Discrimination in Bing Ad Delivery

Jonathan Lo jolo@ucsd.edu
Flynn O'Sullivan fosulliv@ucsd.edu
Elsie Wang e2wang@ucsd.edu
Stuart Geiger sgeiger@ucsd.edu

Abstract

Several studies have reported instances of gender and race discrimination in online ad delivery, whether that be for housing, employment, or public records. One highly-cited study conducted by Sweeney (2013) audited Google AdSense ads and found that searches using black-sounding names were more likely to generate ads suggesting an arrest record compared to searches using white-sounding names. However, Google AdSense and other advertising platforms have since evolved and many of the issues found in the study may not be relevant to today. In this paper, we expand on the scope of this problem by providing a more thorough analysis, auditing ads generated by Microsoft Advertising, specifically on Bing.com, and seeing whether this problem still persists a decade later. To do this, we employ a similar method by using the most common white and black first and last names and searching full names on the search engine to determine whether ads suggesting a criminal record appear.

Code: https://github.com/jonathanlo411/dsc180

1 Introduction

Searching someone's name on the Internet can return a plethora of results, especially if that person is famous, but one thing that might come up is an ad for a public records searching website. There are many such services that will search public databases for a name, showing all types of information, sometimes including a criminal record. One such case where someone might be searching names on the Internet is if that person is in competition for a job, an award, or a rental application. If the top results when this name is searched are about a potential criminal record, this may jeopardize that person's trustworthiness, or make the person searching think they have a criminal background, even if they have none. Disqualification for a job based on information that indicates an arrest record is prohibited by U.S. law (US Equal Employment Opportunity Commission 2012). However, it can be hard to prove that an employer used these ads and even if it can be proven, the company may argue the ads are commercial free speech, thus protected by the law. If these ads indicating a criminal record appear more for one racial group than another, this would be a clear display of structural racism.

In this paper, we set out to determine if personalized ads suggestive of an arrest record appear more for one group of people than another, using searches of racially identifying names. We use the hypothesis that no difference exists. With 16,896 searches and 17,471 ads queried, we find no statistically significant difference in criminal ads by race (8.41% black, 8.22% white), but there is a significant difference in gender with male names having a higher percentage of criminal ads than female names (9.17% male, 7.49% female). Our results build on earlier research, but also demonstrate the need for consistent and larger-scale auditing of online ad delivery of major platforms for discrimination and bias.

1.1 Literature Review

This study combines two topics that have long been an area of interest in auditing: potential discrimination regarding black and white sounding names and online ad delivery. Since even before the Internet became as ubiquitous as it is today, researchers have been conducting studies to determine if white and black sounding names have an effect on things like employment or housing. One such example is Bertrand and Mullainathan (2004) in which the authors responded to help-wanted ads in Boston and Chicago with both black and white sounding names. Overall, they found that the resumes with white-sounding names received 50 percent more callbacks for interviews than the black-sounding names. This level of discrimination was relatively uniform across all occupations and industries. What this paper tells us is that, perhaps rather obviously, racial discrimination definitely exists in the job market.

More recently, Ali et al. (2019) delved into the topic of online ad delivery, auditing ads that appeared on Facebook. They tested whether ads the authors created appeared on Facebook differently depending on the user's gender and race. The ads created were largely for employment and housing opportunities. An important distinction is that this study creates "fake" ads and real people view them, whereas our method will be something of an inverse of this, using real ads that any user can view. For this paper, the content of the ads was consistent and the users' demographics were what they were recording, but it will be the opposite for us. Nonetheless, Ali et. al (2019) found significant skewing in their ad delivery, meaning there was a clear preference in certain ads being delivered to a specific demographic of people, even when the ad targeting parameters were set to be inclusive.

A highly-cited study conducted by Sweeney (2013) also addressed whether there is bias in online ad delivery by doing the inverse of Ali et.al, using real ads any user can view rather than creating "fake" ads that real people can view. This study audited Google AdSense by searching 2,814 white-identifying and black-identifying full names on Google.com and Reuters.com, both hosts of Google AdSense, and determining whether ads suggesting a criminal record appear. The study found that there was a significantly greater percentage of ads containing the word "arrest"in the ad when searching black-identifying names than white-identifying names on both websites. In particular, searching a black-identifying name on Reuters.com had a 25% more likely chance of yielding an ad suggestive of an arrest record.

Together, these examples suggest that there may be a pattern of racial discrimination from how names are presented—that one name more commonly associated with a race may affect one's reputation, employment, housing, or other opportunities. Studies like Ali et. al (2019) and Sweeney (2013) especially showed how ads may be presented differently based on race or gender. This paper expands on such studies by similarly replicating Sweeney's methods with Bing, a Microsoft web search engine, and looks into how this discrimination manifests in online public record ads that appear when you search a name. With a larger dataset, different method, and extensive statistical analysis, this paper explores whether discrimination in online delivery still persists today.

1.2 Microsoft Advertising

Before looking at how this is done, we must first understand how ads are generated. In traditional media such as newspapers and magazines, every reader of the publication encounters the same ad in a fixed space. However, the online advertising space is not limited by physical constraints and instead, can appear anywhere on a website and dynamically change its content based on the reader's preferences, search history, geographical location, and other factors. This means two different readers can get entirely different ads on the same website.

Microsoft Advertising, formerly known as Bing Ads, is one of the largest providers of dynamic online advertising, garnering $12.23 billion in digital ad revenues in 2022 (Lebow 2023). Microsoft Search Network holds 38.1% of the share of the US desktop search market with 109 million unique desktop searchers and 6.4 billion monthly desktop searches, covering 46 million users not reached on Google (Microsoft Advertising 2023b). Their websites include Bing, Yahoo, and MSN search results. For this paper, we only look at ads delivered on Bing.

Similar to Google AdSense, Microsoft Advertising is a pay-per-click (PPC) advertising system that uses bidding based on how much a sponsor is willing to pay per click on their ad. According to their website on Microsoft Advertising (2023a), the position of which an ad appears is determined by several things: how you bid compares to other advertisers' bids in a real-time auction, how relevant your ad and the website fit the terms searched, and past ad performance, including click-through rate. The advertisers themselves can choose how to target their ads, such as to a geographic region, devices, times or days of the weeks, demographics, and keywords to improve their click-through rate.

1.3 Data Description

In this study, we employ data of various names searched on the Bing platform, categorized based on race and gender: black female, black male, white female, and white male. "Black-identifying" and "white-identifying" first names are considered those for which a significant number of people have that name more than any other race.

For our dataset, we use the same first names as the Sweeney (2013), which partly borrows from Bertrand and Mullainathan (2004). This study used names given to black and white babies in Massachusetts between 1974 and 1979, defining black-identifying and white-identifying names as those with the highest ratio of frequency in one racial group to the frequency in the other racial group. Additionally, Sweeney used names from the book *Freakonomics* by Levitt and Dubner (2008) which contains a list of the top 8 white and black-identifying female and male baby names in California, from 1961 to 2000. To maintain an equal dataset for each group, we exclude the first names "Latanya" and "Latisha" which were added based on observation in Sweeney's study. For last names, we use the top 11 most common black and white last names in the United States from Name Census (2010), which sources from the 2010 Census and categorizes race based on self-identification in the Decennial Census survey. Figure 1 shows a list of these names categorized by race and gender, and Figure 2 shows a list of the middle initials and the last names categorized by race.

	White Female	Black Female	White Male	Black Male
a)	Allison Anne Carrie Emily Jill Laurie Kristen Meredith	Aisha Ebony Keisha Kenya Latonya Lakisha Latoya Tamika	Brad Brendan Geoffrey Greg Brett Jay Mathew Neil	Darnell Hakim Jermaine Kareem Jamal Leroy Rasheed Tremayne
b)	Molly Amy Claire Emily Katie Madeline Katelyn Emma	Imani Ebony Shanice Aaliyah Precious Nia Deja Diamond	Jake Connor Tanner Wyatt Cody Dustin Luke Jack	DeShawn DeAndre Marquis Darnell Terrell Malik Trevon Tyrone

Figure 1: Black-identifying and white-identifying first names from (a) Bertrand and Mullainathan (2004) and (b) *Freakonomics* by Levitt and Dubner (2008). Emily, Ebony, and Darnell appear in both (a) and (b), giving a total of 61 distinct first names.

White Last Names	Black Last Names	Middle Initials
Smith	Williams	D
Johnson	Johnson	J
Miller	Smith	R
Brown	Jones	M
Jones	Brown	F
Williams	Jackson	E
Davis	Davis	A
Anderson	Thomas	G
Wilson	Harris	
Martin	Robinson	
Taylor	Taylor	

Figure 2: Black-identifying and white-identifying last names from the Name Census (2010), along with middle initials used. Smith, Johnson, Brown, Jones, Williams, Davis, and Taylor appear in both groups, giving a total of 15 distinct last names and 8 distinct middle initials.

In order to test whether certain ads appear for names, we create full names from a union of our first names and last names list, assigning the most common black surnames to black female and male first names and the most common white surnames to white female and male first names. Additionally, we use 8 initials to assign as a middle initial for each of the full names. Removing duplicates gave a total of 61 distinct first names, 15 distinct last names, and 8 distinct middle initials. With this, the total number of full names in our dataset is 704 full names and 5,623 names including middle initials.

2 Methods

Due to limited resources and time, it is difficult to manually search names and collect all the ads of 5,623 names, compared to Sweeney's dataset of 2,184 names in which the majority was searched manually over the course of a month. Thus, to handle a larger scale of searches, we use Selenium Webdriver, a web-based automation framework that allows us to automatically and efficiently search names on a browser without having to manually search it ourselves (Selenium 2023). Specifically, we utilize Google Chrome and its corresponding driver, ChromeDriver, to execute a written script of commands on Chrome for each of the full names, as follows:

First, an Incognito window is opened to eliminate any cookies or cache stored on the browser. Following this, for each full name, a search formatted as "{*first name*} {*initial*} {*last name*} *public records*" is entered into the search bar. For instance, in the case of Allison D. Smith, the search would be "*Allison D Smith public records*." Each full name that includes a middle initial undergoes 3 searches to record all ads displayed on the website, a total of 24 searches for each first name-last name pair. This amounts to 16,896 searches needed to query the ads. Figure 3 and 4 shows an example of a search of Allison D. Smith, a white female name, and Darnell E. DeShawn, a black male name, and some of the ads delivered.

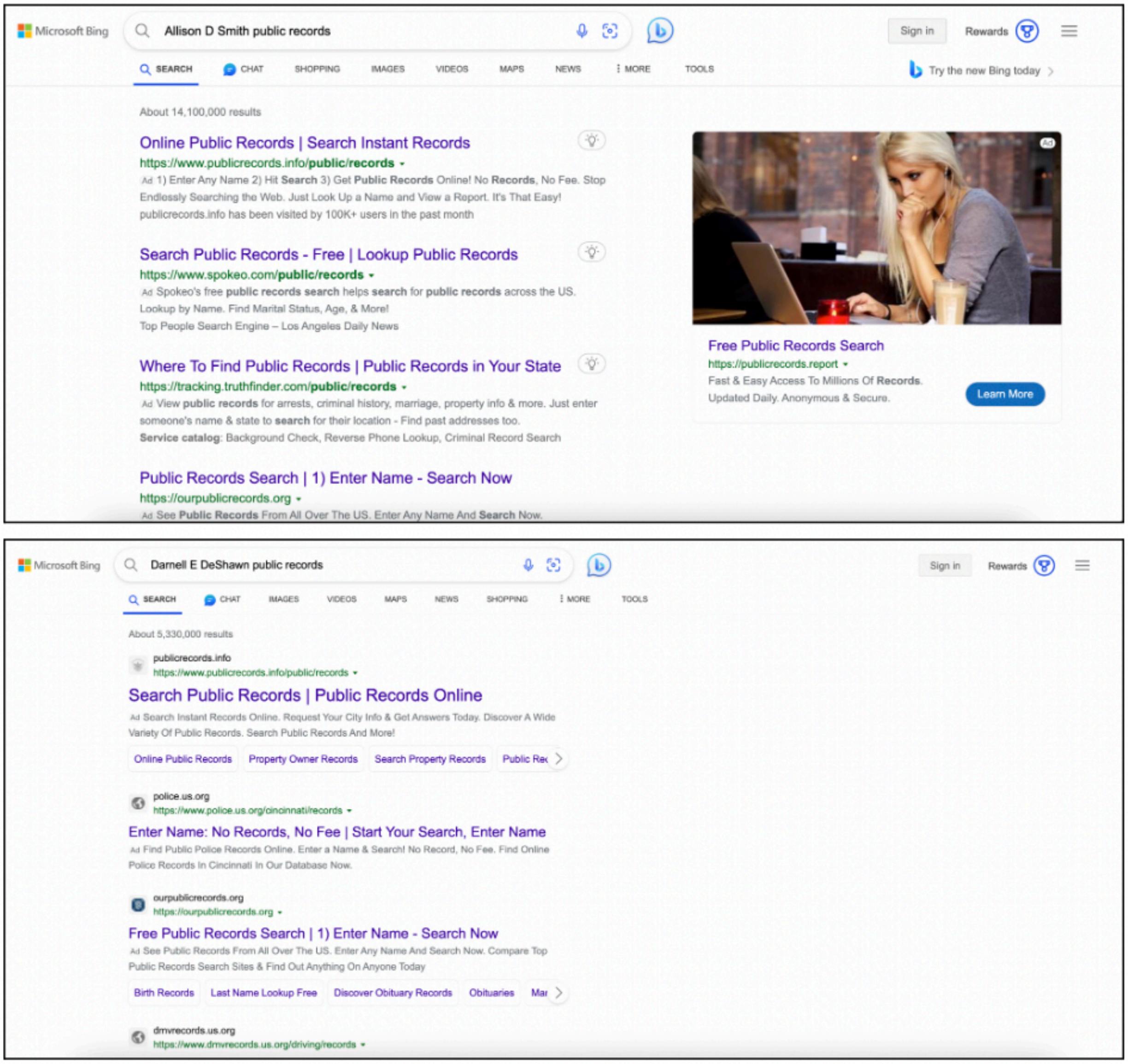

Figure 3: Sample of the webpages shown when searching "*Darnell E DeShawn public records*" and "*Allison D Smith public records*" on Bing.com.

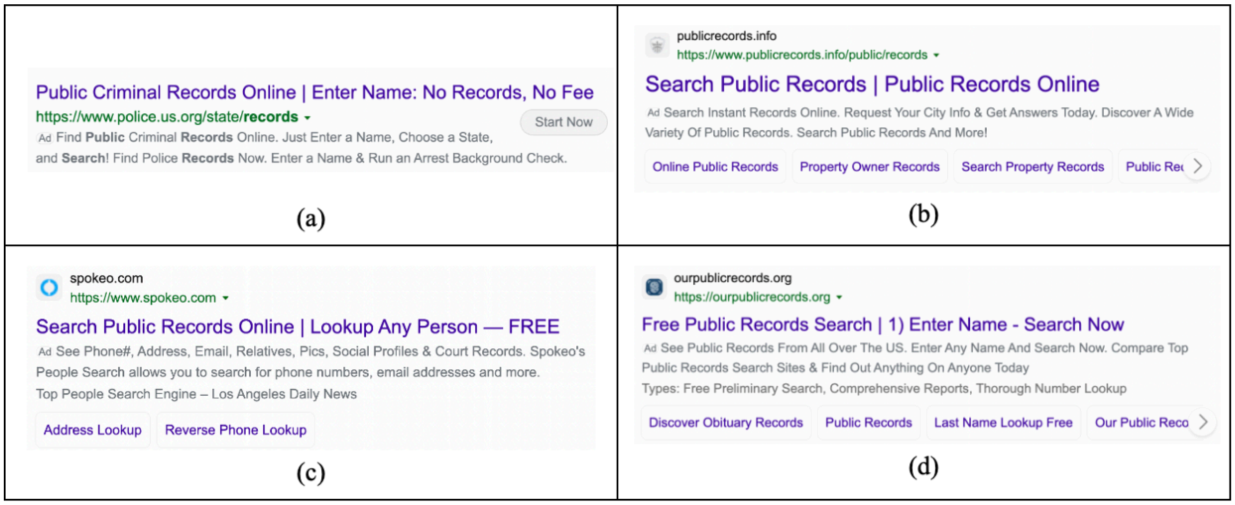

Figure 4: Sample ads from "*Darnell E DeShawn public records*" (a, b) and "*Allison D Smith public records*" (c, d). One ad that appeared when searching "*Darnell E. DeShawn public records*" was a criminal ad with the title containing the word "*criminal*" (a). The rest are examples of neutral ads (b, c, d).

In cataloging our ads, we document what website each ad is from, the title of the ad, the first and last name of the search along with the associated race and gender, and whether the ad is suggestive of a criminal record. To determine this, we use a larger word bank to capture any ads suggestive of a criminal record rather than an arrest record, in which Sweeney only used the word "arrest" for classification. For this study, we consider any ad with a title that contains the words "criminal", "jail", or "arrest" to be a criminal ad.

3 Results

On December 7, 2023, we set out to discover what type of ads surface when first and last names indicative of a particular race are searched. Execution took place during the evening with the same IP address and machine operating in San Diego, California. We compiled a list of ads delivered on Bing.com in response to 16,896 searches in total.

After querying Bing with 704 unique full firt name-last name pairs each queried 24 times (8 initials, each searched 3 times), we collected data from 72,808 ads. Given 16,896 total searches, this means we were served about 4.31 ads on average per search. Of all the ads from which data was collected, 6,054 (8.32%) satisfied our requirement for being a criminal ad, which is to say the main title of the ad contained the words "criminal," "jail," or "arrest." Thus, 8.32% was our criminal ad rate across the entire dataset. When divided by race, we found that 8.41% of ads queried by searching a black-sounding name were criminal ads, as opposed to 8.22% of ads from searching a white-sounding name. With a p-value of 0.351, we did not find sufficient evidence to say that these proportions were taken from different distributions. In other words, there is no statistically significant difference between the rate of criminal ads shown for black-sounding names versus white-sounding names.

When grouped by gender, however, the criminal ad rate for male names was 9.17% versus just 7.49% for female names. This result was statistically significant, with a p-value of 2.2 x 10-16, meaning the odds of these proportions occurring in the data if male and female criminal ads were drawn from the same distribution is a percentage with 14 zeros after the decimal point.

The ad domain that was seen the most in the data was spokeo.com, with a total of 17,471 ads. The most seen domains and their count of ads can be found in Figure 5. The numbers seen are what was observed for all ads, not just criminal ads. Of all 72,808 ads scraped, 39,556 (54.33%) came from the top three most prevalent domains.

Ad Domain	Count
spokeo.com	17,471
publicrecords.info	11,545
tracking.truthfinder.com	10,540
police.us.org	10,184
checksecrets.com	6,080
ancestry.com	5,947
instantcheckmate.com	2,827
peoplelooker.com	1,145
courtrecords.us.org	935
criminalrecords.us.org	727
dmvrecords.us.org	692
peoplefinders.coom	599
publicrecords.report	565
ourpublicrecords.org	503
go.newspapers.com	412
adopted.com	343
ancestorrecords.org	338
persopo.com	289
search.peoplefinders.com	253
top10.com	249

Figure 5: 20 most prevalent ad domains (out of 55 total) and how many times they appeared in the data.

Analysis was done on ad domains that occurred in the data more than 10,000 times. First, we determined if there was any bias in where ads from these domains show up. For example, the top domain, spokeo.com, displayed a statistically significant bias (p = 4.33 x 10-9) towards showing up for black-souding names. To be clear, this is not in relation to criminal ads, it just means that spokeo.com was the domain for 24.9% of ads served from searching black-sounding names, compared to just 23.1% of ads served from searching white-sounding names. The full analysis on each of the domains with over 10,000 ads can be found in Figure 6. Publicrecords.info, the second most occurring domain, had a significant bias towards white, female names. Tracking.truthfinder.com had significant bias towards showing ads when white-sounding names are searched, specifically white female names. Lastly, police.us.org showed bias towards black-sounding male names.

Figure 6: Bias analysis on domains with over 10,000 ads. Red line indicates overall rate of ads with specified ad domain. Each chart represents a two proportion z-test to measure differences in ad rates between groups.

Interestingly, the top ad domain represented in the dataset, spokeo.com, had zero criminal ads show up, no matter what group was being queried. The domain with the most criminal ads was police.us.org, with 5,112 criminal ads being displayed from this site. This means just over half the ads served from police.us.org were criminal ads and the criminal ads from this domain constitute 84.4% of all the criminal ads in the dataset. Overall, there were 55 unique ad domains in the dataset, and just 10 of these had any criminal ads. One last interesting piece of the data is that one of the domains that did technically serve "criminal ads," at least according to our definition, was apple.com, since several ads for an Apple TV+ trailer for a show called Criminal Record were served.

4 Discussion

Our hypothesis stated that no difference existed in the delivery of ads suggestive of a criminal record based on searches of racially associated names. Our findings support this hypothesis. Although we found that a greater percentage of ads having "criminal", "jail" or "arrest" in the title appeared for black-identifying names than for white-identifying names on Bing.com, results of the two proportion z-test were not statistically significant. However, with gender associated names, criminal ads appeared 9.71% for male names versus female names, 7.49%. Results of this two proportion z-test were statistically significant, with a p-value of 2.2 x 10-16. While Sweeney (2013) reported no more than three ads per search, our result of an average of 4.31 ads per search shows a significant increase in the amount of ads displayed. This could also be a result of using Bing as our search engine, instead of Google and Reuters, like Sweeney did. In her study, Sweeney reported that the top four domains accounted for more than half of all ads served. We had similar observations, with our top four domains accounting for 68.3% of ads served and our top three accounting for 54.3%. The ad domain that occurred the most in Sweeney's data, Instant Checkmate, was the seventh most occurring domain in our data.

One interesting feature of our data that could be counted as a limitation is the prevalence of ads from the spokeo.com domain. The reason for this being a limitation is because this domain showed zero ads that contained the words "criminal", "jail", or "arrest." In fact, just over 70% of ads from this domain had the exact same title ("Search Public Records - Free | Lookup Public Records") with the other 30% being just slight variations. While this feature of the data is not exactly a limitation as there was nothing we could have or should have done to prevent it, it does add a static element to our analysis, where almost 17% of the ads we scraped had the same title.

Another constraint is limited time and resources, which prevented us from conducting a more thorough data collection process. Although we had a significantly larger dataset than the Sweeney study, we were only able to verify the association of first and last names to a specific race, but not full names; for example, by searching online images associated with full names as done in Sweeney's paper. Additionally, we were only able to search names at one given time on one machine and IP address. As discussed above, advertisers are able target audiences based on geographical location, device, and time, and changing one or more of these factors may yield different results. Similarly, searching ads on different websites may also yield different results. Microsoft Advertising delivers ads to multiple websites, and outcomes may differ across various platforms, but we conducted searches on Bing only.

This paper is a valuable contribution to a broader ongoing discussion of discrimination in online ad delivery. By introducing a larger dataset and using a web-based automation tool like Selenium for faster querying, we are able to bring attention to another major search engine and advertising system and evaluate the state of online advertising today. However, it is important to recognize that our work only represents a small moment in time and that it is a continuous process of holding companies accountable for algorithms that may produce bias and greatly disadvantage one race or gender over the other, whether that be in employment, housing, or other areas.

More research is therefore required to further expand and adapt to the dynamic nature of online advertising—new algorithms could eliminate some biases but also introduce new biases. To build onto this project, one could use a more up-to-date dataset that better reflects the current population or focus on one socioeconomic issue such as in employment and build a dataset of names of working professionals. Conducting searches on other websites like Yahoo and MSN could also better indicate whether discrimination in Microsoft Advertising exists. Alternatively, with more resources, a more thorough data collection process could be done, using different IP addresses, times, and devices to get rid of any confounding factors and contributing to a deeper understanding of discrimination in online advertising.

References

Ali, Muhammad, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 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

Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American economic review* 94 (4): 991–1013

Lebow, Sara. 2023. "Microsoft's advertising moves in 5 charts." Feb. [Link]

Levitt, Steven D, and Stephen J Dubner. 2008. *Freakonomics*. INCA

Microsoft Advertising. 2023a. "Microsoft Advertising.Great Relationships Start Here." [Link]

Microsoft Advertising. 2023b. "Microsoft Search Network Data." Jul. [Link]

Name Census. 2010. "Name Census: United States Demographic Data." [Link]

Selenium. 2023. . [Link]

Sweeney, Latanya. 2013. "Discrimination in online ad delivery." *Communications of the ACM* 56 (5): 44–54

US Equal Employment Opportunity Commission. 2012. "Enforcement Guidance on the Consideration of Arrest and Conviction Records in Employment Decisions under Title VII of the Civil Rights Act." [Link]