and navigate to a website—say, *The Washington Post*—as I just have. In the first split second the page is loading, right below the name of the paper, I see a wide gray box that says "Advertisement" (see Figure 1). Within less than a second, this box has been replaced with an ad—in my case, an ad for the Press Freedom Partnership, raising awareness about a press freedom issue in Ethiopia.

What happened in that split second that led to this particular ad being loaded onto the page? The answer is an auction. In the milliseconds after I loaded this page in my browser, an auctioning process called real-time bidding (or RTB, for short) occurred, with an ad server putting out a request for bids to fill that space, receiving and evaluating those bids, and placing the winning ad on my page. And while this took long enough today that I was able to capture a screenshot of the unfilled ad slot, in many cases it occurs fast enough that a user would never even notice the interruption.

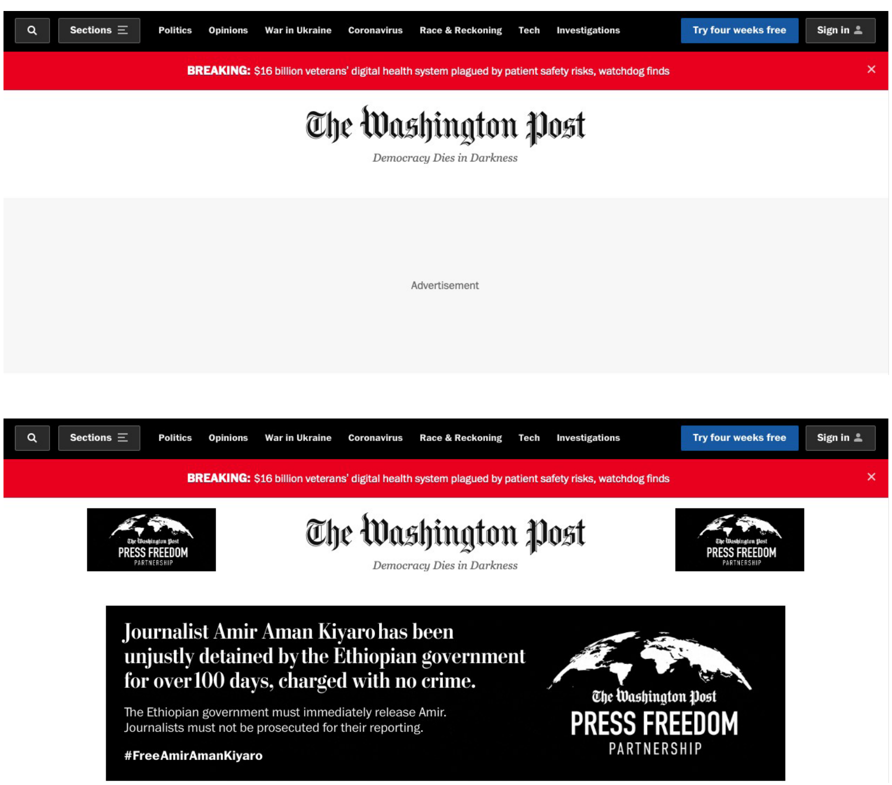

Figure 1: Two screenshots from *The Washington Post*'s website taken in March 2022, (above) in the moments just after the webpage is loaded, and (below) about one second later when the advertisement loads.

Key Players in the Digital Ad Marketplace

As with any auction, in the auction for digital ads, there are buyers and sellers. (Neither of these figures is the person using the web browser.) Advertisers are the buyers (also called bidders, buy-side, or demand side)—those paying money to have their advertisements shown to people online. Publishers are the sellers of online real estate (also called sellers, sell-side, or supply side); these are the websites looking to accept revenue in exchange for showing an advertiser's ads to people visiting their site.

Because these auctions happen in real time, in a matter of milliseconds, the entire process is programmatic, facilitated by software platforms. These platforms fall into two categories: demand-side platforms (DSPs) that service advertisers looking to buy space on websites and supply-side platforms (SSPs) that service the publishers with space to sell. DSPs allow advertisers to specify things like the images they want to show and various aspects of the audiences they are trying to reach, and SSPs allow publishers to do things like set price floors and specify what kinds of ads they are willing to have shown on their pages. Every time a person loads a webpage whose publisher has made space available for ads to appear, an auction occurs automatically between those two parties (the DSP and SSP), and once it resolves (nearly instantaneously), an ad is selected and loaded onto the page for the user to see.

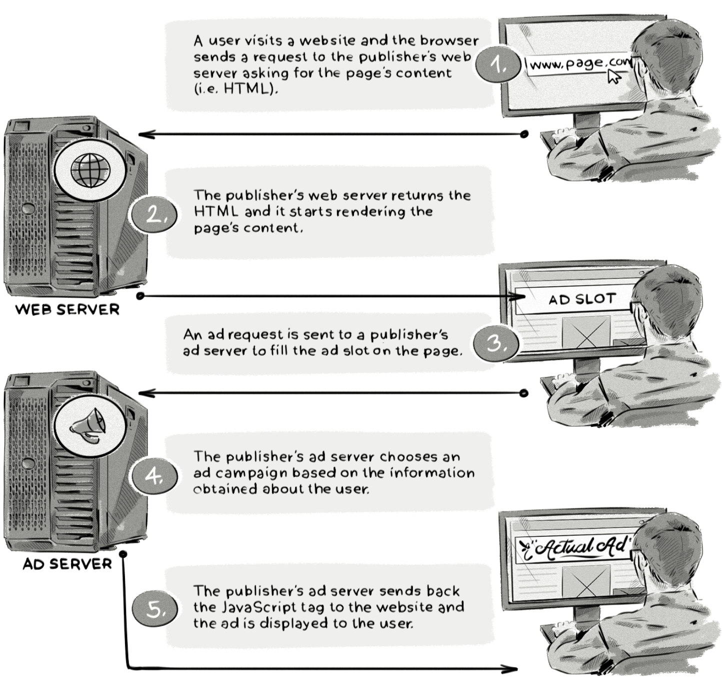

Figure 2: A visual summary of the way publisher-side ad servers work.
Source: [Zawadziński, 2014]

First-Party versus Third-Party Ad Servers

There is another small layer of complexity here—the ad servers that run these auctions to enable the management, serving, and tracking of ads come in two flavors: publisher-side (or sell-side, or first-party) ad servers and advertiser-side (or buy-side, or third-party) ad servers. Publisher-side ad servers allow publishers to manage the ads appearing on their own properties, displaying ads in real estate that was sold directly to advertisers (thus the term “first-party”). In this regime, the publisher’s ad server chooses an ad to serve using information about the open real estate and the user and sends back the ad (see Figure 2).

Advertiser-side ad servers manage ads appearing on someone else’s digital real estate, with the goal of helping advertisers do things like storing and managing their own ads, and tracking views and other performance metrics. In this case, the same auction occurs, but instead of directly placing the ad on the page, the ad publisher’s server instead inserts a reference to the advertiser’s ad server, and that second server provides the ad as the page loads. Since the advertiser’s server is getting directly pinged to provide that content, the advertiser gets additional direct information (like that the image was loaded and displayed to the user) (see Figure 3).

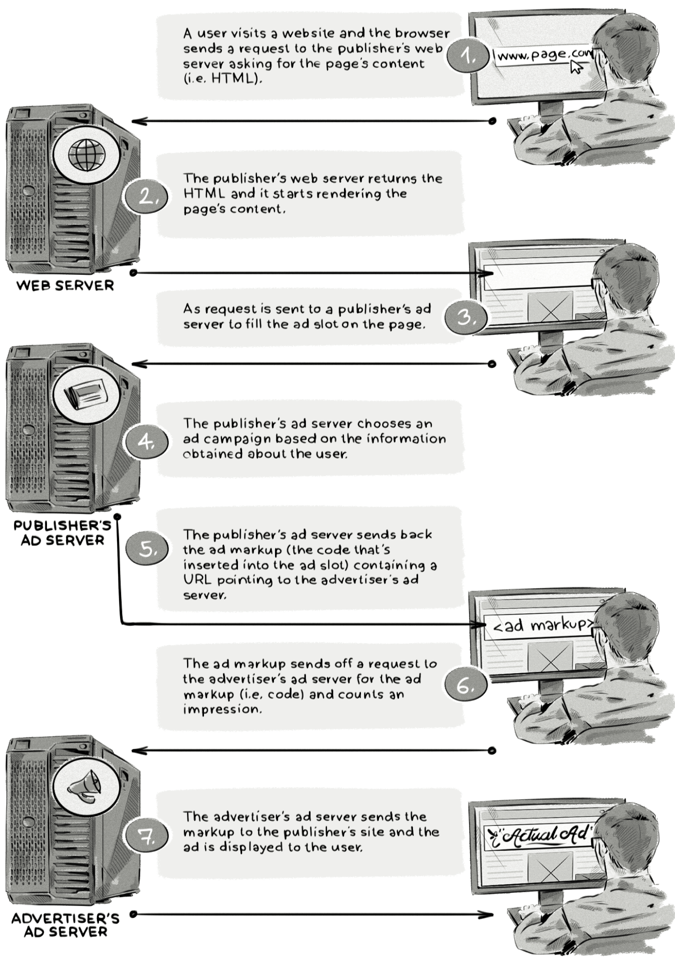

Figure 3: A visual summary of the way advertiser-side ad servers work.
Source: [Zawadziński, 2014]

Ad Exchanges

An ad exchange is a technology platform that facilitates the buying and selling of ads—an intermediate platform that serves neither the buyer nor the seller but instead connects the two sides and runs the auction. Publishers connect to this marketplace through their SSPs, and advertisers connect through their DSPs. Some examples are Google's Ad Manager and Microsoft Advertising.

Some ad exchanges are not open to anyone wanting to participate in the marketplace but are instead a closed group for private trading of ads; these private marketplaces only allow selected advertisers who have a special agreement with a publisher to bid for a given piece of online real estate.

Performance Metrics

As we have seen described above, digital advertising is essentially a marketplace driven by auctions, with advertisers bidding to pay publishers for the space on their webpages. As with any market, it is necessary for both buyers and sellers to have a common understanding of the value of the good being sold—in the case of ads, this good is the attention of website visitors who will see the ad appear in their browsers. A web user's attention, however, is a bit of a nebulous concept, so the ad marketplace needs concrete ways to measure and price it. These ways of measuring and pricing are ad performance metrics.

One main metric is the *impression*. One impression counts an instance of a digital advertisement that is rendered in the webpage—displayed on the user's screen. This is also referred to as "view-through" (as opposed to "click-through"). *Conversion* refers to any goal action taken by the user that the advertiser deems important (a view having been "converted" to something else). This might just be an ad view but could also be a click on the ad, purchase of a product, download of an app, and so on.

If users' exposures and responses to ads are measured in impressions and conversions, there also need to be currencies, that is, units of measurement for the cost of those goods. There are several such units, including cost *per mille (CPM)* which measures the cost to the advertiser per thousand impressions; cost *per click (CPC)*, the cost to the advertiser per click rather than per view of an ad; *click-through rate (CTR)*, a measure of the number of times a user clicked on an ad or link; and *cost per acquisition* or *cost per action (CPA)*, the effective cost to the advertiser for the engagement they see in return or, in other words, a ratio of the amount they paid to place the ad against the number of conversions.

For another detailed examination of the programmatic ad marketplace and a comparison with financial markets that largely work in the same automated way today, see Tim Hwang's 2020 book, *Subprime attention crisis* [Hwang, 2020], or other online resources like that from which Figures 2 and 3 were sourced [Zawadziński, 2014].

Digital Advertising's Strength: Targeting Users

For all the technical innovation that goes into the digital ad marketplace, the key difference about online versus offline advertising is the ability to target consumers more precisely, and much more inexpensively [Goldfarb, 2014].

The ad ecosystem includes several different types of ad targeting, most of which are relatively intuitive to understand by name. They include the following:

- *Contextual targeting:* With contextual advertising, ads match their context, the content on the page on which they are shown. This includes "sponsored" listings when a user searches Amazon for a product or ads that simply thematically match the webpage around them.
- *Geographical targeting:* Location information may be present among the many pieces of information available to a server when a browser makes a request for some content. Geo-targeting involves serving users ads that are relevant to their current location, like when a search for "coffee" displays ads for local coffee roasters. Many other types of information (including nonadvertisement search results) also use geo-targeting.
- *Demographic targeting:* Demographic targeting is one major type of targeting made possible at a much more sophisticated level by digital ads. When buying ads, advertisers can specify demographic groups to which they would like their ads served—attributes like gender, age, and marital status. Some of these attributes might be explicitly provided by the user in question, though not always knowingly; for instance, providing one's birthday or gender when signing up for a service like a Facebook (social media) account gives Facebook (in its capacity as an advertising platform and intermediary) access to that information. But others are inferred attributes that ad platforms guess about the user.
- *Behavioral targeting:* Users are also targeted based on their own past behaviors. Like some demographic attributes, this type of targeting is enabled not by explicit user contribution but rather by inferences driven by user actions online. This can include *retargeting*, when users are shown ads for content they previously saw or searched for (think of browsing a website looking to buy, say, a pair of sneakers, and seeing ads for sneakers on other websites in the subsequent days or weeks). This kind of behavioral targeting is also used by ad platforms to categorize users with nondemographic categorizations, like topics of interest ("motorcycles" or "pets"), categories later used to target future ads.

For more discussion on targeting, see [Goldfarb, 2014]. It is important to note that although those interested in privacy or inequality (e.g., readers of this article) may immediately think of demographic or behavioral targeting when they hear "targeted ads," these are only two of many forms of ad targeting that occur.

Does Digital Advertising Work?

The advertising industry brings in hundreds of billions of dollars annually (nearly $150 billion in the United States alone in 2020) [Statista, 2022]. Nevertheless, the assumptions it rests on are still relatively untested [Marotta et al., 2017]. In fact, this is true of advertising as a whole; as one article notes, "there is little well-identified analysis on the effectiveness of advertising in general" [Goldfarb, 2014]. This is true of the direct goal (changing people's beliefs and behaviors by showing them ads) and of the methods used (targeting people accurately and effectively). It is important to note that "effectiveness" means different things depending on the perspective one takes; what is effective for advertisers may be different than for publishers, intermediaries, and users. Let's explore each of these perspectives in turn.

For Advertisers?

How well does digital advertising work for advertisers? In this context, effectiveness refers to the perspective of advertisers, who are hoping that paying money to place ads in front of users will yield some desired outcome (getting information out to consumers, driving sales, and so on).

The effectiveness of offline ads is difficult to study since researchers, performing studies in lab environments or looking for correlations between ads and sales, have a very limited view of a consumer's actual, multi-faceted experience of the world. In contrast, in the realm of digital ads, many magnitudes more data, and more fine-grained data, are available: what ads a user saw, how long it appeared on their screen, whether they clicked on it, or whether they bought the product.

Still, this availability of data does not solve the problem of understanding whether an ad is effective or how effective it is. Recall the DEEP problems in studying sociotechnical systems outlined in Section 1: dynamicity, ephemerality, embeddedness, and personalization. As with other sociotechnical systems, these challenges trouble the study of targeted advertising. The content users see is constantly changing and hard to get a full picture of, and users are undertaking many actions simultaneously while browsing. As a result, using purely observational data can lead to overestimating the effect of ads [Lewis et al., 2011], but conducting effective experiments to causally demonstrate the impact of different aspects of an ad on users is hard to do at a scale large enough to closely measure the effects [Lewis and Rao, 2013].

Clearly, this uncertainty has not slowed the market for ads; advertisers still spend large budgets on digital advertising every year. This has led some to speculate that the foundation of digital advertising is shaky, and prone to collapse [Hwang, 2020]. Whether this will come to pass remains to be seen, but for the time being, targeted ads seem to work well enough for advertisers.

For Publishers?

How well does digital advertising work for publishers? Advertisers, however, are not the only stakeholders in this environment. We might similarly ask whether targeted advertising works for publishers. For the most part, publishers stand to clearly benefit from this ad marketplace; they are able to sell space on their pages to advertisers relatively easily. However, there are a few potential risks. Publishers might have their reputations or their users harmed by bad advertisements—these come in at least three flavors [Rosen, N.D.]. First, ads that are misplaced: those without any connection to their page or that are bothersome to users. Second, ads that are low-quality: those that advertise for shoddy products or spread misinformation might lead a web user to think poorly of the site on which the ad appears. And finally, those that cause outright harm: ads for scams, those causing users to download malware, and so on. Minimizing these kinds of harms is a priority for ad exchanges who would like publishers to keep coming to them to be matched with ads. Despite these risks, publishers all over the web choose to sell real estate on their sites to advertisers for revenue—because it works well, it works well enough, or they do not have other options for monetizing their content.

For Intermediaries?

How well does digital advertising work for intermediaries? This system undeniably works best for intermediaries, those companies running the ad exchanges we described above. Major players like Google and Facebook, among others, have built infrastructures generating huge revenue—enough to power the rest of their massive companies, in some cases—by selling ads on their own platforms as publishers, but also by acting as intermediaries, helping advertisers place ads using the large swaths of data they have about web users.

Having asked how well targeted digital advertising works for the main economic actors in the industry—advertisers, publishers, and intermediaries—we are still missing one key part of the ecosystem: web users. Accordingly, in the next section we ask: How well does targeted advertising work for web users?

For the Rest of Us?

How well does digital advertising work for the rest of us? Most of the works cited in the section above come from the fields of management science and economics; these disciplines are concerned with the economic and business aspects of industries like advertising. But this focus only tells one part of the story. Social scientists, information scientists, and media scholars have instead studied advertising with a closer eye to the experiences of web users—those being served digital ads, especially targeted ones.

Before getting into the possibility for bias and discrimination against users (the focus of the rest of this paper), it is worth first reviewing people's feelings about ad targeting in general, especially behavioral and other types of targeting that involved tracking users or making inferences about them. In one of the earliest surveys of web users, conducted via phone in 2009, the majority of respondents did not want tailored advertisements at all (66%), and an even larger majority did not want their behavior on other websites to be used to display ads on a given site (84%). The vast majority also felt they should be able to tell companies to delete all their data (92%) [Turow et al., 2009].

Another survey from 2010 asked related questions of users: what kinds of their personal data were involved in targeted advertising, how such personal data were collected, whether they had a right to the privacy of their data, and whether they would exchange it for free access to services (like search engines and social media) [McDonald and Cranor, 2010]. This study found widespread misunderstandings, including a lack of awareness that content like one's email could be used for ad targeting (only 39% believed this to be the case, though the practice did—and does—exist), and widespread rejection of such practices (e.g., only 9% believing such email-based targeting was acceptable as an exchange for free email).

Fast forward nearly a decade to 2019, and surveys from the Pew Research Center found web users to be more informed about advertising practices, with a majority saying they had heard that personal data was used to target them with advertisements (77%) [Pew Research, 2019b]. However, when asked whether they were aware that Facebook kept a list of their traits and interests, the majority were not (74%), and a small majority were not comfortable with this information being collected (51%) [Pew Research, 2019a]. Despite this lack of awareness and discomfort, when asked whether the ads they saw reflected them, a plurality (39%) of U.S. adults reported that they did at least somewhat well [Pew Research, 2019b]. When viewing the specific categories Facebook had placed them in, a majority (59%) said that those categories were reflective of their real lives.

So, does targeted advertising work for users? Taken together, these studies suggest that there may still be widespread ignorance and even misunderstanding of how such technologies work, but users do perceive them as working relatively well in terms of accurate targeting. However, users are still not fully comfortable with these practices, a fact that has remained the case for over a decade.

3.0
Bias and Discrimination in Digital Advertising

An unfortunate recurring issue among the sociotechnical systems we discussed in Section 1 is the potential for bias and harm to people—their proximity to real life is a fundamental benefit of these systems, one that naturally also comes with risks. In the context of digital advertising, bias and discrimination can negatively impact all aspects of the system: the advertisers, publishers, intermediaries, and users. In this section, I begin with some important legal context for these issues; next, I cover a (slightly informal) taxonomy of bias in digital advertising; third, I summarize existing empirical studies of such biases; and, finally, I end by considering cases where bias in digital ads may actually be beneficial.

Legal Context

Bias and discrimination are recurring issues in studies of algorithmic systems [Metaxa et al., 2021]. Whether studies are undertaken by academic researchers, journalists, regulators, or others, the aim of such studies is usually not to merely gain knowledge, but to enact some meaningful change to those systems. As a result, it is helpful and important to understand existing law and policy in these domains. This is not to say that these issues are *only* worth studying in cases where explicit legal protections exist, but rather that such protections can help identify important types of bias to measure, or enable real change to occur as a result of one's research findings.

Accordingly, we begin this section with a short overview of a key aspect of non-discrimination law in the United States.[*] In the U.S., the law recognizes two forms of discrimination: *disparate treatment* and *disparate impact*. Disparate treatment refers to the act of explicitly and intentionally treating different classes of people differently. This can be a legitimate practice; consider a company that provides interpreters and only hires people who are bilingual for that role. Disparate impact describes practices that result in significantly different outcomes for members of different groups in the absence of explicit intent. For instance, if men are promoted more frequently than women at a certain company due to managers' own unconscious biases or other unintentional aspects of the promotion process, this may constitute disparate impact.

Having understood how U.S. law categorizes discrimination, we might ask, what kinds of discrimination count? In the United States, discrimination is disallowed along the lines of a specific set of protected categories. These protected categories include race, religion, national origin, age,

* Readers will, of course, note that these standards are specific to the U.S. context, and may not apply everywhere—when considering how to best address issues of inequality, it is necessary to consider regional, national, cultural, and other dimensions of formal and informal standards pertaining to discrimination.

sex, gender and gender identity, sexual orientation, pregnancy status, familial status, disability, veteran status, and genetic information. While other types of discrimination may be important and worth considering (notably absent, for instance, is socioeconomic or class status), these categories are protected under U.S. law.

In the next section, I draw on this background on illegal discrimination to inform our thinking through different types of bias in digital ads.

Taxonomizing Bias in Digital Ads

In this section, bias in ads is categorized first along the lines of whether it is explicit or implicit (and subsequently attributable to the advertiser, publisher, or intermediary), and then also along a second useful taxonomy, by the type of bias—whether due to the content itself or the targeting systems.

It is important to note throughout that some amount of bias is inevitable—the business of marketing is sending messages to people, and while this was not always done in a highly targeted way (think of billboards and other mass media, as described in Section 2), from the point that any targeting exists, ads are, by definition, biased toward some group of people. This section moves to a summary of research focusing on discrimination—bias resulting in unequal and unjust treatment of different groups.

Explicit Bias and Implicit Bias

One important axis on which we can distinguish types of bias is whether it occurs explicitly or implicitly. *Explicit bias*, analogous to disparate treatment, involves intentional targeting or exclusion of certain groups. This type of bias can be introduced by all members of the system: advertisers, publishers, and intermediaries.

An advertiser might be explicitly biased when trying to actively target particular groups of consumers or explicitly avoid others—for example, an advertiser of women's clothing might intentionally try only to target women; an advertiser of video games (holding rather anachronistic views) might build an ad campaign with the intention of only targeting men or only targeting young people.

A publisher could be explicitly biased toward certain content as well. One industry term for this kind of bias is *brand safety*, which describes publishers' concerns that their content or image could be harmed due to the advertisements that are loaded onto the page. We could imagine the reputational harm that would occur if ads containing sexually explicit content were loaded onto a website focusing on children's education. Publishers' biases against certain content might be appropriate but might also be problematic; for instance, publishers' homophobic biases might make it difficult for companies catering to an LGBTQ audience to get their ads placed. Intermediaries can also be explicitly biased, for instance by allowing advertisers or publishers to target or exclude audiences along the lines of protected category attributes. In one high-profile example from 2019, the U.S. Department of Housing and Urban Development (HUD) charged Facebook with discrimination that violated the Fair Housing Act for enabling housing discrimination along the lines of nationality, race, and other attributes [Booker, 2019].

Digital ads might also suffer from *implicit bias*—bias arising due to inference, automation, correlation, or other less direct reasons. This type of bias, also sometimes called emergent bias, is akin to disparate impact discrimination, and it is similarly harder to identify and measure, since its impacts are only measured in aggregate, as patterns that indirectly lead to the disadvantage of some group. Again, this could be a result of actions of the advertiser, publisher, or intermediary.

At the level of the advertiser, implicit bias might be caused by the active selection of a certain target audience (for example, people in a particular zip code) being correlated with other attributes (like race or socioeconomic status). But the increased use of automation to do more than target ads—for instance, to create them—can also lead to bias. In one now-legendary case, academic and researcher Latanya Sweeney discovered that Google search queries for Black-sounding names were more likely to lead to an ad from a background check company suggesting the person had an arrest record, even when this was not the case [Sweeney, 2013]. This could have been a result of the ads the company created; perhaps the advertiser automatically selected first names that were statistically more likely to have an arrest record, resulting in a list that reflected the racist trends in existing carceral data. Going further, web users' own biases might lead advertisers to create more biased advertising; an advertiser for beauty products might find that customers seem to prefer some marketing images over others, and choose not to run ads with the other images. If a part of consumers' preferences is driven by unconscious racism—for instance, preferring photos of lighter-skinned faces over darker-skinned models—this automated optimization by the advertiser might contribute to implicit bias in ad imagery as a whole.

Publishers might also contribute to implicit bias in digital ads, perhaps driven by the same desire for "brand safety" that leads to explicit bias. Publishers may deem some ad content lower status or less acceptable resulting in implicit bias along other lines—like a publisher who only wants family-friendly content appearing on their site inadvertently precluding family-friendly LGBTQ content. (What counts as family-friendly, however, is likely the decision of an ad intermediary, not the publisher itself, so it may be more correct to attribute most of these biases to the intermediaries.) Intermediaries are perhaps the most likely culprits of implicit bias, since their systems rely heavily on automation and inference. In the aforementioned cases discovered by Dr. Sweeney where

Google searches for Black-sounding names resulted in ads for alleged arrest records, the advertiser in question claimed that they provided Google's ad service with the same text and lists of last names. If this is the case, it suggests that some automated decision internal to Google's system led to the use of problematic text for some names and not others and reproduced existing social biases [Sweeney, 2013].

Content Bias and Targeting Bias

In addition to thinking about discrimination in terms of whether it happened explicitly (directly) or implicitly (indirectly), it is useful to categorize these biases according to whether they are predominantly in the realm of the ad content or the ad targeting.

Content bias includes cases where the ad content itself is biased, discriminatory, or exclusionary. This kind of bias is common in marketing, where messages are often designed differently depending on, for example, the advertiser's expectation of the gender of the receiver [Goffman, 1979]. This has been studied extensively for decades before the advent of digital advertising and persists online. Being flexible with the definition of content bias, this could also include various kinds of malicious or bad content like misinformation.

Targeting bias, meanwhile, involves cases in which bias is caused by the machinery of ad targeting—the intentional selection of a biased audience by an advertiser, or automated creation of such an audience by the intermediary. This includes, for instance, the apparently effective ad campaign run by supporters of Great Britain's exit from the European Union, who ran ads emphasizing different issues (like immigration or animal rights) and used ad targeting to deliver those ads to groups they believed would be most receptive to a specific message [BBC, 2018]. Another example, brought into public attention by ProPublica, showed that Facebook allowed advertisers to exclude people from seeing ads by race [Angwin and Parris Jr., 2016], a discovery that contributed to the lawsuit against them by HUD mentioned above [Booker, 2019].

Content bias and targeting bias are not mutually exclusive; in fact, as several findings detailed in the next section show, the content of an ad can be one contributing factor causing it to be targeted at a biased group of users. For one example, journalists reported in 2020 that certain types of misinformation and polarizing messages (biased content) were targeted specifically at Latinx voters in the U.S. (targeting bias) [Ghaffary, 2020]. This divide is also not exhaustive—some kinds of bias, like those arriving due to publisher's concerns about brand safety, are not included under either umbrella.

Prior Research Findings: Ad Discrimination

Despite advertising's status as a major industry backed by formidable economic motives, there exists a dearth of empirical studies into its biases from the perspective of user impacts. Below I visit seven key papers in this space—most of which were published in the last few years, as evidence of the nascent state of this work.

One of the first studies in this domain, from 2015, addressed bias in Google's ad targeting. Researchers built a tool called AdFisher to run experiments studying the impact of Ad Settings and user behavior on the ads received by (fake) users [Datta et al., 2014]. This method, in which a tool is built to imitate a real user and interact with a system (in this case, digital advertising and Google's Ad Settings) is called a *sock-puppet audit* [Sandvig et al., 2014]. AdFisher revealed some biases in Google's ad targeting. For instance, accounts whose gender was set to male received more ads for high-paying jobs, suggesting gender-based discrimination; meanwhile, accounts with a visit history to substance abuse sites received different ads than other accounts, without an option on the settings page to toggle this, suggesting users could be tagged with some attributes using inferences being made through behavioral targeting without the potential to opt out. This article led to several more from the same author team, including one in 2018 exploring possible avenues for legal recourse [Datta et al., 2018].

In 2016, ProPublica's series on Machine Bias identified (as referenced in the previous section) that Facebook's ad intermediary platform allowed ads to exclude users by race, including in protected domains like housing [Angwin and Parris Jr., 2016]. While this negative publicity led to some changes on the platform, the problem was far from over. Instead, this finding fueled a wave of research on targeted ads, beginning with another audit of Facebook's ad targeting functionality that found that advertisers could accomplish the same discriminatory goals implicitly instead of explicitly [Speicher et al., 2018]. The researchers argued that discrimination from such systems should be measured in terms of impacts on marginalized people, not whether or not sensitive attributes were explicitly used for targeting.

Another paper from the same year also studied Facebook's advertising tools from a computer systems security perspective [Faizullabhoy and Korolova, 2018]. Using various strategies, the authors of that paper were able to discern sensitive attributes of individual people including income, relationship status, home value, age, interests, and frequency of travel (they ran these experiments on a subset of their own Facebook friends, who consented to the experiment). They concluded that Facebook's ad tools could be exploited to enable privacy violations, microtargeting (delivering content at the level of a single person or a single household), and the relatively simple and inexpensive targeting of marginalized groups of people.

After such publicity led to a lawsuit settlement with Facebook (the aforementioned lawsuit brought by HUD [Booker, 2019]), the company changed part of its ad targeting system, renaming the "Lookalike Audiences" tool to "Special Ad Audiences." This led researchers to compare the previous tool with the new one [Sapiezynski et al., 2019]. Unfortunately, they found, again, that even absent demographic features, Facebook's ad tools could produce audiences for advertisers that were biased on protected categories including age, race, and gender. They also ran experiments running ads intended for a neutral audience and found that ads for tech jobs were more likely to be shown to young men than others, and supermarket jobs shown to middle-aged women—findings that suggested, as prior research had, that Facebook's own algorithms could lead to bias even without that intention from advertisers using the system.

The following year, many of the same researchers involved in that study also designed and ran an experiment focused on the ad delivery process [Ali et al., 2019]. They published different sample ads and analyzed the audience to which they were shown, finding more evidence that Facebook's internal algorithms were a driver of bias. This bias seemed to arise due to the system's goal of optimizing for user engagement with ads, and its predictions about the "relevance" of ads to different groups of people. As a result, aspects of an ad campaign like the advertiser's budget and the ad content created skews along gender and racial lines even when advertisers intend for the audience to be inclusive, and even for ads about employment and housing (areas in which discrimination is illegal in the U.S.). Similarly, also in 2019 researchers ran experiments with ads for employment on Facebook and showed that women were disadvantaged relative to men in their exposure to ads for STEM jobs [Lambrecht and Tucker, 2019].

The previous studies nearly all relied on the audit methodology (which is discussed in more detail in the next section), posing as an advertiser and distributing researcher-designed faux advertisements. In 2020, a study instead analyzed ads that were actually distributed on the platform using machine learning to categorize the content of the ads [Kingsley et al., 2020]. Like prior work, these researchers also found evidence of bias, specifically that men were more likely to be shown ads for financial credit than women.

Limitations of Existing Work

The above summary of major past research on the topic of targeted ads and bias was not written with an intentional focus on any particular method or advertising platform. However, readers will note that the majority of this work focuses on Facebook's advertising platform, from the perspective of an advertiser trying to target specific groups, and doing so using an audit method that entails creating faux advertisements. These studies also focus on the same few axes of identity: frequently gender, less frequently race, age, and a few others. Such studies are important contributions to the literature, but they also indicate the system's current weaknesses—easier entry points for researchers than studying other advertising platforms, or studying ads from the perspective of the user (e.g., collecting all the ads a group of users sees across many different platforms), or bias along axes like disability or socioeconomic status.

The next section discusses how to address some of these limitations—current strategies that are pushing the state of the art and areas still in need of improvement.

Is Biased Advertising All Bad?

While this section focuses on harms, it is important to consider this issue with sufficient nuance, and not paint all ad targeting with too broad a brush. There are certainly some cases where this tech technology can be used for good; civil rights organizations, minority-focused healthcare providers, academics looking to recruit minority populations for research, and others can and do use targeted advertising to get in contact and spread information through marginalized populations. Such targeting can make a huge difference for outreach, since such publicity is very expensive and communicating with very small segments of the population would be infeasible without targeting. If we agree that the mission of these organizations is important and socially beneficial, we could say that *some* groups being able to create biased advertisements is actually in the public good. This leads us to the next section, in which I discuss changes and improvements to this system. Our goal, as individuals and organizations invested in improving the welfare of marginalized and oppressed groups and people, is not necessarily to do away with this whole infrastructure of digital ads, but rather to create tools and frameworks for evaluation and decision-making regarding such technologies across a wide spectrum of possible uses. In the next section, I discuss strategies that bring us closer to that outcome.

4.0
Changing the System

Having covered the current state of digital advertising and its potential for bias and discrimination, in this final section, I turn to potential improvements.

Why Do These Problems Persist?

As the previous sections demonstrate, the issues with digital ads—from people's privacy concerns to journalists' and researchers' findings of discrimination—have been well documented in the past decade. So why do these problems persist? I propose three major reasons: financial interests, system opacity, and underlying social inequality.

One major and daunting issue with digital advertising is, of course, the high-value economy underlying it. From publishers monetizing their web presence, to advertisers mobilizing huge budgets to increase their business, to intermediaries whose entire business relies on facilitating this exchange, hundreds of billions of dollars a year are tied up in digital ads. Shifting such major financial interests is a massive-scale challenge.

A second reason is the opacity of the system. As discussed in Section 2, it can be difficult for advertisers to know whether their ad spend is being used optimally, and even harder for web users to know why or how they are being targeted and served ads online. A large part of this opacity is intentional on the part of intermediaries, whose control of the marketplace is strengthened by a lack of transparency [Hwang, 2020]. Even while touting moves toward greater transparency in the wake of federal lawsuits alleging discrimination, in late 2021 Facebook suddenly blocked access to the platform for New York University researchers (disabling their private Facebook accounts) who were studying political ads distributed on the platform [Bobrowsky, 2021]. Moreover, these systems require constant monitoring; a one-time effort will not suffice, as evidenced by findings in 2020 that discriminatory ads were still a problem on Facebook, years after they settled similar lawsuits [Merrill, 2020]. And, on a more meta level (pun intended), the focus on Facebook's platform reveals the relative difficulty of studying other ad intermediaries or taking more user-centered approaches to understanding digital ads.

Finally, many of these biases mirror society's existing social inequities. From the creation of biased ad content (long effective due to people's own factional identities) to the biases that emerge from machine learning-powered ad systems trained on biased data, addressing this issue involves understanding and, to some degree, reforming real-world biases. A serious overhaul of the digital ad marketplace to address

bias and discrimination would likely need to include changes on each of these fronts.
In the rest of this section, I focus on two different branches of efforts: research to better understand the issue and uncover problems and transformative efforts aiming at changing the status quo.

Research

Research—passive measurement and active experiment—has an important role to play in understanding the advertising ecosystem. This research can be done by academics, private companies, grassroots user-led groups, journalists, and others. Below some of those strategies are summarized.

Typical Algorithm Audits

Algorithm audits are a method commonly used in the study of sociotechnical systems, especially in cases where researchers want to identify possible biases. Auditing, a method developed from audit studies conducted in the social sciences, involves interacting with a given system in order to collect data about the outputs resulting from those interactions, with the end goal of statistically analyzing the system without actually having direct access to its inner workings [Metaxa et al., 2021]. For one of the first articles on algorithm audits, see [Sandvig et al., 2014]; for a recent overview of the method, including its history, best practices, and notable examples of its use, see [Metaxa et al., 2021].

In the domain of advertising, this includes several studies mentioned in Section 3, such as the AdFisher browser extension that ran experiments connecting browser behavior with Google Ad Settings and ads received by creating fake accounts and computationally managing them to appear like real users browsing the web [Datta et al., 2014]. A more recent study covered in the last section also used auditing to study discriminatory advertising using Facebook's advertising tools [Speicher et al., 2018]. This method is powerful for its ability to first prove implicit bias or disparate impact discrimination by using computational and statistical techniques. It can also allow the researcher to estimate the magnitude of the problem, and to hypothesize (but not necessarily causally prove) reasons for that problem. As such, it is a strategy that will continue proving useful in this space. However, as we saw in the previous section, there is a huge dearth of research about many marginalized people's experiences with ads. While most studies focus on gender and race, work remains to be done to audit digital ad systems with respect to low-income people, LGBTQ people, those with disabilities, and so on.

Community-Centered Technical Tools

Expanding on the idea of an audit, several groups in industry and academia have been involved in the creation of tools for studying these issues longitudinally, and in direct collaboration with users. A few of these are New York University (NYU) Ad Observer, Who Targets Me?, Mozilla's dual efforts Regrets Reporter and Rally, and, from my own research group, Intervenr.

Ad Observer and Who Targets Me? are both efforts focusing on political advertising. The former, run by an academic team at NYU, was rather unceremoniously blocked by Facebook in 2021, without much explanation. At the time, Facebook claimed they took this step in order to be in compliance with Federal Trade Commission (FTC) guidelines; however, popular press subsequently reported that the FTC rejected this assertion [DeLong, 2021]. While this led to much outcry from researchers, civil society groups, and even some politicians, as of March 2022, the project was still frozen. Who Targets Me?, similarly, is a browser tool focusing on U.K. politics with the goal of understanding political advertisements by collecting the ads shown to users who have installed the tool in collaboration with academic researchers [Who Targets Me?, 2022; Booth, 2017].

Mozilla has also been involved in this space, joining with efforts headed through both its nonprofit and for-profit arms. These efforts include Mozilla Rally, a platform for allowing researchers to get access to browsing data from some Firefox browser users who have explicitly volunteered their information [Mozilla, 2021b]. Their other effort, Regrets Reporter, is a parallel project that allows users to report complaints with the recommendations served to them on YouTube; this system could be expanded to collect user feedback on a wide variety of content [Mozilla, 2021a].

A final such project, from my own research group, is the Intervenr system [Team, 2022]. Intervenr is a system for academic researchers to conduct audits, providing infrastructure for onboarding participants, collecting their data, running in-browser experiments with them, and compensating them for participating [Metaxa, 2021]. While Intervenr is multi-purpose, allowing researchers to answer questions about content ranging from the news media people find through social media to the searches they conduct, we are currently using it to study digital advertising from the user's perspective. Unlike the studies mentioned in Section 3, Intervenr does not collect information from the perspective of a potential advertiser, but rather, in line with the other community-focused tools just mentioned, it instead allows us to study the experience of a web user. While this work is ongoing and findings are only preliminary, our infrastructure is designed to identify both content and targeting bias. We study content bias by analyzing the ad images participants are exposed to across all users for things like the topic of the ad and the gender and race of people appearing in ad images. Then, by aggregating according to demographic information about our users (including race, gender, socioeconomic status, political leaning, and other information), we can analyze differences between different groups of people to draw inferences about targeting bias. Importantly, this infrastructure goes beyond most of the other efforts mentioned by providing the capability to perform interventions on users—for instance, showing them someone else's targeted ads instead of their own to gauge how much more effective user-targeted ads are than a random match. Our first study using this methodology finds that users' own targeted ads do initially perform better on metrics like user interest in and feeling of representation by the ad. However, after just a week of exposure to another randomly assigned user's ads, preference for those other ads begins to increase while preference for one's own ads begins to decrease suggesting that the power of personalization may be weaker than purported [Lam et al., 2023].

Public Pressure

In addition to academic research, efforts led by journalists have proven exceptionally impactful on the topic of understanding algorithmic content and mobilizing change. Investigative journalist Julia Angwin's team at *ProPublica* is the quintessential example of this work, uncovering bias in algorithm-assisted bail and sentencing systems, online shopping giant Amazon's pricing algorithms, and Facebook's advertising systems [Angwin et al., 2016; Angwin and Mattu, 2016; Angwin and Parris Jr., 2016]. This kind of general audience work combining solid research methodologies with public-facing narratives has been uniquely effective, suggesting that further efforts from investigative journalists, data journalists, and others reporting direct to the public can be a strong vehicle for social change.

Digital Literacy

A final strategy worth mentioning, if only for its repeated appearances in this space, is digital literacy. Articles informing us about how to protect our privacy while using online platforms abound, and in the domain of advertising focus on things like turning off ad personalization or viewing the ad categories intermediaries like Google have tagged us with based on our provided data and past behaviors [Gralla, 2019].

Academic scholars measuring the public's perceptions of online advertising have repeatedly found that people have "shallow awareness of behavioral targeting capabilities and the potential for discrimination" [Plane et al., 2017]. Preeminent scholar of advertising Joseph Turow lists digital literacy as the first of a few possible solutions to issues with digital advertising: "teach our children well—early and often" [Turow, 2012].

However, the limits of this strategy are significant, and obvious: not everyone has the time or interest to become knowledgeable about this complex infrastructure and its possible discriminatory effects on their own lives. And even if we were all fully empowered, individual-level efforts are no match for the massive and powerful industry behind digital ads. That is not to say we should not try—understanding the information ecosystem of the web is more important now than ever. But digital literacy cannot be the headlining effort to address this problem.

Law and Policy

To mobilize the research measuring the extent of these issues and experiments causally demonstrating it, many efforts in law and policy are underway. A note on nuance: as we touched upon at the end of Section 3, not all targeting or bias is discriminatory in a negative sense—some of it can even be helpful in uplifting marginalized communities. As a result, efforts to legislate correct behavior or penalize faults must be nuanced and avoid hindering legitimate and socially progressive uses of this technology.

Regulation

The elephant in the room whenever one speaks of policy regarding tech platforms is regulation; in private industry, the concept is alternately derided (as out-of-touch and unduly burdensome) or feared. There is no denying, however, that legal regulation from government can have significant impacts. Just a few years ago, the now-ubiquitous popups on websites asking for user consent to store cookies were nonexistent; these are, in part, the result of the European Union's May 2019 General Data Protection Regulation (GDPR), a law requiring that cookies qualify as personal user data and, as such, require user consent to store [Stewart, 2019].

Scholars of auditing, including myself, have begun to call for a federal organization or department, or other neutral third party, charged with conducting regular audits [Metaxa et al., 2021]. While some of this mission is being taken up by private ventures (for instance by companies like ORCAA and Mozilla), I see value in this functionality falling under the umbrella of governmental agencies, not-for-profit companies, or other groups that will act in the public's best interest as part of their mission. To enact such change would require savvy legislation and regulation to create real consequences and incentivize responsible, transparent auditing [Metaxa, 2021].

Specifically pertaining to advertising regulation, a study from 2011 surveyed 3.3 million people who had been randomly exposed to over 9,500 banner ads, and found that such display ads became less effective in European Union countries where privacy regulation preventing advertisers from collecting some user data for ad targeting had recently been enacted [Goldfarb and Tucker, 2011].

Advertising scholar Turow also recommends regulation in his books [Turow, 2012]. His ideas include making some sensitive user data fully off-limits to advertisers rather than opt-in and for other such data to become opt-in rather than fully acceptable for advertisers and tech companies to collect and use.

Regulation need not focus solely on whether or not targeting can occur, or what data ad targeters can use. Other strategies might include the requirement that algorithmic models meet some functionality standard, or that the government curate and safeguard a representative data set so they can be trained on nationally representative data and eliminate some biases due to skewed training sets [Blass, 2019].

Notably, regulation can go both ways. Even while HUD was suing Facebook for violating the Fair Housing Act [Booker, 2019], the Trump Administration in 2019 was trying to create a loophole for discrimination in cases where algorithms or models were used (as in algorithm-assisted risk assessment and sentencing decisions) without protected category attributes explicitly included [Meckler and Barrett, 2019]. Given increasing popular attention to this issue, regulation is becoming increasingly less of a question and more of a certainty—though the specifics of what will be regulated and the goals of that regulation are still very much up in the air.

Conclusion

The solutions, or steps toward solutions, for discriminatory online advertising detailed above range widely, from academic research projects to potential startup companies to regulatory measures to grassroots citizen-led efforts. Up against an industry as massive and powerful as advertising, no one effort or even single style of effort is likely to succeed; none of these options are complete solutions on their own. To put it succinctly, my hope is that, through a combination of these different strategies, a more equitable and socially just marketplace might emerge, in tandem with the broader conversation about algorithmic content and sociotechnical systems—one that prioritizes, above technical functionality and above financial profit, the health of our online information ecosystem and the well-being of the people moving through it.

5.0
References

Ackerman, M. S. (2000). The intellectual challenge of CSCW: the gap between social requirements and technical feasibility. *Human–Computer Interaction,* 15(2-3):179–203.

Ali, M., Sapiezynski, P., Bogen, M., Korolova, A., Mislove, A., and Rieke, A. (2019). Discrimination through optimization: How Facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on Human-Computer Interaction,* 3(CSCW):1–30.

Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. And it's biased against blacks. *ProPublica*, May 23.

Angwin, J. and Mattu, S. (2016). Amazon says it puts customers first. But its pricing algorithm doesn't. *ProPublica* September 20.

Angwin, J. and Parris Jr., T. (2016). Facebook lets advertisers exclude users by race. *ProPublica* October 28.

BBC (2018). Vote leave's targeted Brexit ads released by Facebook, July 26, 2018. https://www.bbc.com/news/uk-politics-44966969

Bentley, T. A., Teo, S. T., McLeod, L., Tan, F., Bosua, R., and Gloet, M. (2016). The role of organisational support in teleworker wellbeing: A socio-technical systems approach. *Applied Ergonomics,* 52:207–215.

Blass, J. (2019). Algorithmic advertising discrimination. *Northwestern University Law Review*, 114:415.

Bobrowsky, M. (2021). Facebook disables access for NYU research into political-ad targeting. *The Wall Street Journal,* August 4, 2021. https://www.wsj.com/articles/facebook-cuts-off-access-for-nyu-research-into-political-ad-targeting-11628052204

Booker, B. (2019). Housing department slaps Facebook with discrimination charge.

Booth, R. (2017). Free software to reveal how Facebook election posts are targeted.

Buolamwini, J. and Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency,* pages 77–91. PMLR.

Chen, L., Ma, R., Hann´ak, A., and Wilson, C. (2018). Investigating the Impact of Gender on Rank in Resume Search Engines. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–14, Montreal QC, Canada. ACM Press.

Datta, A., Datta, A., Makagon, J., Mulligan, D. K., and Tschantz, M. C. (2018). Discrimination in online advertising: A multidisciplinary inquiry. In Friedler, S. A. and Wilson, C., editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency,* volume 81 of *Proceedings of Machine Learning Research*, pages 20–34. PMLR.

Datta, A., Tschantz, M. C., and Datta, A. (2014). Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. *arXiv preprint arXiv:1408.6491*.

DeLong, L. A. (2021). Facebook disables ad observatory; academicians and journalists fire back.

Faizullabhoy, I. and Korolova, A. (2018). Facebook's advertising platform: New attack vectors and the need for interventions. *arXiv preprint arXiv:1803.10099*.

Ghaffary, S. (2020). How fake news aimed at Latinos thrives on social media, *Vox*, November 19, 2020. https://www.vox.com/recode/21574293/social-media-latino-voters-2020-election

Goffman, E. (1979). *Gender advertisements.* Macmillan International Higher Education.

Goldfarb, A. (2014). What is different about online advertising? *Review of Industrial Organization,* 44(2):115–129.

Goldfarb, A. and Tucker, C. E. (2011). Privacy regulation and online advertising. *Management Science,* 57(1):57–71.

Gralla, P. (2019). How to protect your privacy on YouTube, *The Verge,* April 19, 2019. https://www.theverge.com/2019/4/19/18484802/youtube-privacy-protect-how-to-video-preferences-web-ad-personalization

Hancock, J. T., Naaman, M., and .Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication,* 25(1):89–100.

Hwang, T. (2020). *Subprime attention crisis: advertising and the time bomb at the heart of the Internet.* FSG originals.

Hyer, N. L., Brown, K. A., and Zimmerman, S. (1999). A socio-technical systems approach to cell design: Case study and analysis. *Journal of Operations Management,* 17(2):179–203.

Kingsley, S., Wang, C., Mikhalenko, A., Sinha, P., and Kulkarni, C. (2020). Auditing digital platforms for discrimination in economic opportunity advertising. *arXiv preprint arXiv:2008.09656.*

Lam, M.S., Pandit, A., Kalicki, C.H., Gupta, R., Sahoo, P., Metaxa, D. (2023). "Sociotechnical audits: Broadening the auditing lens to investigate targeted advertising." *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW) (conditionally accepted).

Lambrecht, A. and Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science,* 65(7):2966–2981.

Lewis, R. A. and Rao, J. M. (2013). On the near impossibility of measuring the returns to advertising. Unpublished paper, Google, Inc. and Microsoft Research. *http://justinmrao. com/ lewis rao nearimpossibility. pdf*.

Lewis, R. A., Rao, J. M., and Reiley, D. H. (2011). Here, there, and everywhere: Correlated online behaviors can lead to overestimates of the effects of advertising. In *Proceedings of the 20th International Conference on World Wide Web*, pages 157–166.

Marotta, V., Zhang, K., and Acquisti, A. (2017). The welfare impact of targeted advertising. Available at SSRN.

McDonald, A. M. and Cranor, L. F. (2010). Americans' attitudes about internet behavioral advertising practices. In *Proceedings of the 9th annual ACM workshop on Privacy in the Electronic Society*, pages 63–72.

Meckler, L. and Barrett, D. (2019). Trump administration considers roll-back of anti-discrimination rules, *The Washington Post,* January 3, 2019.

Merrill, J. (2020). Does Facebook still sell discriminatory ads? The Markup, August 25, 2020. https://themarkup.org/the-breakdown/2020/08/25/does-facebook-still-sell-discriminatory-ads

Metaxa, D. (2021). *Auditing Bias and Representation in Sociotechnical Systems*. PhD thesis, Stanford University.

Metaxa, D., Park, J. S., Robertson, R. E., Karahalios, K., Wilson, C., Hancock, J., & Sandvig, C. (2021). Auditing algorithms: Understanding algorithmic systems from the outside in. Foundations and Trends® in Human–Computer Interaction, 14(4), 272-344.

Mozilla (2021a). Take back control from the algorithm. https://foundation.mozilla.org/en/youtube/regretsreporter/.

Mozilla (2021b). Take control over your data with rally, a novel privacy-first data sharing platform. https://blog.mozilla.org/en/mozilla/take-control-over-your-data-with-rally-a-novel-privacy-first-data-sharing-platform/.

O'Barr, W. M. (2010). A brief history of advertising in America. *Advertising & Society Review*, 11(1).

Page, L., Brin, S., Motwani, R., and Winograd, T. (1999). The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab.

Pew Research (2014). World wide web timeline, March 11, 2014. https://www.pewresearch.org/internet/2014/03/11/world-wide-web-timeline/

Pew Research (2019a). Facebook algorithms and personal data, January 16, 2019.

Pew Research (2019b). Public knowledge and experiences with data-driven ads, November 15, 2019. https://www.pewresearch.org/internet/2019/11/15/public-knowledge-and-experiences-with-data-driven-ads/

Plane, A. C., Redmiles, E. M., Mazurek, M. L., and Tschantz, M. C. (2017). Exploring user perceptions of discrimination in online targeted advertising. In *Proceedings of the 26th USENIX Conference on Security Symposium*, SEC'17, page 935–951, USA. USENIX Association.

Robertson, R. E., Jiang, S., Joseph, K., Friedland, L., Lazer, D., and Wilson, C. (2018). Auditing partisan audience bias within Google Search. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–22.

Rosen, A. (N.D.). A publishers guide to blocking bad advertising. GeoEdge.

Sandvig, C., Hamilton, K., Karahalios, K., and Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and Discrimination: Converting Critical Concerns into Productive Inquiry*, 22.

Sapiezynski, P., Ghosh, A., Kaplan, L., Mislove, A., and Rieke, A. (2019). Algorithms that "don't see color": Comparing biases in lookalike and special ad audiences. arXiv preprint arXiv:1912.07579.

Speicher, T., Ali, M., Venkatadri, G., Ribeiro, F., Arvanitakis, G., Benevenuto, F., Gummadi, K., Loiseau, P., and Mislove, A. (2018). Potential for discrimination in online targeted advertising. In *Conference on Fairness, Accountability, and Transparency (FAT '18)*.

Statista (2022). Estimated aggregate revenue of the advertising, public relations, and related service industry in the United States from 2004 to 2020. https://www.statista.com/statistics/183932/estimated-revenue-in-advertising-and-related-services-since-2000/

Stewart, E. (2019). Why every website wants you to accept its cookies, Vox, December 10, 2019. https://www.vox.com/recode/2019/12/10/18656519/what-are-cookies-website-tracking-gdpr-privacy

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Sweeney, L. (2013). Discrimination in online ad delivery. Queue, 11(3):10:10–10:29.

Team, I. (2022). Welcome to intervenr ads. https://intervenr.stanford.edu.

Trist, E. L. (1981). *The evolution of socio-technical systems*, volume 2. Ontario Quality of Working Life Centre Toronto.

Turow, J. (2012). *The daily you.* Yale University Press.

Turow, J., King, J., Hoofnagle, C. J., Bleakley, A., and Hennessy, M. (2009). Americans reject tailored advertising and three activities that enable it. Available at SSRN 1478214.

Who Targets Me? (2022). Our research. https://whotargets.me/en/our- research/.

Zawadziński, M. (2014). What is an ad server and how does it work? Clearcode, last modified May 4, 2022. https://clearcode.cc/blog/what-is-an-ad-server/.