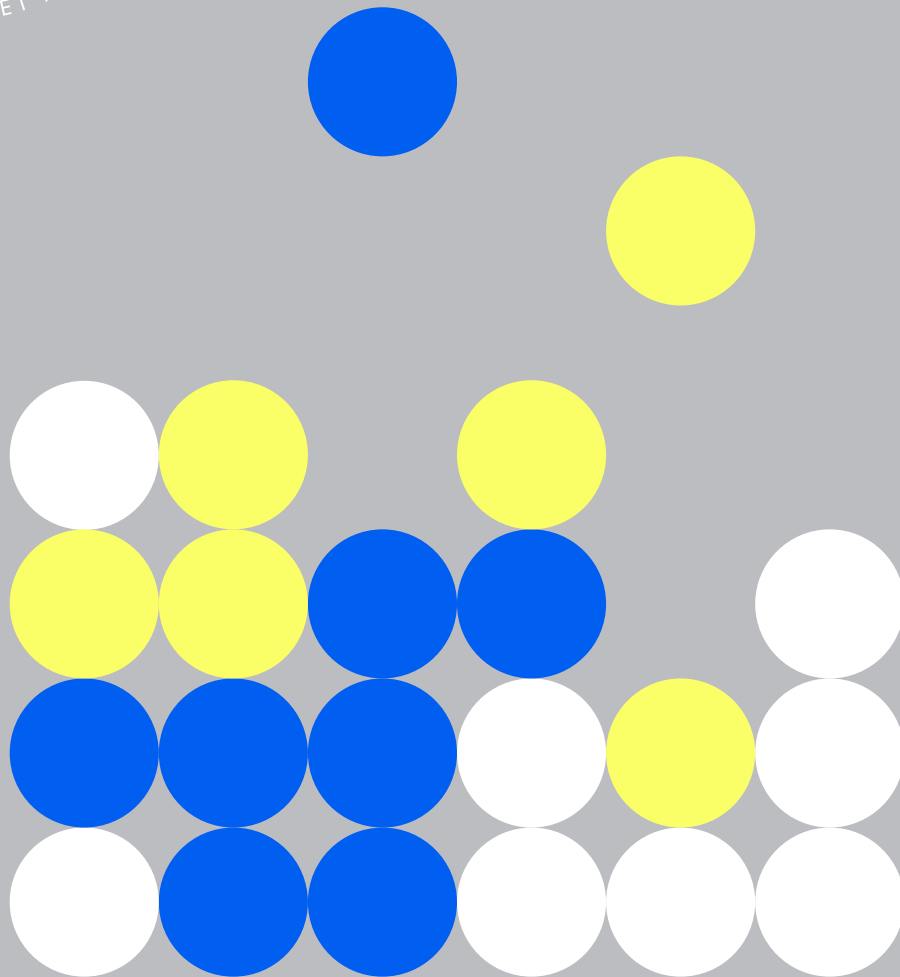




FUNDED BY THE EUROPEAN UNION'S
RIGHTS, EQUALITY AND CITIZENSHIP
PROGRAMME (2014-2020)



Gender and race biases in the AI assisted recruitment process

State of the art and future directions

Filippo Chiarello, Silvia Fareri e Miriam Crudelini



Disclaimer

The content of this Report *Gender and race biases in the AI assisted recruitment process: State of the art and future directions* represents the views of the authors only and are their sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.

The authors shared Report ideas and contents with the Steering Group and the Board of Expert of the Project GRASE that approved it, but they are the only responsible for the final content, expressions and words used.

Report realized within the GRASE Project Funded by the European Union's Rights, Equality and Citizenship Programme (2014 - 2020)

Fondazione Iniziative e Studi
sulla Multietnicità - ISMU
Via Copernico, 1 - 20125 Milano
Tel. 02.678779.1
www.ismu.org

Graphic Project

Alexandra Gredler
© Copyright Fondazione ISMU, Milano 2021

Suggested citation

Chiarello F., Fareri S., Crudelini M. (2021),
Gender and race biases in the AI assisted recruitment process: State of the art and future directions
ISMU, Milano.

ISBN: 9788831443197

Partner

This research was carried out as a part of the GRASE project, coordinated by ISMU Foundation (Italy) in partnership with Fondazione Giacomo Brodolini (Italy), the Adecco Group (Italy), Fundación Adecco (Spain), Asociación AMIGA por los Derechos Humanos de las Mujeres (Spain), APROXIMAR Cooperativa de Solidarieda de Social (Portugal).

Board of Expert

Emanuela Bonini (Coordinator),
Barbara De Micheli, Abel Guio, Tiago Leitão,
Nicola Pasini, Gemma Torre Ruiz, Sara Toticchi,
Cristina Zamora, Laura Zanfrini.

Research Team

The report was conducted in collaboration with the research group B4DS (<http://b4ds.unipi.it/>) of University of Pisa, with Prof. Filippo Chiarello and Miriam Crudelini.

Table of Contents

Executive summary	5
1. State of the art	8
1.1 Data Collection & Analytics	9
1.2 Bias in Traditional Recruiting Processes	12
1.3 Biases in Machine Learning	13
1.4 Gender & Ethnicity Biases in Digital Recruitment Processes	16
2. Adecco Data Description and Tool	18
References	20



Executive Summary

Scholars have dedicated a massive effort towards understanding gender equality, its determinants, its consequences for women and society, and the appropriate actions and policies to advance women's equality. Many topics have been covered, ranging from women's education and human capital (Deressa et al., 2009; Dumais, 2002) and their role in society (e.g., Kamin & Vezovnik, 2017; Kang et al., 2020), to their appointment in firms' top ranked positions (e.g., Smith & Parrotta, 2018; Smith et al., 2013) and performance implications (e.g., Adams & Ferreira, 2009; Campbell & Mínguez-Vera, 2008). Despite some attempts, extant literature reviews provide a narrow view on these issues, restricted to specific topics – e.g., female students' presence in STEM fields (Yazilias et al., 2013), educational gender inequality (Minasyan et al., 2019), the gender pay gap (Bishu & Alkadry, 2017), the glass ceiling effect (Jackson & O'Callaghan, 2009), leadership (Bark et al., 2014), entrepreneurship (Prashar et al., 2018), women's presence on the board of directors (Kirsch, 2018; Terjesen et al., 2009), diversity management (Köllen, 2019), gender stereotypes in advertisement (Grau & Zotos, 2016), or specific professions (Ahuja, 2002). Extant literature has also highlighted that gender issues, and their economic and social ramifications, are complex topics that involve a large number of possible antecedents and outcomes (Cuberes & Teignier, 2014). Indeed, gender equality actions are most effective when implemented in unison with other SDGs (e.g., with SDG 8, see Rai et al., 2019) in a synergetic perspective (Asadikia et al., 2020). Moreover, many bodies of literature (e.g., business, economics, development studies, sociology and psychology) all contribute their unique point of view to achieving gender equality. As extant research often tackles specific and narrow aspects, there is a lack of clarity on how different issues, circumstances and solutions may be related in precipitating or mitigating gender inequality or its effects.

The combination of ethnicity and gender discrimination are, undoubtedly, ever more recognized as a sensitive issue. Unfortunately, our society is widely affected by race and gender biases in multiple environments, especially the professional one (Snizek & Neil, 1993; Watts & Carter, 1991). This is a strong issue for economics and business, considering that hiring is the entrance gateway for women into the workforce. In this context, some literature exists, that is focused on gender only issue.

A first stream of literature focuses on the process leading up to candidates' job applications, demonstrating that bias exists before positions are even opened, and it is perpetuated both by men and women through networking and gatekeeping practices (e.g., Milkman et al., 2015; Van den Brink et al., 2010). The hiring process itself is also subject to biases (Hardy III et al., 2020), for example gender-congruity bias that leads to men being preferred candidates in male-dominated sectors (e.g., Koch et al., 2015), women being hired in positions with higher risk of failure (e.g., Haslam & Ryan, 2008) and limited transparency and accountability afforded by written processes and procedures (e.g., Van den Brink et al., 2010) that all contribute to ascriptive inequality. In addition, providing incentives for evaluators to hire women may actually work to this end; however, this is not the case when supporting female candidates endangers higher-ranking male ones (Lee & Waddell, 2021). Another interesting perspective, instead, looks at top management teams' composition and the effects on hiring practices, indicating that firms with more women in top management are less likely to lay off staff (e.g., Matsa & Miller, 2013).

At the same time, technological evolution is helping people and firms to be more efficient and objective in a huge variety of activities (Sung, 2018; Arnold et al., 2016; Porter

& Heppelmann, 2014). This is due to the emphasis given to new technologies, which has seen an increasing application of the latest tools provided by machine learning and A.I. Nevertheless, algorithms also suffer from intrinsic biases that are related to the datasets used to train them (Barocas & Selbst, 2016). The latter issue is quite complex to be managed and could also influence the process of decision making when, for instance, the technology is used to support the recruitment process (Kochling et al., 2021).

Therefore, the topic discussed in this report concerns biases existing, intentionally or not, thus influencing final HR decisions, with a particular focus on gender biases, race biases and the conjunction of the two. As well as other business functions, even in human resources (HR) there was the introduction of new technology with the aim of facilitating and improving their work (Fareri et al., 2020). What is not comforting is that gender and ethnicities biases occurred in the traditional recruiting process, but they are still detectable when applying machine learning algorithms. This theme is not yet much exploited in literature and, for this reason, it was approached gradually.

Starting from a general overview about biases in the recruitment process, why and how they exist, the study goes into detail investigating gender and ethnic bias in machine learning. Then, the study explores how these two kinds of biases negatively influence the digitized recruitment process, presenting a framework of the main criticalities detected.

Finally, considering the analysed literature, the authors propose an approach to measure race and gender biases, on a tool developed by Adecco, that supports the automatic pre-screening of candidates.

1. State of the art

1. State of the art

In this section, the authors propose a deep literature review on the gender and ethnicity biases related to the use of artificial intelligence, with a specific focus on recruitment activities.

To get a complete overview and to capture the changes over time, it was necessary to start from the evaluation of traditional processes, which do not make use of machine learning algorithms. The biases documented in literature, voluntarily or not, were expected to be related to prejudices against candidates. After that, the authors investigated the field of artificial intelligence, focusing on presenting the biases that can be detected in the use of these algorithms, to understand why they still exist and how to improve them. Finally, the main purpose of the paper was presented, analyzing if gender and ethnicity biases in recruitments are detectable today.

We are aware of the fact that gender and ethnicity biases, if they occur together, are stronger than the sum of the two. For these reasons, the focus of the present document is on works that study the phenomena together. Anyway, as will be evident from the present section, it is hard to find works that focus on the two problems conjunctively. For these reasons, to give a wide view on the problem, we decided to analyse the two bases separately.

The paragraph is structured as follows: firstly, the methodology used to collect relevant papers is described (paragraph 2.1); secondly, an analysis of the biases that occur in the traditional recruitment process are presented (paragraph 2.2). Thus, the gender and ethnicity biases that could come from the use of algorithms are described, together with their principal causes (paragraph 2.3). Finally, the authors report the main issues concerning gender and ethnicity biases that are observed in the digitized recruitment process (paragraph 2.4).

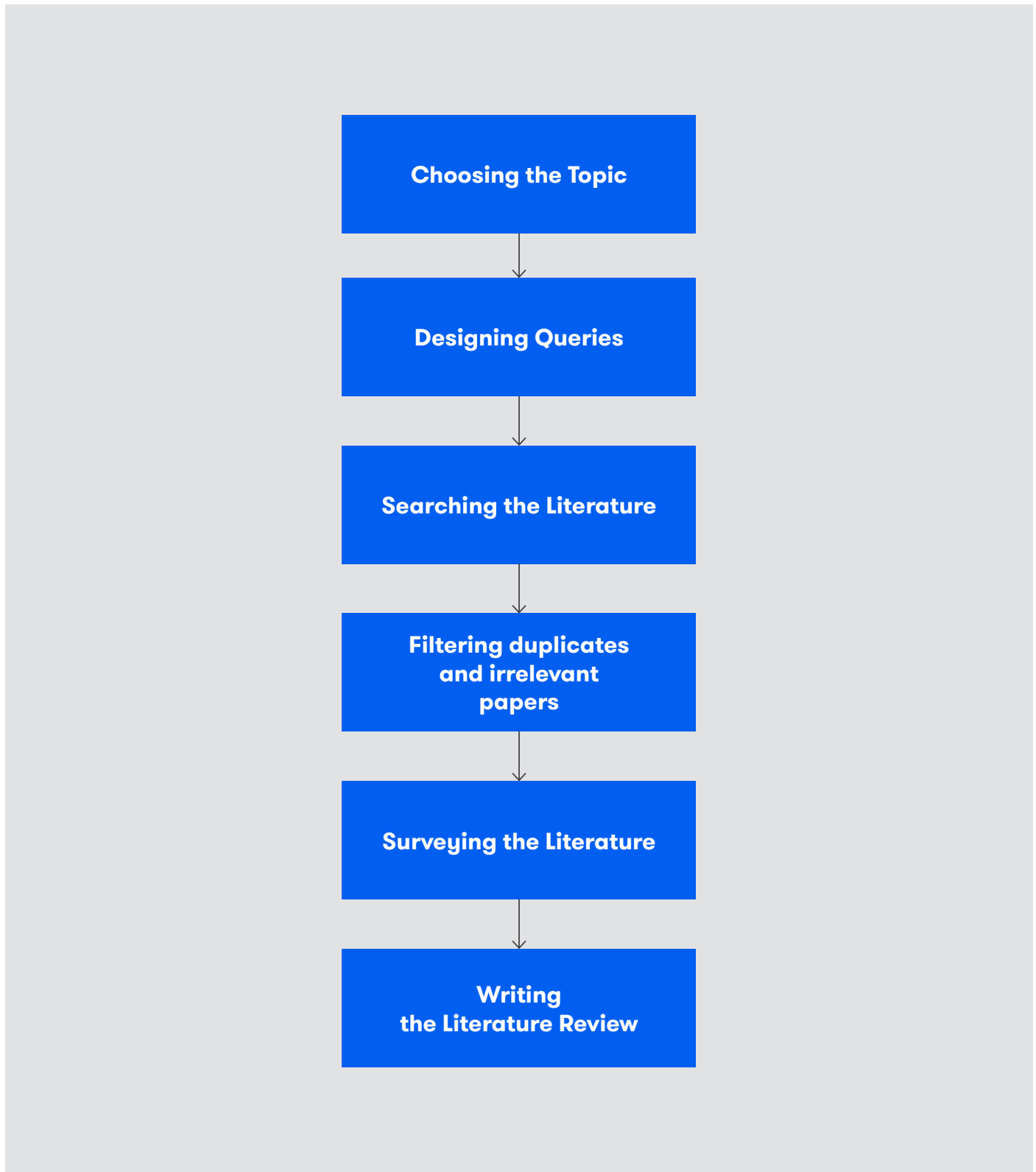
1.1 Data Collection & Analytics

In order to gain an overview of the main biases related to artificial intelligence in the context of the recruitment process, the authors collected all the scientific papers that addressed this issue. To reach the goal, we followed the workflow represented in Figure 1, which is partially based on Machi et al. (2012).

Once the topic was chosen, the authors

searched for pertinent publications on scientific databases, designing elaborated queries to capture the most valuable papers. Once detected, we filtered duplicates and irrelevant results. After that, we surveyed the literature, identifying the major studies that have been published to date, and how they were related to each other. Finally, we wrote the literature review.

Figure 1. Methodology applied to conduct the literature review



1. State of the art

In more detail, the research was performed using Scopus, designing four research queries focused on gender and ethnicity biases when applying AI (first and second query), and

the ones that appear specifically in digitized recruiting processes (third and fourth query).

The queries are reported in **Table 1**.

Table 1. Queries designed to detect relevant papers in Scopus

Topic	Query	Occurrence
Gender bias in data analysis	TITLE-ABS-KEY (("gender" OR "woman") AND ("discriminative" OR "discriminatory" OR "bias") AND ("artificial intelligence" OR "algorithm" OR "data analysis" OR "machine learning")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "DECI"))	475
Ethnicity biases in data analysis	TITLE-ABS-KEY (("race" OR "racial" OR "ethnicity") AND ("discriminative" OR "discriminatory" OR "bias") AND ("algorithm" OR "data analysis" OR "machine learning")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "DECI"))	219
Gender bias in digitized recruitment process	TITLE-ABS-KEY (("gender" OR "female" OR "woman") AND ("recruitment" OR "employment" OR "recruiting" OR "hiring" OR "job offer" OR "job vacancies") AND ("artificial intelligence" OR "data science" OR "chatbot" OR "chat-bot" OR "conversational bot")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "SOC") OR LIMIT-TO (SUBJAREA, "BUS") OR LIMIT-TO (SUBJAREA, "ENG"))	64
Ethnicity bias in digitized recruitment process	TITLE-ABS-KEY (("race" OR "ethnicity") AND ("recruitment" OR "employment" OR "recruiting" OR "hiring" OR "job offer" OR "job vacancies") AND ("artificial intelligence" OR "data science" OR "chatbot" OR "chat-bot" OR "conversational bot")) AND (LIMIT-TO (SUBJAREA, "COMP") OR LIMIT-TO (SUBJAREA, "SOC") OR LIMIT-TO (SUBJAREA, "BUS") OR LIMIT-TO (SUBJAREA, "ENG"))	24

To obtain a sample of documents focused on the theme of gender biases in artificial intelligence, we kept only the documents that cite, in their title, abstract or keywords gender (or woman), terms referring to prejudices and equality issues (discrimination, discriminatory, bias) and words related to artificial intelligence (artificial intelligence, algorithm, data analysis, machine learning). Finally, the documents were limited to the computer science and decision science area, to delete those documents that created noise. The full query led to 475 papers.

In a similar way, we built a query concerning the biases of ethnicity. We then considered documents containing, in the title abstract or keywords, terms concerning ethnicity

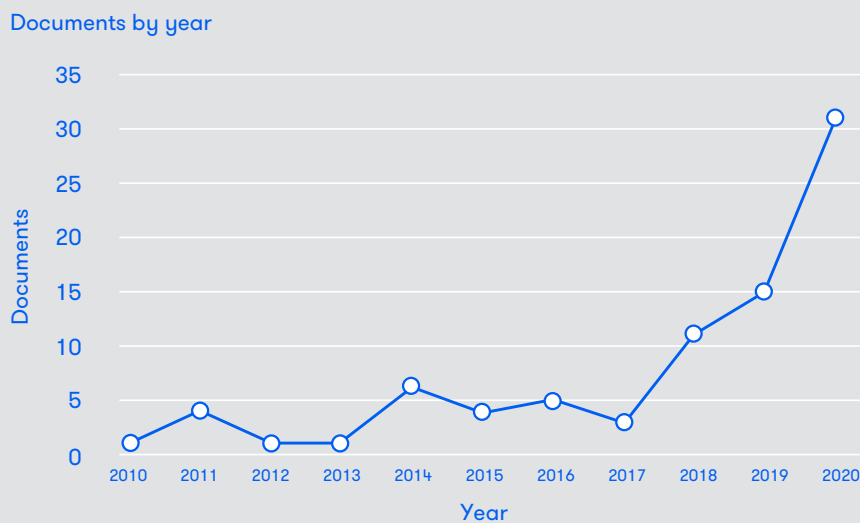
(race, racial), terms related to prejudice (discriminatory, discriminatory, bias) and, finally, words related to data (algorithm, data analysis, artificial intelligence, machine learning). Also in this case the sample was reduced by limiting the results to the computer science and decision science areas, obtaining 219 documents.

Concerning biases in the recruitment process, queries were constructed by including documents that contain in the title-abstract-keywords, terms such as recruitment, employment or other synonyms, while it was not necessary to insert terms related to differences, since the set of documents obtained was already focused on the subject of interest. In particular, in the two queries were inserted terms that allowed to

1. State of the art

infer research on gender biases (such as gender, woman, female) and ethnicity biases (race, ethnicity) separately. In addition, documents were limited to computer science, social sciences, business management and accounting, engineering. The two queries led

to 64 documents for gender bias and 24 for ethnicity. We also represented the cumulative trend of interest over time for the two topics. As we can infer from Figure 2, the attention to the topic is constantly increasing, especially in the last five years.



1.2 Bias in Traditional Recruiting Processes

This section presents the biases that can occur in a recruitment process where machine learning algorithms are not used.

The job interview is one of the most used selection methods, so that it is rare to be hired without any kind of interview (Huffcutt et al., 2010). Despite decades of research, questioning the validity of the interview and the reliability of interviewers, most organizations still include some type of interview in their selection process (Judge et al., 2000). In fact, there are many aspects that could influence the assessment of a candidate

during an interview, such as appearance, gender, ethnicity or age (Lepri et al. 2018; Levashina et al., 2014).

Some research has investigated a number of different aspects of the interview in an attempt to elucidate the reasons behind their continued use. The candidate's appearance, face, clothing and way of presenting affect the formation of a prejudice that then affects the final choice (Pingatore et al., 1994). Eye contact, and body language in general, also influence the interviewer's assessments (Dipboye, 1992).

1. State of the art

Schmid Mast et al. (2011) examine the importance of the first impressions of candidates on recruiters. In their study they test the accuracy of the interviewers in evaluating the candidates, comparing these assessments with those made by students, therefore without experience. The observed result is that professionals have been able to assess candidates' profiles more accurately. The effects of the first impression dominate the interviewer's judgments of the candidate; on average in the first 4 minutes of the interview the recruiter has already reached the final decision (Dougherty et al., 1994). Moreover,

when the candidates are similar in appearance to the interviewer, for example in terms of ethnicity and gender, the assessment is generally higher (Judge et al., 2000).

Job interviews are an important area of research where it is possible and necessary to improve decision-making. The experience of the recruiter can be a disadvantage when the prejudice on a candidate particularly affects the final decision (Levashina et al. 2014). An improvement in this field, in addition to providing benefits in terms of time and costs, can make hiring a fairer process.

1.3 Biases in Machine Learning

In the present paragraph, Gender and Ethnicities biases in machine learning are described. Firstly, the authors propose an overview of the main criticalities concerning gender discriminations; secondly, we define the ethnicities' biases retrievable from literature and the critical factors from which they could derive.

a. Gender Biases in Machine Learning

Gender discrimination is a widely analysed problem, which affect different countries and cultures and it is frequently treated in literature (Alatrasta-Salas et al., 2017). Gender biases are also foundable in machine learning processes as confirmed by the studies (Kamiran et al., 2009; Kamiran et al., 2012), and they seem to be closely related to the algorithms (Feldman et al., 2015) and the datasets used to train them (Lee et al. 2021).

Sweeney (2013) and Datta et al. (2015), have studied the behavior of online advertising and Google Ads compared to gender, confirming a discrimination in advertising publication. In particular, both studies confirm

that jobs with a higher salary are shown in a lower percentage to a female audience. Therefore, recommendation systems targeting employment-related advertisements were found to demonstrate gender bias (Lambrecht et al., 2019). Similar studies have been done in several other areas, such as vision and language models (Zhao et al., 2017), online news (Ross et al., 2011) and web research (Kay et al., 2015).

The biases could also be found in different kinds of technologies applied. For instance, Adams and Ni Loideain (2019) studied personal voice assistants and criticized the reproduction of negative gender stereotypes in virtual personal assistants. This research also examines the provisions and results of the International Law on Women's Rights to assess how this constitutes indirect discrimination. In addition, gender bias is also present in image search results, in relation to professions.

Kay et al. (2015), analyzes the effects of prejudice in this type of research, focusing on the choice of images that are made to represent certain professions and on people's

1. State of the art

perception of the prevalence of men and women in every occupation. The observed result is both the exaggeration of stereotypes, and the under-representation of women in research results (Kay et al., 2015).

Other studies focus not so much on the different ways in which such gender biases can be observed, but more on machine learning algorithms that reproduce stereotypes and amplify them (Bolukbasi et al. 2016). In 2016, Bolukbasi et al. study changed the embedding of gender-neutral words by removing their gender associations in order to reduce biases. The latter process allowed to reduce the stereotyped associations between gender and types of employment (Bolukbasi et al., 2016).

However, studies have shown that implicit gender bias persists despite these de-biasing methods (Gonen et al., 2019). The modification of training corpora prior to learning of gender bias has been explored through the provision of training data where the gender of entities in the corpora are swapped, and has been proven to reduce gender bias in predictions (Zhao et al., 2018).

In 2018, Leavy showed that identifying gender bias in the data set from which algorithms are learned, is a concept that can be captured in language characteristics. Following word-embedding approaches, it has been observed that machine learning methods, used in a framework informed by feminist linguistics and gender theory, can be used to assess levels of gender bias within natural language training bodies (Leavy et al., 2020). In this case, 19th-century texts and a decade of articles by The Guardian have been studied. Word-embedding was then applied to this corpus which led to the results.

Other researches focused on different methods to reduce gender bias, for instance, in different

types of classification (Kamishima et al. (2012, 2011)) and in predictions (Zhao et al., 2017). In Zaho et al., 2018, they suggest generic methods to reduce prejudices, while other approaches are based on equi-classification models (Calders et al., 2009; Feldman et al., 2015; Misra et al., 2016). Concerning data, machine learning algorithms used in A.I., exploit data to build models capable of assessing the labels and properties of novel data.

Unfortunately, the available training data frequently contains biases with respect to things that we would rather not use for decision making. Machine learning builds models faithful to training data and can lead to perpetuating these unwanted prejudices (Zhang et al., 2018). In Torralba et al. (2011) the authors show that the quality of the dataset on which the models are built is important to ensure that they are able to represent reality.

In addition, biases in this field are closely related to who designed the algorithms, reflecting a working context in which the female presence is limited (West et al., 2019). To conclude, the data analysis process therefore reports imbalances that depend mainly on the learning algorithms used and the often-unbalanced data sets on which the analysis is carried out (W Flores, Bechtel, and Lowenkamp 2016). In this context, therefore, it is essential to update and rethink some models that are the basis of applications used for decision-making processes (Barocas and Selbst 2016). Similarly, it is essential to try to have a gender balance between data scientists that process data, so as to avoid any imbalance at any stage of analysis.

b. Ethnicity Biases in Machine Learning

Even if ethnicity discrimination is ever more recognized as a sensitive issue, the topic of race biases has (unfortunately)

1. State of the art

not been solved yet. Furthermore, the world of Artificial Intelligence is not exempt from it and algorithms frequently suffer from ethnicity biases (Hajian et al., 2016). Even if it is essential avoiding these algorithms to be discriminative, several kinds of race biased A.I. have been detected in the past.

In 2013, Lee (2013) reported that search results of “black sounding names” were more likely to be linked to arrests (even when false), while Google automatically labelled as “gorillas” two Africans after a word search (Kasperkevic, 2015). Also Baker & Potts (2013) highlighted the biases behind the auto-complete search algorithm offered by Google; the latter was interrogated asking “why are blacks...” with the aim of eliciting auto-completed questions. The system provided 2690 questions that were categorized and analyzed. At the end of the process, several ethnicity groups seemed to attract particular stereotypes (Muslim were linked to appearance, white people to sexual attitudes, while homosexuals and black people to negative stereotypes). Also health systems, which frequently rely on A.I. predictions to identify patients, seem to be affected. Obermeyer et al. (2019) showed that algorithms exhibited significant racial biases, since they predicted health care costs rather than illness. The unequal access to health for Black patients made racial biases arise, demonstrating that health costs could not be the only proxy to be considered. Above all, one of the most affected processes is represented by Face Recognition, which is an easy task for humans (Loo et al., 2018) but quite difficult for machines (Alhindi et al., 2018). The latter training process was found to light the skin of African Users, since the European face color was recognized as a standard of beauty for the machine (Morse, 2017).

Generally, existing public face image datasets are strongly biased toward Caucasian

faces, while other ethnicities are definitely underrepresented. For this reason, the models that are trained from such datasets are not accurate when the technology is applied to a face belonging to a non-White race groups.

To mitigate the problem, Karkkainen & Joo (2021) built a new image dataset containing more than 100.000 images equally distributed on race. In this way, they found that the model was substantially more precise and its accuracy significantly increased. Similarly, RamFace, a race adaptive margin based face recognition model was trained to learn the racial features and improve the discriminability of the extracted features. Following the previous methodological framework, it was possible to increase face recognition accuracy and to mitigate race biases (Yang et al., 2021). Also Muhammad et al. (2021) designed a way to avoid the risk of ethnicity biases, collecting a face image database with 38546 images of African subjects made by multispectral cameras to capture expressions under various illumination settings. The previous processes significantly decreased biases.

In the end, what seem evident is that race bias could be detected even if there is no discrimination intent in the developer of the algorithm; the drawback is often related to the data sources used (Barocas & Selbst, 2016), even when the sensitive attributes have been suppressed from the input (Hajian et al., 2016). In fact, training data could be affected by unfairness for reasons related to historical prejudices or other factors that are outside an individual’s control (Kusner et al., 2017).

Since those algorithms are widely used, also during recruitment, it is definitely necessary to avoid under-representation concerning ethnicity in the training data set as biases occur if the underlying training data set is unbalanced (Kochling et al., 2021).

1.4 Gender & Ethnicity Biases in Digital Recruitment Processes

Over the past decade, artificial intelligence has strongly influenced society and the way we live. Moreover, it had an important impact on firms across different business functions. Also HR management could undoubtedly take advantage of it (Fareri et al., 2020), in particular in the field of recruitment, where A.I. offers a high potential, both in terms of efficiency increasing and augmentation of skills knowledge bases (Fareri et al., 2021).

It is therefore in the interests of companies to use algorithms that allow them to make objective and fair decisions, thus avoiding being influenced by subjective and discriminatory prejudices (Langer et al., 2019; Persson 2016).

For HR management, the evaluation of application and the design of new vacancies is still a time consuming activity (Leicht-Deobald et al., 2019); therefore, there is a significant difficulty in evaluating each applicant with the same attention and objectivity (Wilson & Daugherty, 2018). Using algorithms, the processes could be almost totally automated and the interviews could also be conducted asynchronously through video recording and automatic evaluation (Dahm & Dregger 2019; Lee & Baykal 2017; Brenner et al., 2016).

However, it is still an area of improvement because, as studies have shown (Pedreschi et al., 2008; Hajian et al., 2013; Mehrabi et al., 2019), the algorithms themselves can be discriminatory with respect to gender, ethnicity and civil status.

These biases exist even when there is no real discriminatory intention in the development of the same: the data sources used can influence, software that as a result amplify certain historical discrimination in the data,

or a trained algorithm can discriminate based on sensitive attributes due to the correlations between the data themselves (Caliskan et al., 2017). Biases can sometimes occur when an algorithm is trained on unbalanced datasets, which therefore do not represent a population well enough, and are then incorporated into the design of the algorithm (Zhang et al., 2018). The main criticalities concerning the use of A.I. in the recruiting process are summarized in **Table 2**.

The designed framework allows the visualization of the “contact points” between scouting steps, biases and criticalities. A.I. is mainly used to automate actions that are intentionally augmented in bold in the text: the search of occupations, the ranking/scoring/filtering of candidates and the sharing of job posts are the key actions detected. What seems evident from the state-of-art analysis is that an underrepresentation of specific categories (women and ethnicities) in training datasets, involves underestimating scores of interviews and underrepresenting them in search results.

Moreover, women are less likely to visualize job postings related to high paid jobs, which is unfortunately in line with the persistent issue of the gender pay gap (Blau & Khan, 2007), and probably helps to sharpen it. Finally, technological innovations also give the opportunity of filtering data in relation to specific requirements or desires, even if they are driven by racism or misogyny. Frequently, the injustices that could be detected in traditional recruitment, could also be found in automated processes, since filtering instruction could be established to discard candidates in relation to their gender or the colour of their skin. Once again, educating people to equality is considerably more necessary than training machines correctly.

1. State of the art

Table 2. Literature Framework on gender and ethnicities biases in digital recruiting processes. Biases are represented with two colours: grey for ethnicities biases (ETHN) and blu for gender biases (GEN). If both the biases are retrieved from the paper, the cell is filled with yellow.

Scouting Step	Algorithmic Process	Criticality	Authors	Bias
Step 1 / Occupations or Talents search	Search for occupations on Google image	Underrepresentation of Black people in search results for common occupations (e.g., searching for “engineer” or “author”) in Google	(Metaxa et al., 2020)	ETHN
		Underrepresentation of women in search results for common occupations (e.g., searching for “engineer” or “author”) in Google	(Kay, 2015)	GEN
	Automatic ranking of candidates	Reduced visibility for women	(Geyik et al., 2019)	GEN
Step 2 / Job Posting	Automatic sharing of Job advertisement	Advertisements for a high- paying executive position shown six times more to men than to women	(Amit et al., 2015)	GEN
	Automatic filtering of Job advertisement	Targeting of Facebook used to exclude specific candidates	(Dalenberg, 2017)	ETHN
Step 3 / Job Interview	Automatic scoring based on a recorded interview	Underestimation of job interview scores due to underrepresentation of ethnicities in the data sets	(Kochling et al., 2021)	ETHN
	Automatic filtering of Job candidates	Sex and Race easily detectable from video and used to filter candidates by appearance reasons	(Fernandéz Martínez & Fernandéz, 2020)	BOTH

2. Adecco Data Description **and Tool**

2. Adecco Data Description and Tool

In general, we can consider four distinct stages of the recruiting process: sourcing, screening, interviewing, and selection. The Adecco Group tools, focuses on screening, and in particular, pre-employment assessments that evaluate candidates.

The aim of this case study is to target the cognitive and structural bias that might be associated with the design and deployment of a software-assisted hiring system. This case study will show how to use an automatic tool to identify gender and race biases on a dataset analysis deriving from a specific software used at the Adecco Group for the outsourcing of candidates. In this regard, the software in use at the Adecco Group covers the recruiting process from the collection of job applications to the screening.

Through the software, it is possible to access a dataset in the form of aggregated data reflecting data such as: age, gender, geographic location, experience, job for which the candidate applied and status of the application. The provided data are shared as aggregated and anonymous, in the respect of the privacy obligation of Adecco and the rights of the candidates.

The analysis of the dataset will make it possible to detect and assess possible bias put in place by the software in the selection of candidates. The next table describes some characteristics of the data that will be studied in the case study.

Table 3: Data volume, type of available data and possible biases

Data volume	10.000
Type of data	Text file, Excel, Csv
Covered countries	Italy
Possible biases	Unbalanced dataset in favour of the election of a specific gender, race thus against a selected group of people or minorities
Validated dimensions	Data bias

The tool used has the aim of automatically detecting gender and ethnicity biases in digitized recruiting processes. To achieve the goal, the analysis of the datasets is conducted to evaluate the possible underrepresentation of specific categories of candidates (such as Women and Ethnicities different from the Caucasian).

References

- Ahuja, M. K.** (2002). Women in the information technology profession: A literature review, synthesis and research agenda. *European Journal of Information Systems*, 11(1), 20-34.
- Alatrasta-Salas, Hugo, Bruno Esposito, Miguel Nunez-del-Prado, and Maria Valdivieso.** 2017. "Measuring the Gender Discrimination: A Machine Learning Approach." In 2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI), 1-6. Arequipa: IEEE. <https://doi.org/10.1109/LA-CCI.2017.8285682>.
- Alhindi, T.J.; Kalra, S.; Ng, K.H.; Afrin, A.; Tizhoosh, H.R.** Comparing LBP, HOG and Deep Features for Classification of Histopathology Images. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, 8-13 July 2018; pp. 1-7.
- Amit Datta, Michael Carl Tschantz and Anupam Datta,** 'Auto- mated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination' (2015) 2015 Proceedings on Privacy Enhanc- ing Technologies 92
- Arnold, C., Kiel, D., Voigt, K.I.,** 2016. How the industrial internet of things changes business models in different manufacturing industries. *Int. J. Innov. Manag.* 20 (8).
- Asadikia, A., Rajabifard, A., & Kalantari, M.** (2020). Systematic prioritisation of SDGs: Machine learning approach. *World Development*.
- Baker, P., Potts, A.** 'Why do white people have thin lips?' Google and the perpetuation of stereotypes via auto-complete search forms (2013) *Critical Discourse Studies*, 10 (2), pp. 187-204. Cited 51 times. DOI: 10.1080/17405904.2012.744320
- Barocas S, Selbst AD** (2016) Big data's disparate impact. *Calif Law. Rev* 104:671
- Bark, A. S. H., Escartín, J., & van Dick, R.** (2014). Gender and leadership in Spain: A systematic review of some key aspects. *Sex Roles*, 70(11-12), 522-537.
- Beigman, Eyal, and Beata Beigman Klebanov.** 2009. "Learning with Annotation Noise." In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1 - ACL-IJCNLP '09, 1:280. Suntec, Singapore: Association for Computational Linguistics. <https://doi.org/10.3115/1687878.1687919>.
- Bishu, S. G., & Alkadry, M. G.** (2017). A systematic review of the gender pay gap and factors that predict it. *Administration & Society*, 49(1), 65-104.
- Blau, F.D., Kahn, L.M.** The gender pay gap: Have women gone as far as they can? (2007) *Academy of Management Perspectives*, 21 (1), pp. 7-23. DOI: 10.5465/AMP.2007.24286161
- Bolukbasi, Tolga, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai.** 2016. "Man Is to Computer Programmer as Woman Is to Homemaker? Debiasing Word Embeddings." In *Advances in Neural Information Processing Systems*. Vol. 29. Curran Associates, Inc.
- Brenner FS, Ortner TM, Fay D** (2016) Asynchronous video interviewing as a new technology in personnel selection: the applicant's point of view. *Front Psychol* 7:863
- Brian Hu, Blake Lemoine, and Margaret Mitchell.** 2018. "Mitigating Unwanted Biases with Adversarial Learning." In Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 335-40. New Orleans LA USA: ACM. <https://doi.org/10.1145/3278721.3278779>.

- Campbell, K., & Mínguez-Vera, A.** (2008). Gender diversity in the boardroom and firm financial performance. *Journal of Business Ethics*, 83(3), 435-451.
- Caliskan, Aylin, Joanna J. Bryson, and Arvind Narayanan.** 2017. "Semantics Derived Automatically from Language Corpora Contain Human-like Biases." *Science* 356 (6334): 183-86. <https://doi.org/10.1126/science.aal4230>.
- Coe, J., Atay, M.** Evaluating impact of race in facial recognition across machine learning and deep learning algorithms (2021) *Computers*, 10 (9), art. no. 113, . DOI: 10.3390/computers10090113
- Cuberes, D., & Teignier, M.** (2014). Gender inequality and economic growth: A critical review. *Journal of International Development*, 26(2), 260-276.
- Dahm M., Dregger A.** (2019) Der Einsatz von künstlicher Intelligenz im HR: Die Wirkung und Förderung der Akzeptanz von KI-basierten Recruiting-Tools bei potenziellen Nutzern. *Arbeitswelten der Zukunft*. Springer, Heidelberg, pp 249-271
- Dalenberg, David Jacobus** (2017). Preventing discrimination in the automated targeting of job advertisements. *Computer Law & Security Review*, doi:10.1016/j.clsr.2017.11.009
- Deressa, T. T., Hassan, R. M., Ringler, C., Alemu, T., & Yesuf, M.** (2009). Determinants of farmers' choice of adaptation methods to climate change in the Nile Basin of Ethiopia. *Global Environmental Change*, 19(2), 248-255.
- Dipboye, R. L.** (1992). *Selection interviews: Process perspectives*. Cincinnati, OH: SouthWestern.
- Dumais, S. A.** (2002). Cultural capital, gender, and school success: The role of habitus. *Sociology of Education*, 75(1), 44-68.
- Fareri S., Melluso N., Chiarello F., Fantoni G.,** (2021). SkillNER: Mining and mapping soft skills from any text. *Expert Systems with Applications*, Volume 184, 115544, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2021.115544>.
- Fareri S., Fantoni G., Chiarello F., Coli E., Binda A.,** (2020). Estimating Industry 4.0 impact on job profiles and skills using text mining. *Computers in Industry*, Volume 118, 103222, ISSN 0166-3615, <https://doi.org/10.1016/j.compind.2020.103222>.
- Fernández-Martínez, Carmen, and Alberto Fernández.** 2020. "AI and Recruiting Software: Ethical and Legal Implications." *Paladyn, Journal of Behavioral Robotics* 11 (1): 199-216. <https://doi.org/10.1515/pjbr-2020-0030>.
- Feldman, Michael, Sorelle Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian.** 2015. "Certifying and Removing Disparate Impact." *ArXiv:1412.3756 [Cs, Stat]*, July. <http://arxiv.org/abs/1412.3756>.
- Friedman B, Nissenbaum H.** Bias in computer systems. *ACM T Inform Syst* 1996; 14(3): 330-347. Geyik, Sahin Cem; Ambler, Stuart; Kenthapadi, Krishnaram (2019). [ACM Press the 25th ACM SIGKDD International Conference - Anchorage, AK, USA (2019.08.04-2019.08.08)] *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining - KDD '19 - Fairness-Aware Ranking in Search & Recommendation Systems with Application to LinkedIn Talent Search*. doi:10.1145/3292500.3330691

- Gonen, Hila, and Yoav Goldberg.** 2019. "Lipstick on a Pig: Debiasing Methods Cover up Systematic Gender Biases in Word Embeddings But Do Not Remove Them." ArXiv:1903.03862 [Cs], September. <http://arxiv.org/abs/1903.03862>.
- Grau, S. L., & Zotos, Y. C.** (2016). Gender stereotypes in advertising: a review of current research. *International Journal of Advertising*, 35(5), 761-770.
- Hajian, Sara; Bonchi, Francesco; Castillo, Carlos (2016). [ACM Press the 22nd ACM SIGKDD International Conference - San Francisco, California, USA (2016.08.13-2016.08.17)] Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16 - Algorithmic Bias, [], 2125-2126.doi:10.1145/2939672.2945386
- Hajian, Sara, and Josep Domingo-Ferrer.** 2013. "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining." *IEEE Transactions on Knowledge and Data Engineering* 25 (7): 1445-59. <https://doi.org/10.1109/TKDE.2012.72>.
- Hardy III, J. H., Tey, K. S., Cyrus-Lai, W., Martell, R. F., Olstad, A., & Uhlmann, E. L.** (2020). Bias in context: Small biases in hiring evaluations have big consequences. *Journal of Management*.
- Haslam, S. A., & Ryan, M. K.** (2008). The road to the glass cliff: Differences in the perceived suitability of men and women for leadership positions in succeeding and failing organizations. *The Leadership Quarterly*, 19(5), 530-546.
- Huffcutt AI, Culbertson SS.** (2010). Interviews. In Zedeck S (Ed.), *APA handbook of industrial and organizational psychology* (pp. 185-203). Washington, DC: American Psychological Association.
- Jackson, J. F., & O'Callaghan, E. M.** (2009). What do we know about glass ceiling effects? A taxonomy and critical review to inform higher education research. *Research in Higher Education*, 50(5), 460-482.
- Judge, Timothy A, Daniel M Cable, and Chad A Higgins.** 2000. "The Employment Interview: A Review of Recent Research and Recommendations for Future Research." *Human Resource Management Review* 10 (4): 383-406. [https://doi.org/10.1016/S1053-4822\(00\)00033-4](https://doi.org/10.1016/S1053-4822(00)00033-4).
- Kamin, T., & Vezovnik, A.** (2017). Slovenia's socialist superwoman: feeding the family, nourishing the nation. *Feminist Review*, 117(1), 79-96.
- Kamiran, Faisal and Toon Calders.** "Classifying without discriminating." 2009 2nd International Conference on Computer, Control and Communication (2009): 1-6.
- Kamiran, Faisal, and Toon Calders.** 2012. "Data Preprocessing Techniques for Classification without Discrimination." *Knowledge and Information Systems* 33 (1): 1-33. <https://doi.org/10.1007/s10115-011-0463-8>.
- Kang, M., Park, H. J., & Park, J.** (2020). Teachers as good mothers, mothers as good teachers: Functional and ideological work-family alignment in the South Korean teaching profession. *Gender, Work and Organization*, 27(3), 395-413.
- Kay, Matthew, Cynthia Matuszek, and Sean A. Munson.** 2015. "Unequal Representation and Gender Stereotypes in Image Search Results for Occupations." In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, 3819-28. CHI '15. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2702123.2702520>.

- Karkkainen, K., Joo, J. FairFace:** Face attribute dataset for balanced race, gender, and age for bias measurement and mitigation (2021) Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021, pp. 1547-1557. DOI: 10.1109/WACV48630.2021.00159
- Kasperkevic, J.** (2015), "Google says sorry for racist auto-tag in photo app", The Guardian, available at: <https://goo.gl/ZEJYng> (accessed 5 March 2018).
- Kirsch, A.** (2018). The gender composition of corporate boards: A review and research agenda. *The Leadership Quarterly*, 29(2), 346-364.
- Koch, A. J., D'Mello, S. D., & Sackett, P. R.** (2015). A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *Journal of Applied Psychology*, 100(1), 128-161.
- Köchling, A., Riazzy, S., Wehner, M.C., Simbeck, K.** Highly Accurate, But Still Discriminatory: A Fairness Evaluation of Algorithmic Video Analysis in the Recruitment Context (2021) *Business and Information Systems Engineering*, 63 (1), pp. 39-54. Cited 1 time. DOI: 10.1007/s12599-020-00673-w
- Kusner, Matt & Loftus, Joshua & Russell, Chris & Silva, Ricardo.** (2017). Counterfactual Fairness. Lambrecht, Anja, and Catherine E. Tucker. 2016. "Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads." SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.2852260>.
- Leavy, Susan.** 2018. "Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning." In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering*, 14–16. Gothenburg Sweden: ACM. <https://doi.org/10.1145/3195570.3195580>.
- Leavy, Susan, Gerardine Meaney, Karen Wade, and Derek Greene.** 2020. "Mitigating Gender Bias in Machine Learning Data Sets." In *Bias and Social Aspects in Search and Recommendation*, edited by Ludovico Boratto, Stefano Faralli, Mirko Marras, and Giovanni Stilo, 1245:12–26. Communications in Computer and Information Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-52485-2_2.
- Lee, D.** (2013), "Google searches expose racial bias, says study of names", BBC, available at: <https://goo.gl/P8oodF> (accessed 5 March 2018).
- Lee, L. M., & Waddell, G. R.** (2021). Diversity and the timing of preference in hiring decisions. *Journal of Economic Behavior & Organization*, 184, 432-459.
- Lee MK., Baykal S.** (2017) Algorithmic mediation in group decisions: Fairness perceptions of algorithmically mediated vs. discussion-based social division. In: *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*. ACM, Portland, pp 1035-1048
- Lee, Woojin, Hyungjin Ko, Junyoung Byun, Taeho Yoon, and Jaewook Lee.** 2021. "Fair Clustering with Fair Correspondence Distribution." *Information Sciences* 581 (December): 155–78. <https://doi.org/10.1016/j.ins.2021.09.010>.
- Leicht-Deobald U., Busch T., Schank C., Weibel A., Schafheitle S., Wildhaber I., Kasper G.** (2019) The challenges of algorithm-based HR decision-making for personal integrity. *J Bus Ethics* 160(2):377–392

- Lepri, Bruno, Nuria Oliver, Emmanuel Letouzé, Alex Pentland, and Patrick Vinck.** 2018. "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges." *Philosophy & Technology* 31 (4): 611–27. <https://doi.org/10.1007/s13347-017-0279-x>.
- Levashina, Julia, Christopher J. Hartwell, Frederick P. Morgeson, and Michael A. Campion.** 2014. "The Structured Employment Interview: Narrative and Quantitative Review of the Research Literature." *Personnel Psychology* 67 (1): 241–93. <https://doi.org/10.1111/peps.12052>.
- Loo, E.K.; Lim, T.S.; Ong, L.Y.; Lim, C.H.** The Influence of Ethnicity in Facial Gender Estimation. In *Proceedings of the 2018 IEEE 14th International Colloquium on Signal Processing & Its Applications (CSPA)*, Penang, Malaysia, 9–10 March 2018; pp. 187–192.
- Machi, L. A., & McEvoy, B. T.** (2012). *The Literature Review: Six Steps to Success* (Second ed.). Corwin.
- Matsa, D. A., & Miller, A. R.** (2013). A female style in corporate leadership? Evidence from quotas. *American Economic Journal: Applied Economics*, 5(3), 136-169.
- Matthew Kay, Cynthia Matuszek, and Sean A. Munson.** 2015. Unequal Representation and Gender Stereotypes Image Search Results for Occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing. Systems - CHI '15*. ACM Press, Seoul, Republic of Korea, 3819–3828. <https://doi.org/10.1145/2702123.2702520>
- Mehrabi, Ninareh, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan.** 2019. "A Survey on Bias and Fairness in Machine Learning." *ArXiv:1908.09635 [Cs]*, September. <http://arxiv.org/abs/1908.09635>.
- Metaxa, D., Gan, M. A., Goh, S., Hancock, J., & Landay, J. A.** (2021). An Image of Society. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), 1–23. doi:10.1145/3449100
- Milkman, K. L., Akinola, M., & Chugh, D.** (2015). What happens before? A field experiment exploring how pay and representation differentially shape bias on the pathway into organizations. *Journal of Applied Psychology*, 100(6), 1678-1712.
- Minasyan, A., Zenker, J., Klasen, S., & Vollmer, S.** (2019). Educational gender gaps and economic growth: A systematic review and meta-regression analysis.
- Morse, J.** (2017), "App creator apologizes for 'racist' filter that lightens skin tones", Mashable, available at: <https://mashable.com/2017/04/24/faceapp-racism-selfie/#zeUltoQB5iqI> [accessed 5 March 2018].
- Muhammad, J., Wang, Y., Wang, C., Zhang, K., Sun, Z.** CASIA-Face-Africa: A Large-Scale African Face Image Database (2021) *IEEE Transactions on Information Forensics and Security*, 16, art. no. 9456939, pp. 3634-3646. DOI: 10.1109/TIFS.2021.3080496
- Obermeyer, Z., Powers, B., Vogeli, C., Mullainathan, S.** Dissecting racial bias in an algorithm used to manage the health of populations (2019) *Science*, 366 (6464), pp. 447-453. Cited 482 times. DOI: 10.1126/science.aax2342

- Prashar, S., Vijay, T. S., & Parsad, C.** (2018). Women entrepreneurship in India: a review of barriers and motivational factors. *International Journal of Entrepreneurship and Innovation Management*, 22(3), 206-219.
- Pedreschi, Dino, Salvatore Ruggieri, and Franco Turini.** n.d. "Discrimination-Aware Data Mining," 9.
- Pingatore, R., Dugoni, B. L., Tindale, R. S., & Spring, B.** (1994). Bias against overweight job applicants in a simulated employment interview. *Journal of Applied Psychology*, 79, 909±917.
- Porter, M.E., Heppelmann, J.E., 2014. How smart, connected products are transforming competition. *Harv. Bus. Rev.* 92 [11], 64–88.
- Rai, S. M., Brown, B. D., & Ruwanpura, K. N.** (2019). SDG 8: Decent work and economic growth-A gendered analysis. *World Development*, 113, 368-380.
- Ross, Karen, and Cynthia Carter.** 2011. "Women and News: A Long and Winding Road." *Media, Culture & Society* 33 (8): 1148–65. <https://doi.org/10.1177/0163443711418272>.
- Schmid Mast, Marianne, Adrian Bangerter, Céline Bulliard, and Gaëlle Aerni.** 2011. "How Accurate Are Recruiters' First Impressions of Applicants in Employment Interviews?: Recruiters and Applicant Assessment." *International Journal of Selection and Assessment* 19 (2): 198–208. <https://doi.org/10.1111/j.1468-2389.2011.00547.x>.
- Smith, N., & Parrotta, P.** (2018). Why so few women on boards of directors? Empirical evidence from danish companies in 1998–2010. *Journal of Business Ethics*, 147(2), 445-467.
- Smith, N., Smith, V., & Verner, M.** (2013). Why are so few females promoted into CEO and vice president positions? Danish empirical evidence, 1997–2007. *ILR Review*, 66(2), 380-408.
- Snizek, W.E., Neil, C.C.** Job Characteristics, Gender Stereotypes and Perceived Gender Discrimination in the Workplace (1992) *Organization Studies*, 13 (3), pp. 403-427. DOI: 10.1177/017084069201300305
- Sung, T. K.,** 2018. Industry 4.0: A Korea perspective. *Technological Forecasting and Social Change*, 132, 40-45. doi:10.1016/j.techfore.2017.11.005
- Sweeney, Latanya.** 2013. "Discrimination in Online Ad Delivery." ArXiv:1301.6822 [Cs], January. <http://arxiv.org/abs/1301.6822>.
- Terjesen, S., Sealy, R., & Singh, V.** (2009). Women directors on corporate boards: A review and research agenda. *Corporate Governance: An International Review*, 17(3), 320-337.
- Torralba, Antonio, and Alexei A. Efros. 2011. "Unbiased Look at Dataset Bias." In *CVPR 2011*, 1521–28. Colorado Springs, CO, USA: IEEE. <https://doi.org/10.1109/CVPR.2011.5995347>.
- Van den Brink, M., Benschop, Y., & Jansen, W.** (2010). Transparency in academic recruitment: a problematic tool for gender equality? *Organization Studies*, 31(11), 1459-1483
- Watts, R.J., Carter, R.T. *Psychological Aspects of Racism in Organizations* (1991) *Group & Organization Management*, 16 (3), pp. 328-344. 10.1177/105960119101600307
- Wilson HJ, Daugherty PR** (2018) Collaborative intelligence: humans and AI are joining forces. *Harvard Bus Rev* 96(4):114–123

- Yang, Z., Zhu, X., Jiang, C., Liu, W., Shen, L.** RamFace: Race Adaptive Margin Based Face Recognition for Racial Bias Mitigation. (2021) 2021 IEEE International Joint Conference on Biometrics, IJCB 2021, art. no. 9484352, DOI: 10.1109/IJCB52358.2021.9484352
- Zhang, Yazilitas, D., Svensson, J., de Vries, G., & Saharso, S. (2013). Gendered study choice: A literature review. A review of theory and research into the unequal representation of male and female students in mathematics, science, and technology. *Educational Research and Evaluation*, 19(6), 525-545.
- Zhao, Jieyu, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang.** 2018. "Gender Bias in Coreference Resolution: Evaluation and Debiasing Methods." ArXiv:1804.06876 [Cs], April. <http://arxiv.org/abs/1804.06876>.

