Algorithmic Gender Bias and Audiovisual Data:
A Research Agenda

MIREN GUTIERREZ
University of Deusto, Spain

Algorithms are increasingly used to offer jobs, loans, medical care, and other services, as well as to influence behavior. Decisions that create the algorithms, the data sets that feed them, and the outputs that result from algorithmic decision making, can be biased, potentially harming women and perpetuating discrimination. Audiovisual data are especially challenging because of online growth and their impact on women's lives. While scholarship has acknowledged data divides and cases of algorithmic bias mostly in online texts, it has yet to explore the relevance of audiovisual content for gender algorithmic bias. Based on previous guidelines and literature on algorithmic bias, this article (a) connects different types of bias with factors and harmful outcomes for women; (b) examines challenges around the lack of clarity about which data are necessary for fair algorithmic decision making, the lack of understanding of how machine learning algorithms work, and the lack of incentives for corporations to correct bias; and (c) offers a research agenda to address algorithmic gender discrimination prevalent in audiovisual data.

Keywords: algorithmic bias, big data, women, equality, audiovisual data, media

An artificial intelligence (AI) algorithm learned to associate women with kitchen images, based on tens of thousands of Internet photographs, because more women than men appear photographed in kitchens on the Web (Zhao, Wang, Yatskar, Ordonez, & Chang, 2017). While learning, the algorithm multiplied the bias present in the data set on which it was based, amplifying—not simply replicating—the biased association between kitchens and women (Zhao et al., 2017). This is one of several case studies that have shown how machine learning (ML) systems incorporate and increase gender biases. The association of women with kitchens and kitchens with domestic work—typically vulnerable (International Labour Organization, 2020)—is problematic because of the stereotypical influences of such pictures. Disseminated by digital platforms, these images are seen by anyone with access to online content, perpetuating, and intensifying unbalanced gender roles (Collins, 2011).

ML is an automatic computing system containing packages of algorithms that, fed by big data—including audiovisual data— make AI possible. Thanks to statistical techniques, these algorithms learn—that is, they use data to perfect tasks. "Learning capacities grant algorithms some degree of autonomy," making ML tasks "difficult to predict" and inhibiting "the identification and redress of ethical challenges in the design

Miren Gutierrez: m.gutierrez@deusto.es

and operation of algorithms" (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016, p. 3). The real challenges here are the adverse effects we are witnessing based on biased algorithms concerning the presentation of women online (e.g., Buolamwini & Gebru, 2018).

This article reveals the relevance of audiovisual data—including images and sound from cinema, music videos, social sharing platforms, and advertising—in algorithmic gender bias. It offers a taxonomy of biases, connecting feminist and media theory with critical data studies. Algorithms are prevalent; attention to gender bias in texts is well-established (e.g., Hitti, Jang, Moreno, & Pelletier, 2019). However, this analysis is focused on audiovisual content, including images and speech, as there are relatively fewer studies on them despite their increasing relevance.

There has been a global upsurge of audiovisual data (Musik & Zeppelzauer, 2018). As online content becomes more visual and ML grows omnipresent, data and algorithmic bias linked with images and sound are increasingly significant. Women and girls have been stereotyped and sexualized through their images (Langton, 2009; Nussbaum, 1995; Szymanski, Moffitt, & Carr, 2011). Meanwhile, women's objectification in media can lead to psychological problems, discrimination, and gender-based violence (e.g., Fredrickson & Roberts, 1997; Meyers, 2007). Besides, the misrepresentation of gender violence by media (e.g., videogames and advertising), portraying abuse as something acceptable, irrelevant, or absurd is a well-studied phenomenon (Galarza Fernández, Sosa Valcarcel, & Castro Martínez, 2018; Lloyd & Ramon, 2016; Meyers, 2007). The profusion of images and audiovisual material channeling stereotypes and misrepresentation is then poured into the Web, along with other content. Prejudiced role models matter—not only do they influence both girls' and boys' perceptions of their capabilities (Prpic, Sabbati, & Shreeves, 2018), but they can also provide the foundation for bad algorithmic decision making, harming girls' and women's rights.

The platformization of media content—the upsurge of the platform as the "dominant infrastructural and economic model of the social web" (Helmond, 2015, p. 1)—changes everything. Inequality embedded in advertising, film, music videos, and television take on a new life when the platforms make biased algorithmic decisions, potentially multiplying prejudice, and establishing a vicious cycle that is not apparent. Namely, an examination of audiovisual content from a gender perspective cannot be undertaken today without tackling algorithmic bias.

ML enables the interactions we have with platforms such as Amazon, Facebook, Google, and Netflix. ML-based AI makes decisions about what to offer by customizing the outputs to display individual interests, backgrounds, and locations. But algorithms are used to grant loans, offer medical care, deliver sentences, award aid, and design policies and campaigns, among other tasks. However, algorithmic biases are found on all platforms (Ciampaglia, 2018; Ciampaglia & Menczer, 2018). Operating in the dark, this technology can be an amplifier of inequalities, as these systems are increasingly employed to make decisions about, for instance, who is eligible for health-care support, how civic dialogue is mediated, and in what way elections are influenced (Naughton, 2018).

Discrimination can involve different dimensions, including color, status, postal code, and income, which are considered intersecting forms of discrimination (Human Rights Council, 2017). These dimensions

serve as vectors affecting women (Hill Collins, 1998). Similarly, in technology, gender bias is a multifaceted phenomenon (Michelfelder, Wellner, & Wiltse, 2017). This article examines unintentional bias, not the kind of deliberate discrimination that occurs in the informational state (Braman, 2009), computational politics (Tufekci, 2014), or dataveillance practices (van Dijck, 2014). The idea is to contribute to feminist theory and audiovisual and critical data studies by linking perspectives that rarely intersect.

This study addresses the following questions:

RQ1:   *How does gender bias happen when ML algorithms are fed with audiovisual data?*

RQ2:   *Although the term "gender" is employed across the article, the focus here is bias harming women, not gender in general. Women are often ignored or singled out in data sets. Which aspects should a gender-specific research agenda tackle when considering algorithmic decision making based on audiovisual data?*

This article is organized as follows: First, it offers some notions on algorithmic bias relevant to audiovisual content; second, it deals with how audiovisual data fuel algorithmic gender bias; third, it tackles the factors that create gender-biased audiovisual content, code, and data; fourth, it lists three significant challenges as a result of gender algorithmic bias; and finally, it offers a discussion and presents recommendations for a research agenda.

Background on Algorithmic Bias

The algorithmic biases embedded in platforms (Ciampaglia, 2018; Ciampaglia & Menczer, 2018) are not necessarily unfair. Bias in ML is a prior condition for intelligent action (Bishop, 2006). Yet bias can be problematic when it "is derived from aspects of a culture known to lead to harmful behavior," resulting in stereotyping and prejudiced action (Caliskan, Bryson, & Narayanan, 2017, p. 2). There is no standard definition of algorithmic bias. For Mehrabi, Morstatter, Saxena, Lerman, and Galstyan (2019), it refers to the kind of preferences incorporated by the algorithm as opposed to other biases entrenched in data sets or caused by the people handling the data. For brevity's sake, here I use algorithmic bias to refer to biases embedded in both audiovisual data sets and related algorithms.

A hindrance to tackling prejudices is that algorithms are perceived as fairer than humans. Algorithms' capacities encourage the tendency to project an anthropomorphic agency onto them (Prey, 2017, p. 1086; Wellner & Rothman, 2019, p. 7). Taddeo and Floridi (2018) speak of AI as a "new form of agency" (p. 751), since this technology determines which services, jobs, and rulings are available to whom. Besides, code is discussed "in almost biblical terms," inspired by corporations, which "have sold and defended their algorithms on the promise of objectivity" (A. Smith, 2018b, para. 8). Some studies support this idea. A survey published by the Pew Research Center indicates that 40% of people in the U.S. think that computer programs can make decisions without bias (A. Smith, 2018a).

However, raw data do not exist outside the imagination (Boellstorff, 2013), and algorithms do not decide or learn. Only humans can act with data agency, which entails "the processes of action based on

reflection" (Couldry, 2013, p. 13). This article is based on this idea. Algorithmic biases do not happen spontaneously, independently of people, in the same way that data are not natural (Gitelman, 2013). Data are always "cooked" in equally "cooked" processes (Boellstorff, 2013). Both data sets and computing decisions offer an imperfect representation of the world; they constitute human judgments that reflect a vision about how the world is (Niklas & Peña Gangadharan, 2018, p. 10). Mehrabi et al. (2019) identify different types of bias in ML; the most relevant for gender and audiovisual data are summarized in the following sections.

Presentation Bias

Presentation bias derives from the location on the webpage of particular content because users can only click on the things they see (Bar-Ilan, Keenoy, Levene, & Yaari, 2008; Coleman, 2013). There are no studies on algorithmic presentation bias in audiovisual content from a gender perspective; however, women's images are affected by presentation biases in media. For example, women make the front-page on mainstream UK newspapers only when they are celebrities or portrayed as victims ("How Women," 2012). Until 2019, British tabloids would show an image of a topless woman on the third page (Waterson, 2019).

Filter Bias

Filter bias occurs when users only find content that strengthens their preconceptions (Nagulendra & Vassileva, 2014; Pariser, 2011). That is, filter or information bubbles do not just reflect the user's identities as defined by the algorithms but also what audiovisual choices they have online (Pariser, 2011, p. 63).

Selection Bias or Sampling Bias

Selection bias or sampling bias arises because of nonrandom sampling of subcategories (Mehrabi et al., 2019). It can happen both at the design phase when, for example, facial recognition training data are based on White men's features, and at the user end, when users search online. As a minority of agents generate most online content (Lunenfeld, 2011), one is not "sampling at random from an entire 'population' of data" (Coleman, 2013, p. 14), but choosing from predetermined content. Selection bias hinders generalization.

Historical Bias

Historical bias is algorithmic decision making based on historical data entrenching undesirable trends (Coleman, 2013, p. 14). For example, algorithms trained on images of former (male) U.S. presidents forecasted that Donald Trump would win the elections in 2016 (Polonski, 2016). Thus, if histories of discrimination go unquestioned, they become part of routine algorithmic systems (Crawford, 2016).

Aggregation Bias

Aggregation bias is typical in discrimination against women; it happens when algorithmic conclusions for women are made drawing from data sets that aggregate information about general populations (Mehrabi et al., 2019).

Interaction Bias

Interaction bias is introduced by users when they search online guided by their audiovisual prejudices (Mehrabi et al., 2019). It can be influenced by presentation bias and others.

Some of these biases act together. Mixing interaction, selection, and presentation biases, Google's online advertising system proposes high-paying jobs to men more recurrently than to women (Datta, Tschantz, & Datta, 2015). Google defended the result, saying that it allows its customers to target their ads according to sex. Still, its algorithm may have determined that men were more apt to hold executive positions, having learned from the behavior of their users (Datta et al., 2015). If the only people who see and click on ads for high-paying jobs are men, the algorithm learns to show those ads only to them.

The companies producing the algorithms may be victims of their own biases. In 2015, Amazon discovered that their recruiting engine disliked women (Dastin, 2018). Table 1 summarizes their impacts.

Understanding How Audiovisual Data Encourage Algorithmic Gender Bias

Feminist scholarship has produced thought-provoking literature on how data-based machines can be biased. For instance, Bivens (2017) reveals that the Facebook software that allows users and advertisers to choose from various nonbinary gender identification options is a façade since nonbinary users are later reconfigured back into a binary system. Although Bivens does not refer to women, she says gender prejudice could play out in actual discrimination practices (e.g., making women invisible). Meanwhile, Williams, Brooks, and Shmargad (2018) denounce how "censoring" information on sex and race in data-driven automatic decision-making systems can lead to the invisibilization of women (p. 78).

Looking at online visual content, Otterbacher, Bates, and Clough (2017) compare the gender distribution in photos retrieved by Bing (a search engine) for the query "person" concerning different qualities. These authors conclude that "pictures of women are more often retrieved for *warm* traits (e.g., 'emotional') whereas *agentic* characteristics (e.g., 'rational') are represented by photos of men" (p. 1). Similar forms of stereotyping are found in online text awarding women feeble and submissive roles (e.g., Caliskan et al., 2017).

An added problem with audiovisual content is that it is impactful; nonverbal discrimination through image or voice pitch, for example, gives women less regard in job interviews (Hess, 2013). Gender stereotyping takes place when women's voices are associated with strident and abnormal noise (Tallon, 2019) or docile and accommodative functions (West, Kraut, & Chew, 2019, pp. 97–98). Visual cues are striking and distinct inputs in people's minds (Raghubir, 2008). Gender stereotypes in music videos influence the normalization of gender-based violence, concludes a report looking at the attitudes of boys and girls in six European countries (Kaili, 2018). If algorithms can multiply existing biases, prejudiced algorithmic decision making linked with audiovisual content is a cause for concern.

Audio data are the basis for different forms of gender algorithmic bias. Speech recognition systems and voice user interfaces (VUIs)—also known as smart speakers, robot speech, and voice assistants—have generated significant scholarship. In speech detection, gender biases are pervasive; they are found, for

instance, in Google's speech recognition systems (Tatman, 2016) and YouTube's automatically generated captions (Tatman, 2017), as well as in automatized broadcast news transcription (Woodland et al., 1998).

Meanwhile, robot speech shows a high degree of gender bias (Hannon, 2018). For example, UNESCO's report "I'd Blush if I Could" derives its title from the reply given by Siri when a user tells her, "You're a bi***!" (West et al., 2019, p. 5). The report looks at how AI voice assistants—Amazon's Alexa and Apple's Siri—projected as young women, propagate gender biases (West et al., 2019). *The Guardian* criticizes automated speakers, noting that call centers for brokerage firms in Japan use female voices to give stock quotes, but deploy a male voice to confirm transactions ("*The Guardian* View," 2019). Habler and Henze (2019) explore whether the usability of smart speakers is affected by users' gender. Gorrostieta, Lotfian, Taylor, Brutti, and Kane (2019) talk about biased speech emotion recognition systems. The companies' justification for the choice of VIUs is that people prefer female voices. But UNESCO finds that the situation is complicated. Though women usually change a default female voice to a male voice when this option is available, researchers found no references of men changing VIUs to male voices (West et al., 2019). The stereotyped perception of women's voices as pleasing and helpful—when not shrill and dismissible—seems to have driven corporations to make their voice assistants young and female (West et al., 2019, pp. 97–98).

The exploration of visual algorithmic gender bias has focused on image processing algorithms (IPAs), and concretely on facial recognition systems (Buolamwini & Gebru, 2018; Lohr, 2018; Tucker, 2017). IPAs are behind the automatization of analyzing visual materials. For example, a study reveals that three of the latest facial recognition AIs—IBM, Microsoft, and Megvii—could identify a person's gender in a photograph 99% of the time, but only in the case of White men; for dark-skinned women, accuracy was reduced to 35% (Lohr, 2018). Another project, Gender Shades, evaluates three commercial classification systems—IBM, Microsoft, and Face—with a data set including pictures of people of all skin shades to conclude that "darker-skinned females are the most misclassified group (with error rates of up to 34.7%)" (Buolamwini & Gebru, 2018, p. 77). White men are only 31% of the population in the U.S., but they account for 65% of public positions (Henderson, 2014) and get all the (facial) "recognition." Visual bias bears the risk for another digital divide that requires attention (Introna & Wood, 2004).

Mittelstadt et al. (2016) identify six types of ethical concerns connected with digital computing. "Inconclusive evidence" is the result of ML techniques that produce probably "but inevitably uncertain" knowledge; "inscrutable evidence" stems from the disconnection between the data and the outcome as data processes are indecipherable; "misguided evidence" derives from the fact that results can only be as reliable as the data they are based on; "unfair outcomes" can surface even from conclusive, transparent, and robust evidence; algorithmic activities that remythologize the world translating into "transformative effects," and the difficulties in detecting the harm, finding its cause and identifying who is responsible (Mittelstadt et al., 2016, pp. 4–5). Table 2 summarizes them.

Understanding the Factors

Women create abundant audiovisual content and are vocal about algorithmic issues. However, they are still a minority in the generation of audiovisual data and ML algorithms, the testing and usage of algorithmic solutions, and in supervisory roles.

Looking at the film industry in the U.S., for example, Lauzen (2019) concludes that only 8% of the directors working on the top 250 domestic grossing films in 2018 were women. A study on female characters in popular movies showed that 31% of all speaking characters were female, and that sexualization was the standard (S. L. Smith, Choueiti, & Pieper, 2014). Gender bias is found in music videos, TV series, ads, video games, and other audiovisual content (Hansen, 1989; Karsay, Matthes, Platzer, & Plinke, 2017; Langston, 2015; S. L. Smith et al., 2014; West et al., 2019).

Meanwhile, less than one in five people who graduate in computer science are women (Girls Who Code, 2018), and the ML industry employs an even smaller proportion of women than the rest of the technology sector (Simonite, 2018). Women occupy only 5% of leadership positions in the technology industry (Mylavarapu, 2016). More than two-thirds of emerging companies in the U.S. have no women on their boards of directors (Ash, Gann, & Dodgson, 2018). Computer science is one of the few disciplines related to science, technology, engineering, and mathematics in which the number of women has been decreasing (Pickett, 2018). Once inside tech companies, women are also are more likely to name discrimination as their reason for leaving this career path (West et al., 2019).

A clue to explain what happens could come from news media organizations, also generators of audiovisual content. Djerf-Pierre's (2007) perverse circle observed in Swedish newsrooms works like this: Male directors rely on male reporters for the most important news, who count on male sources for their stories, which populate the front pages and open news programs. These male reporters are then promoted, and the wheel turns again. A correlation exists between gender balance in the media workforce and fairer content (Prpic et al., 2018). A similar mechanism could be in place here.

The lack of women decision-makers and creators is only one factor. Wachter-Boettcher (2017) discusses two more problems. First, the lack of women testing technologies. When researchers do not include all kinds of people in the product tests, they can introduce asymmetries. For instance, a prevalent bias of speech recognition systems may have to do with the men who develop them (Musik & Zeppelzauer, 2018). Second, women are missing as clients of algorithmic-based products. The typical product design incorporates the clients' objectives. For example, male clients procuring algorithmic hiring services may introduce biases when describing a job's requirements, including data and images. "Operational parameters are specified by developers and configured by users (clients) with desired outcomes in mind that privilege some values and interests over others" (Mittelstadt et al., 2016, p. 1). By introducing demographic data and pictures, stereotypes can be added (Wachter-Boettcher, 2017). Contrary to what Williams and colleagues (2018) say, Wachter-Boettcher recommends eliminating them from algorithmic processes to avoid discrimination. As seen later, this matter poses new challenges.

The question of locating the bias "becomes complex with the duo of algorithm-data set" (Wellner & Rothman, 2019, p. 13). Hotspots and blind spots in data sets constitute a significant factor in gender bias. The maxim garbage in, garbage out applies to ML (Naughton, 2018). Becoming overly reliant on data is part of the problem here (Wronkiewicz, 2018). If the data are biased, as audiovisual content often is, the garbage-in-garbage-out effect occurs. Faulty data have always existed. The problem worsens when algorithms learn through data sets that put the accent on sections of the population based on prejudice or convenience. Waardenburg, Sergeeva, and Huysman (2018) talk about hotspots and blind spots in data-

based policing; this idea applies to more data sets. Blind spots happen when women are absent in audiovisual content, leading to invisibilization; meanwhile, hotspots ensue when women are differentiated, leading to discrimination. Namely, there can be too much and too little data presence.

The first 112 findings when I search for "CEO" (chief executive officer) on Google Images using Google's Chrome Web browser are mostly men's images. Chrome's algorithm incorporates my own biases. This search took place on February 28, 2020. The findings include 90 men appearing alone, nine women, five groups of only men, and eight mixed groups. Although this selection is not representative, it is telling that only 8% of the pictures representing CEOs display unaccompanied women when in the U.S.; for instance, 27% of the people in management are women (Langston, 2015). This exercise shows how algorithmic bias and inequality interact. There is a real-life gender gap in management jobs, and Google Images' algorithm associates the word CEO predominantly with men's images, producing a more significant gap than the reality. It also shows how historical bias operates. Currently, a minority of women occupy management positions, and there are even fewer pictures of women at the helm; job hunting applications based on these historical data could perpetuate the situation or make it worse.

There can be too little and too much data presence. First, insufficient data about women conceal them and their struggles; when data sets are not available, no informed policy can be devised (D'Ignazio & Klein, 2019). Women CEOs not appearing on Google searches are, after all, privileged groups, but there is "little or no signal coming from particular communities" (Crawford, 2013, para. 3). Millions of people have no access to smartphones, online banking, or CCTV-monitored cities; of them, women and girls are more vulnerable (House Foreign Affairs Committee Republicans, 2015) and have less access to technology (Taylor, 2018). As seen with selection biases, extrapolating from digital users' data sets to entire populations leads to prejudice (Trevisan, 2013). Some organizations are filling gaps, visualizing hidden stories about women. ProPublica scanned Facebook groups and crowdfunding sites to collect 5,000 visual narratives of maternal harm that would have gone unnoticed (Gallardo, 2018). Meanwhile, María Salguero (2019) maps stories of assassinated women in Mexico. Looking at the lack of data about women who die in childbirth in the U.S., D'Ignazio and Klein (2019) note that "bodies" (i.e., real people's images) are missing from data science (para. 1). Maternal harm and femicides are not recorded often, even if they affect millions of women.

And second, only integrating data without examining its consequences can make women more vulnerable. There is a long tradition of creating registers of marginalized communities. NGOs specialized in the rights of transgender people challenge the idea of disaggregation because gender data can result in misclassification (Niklas & Peña Gangadharan, 2018, p. 15). Misrepresentation is a problem in other areas. "[Facebook] employees would sometimes mischaracterize ads based on their own inherent biases. . . [Later] these mischaracterizations are incorporated into manuals that train both human reviewers and machines," notes Eisenstat (2019, para. 10). Presence in data sets can be due to recognition, too. Taylor (1994) says that some strands in contemporary policies "turn on the need, sometimes the demand, for *recognition*" (p. 25). In contrast, "*mis*recognition" and lack of recognition makes people imperceptible in policies (Taylor, 1994, p. 25, emphasis in original). Boellstorff (2013) talks about an "incitement to disclose" (para. 48) as the technologically mediated compulsion to reveal one's identity and image to demand recognition.

Because of the silencing or emphasizing of women in data sets, the organized society fighting for digital rights is divided into two groups: those defending the privacy of vulnerable communities in the face of the invasive employment of data technologies and those advocating inclusion in the very same technologies, so women have a voice in their matters. Table 1 connects the factors influencing algorithmic gender bias based on audiovisual content, the types of algorithmic bias, and their consequences.

Table 1. Main Types of Algorithmic Gender Bias, Factors, and Effects.

Type	Related factors	Effects
Interaction bias introduced by users' prejudices	Real-life discrimination of women and gender gaps	Algorithmic discrimination; invisibilization of women
Presentation bias when content that is better positioned on a webpage is easier to find	A minority of agents generate and present the most online content; few women are coders or in charge of audiovisual content generation	Content available online is limited to those made by people with the opportunity to post and position content; invisibilization and stereotyping of women
Filter bias creating information bubbles	Few women are coders and testers; interaction bias	People have access only to content they agree with or like in the first place; discrimination and invisibilization of women
Selection bias	Few women are coders; perverse gender circle in firms creating algorithms/generating audiovisual content; narratives that consider data and algorithms neutral	Nonrandom data lead to faulty generalizations; bias gets introduced and goes unnoticed; bias is obscured or denied
Historical bias and demographic data and images from clients	General discrimination of women; gender gaps; sexualization and objectification of women; perverse gender circle in firms creating algorithms/generating audiovisual content; few women are testers	The perpetuation of situations that may not be desirable; bias gets introduced and goes unnoticed
Aggregation bias	Few women are coders or testers	Potential misclassification; invisibilization

Table 1 connects the types of algorithmic bias identified earlier with the factors that make them possible and their potential consequences in real life, linking the previous sections. Next, the main three obstacles for fair gender algorithmic decision making are examined and integrated into the recommendations at the end.

Challenges in Addressing Biased Audiovisual Content

This section connects issues around algorithmic processes to produce a catalog of three challenges: (a) lack of clarity about data presence issues, (b) lack of understanding of how ML works, and (c) lack of incentives for corporations to prevent and correct bias. These challenges are present in all algorithmic

processes, not just those affecting women or related to audiovisual data; however, they can help understand how women end up being discriminated against in algorithmic processes based on audiovisual data.

First, women can be either absent or emphasized in data sets. That is why the lack of clarity about data presence can become a problem. Wang (2018) asks whether an algorithm should be debiased when this could mean removing specific dimensions of the data that could lead to eliminating valuable information. Narayanan concurs: "What constitutes a terrible bias or prejudice in one application might end up being exactly the meaning you want to get out of the data in another application" (Knight, 2016, para. 9). Because there is no clarity about the correct data presence, some authors refer to fairness as a gauge. However, Narayanan (2018) notes that any definition of "fair" is loaded with prejudices. Mehrabi and associates (2019) speak of fairness in technical terms. They provide descriptions that can be grouped in three types: individual fairness gives similar predictions to similar individuals, group fairness treats different groups equally, and subgroup fairness picks a group fairness constraint and asks whether it is maintained for other subgroups (Mehrabi et al., 2019). Others say transparent algorithms are the solution so that they can be judged on a case-by-case basis. Matsakis (2018) reports that several public and private entities are looking at ML models that can explain how they make decisions. Nevertheless, transparency alone might be inadequate to look at how algorithms work (Ananny & Crawford, 2019).

Second, not all problematic outcomes in ML processes are down to flaws in the input data; the code can also be obscure or inaccessible. These two issues present different challenges. Corporate secrecy laws and intellectual property regimes can contribute to making code unreachable to auditing, generating the "black box effect" (Whittaker et al., 2018). Edionwe (2017) notes that algorithms are opaque "for proprietary reasons or by deliberate design" (para. 16). This problem is so pervasive that Malgieri and Comandé (2017) call ours the "black box society." The difficulty is accentuated by market concentration as it puts growing power with a handful of multinational private players (Reichman & Pollicino, 2018). The lack of transparency makes it unfeasible to contest results. That is why the public sector should stipulate that providers waive code protection mechanisms (Whittaker et al., 2018). This recommendation suggests that, if there is transparency, bias can be removed. A different matter is the "unintelligibility" of some algorithms, which hamper the detection of problems (Niklas & Peña Gangadharan, 2018, p. 14). Concurring, Wronkiewicz (2018), Naughton (2018), Edionwe (2017), and Rahimi (ObservantVids, 2017), among others, suggest even that coders do not understand how their algorithms work. That is, if some algorithms are incomprehensive, there is nothing anyone can do about them. ML models are so intricate that they can be compared with alchemy (ObservantVids, 2017). Data scientists often do not comprehend why algorithms output an answer from a data set and lack a robust theoretical grasp of their tools (ObservantVids, 2017). Rahimi, from Google, concludes that, since ML has gone beyond photo-sharing systems to systems that govern health care, civic dialogue, and elections, it should rely on verifiable knowledge (ObservantVids, 2017). But not everybody agrees; LeCun (2017), from Facebook, believes that criticizing coders "for practicing 'alchemy' because our current theoretical tools haven't caught up with our practice is dangerous" (para. 6). Peng (2017) summarizes the discussion as a choice between "more effective ML models without clear theoretical explanations, or simpler, transparent models that are less effective in solving specific tasks" (para. 15).

And third, the tech industry is not doing enough to address these biases (Eisenstat, 2019). The problem is not only that they are hidden but also that "most companies using similar algorithms don't even want to know" (O'Neil, 2018, para. 1), or need to know, as there are no incentives to avoid bias. There is no "universal standard for tagging algorithmic bias" and "it is unlikely that AI developers will voluntarily disclose their algorithmic bias profile;" thus, users deal with AI applications that have "zero accountability" (Kahana, 2018, para. 2). The cases mentioned here were uncovered by investigative units or scholars, not by the corporations that created the algorithms. The AI Now Institute notes that internal governance structures within companies are failing to ensure accountability (Whittaker et al., 2018). Without effective mitigation, this is the "era of plausible deniability in big data" (O'Neil, 2018, para. 2). Namely, companies can plead ignorance. One factor is how fast the industry is moving. For example, although ML software to eradicate the prejudice of hiring top jobs has produced "encouraging" results, "tech executives with experience at Google, Microsoft, and Facebook say the algorithmic revolution in hiring is moving too fast (to make corrections)" (Rosenbaum, 2018, para. 2). Byrne advocates for an antidiscrimination legal framework for the technology industry to force companies to invest in correcting biases since they will not do so voluntarily (Kuchler, 2018). O'Neil (2016) suggests that data scientists should pledge a Hippocratic Oath that "focuses on the possible misuses and misinterpretations of their models" (p. 386). Most hope that regulation can strengthen algorithmic governance. The AI Now Institute issued a set of recommendations that refer to stringent law, transparency, oversight, and limitations of public algorithmic decision making, including the need to regulate facial recognition "by expanding the powers of sector-specific agencies to oversee, audit, and monitor these technologies by domain" (Whittaker et al., 2018, p. 4). The AI Now Institute notes that affect recognition deserves "particular attention" since it "claims to detect things such as personality, inner feelings, mental health, and 'worker engagement' based on images or video of faces," despite not being based on robust science (Whittaker et al., 2018, p. 8). Besides, there is a long tradition of discriminating against women based on image and sentiment. The era of deniability may pass if lawsuits start to be waged (Rosenbaum, 2018). The AI Now Institute anticipates five areas of future legal activity, including medical aid and government benefits (AI Now Institute, 2018). Table 2 combines the gaps and outcomes identified earlier with these ethical concerns.

Table 2. Challenges in Gender Algorithmic Bias.

Gaps	Results	Ethical concerns
Clarity about data presence and fairness	Imbalances; disparities; unfairness	Misguided evidence; transformative effects
Understanding of machine learning	Opacity; inability to correct bias	Inconclusive; inscrutable or misguided evidence; problematic traceability
Incentives to prevent and correct bias	Lack of accountability; plausible deniability	Inscrutable evidence; problematic traceability

Table 2 relates both clarity about audiovisual data presence and understanding of ML to lack of transparency, and incentives for change. This table connects the specific question of women's representation in sound and image with the broader issue of how ML works and what conditions make algorithmic bias challenging to trace and tackle. The lack of clarity about women's data presence, ML's

intelligibility, and the impediments for change can result in inequalities, opaqueness, and irresponsibility that are cause for ethical concerns.

Discussion and Recommendations

How does gender bias happen when ML algorithms are fed with audiovisual data? Gender algorithmic bias can be defined, in simple terms, as algorithmic computation based on audiovisual data, which results in misrepresentation, discrimination, or invisibilization of women, creating adverse outcomes for them. These biases can be introduced by the designers and users of the ML-based systems, the selection of training data, or historical data representing unfair situations. Other factors include how and where women's representation appears online, how audiovisual data are aggregated, and the way algorithms are designed to search for relevant content, among others. Both data sets and algorithms can be biased. The lack of clarity about women's data presence and the lack of incentives to prevent bias can cause imbalances, disparities, and unfairness in the real world; when these outcomes affect women, then algorithmic gender bias occurs. That is, real-life gender discrimination embedded in audiovisual data can result in bias in the digital world. Likewise, algorithmic gender bias translates into more discrimination against women in the real world, establishing a perverse circle. Figure 1 summarizes this connection.
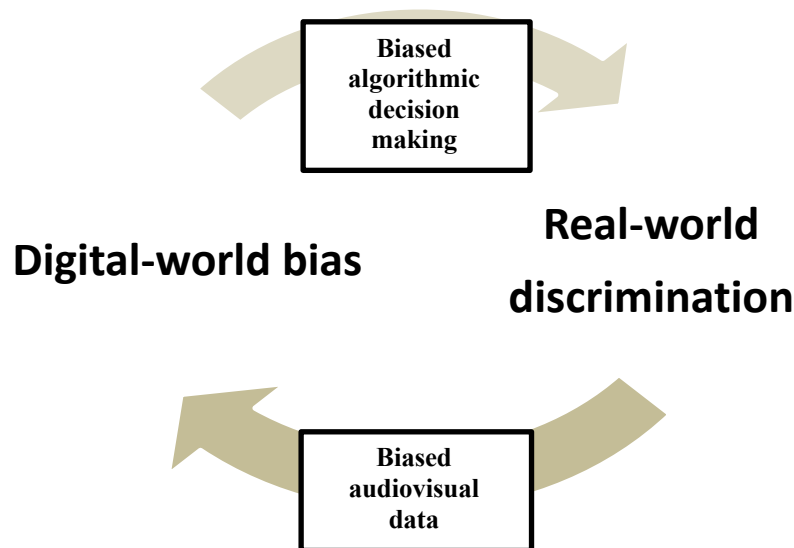

Figure 1. The perverse circle of algorithmic gender bias.

Which aspects should a gender-specific research agenda tackle when considering algorithmic decision making based on audiovisual data? The recommendations included below emerge from the analysis of audiovisual data and algorithmic biases. This article proposes six specific aspects, drawing from Floridi (2013) and Colman, van der Tuin, O'Donnell, and Bühlmann (2018, p. 6) advise to go beyond the current

emphasis on privacy, the security of data, copyright issues, and existing tick-box models. A robust body of research around these areas is needed to expose systematic breaches and mitigate unfairness.

Objectives First

Focusing on the objectives of any data analysis is essential to establishing women's data presence. Data should be collected with a declared aim, affording a degree of flexibility for genuine research projects that operate within an ethical framework (Wiewiorówski, 2020, p. 3). That is, the intention of any automated system using personal data–either public or private—must be transparent before embarking in designing it, and determining what data sets are necessary to avoid selection and aggregation biases. By-default black boxes should be the object of intense scrutiny on a case-by-case basis.

User-Centered Approaches

Musik and Zeppelzauer (2018) recommend an adaptative and user-centered approach. Namely, women's data presence—whether to include specific data sets or exclude data dimensions—is to be decided together with the women as data providers and beneficiaries. Whether the women targeted as users are integrated into algorithmic decision making from the outset should also be investigated in concrete cases.

More Than Historical Data

Given the past and present of real-life discrimination against women, and the perverse circle of algorithmic gender bias, historical data cannot be the sole basis for algorithmic policy affecting women. Clarity about data presence is essential; it must be fair. Audiovisual content employed as training data should be adequate to represent women justly, even integrating equality targets (e.g., the UN's sustainable development goal to achieve gender equality and empower all women and girls). Whether this happens in specific projects with social impacts should be investigated.

Social Sciences Integration

Research centers should incorporate scholars from social domains to pay attention to the risks of algorithmic decision making (Whittaker et al., 2018). Eisenstat (2019) also advocates for tech companies to train their data scientists on understanding cognitive bias. Whether data teams include social science scholars or data scientists who are trained in managing cognitive biases in particular projects is another venue for analysis.

More Women

ML development should include more women as designers of algorithms and as testers to detect possible problems that may not be obvious for a White man who has never felt discriminated against, says Byrne (Kuchler, 2018). Another area to examine is who the coder of specific projects is and whether the testers of algorithmic applications include women, endorsing women's needs and viewpoints.

Regulation

Regulation has a role in creating incentives, whether negative or positive, for corporations to stop, identify, and mend unjust systems, and to create mechanisms that repair problems and compensate victims of poor algorithmic decision making. Another research focus should be the law systems that enable algorithmic decision making and the incentives that governments offer or impose on the commercial exploitation of big data. Corporate secrecy laws can be investigated, vis-à-vis their social impacts.

The "algorithmic condition" is altering human rights (Colman et al., 2018, p. 7), which ban discrimination based on sex or gender. In a society where everything we do is datafied, crunched, digested, and mediated by algorithms that contribute to critical decisions, women's rights hinge on an algorithmic ethos and research agenda that promotes equality proactively.

References

AI Now Institute. (2018). *Litigating algorithms: Challenging government use of algorithmic decision systems*. Retrieved from https://ainowinstitute.org/litigatingalgorithms.pdf

Ananny, M., & Crawford, K. (2019). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, *20*(3), 973–989. doi:10.1177/1461444816676645

Ash, S., Gann, D., & Dodgson, M. (2018). *The tech industry needs more women. Here's how to make it happen.* Retrieved from https://www.weforum.org/agenda/2017/11/the-tech-industry-needs-more-women-heres-how-to-make-it-happen-d1eec473-48cc-4801-bd40-dfba040b0e4a/

Bar-Ilan, J., Keenoy, K., Levene, M., & Yaari, E. (2008). Presentation bias is significant in determining user preference for search results—A user study. *Journal of the Association for Information Science and Technology, 6*(1), 135–149. doi:10.1002/asi.20941

Bishop, C. M. (2006). *Pattern recognition and machine learning*. London, UK: Springer.

Bivens, R. (2017). The gender binary will not be deprogrammed: Ten years of coding gender on Facebook. *New Media & Society, 19*(6), 880–898. doi:10.1177/1461444815621527

Boellstorff, T. (2013). Making big data, in theory. *First Monday*, *18*(10). doi:10.5210/fm.v18i10.4869

Braman, S. (2009). *Change of state: Information, policy, and power*. Cambridge, MA: MIT Press.

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities on commercial gender classification. In S. A. Friedler & C. Wilson (Eds.), *Conference on Fairness, Accountability, and Transparency (Proceedings of Machine Learning Research)* (Vol. 81, pp. 77–91). New York: New York University Press. Retrieved from http://proceedings.mlr.press/v81/buolamwini18a.html

Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, *356*(6334), 183–186. doi:10.1126/science.aal4230

Ciampaglia, G. L. (2018). Fighting fake news: A role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science, 1*(1), 147–153. doi:10.1007/s42001-017-0005-6

Ciampaglia, G. L., & Menczer, F. (2018, June 20). *Misinformation and biases infect social media, both intentionally and accidentally*. Retrieved from https://theconversation.com/misinformation-and-biases-infect-social-media-both-intentionally-and-accidentally-97148

Coleman, G. E. (2013). C*oding freedom: The ethics and aesthetics of hacking*. Princeton, NJ: Princeton University Press.

Collins, R. L. (2011). Content analysis of gender roles in media: Where are we now and where should we go? *Sex Roles*, *64*, 290–298. doi:10.1007/s11199-010-9929-5

Colman, F., van der Tuin, I., O'Donnell, A., & Bühlmann, V. (2018). *Ethics of coding: A report on the algorithmic condition*. Brussels, Belgium: European Commission.

Couldry, N. (2013). A necessary disenchantment: Myth, agency, and injustice in a digital world. *The Sociological Review, 62*(4), 880–897. doi:10.1111/1467-954X.12158

Crawford, K. (2013, April 1). *The hidden biases in big data*. Retrieved from https://hbr.org/2013/04/the-hidden-biases-in-big-data

Crawford, K. (2016, June 25). Artificial intelligence's white guy problem. *The New York Times*. Retrieved from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html

Dastin, J. (2018, October 11). *Amazon scraps secret AI recruiting tool that showed bias against women.* Retrieved from https://medium.com/datadriveninvestor/amazon-scraps-secret-ai-recruiting-engine-that-showed-biases-against-women-995c505f5c6f

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies, 2015*(1), 92–112. Philadelphia, PA. doi:10.1515/popets-2015-0007

D'Ignazio, C., & Klein, L. F. (2019). Chapter one: Bring back the bodies. In C. D'Ignazio & L. F. Klein (Eds.), *Data feminism* (pp. 1–22). Cambridge, MA: MIT Press Open. Retrieved from https://mitpressonpubpub.mitpress.mit.edu/pub/zrlj0jqb

Dijck, J. van. (2014). Datafication, dataism, and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, *12, 2*197–208. doi:10.24908/ss.v12i2.4776

Djerf-Pierre, M. (2007). The gender of journalism: The structure and logic of the field in the twentieth century. *Nordicom Review, 28,* 81–104. Retrieved from https://www.gu.se/english/research/publication?publicationId=49534

Edionwe, T. (2017, May 24). *The fight against racist algorithms: Can we teach our machines to unlearn racism?* Retrieved from https://theoutline.com/post/1571/the-fight-against-racist-algorithms?zd=1&zi=3hudf2l2

Eisenstat, Y. (2019, February 12). *The real reason tech struggles with algorithmic bias*. Retrieved from https://www.wired.com/story/the-real-reason-tech-struggles-with-algorithmic-bias/

Floridi, L. (2013). *The ethics of information*. Oxford, UK: Oxford University Press.

Fredrickson, B. L., & Roberts, T. A. (1997). Objectification theory: Toward understanding women's lived experiences and mental health risks. *Psychology of Women Quarterly*, *21*(2), 173–206. doi:10.1111/j.1471-6402.1997.tb00108.x

Galarza Fernández, E., Sosa Valcarcel, A., & Castro Martínez, A. (2018). The (un)protection of women in the audiovisual communication law: A legal reality in the Spanish scenario. *Congreso Universitario Internacional sobre la Comunicación en la Profesión y en la Universidad de Hoy [International Congress on Today's Professional and University Communication]: Vol. 1. Vivat Academia* (pp. 227–230). Madrid, Spain. doi:10.15178/CUICIID.2018

Gallardo, A. (2018, March 20). How we collected nearly 5,000 stories of maternal harm. *ProPublica*. Retrieved from https://www.propublica.org/article/how-we-collected-nearly-5-000-stories-of-maternal-harm

Girls Who Code. (2018). *About us.* Retrieved from https://girlswhocode.com/about-us/

Gitelman, L. (Ed.). (2013). *Raw data is an oxymoron*. Cambridge, MA: MIT Press.

Gorrostieta, C., Lotfian, R., Taylor, K., Brutti, R., & Kane, J. (2019). Gender de-biasing in speech emotion recognition. *Proceedings of Interspeech 2019,* 823–2827. doi:10.21437/Interspeech.2019-1708

The Guardian view on female voice assistants: Not OK, Google [Editorial]. (2019, June 24). *The Guardian*. Retrieved from https://www.theguardian.com/commentisfree/2019/jun/24/the-guardian-view-on-female-voice-assistants-not-ok-google

Habler, F., & Henze, N. (2019). Differences between smart speakers and graphical user interfaces for music search considering gender effects. *Proceedings of the 18th International Conference on Mobile and Ubiquitous Multimedia, 30,* 1–7. Pisa, Italy. doi:10.1145/3365610.3365627

Hannon, C. P. (2018). Avoiding bias in robot speech. *Interactions, 25*(5), 34–37. doi:10.1145/3236671

Hansen, C. H. (1989). Priming sex-role stereotypic event schemas with rock music videos: Effects on impression favorability, trait inferences, and recall of a subsequent male-female interaction. *Basic and Applied Social Psychology, 10*(4), 371–391. doi:10.1207/s15324834basp1004_6

Helmond, A. (2015). The platformization of the Web: Making Web data platform ready. *Social Media + Society*, 1–11. doi:10.1177/2056305115603080

Henderson, N. M. (2014, October 8). White men are 31 percent of the American population. They hold 65 percent of all elected offices. *The Washington Post*. Retrieved from https://www.washingtonpost.com/gdpr-consent/?next_url=https%3a%2f%2fwww.washingtonpost.com%2fnews%2fthe-fix%2fwp%2f2014%2f10%2f08%2f65-percent-of-all-american-elected-officials-are-white-men%2f

Hess, K. P. (2013). Investigation of nonverbal discrimination against women in simulated initial job interviews. *Journal of Applied Social Psychology, 43*(3), 544–555. doi:10.1111/j.1559-1816.2013.01034.x

Hill Collins, P. (1998). It's all in the family: Intersections of gender, race, and nation. *Hypatia, 13*(3), 62–82. doi:10.1111/j.1527-2001.1998.tb01370.x

Hitti, Y., Jang, E., Moreno, I., & Pelletier, C. (2019). Proposed taxonomy for gender bias in text; a filtering methodology for the gender generalization subtype. *Proceedings of the First Workshop on Gender Bias in Natural Language Processing,* 8–17. Florence, Italy. doi:10.18653/v1/W19-3802

House Foreign Affairs Committee Republicans. [HouseForeign]. (2015, November 17). *Chairman Royce opening statement at hearing on women & technology* [Video file]. Retrieved from https://www.youtube.com/watch?v=5gSiatLehZY

How women make the front page in Britain: Get the full data. (2012, October 15). *The Guardian*. Retrieved from https://www.theguardian.com/news/datablog/2012/oct/15/women-newspapers-front-pages

Human Rights Council. (2017, September 25). *Council holds a panel discussion on the impact of multiple and intersecting forms of discrimination on women and girls.* Retrieved from https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=22146&LangID=E

International Labour Organization. (2020). *Who are domestic workers?* Retrieved from https://www.ilo.org/global/topics/domestic-workers/who/lang--en/index.htm

Introna, L., & Wood, D. (2004). Picturing algorithmic surveillance: The politics of facial recognition systems. *Surveillance & Society, 2,* 2–3. doi:10.24908/ss.v2i2/3.3373

Kahana, E. (2018, December 22). *Solving algorithmic bias via artificial intelligence computational law applications.* Retrieved from https://law.stanford.edu/2018/12/22/artificial-intelligence-computational-law-applications-and-algorithmic-bias/

Kaili, C. (2018). *Play it for change*. Ljubljana, Slovenia: Mirovni Institut & The Peace Institute.

Karsay, K., Matthes, J., Platzer, P., & Plinke, M. (2017). Adopting the objectifying gaze: Exposure to sexually objectifying music videos and subsequent gazing behavior. *Media Psychology, 21*(1), 27–49. doi:10.1080/15213269.2017.1378110

Knight, W. (2016, November 23). *How to fix Silicon Valley's sexist algorithms: Computers are inheriting gender bias implanted in language data sets—And not everyone thinks we should correct it*. Retrieved from https://www.technologyreview.com/2016/11/23/155858/how-to-fix-silicon-valleys-sexist-algorithms/

Kuchler, H. (2018, March 9). Tech's sexist algorithms and how to fix them. *Financial Times.* Retrieved from https://www.ft.com/content/d2a1ab08-f63e-11e7-a4c9-bbdefa4f210b

Langton, R. (2009). *Sexual solipsism: Philosophical essays on pornography and objectification.* Oxford, UK: Oxford University Press.

Langston, J. (2015). *Who's a CEO? Google Images results can shift gender biases*. Seattle, WA: University of Washington.

Lauzen, M. M. (2019). *The celluloid ceiling: Behind-the-scenes employment of women on the top 100, 250, and 500 films of 2018*. San Diego, CA: Center for the Study of Women in Television and Film.

LeCun, Y. (2017). *My take on Ali Rahimi's "test of time" award talk at NIPS.* Retrieved from https://www.reddit.com/r/MachineLearning/comments/7i1uer/n_yann_lecun_response_to_ali_rahimis_nips_lecture/

Lloyd, M., & Ramon, S. (2016). Smoke and mirrors: UK newspaper representations of intimate partner domestic violence. *Violence Against Women, 23*(1), 1–26. doi:10.1177/1077801216634468

Lohr, S. (2018, February 9). Facial recognition is accurate, if you're a White guy. *The New York Times*. Retrieved from https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

Lunenfeld, P. (2011). *The secret war between downloading and uploading: Tales of the computer as culture machine*. Cambridge, MA: MIT Press.

Malgieri, G., & Comandé, G. (2017). Why a right to legibility of automated decision making exists in the general data protection regulation. *International Data Privacy Law, 7*(4), 243–265. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3088976

Matsakis, L. (2018, October 11). *What does a fair algorithm actually look like?* Retrieved from https://www.wired.com/story/what-does-a-fair-algorithm-look-like/

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). *A survey on bias and fairness in machine learning.* Retrieved from https://arxiv.org/abs/1908.09635

Meyers, M. (2007). African American women and violence: Gender, race, and class in the news. *Critical Studies in Media Communication, 21*(2), 95–118. doi:10.1080/07393180410001688029

Michelfelder, D. P., Wellner, G., & Wiltse, H. (2017). Designing differently: Toward a methodology for an ethics of feminist technology design. In S. O. Hansson (Ed.), *The ethics of technology* (pp. 193–219). Lanham, MD: Rowman & Littlefield.

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society,* 1–21. doi:10.1177/2053951716679679

Musik, C., & Zeppelzauer, M. (2018). Computer vision and the digital humanities: Adapting image processing algorithms and ground truth through active learning. *VIEW Journal of European Television History and Culture, 7*(14), 59–72. doi:10.18146/2213-0969.2018.jethc153

Mylavarapu, S. (2016, May 10). *The lack of women in tech is more than a pipeline problem*. Retrieved from https://techcrunch.com/2016/05/10/the-lack-of-women-in-tech-is-more-than-a-pipeline-problem/

Nagulendra, S., & Vassileva, J. (2014). Understanding and controlling the filter bubble through interactive visualization: A user study. *Proceedings of the 25th ACM Conference on Hypertext and Social Media,* 107–115. New York, NY. doi:10.1145/2631775.2631811

Narayanan, A. (2018, March 1). *Tutorial: 21 fairness definitions and their politics* [Video file]. Retrieved from https://www.youtube.com/watch?v=jIXIuYdnyyk

Naughton, J. (2018, August 5). Magical thinking about machine learning won't bring the reality of AI any closer. *The Guardian*. Retrieved from

https://www.theguardian.com/commentisfree/2018/aug/05/magical-thinking-about-machine-learning-will-not-bring-artificial-intelligence-any-closer

Niklas, J., & Peña Gangadharan, S. (2018). *Between antidiscrimination and data*. London, UK: London School of Economics and Political Science.

Nussbaum, M. (1995). Objectification. *Philosophy and Public Affairs, 24*(4), 249–291. doi:10.1111/j.1088-4963.1995.tb00032.x

ObservantVids. (2017, December 5). *Ali Rahimi's talk at NIPS (NIPS 2017 Test-of-time award presentation)* [Video file]. Retrieved from https://www.youtube.com/watch?v=Qi1Yry33TQE

O'Neil, C. (2016). *Weapons of math destruction*. New York, NY: Crown.

O'Neil, C. (2018, October 16). Amazon's gender-biased algorithm is not alone. *Bloomberg*. Retrieved from https://www.bloomberg.com/opinion/articles/2018-10-16/amazon-s-gender-biased-algorithm-is-not-alone

Otterbacher, J., Bates, J., & Clough, P. (2017). Competent men and warm women: Gender stereotypes and backlash in image search results. *Proceeding of the CHI Conference on Human Factors in Computing Systems,* 6620–6631. New York, NY. doi:10.1145/3025453.3025727

Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York, NY: Penguin.

Peng, T. (2017, December 13). *LeCun vs. Rahimi: Has machine learning become alchemy?* Retrieved from https://medium.com/@Synced/lecun-vs-rahimi-has-machine-learning-become-alchemy-21cb1557920d

Pickett, A. (2018, January 26). *The dangers of keeping women out of tech*. Retrieved from https://www.wired.com/story/dangers-keeping-women-out-of-tech/

Polonski, V. (2016, November 6). Would you let an algorithm choose the next president? *Tech Crunch*. Retrieved from https://techcrunch.com/2016/11/06/would-you-let-an-algorithm-choose-the-next-u-s-president/

Prey, R. (2017). Nothing personal: Algorithmic individuation on music streaming platforms. *Media, Culture & Society, 40*(7), 1086–1100. doi:10.1177/0163443717745147

Prpic, M., Sabbati, G., & Shreeves, R. (2018). *Gender equality in the digital and media sectors*. Brussels, Belgium: European Parliament.

Raghubir, P. (2008). Are visual perception biases hard-wired? In M. Wedel & R. Pieters (Eds.), *Visual marketing: From attention to action* (pp. 143–166). Mahwah, NJ: Erlbaum.

Reichman, A., & Pollicino, O. (2018, November 6). New research group on algorithmic state, society, and market–constitutional dimensions. *AICL-AIDC.* Retrieved from https://blog-iacl-aidc.org/iacl-news/2018/11/4/new-research-group-algorithmic-state-society-and-market-constitutional-dimensions

Rosenbaum, E. (2018, April 29). Silicon Valley is stumped: Even AI cannot always remove bias from hiring. *CNBC*. Retrieved from https://www.cnbc.com/2018/05/30/silicon-valley-is-stumped-even-a-i-cannot-remove-bias-from-hiring.html

Salguero, M. (2019). *Los feminicidios en México* [Femicides in Mexico]. Retrieved from https://feminicidiosmx.crowdmap.com/

Simonite, T. (2018, August 17). AI is the future—But where are the women? *Wired*. Retrieved from https://www.wired.com/story/artificial-intelligence-researchers-gender-imbalance/

Smith, A. (2018a). *Attitudes toward algorithmic decision making*. Washington, DC: Pew Research Center.

Smith, A. (2018b, August 30). Franken-algorithms: The deadly consequences of unpredictable code. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger

Smith, S. L., Choueiti, M., & Pieper, K. (2014). *Gender bias without borders*. Los Angeles, CA: Geena Davis Institute on Gender in Media.

Szymanski, D. M., Moffitt, L. B., & Carr, E. R. (2011). Sexual objectification of women: Advances to theory and research. *The Counseling Psychologist*, *39*(1)*, 6–38. doi:10.1177/0011000010378402

Taddeo, M., & Floridi, L. (2018). How AI can be a force for good. *Science, 361*(6404), 751–752. doi:10.1126/science.aat5991

Tallon, T. (2019, September 3). Century of "shrill": How bias in technology has hurt women's voices. *The New Yorker*. Retrieved from https://www.newyorker.com/culture/cultural-comment/a-century-of-shrill-how-bias-in-technology-has-hurt-womens-voices

Tatman, R. (2016, July 12). Google's speech recognition has a gender bias. *Making Noise & Hearing Things*. Retrieved from https://makingnoiseandhearingthings.com/2016/07/12/googles-speech-recognition-has-a-gender-bias/

Tatman, R. (2017). Gender and dialect bias in YouTube's automatic captions. *Proceedings of the First Workshop on Ethics in Natural Language Processing,* 53–59. Valencia, Spain. doi:10.18653/v1/W17-1606

Taylor, C. (1994). The politics of recognition. In A. Gutmann (Ed.), *Multiculturalism, and the politics of recognition* (pp. 25–73). Princeton, NJ: Princeton University Press. Retrieved from http://elplandehiram.org/documentos/JoustingNYC/Politics_of_Recognition.pdf

Taylor, L. (2018, July 10). As technology advances, women are left behind in digital divide. *Reuters*. Retrieved from https://www.reuters.com/article/us-britain-women-digital/as-technology-advances-women-are-left-behind-in-digital-divide-idUSKBN1K02NT

Trevisan, F. (2013, May 15). Social engines and social science: A revolution in the making. *Economic and Social Research Council.* doi:10.2139/ssrn.2265348

Tucker, I. (2017, May 28). A white mask worked better: Why algorithms are not color blind. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2017/may/28/joy-buolamwini-when-algorithms-are-racist-facial-recognition-bias

Tufekci, Z. (2014). Engineering the public: Internet, surveillance, and computational politics. *First Monday, 19*(7). doi:10.5210/fm.v19i7.4901

Waardenburg, L., Sergeeva, A., & Huysman, M. (2018). Hotspots and blind spots: A case of predictive policing in practice. In U. Schultze, M. Aanestad, M. Mähring, C. Østerlund, & K. Riemer (Eds.), *IFIP WG 8.2 Working Conference on the Interaction of Information Systems and the Organization* (pp. 96–109). San Francisco, CA: IFIP Advances in Information and Communication Technology. doi:10.1007/978-3-030-04091-8_8

Wachter-Boettcher, S. (2017). *Technically wrong: Sexist apps, biased algorithms, and other threats of toxic tech*. New York, NY: Norton & Company.

Wang, E. (2018, May 1). What does it really mean for an algorithm to be biased? *The Gradient*. Retrieved from https://thegradient.pub/ai-bias/

Waterson, J. (2019, April 12). *Daily Star* covers up its page 3 girls, signaling end of tabloid tradition. *The Guardian*. Retrieved from https://www.theguardian.com/media/2019/apr/12/daily-star-covers-up-its-page-3-girls-signalling-end-of-tabloid-tradition

Wellner, G., & Rothman, T. (2019). Feminist AI: Can we expect our AI systems to become feminist? *Philosophy & Technology, 33*, 191–205. doi:10.1007/s13347-019-00352-z

West, M., Kraut, R., & Chew, H. E. (2019). *"I'd blush if I could": Closing gender divides in digital skills through education*. Paris, France: UNESCO.

Whittaker, M., Crawford, K., Dobbe, R., Fried, G., Kaziunas, E., Mathur, V., Myers West, S., . . . & Schwartz, O. (2018). *AI Now report*. New York, NY: AI Now Institute.

Wiewiorówski, W. (2020). *A preliminary opinion on data protection and scientific research*. Brussels, Belgium: European Data Protection Supervisor.

Williams, B. A., Brooks, C. F., & Shmargad, Y. (2018). How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. *Journal of Information Policy, 8,* 78–115. doi:10.5325/jinfopoli.8.2018.0078

Woodland, P. C., Hain, T., Johnson, S. E., Niesler, T. R., Tuerk, A., & Young, S. J. (1998). Experiments in broadcast news transcription. *International Conference on Acoustics, Speech, and Signal Processing*. Piscataway, NJ: IEEE Signal Processing Society. doi:10.1109/ICASSP.1998.675413

Wronkiewicz, M. (2018, June 4). *Realistic expectations for applied machine learning*. Retrieved from https://medium.com/devseed/realistic-expectations-for-applied-machine-learning-a8171250db28

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, V. (2017). *Men also like shopping: Reducing gender bias amplification using corpus-level constraints*. Retrieved from https://arxiv.org/pdf/1707.09457.pdf