Gender Bias in Data-Driven AI Systems

List of Authors

Panagiota Fatourou, Foundation for Research and Technology-Hellas (FORTH) & University of Crete, Greece (faturu@csd.uoc.gr)

Chris Hankin, Imperial College London, UK (c.hankin@imperial.ac.uk)

Bran Knowles, Lancaster University, UK (b.h.knowles1@lancaster.ac.uk)

Table of contents

1. Executive Summary 3
   Recommendations 4
2. Introduction 6
3. Implications of Artificial Intelligence in achieving gender-equality in STEM 7
   3.1. Encoding Gender Stereotypes in the Data Training Process 7
   3.2. Gender Bias in job advertisements and online news 9
   3.3. Gender Bias in Software as an obstacle to career progression 11
4. Recommendations 11
   4.1. Technical 11
   4.2. Ethical 13
   4.3. Legal 14
   4.4. Economic 14
   4.5. Societal 16
   4.6. Educational 17
5. Conclusion 19
6. Bibliography 20

1. Executive Summary

The impact that AI systems have in people's daily life is growing. Predictions and suggestions made by AI algorithms have the power to shape individuals' future, as data-driven AI systems governed by such algorithms may play a crucial role in determining the job postings communicated to them, the employment offers they get, the news they are informed about, the professional networks they create, the friends they make, the products they buy, and many other aspects of their lives.

As beneficial as these technologies are, data-driven AI systems may discriminate against certain groups or individuals by reflecting or reinforcing human or society structural bias, or by even introducing new bias (e.g., through unintended consequences of complex interactions between multiple technical components). Such discrimination may have serious negative effects in fields where certain society groups are already not sufficiently represented. An example are women in Science, Technology, Engineering and Mathematics (STEM) professions who are persistently underrepresented despite the good employment opportunities that the field provides and the labor expansion that is foreseen for the STEM market in the near future.

Bias caused by data-driven AI systems and their training processes, including the challenges to be addressed, the implications to be considered, and potential prevention or "remedy" mechanisms, have provoked extensive research, and have given rise to thorough debates and in-depth discussions in academia, industry, public policy circles, and the media.

This paper focuses on how bias caused by data-driven AI systems may impact gender equality in STEM. Specifically, it discusses gender bias in AI technology used for initial screening and recruitment, gender bias in the targeting of job advertisements, and more generally bias in software that could introduce obstacles to the career progression of women. Such bias reflects and reinforces gender stereotypes, thus having the power to contribute to known gender imbalances in STEM. The paper also provides a collection of recommendations regarding such gender bias from the technical, ethical, legal, economic, societal and educational point of view. These recommendations are summarized below.

Recommendations

Technical

1. **Establish means, measures and standards to define, measure, and address gender bias in data-driven AI systems.**

   This will require a multi-stakeholder, multi-disciplinary collaboration to develop a theoretical framework, standards and practices to measure and prevent discrimination resulting from the use of data-driven AI technologies. The work will draw on expertise from computer science, law and ethics amongst others.

2. **Ensure that the design of data-driven AI technology takes into consideration principles and practices for ensuring gender fairness and the avoidance of bias.**

   In particular, data-driven AI technology should be designed on the basis of principles such as openness and algorithmic transparency[1]. Best practices and policies for avoiding discriminative behavior should be taken into consideration. It should also undergo rigorous validation and frequent re-assessment.

Ethical

3. **Ensure that the focus in the development of data-driven AI systems and processes is ethics-, rights-, and values-based.**

   In particular, 1) ensure that data-driven algorithmic assessments and automated screening and recruitment recommendation processes are objective and gender fair, 2) provide a development framework (e.g., through appropriate standardization efforts) to avoid unforeseen and unsubstantiated differential outcomes for applicants, and 3) ensure that a human remains in the loop to assist in determining whether a data-driven AI system could result in bias.

Legal

4. **Develop a clear legal framework for ensuring gender fairness in data-driven AI systems and for promoting accountability in whatever concerns their use and impact.**

5. **Ensure that people have the right to know the impact that AI systems may have in their lives.**

Economic

6. **Ensure adequate consideration of the consequences of data-driven AI technologies on job markets with particular emphasis on the effects on the gender gap in STEM-related professions.**

Societal

7. **Strive for gender-fair research and gender inclusion in the development of software.**

   Development teams of software (and in particular of data-driven AI systems and tools) should follow best practice in diversity and inclusion. This may require the involvement of non-profit

[1] ACM Statement on Algorithmic Transparency and Accountability, January 2017, https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

organizations to formulate gender-inclusiveness guidelines and assign roles and responsibilities to appropriate bodies for their application. Gender-awareness should be considered necessary and basic knowledge for computer scientists and engineers.

8. **Ensure that gender-sensitive language is used in advertisements, news, and all other Internet material. Ensure also that linguistic bias is avoided in machine translation systems and other Natural Language Processing (NLP) software.**
9. **Ensure the provision of ways and means that will make data-driven AI systems and their consequences explainable to society.**

   People should be given effective ways and means to get informed about the impact that data-driven AI systems may have in their lives. It should also be easily possible to obtain information about privacy, data acquisition and processing practices, fairness and anti-discrimination policies governing such systems. Finally, people should be informed about the accountability implications of the operation of such systems.

Educational

10. **Ensure that gender fairness is taken into consideration in all levels of education.**
11. **Develop a framework for enhancing awareness of ethics and social responsibilities in all educational levels.**
12. **Stimulate AI-related technical education in university programs curricula.**

2. Introduction

The proliferation of AI systems and the data-driven automation processes met in various domains of every-day life revealed their discriminatory behavior in many cases. One important dimension of this discrimination is related to gender. Various examples of gender-discriminatory AI systems, developed even by companies that are considered technological giants, have been reported in the past. This paper discusses the implications of such bias in the long-standing underrepresentation of women in STEM.

From a technical perspective, researchers still struggle to answer various questions regarding the discrimination resulting from the use of AI systems, as there is lack of consensus in defining fairness in whatever regards the use of data-driven AI systems, let alone in measuring the fairness performance exhibited by such systems. The data-driven nature of learning algorithms has led to what seems to be today the data paradox: the behavior of data-driven AI systems is determined mainly by their training data.

From an ethical and legal perspective, several issues arise as data-driven AI technology is making its way in hiring practices through automated employment assistant software. People should have the right to know the impact that such technology may have in their lives and to protect their personal data.

The economic and social consequences of the gender gap in STEM have been extensively acknowledged. As discussed later on, AI technology can have a major role in enhancing or reducing this gap. It is thus important to ensure that AI systems comply with gender fairness policies, to achieve better inclusion of women in STEM education, research, and professions. This will have tremendous impact in economic and social cohesion, as well as in improving the professional lives of women.

The remainder of this paper is divided into two main sections. The first presents background evidence of the impact of gender bias on the whole employment lifecycle from job advertisements through recruitment interviews to career progression. The second provides a detailed rationale for each of our twelve recommendations.

Although we focus mainly on gender bias on data-driven AI systems, discrimination in AI systems comes as a more general issue and concerns several minority groups. Most of our recommendations are meaningful or can be easily generalized to address broader forms of discrimination in AI systems generally. We also do not specifically address gender beyond binary, despite the growing interest on this issue.

Our recommendations are structured using a variant of a standard approach to considering macro-environmental factors in strategic planning. We consider the technical, ethical, legal, economic, societal and educational dimensions. We did not consider the political or environmental factors. In consideration of each of the factors, we give a detailed rationale for the accompanying recommendations. Our findings are based on our own research, a review of relevant literature and consultation with other experts.

Whilst the issues addressed in this paper are of global significance, we have approached the work through a European lens. Much of our evidence base, particularly job market intelligence, is from the European Union. Notwithstanding this, our recommendations

are of a global nature and we commend them to policy and law makers wherever they are.

3. Implications of Artificial Intelligence in achieving gender-equality in STEM

3.1. Encoding Gender Stereotypes in the Data Training Process

The use of Artificial Intelligence (AI) in systems and applications to automate processes is fast evolving. For instance, there is an increasing trend of using data-driven AI systems for initial screening in the admission process. On the positive side, when trained appropriately, such systems can remove human bias from some parts of the process [P19, M18-II, GG18]. They can also automate time-consuming tasks such as resume screening and short-listing of applicants, making the recruitment process less tedious and faster[2]. Moreover, they can be used to achieve deeper understanding of the applicant's skills with the goal of improving the quality of the hiring process [ES+18].

The design of such AI systems is usually based on predictive models that are trained using data. Unfortunately, these data may encode gender and other social stereotypes [3], [4] [BCZ+16, BLG+20], thus exhibiting social bias with severe consequences in gender fairness in admission and recruitment processes [BM13, CD19, CPF+16, CTY18, H18, P19, SDE20, WH16].

According to Catalyst[5], less than 33% of all employees in scientific research and development across the world in 2014 were women (see also [FPP19]). In particular, less than 2% of all women in the European labor market chose careers in ICT [EC14]. Women's involvement in innovation and entrepreneurship is also discouraging. In the European Union [BH+14], women account for less than 25% of science and engineering professionals [ISCO08] (major group code 21[6]); also, they constitute only 14% of science and engineering associate professionals [ISCO08] (major group code 31[7]).

This persistent underrepresentation of women in the ICT domain may have negative implications on gender fairness in automated recruitment processes, as illustrated in several examples provided in the literature [G16, M18, BMJ88] and in

[2] https://medium.com/the-research-nest/the-pros-and-cons-of-ai-in-recruitment-19c141d1c4b7

[3] https://www.wired.com/story/opinion-conversational-ai-can-propel-social-stereotypes/

[4] https://edition.cnn.com/2019/11/21/tech/ai-gender-recognition-problem/index.html

[5] https://www.catalyst.org/knowledge/women-science-technology-engineering-and-mathematics-stem

[6] https://ec.europa.eu/esco/portal/occupation?uri=http%3A%2F%2Fdata.europa.eu%2Fesco%2Fisco%2FC21&conceptLanguage=en&full=true#&uri=http://data.europa.eu/esco/isco/C21

[7] https://ec.europa.eu/esco/portal/occupation?uri=http%3A%2F%2Fdata.europa.eu%2Fesco%2Fisco%2FC31&conceptLanguage=en&full=true#&uri=http://data.europa.eu/esco/isco/C31

online sources[8,9,10,11]. In these examples, historical data (e.g., resumes of previous successful applicants and past hiring decisions) were used to train screening and recruitment recommendation systems. However, the fact that the ICT field is men-dominated and thus historical data refer mainly to male applicants, resulted in discrimination against women. Without special care for their appropriate training, such systems simply reproduce bias embedded in the admission processes, even if the computing process is fair [BS16, CP14, CZ13, Z17]. This could contribute to the gender-disparity crisis within the technological and ICT ecosystem.

Interview processes that employ video may also use learning algorithms to assist the recruitment decision. The goal of such algorithms is usually to assess candidates based on the keywords, facial expressions and tones they use in video interviews[12]. However, studies [H18, P19] show that AI technology for face-analysis[13] also exhibits gender bias. So, if not thoroughly tested against gender discrimination, AI models used to analyze interview responses may result in bias.

From the above, it becomes apparent that data-driven AI systems may exhibit gender bias by encoding stereotypes in their training data or when their gender distribution is strongly imbalanced. Gender fairness should be, therefore, one of the key issues when defining the quality of the training data set.

The underrepresentation of women in ICT is also present in professions related with artificial intelligence [WWC19]. According to the World Economic Forum[14], only 22% of AI professionals are women. Specifically, technical skills where men are reported to outnumber women in AI specialization fields include deep learning (66%), artificial neural networks (66%), pattern recognition (98%), and computer vision (67%). Additionally, other reports[15,16,17] provide evidence on women underrepresentation in AI-related jobs in big technology companies.

This gender imbalance in AI jobs is important[18] as it implies that AI technologies are largely designed and developed by a male-dominated industry. Apparently, any unconscious biases baked into the decisions taken by these people could have serious consequences. It is thus essential to have gender-balanced teams designing AI technologies, as gender inclusion adds additional perspectives and levels of analysis when designing software [A16]. This is indeed essential as AI technology is meant to be used by the entire society. It also results

[8] https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G

[9] https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

[10] http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf

[11] https://www.catalyst.org/research/trend-brief-gender-bias-in-ai/#easy-footnote-bottom-15-13438

[12] https://www.hirevue.com/

[13] https://www.catalyst.org/research/trend-brief-gender-bias-in-ai/#easy-footnote-bottom-17-13438

[14] https://www.weforum.org/agenda/2019/01/ai-artificial-intelligence-failing-next-generation-women-bias/

[15] https://diversity.google/annual-report/

[16] https://www.vox.com/2017/11/9/16628286/apple-2017-diversity-report-black-asian-white-latino-women-minority

[17] https://www.weforum.org/agenda/2019/05/ai-assisted-recruitment-is-biased-heres-how-to-beat-it/

[18] https://hbr.org/2019/11/as-jobs-are-automated-will-men-and-women-be-affected-equally

in financial benefits to the companies, and thus also to the economy at large.

3.2. Gender Bias in job advertisements

It is increasingly commonplace for AI to drive labor market decision making, with an estimated 98% of Fortune 500 companies using some form of Applicant Tracking System [19]. Bias can enter into decision making at various stages in the hiring funnel [SDE20]. Described below are the different forms of bias that can affect gender equality at these stages.

Structural bias occurs in shaping the pool of candidates who have seen or intend to respond to a given job advertisement. Unfairness can be introduced simply by an employer choosing to use one platform or hiring algorithm over another, as the available candidates in these sets are not only like to differ but may differ *structurally* (e.g. including more men than women[20]). The candidate pool is also a function of the reach and coverage of advertisements. AI is often used to optimize the display of job advertisements in order to accelerate the process of finding candidates, maximize the number of applications, and reach good quality candidates. The logic underpinning the forwarding of job advertisements to potential applicants can structurally bias the resulting candidate pool [K18, P19, S18].

Previous work [LT18, S13, DTD15] has provided evidence that groups against which discrimination has historically occurred are more likely to be associated with advertisements that are not of interest to them. Web crawlers often try to find appropriate applicants for job openings by retrieving information from social media platforms and other publicly available online sources. Job openings advertised in this way often result in gender bias [M18, P19, S18], as historical data and statistics may illustrate that forwarding the openings to men will optimize the number of applications, and potentially, also the quality of the applicants.

Such advertising techniques have an apparent negative effect on female participation in the STEM labor market. Previous work [CSV11, LT18, SW12] points out that the dissemination of information about STEM careers to women is indeed essential for attracting more women into STEM education, research and related professions. Therefore, structural bias that affects the reach of advertisements may contribute to (and explain), to some degree, why women do not apply for STEM positions [DB+10], despite the fact that when they do apply they have good chances of success [WC15].

Another study [LT18] shows that even advertising behavior that is not intended to be discriminatory can end up being gender biased because women may be considered a more costly advertising target group than men. [LT18] argues that the way information is distributed to different target groups may seriously depend on "the return on investment on advertising across all industry sectors".

Representational bias occurs when linguistic features of advertisements carry and thus reinforce a gender bias [CBN+16, PAL19, PHL18, FCL12]. This can occur in the original

[19] https://www.jobscan.co/blog/fortune-500-use-applicanttracking-systems/

[20] https://www.statista.com/statistics/933964/distribution-of-users-on-linkedin-worldwide-gender/

writing of the job advertisement, and indeed in developing criteria used to define the ideal candidate, which the AI then seeks to optimize. When AI is used to derive the criteria for a candidate's fit for a job using historical data, this can produce spurious correlations that perpetuate longstanding labor market inequalities (similar to those discussed in Section 3.1) [SDE20]. This bias is particularly insidious at the point of promotion, but can also preclude female candidates from selection for high-income jobs. Previous studies [21], [22] show that advertising systems often favor men when it comes to high-income jobs and leading positions by displaying to them advertisements for such jobs more often than to women [G16]. This tendency negatively impacts gender balance in senior leadership and thus affects gender imbalances in broader decision-making processes.

Additionally, machine translation may be used to translate advertisements into other languages in an effort to disseminate them more broadly. Well-known machine translators (MT) [PAL19, SSZ19] have default behaviors that have been criticized as being gender-biased, reproducing stereotypes in translated expressions. For instance, such translators may provide translations which (implicitly) imply that programmers or people in leadership positions or in specific domains are male [AJ17, FC19, RN+18, SSZ19, ZWY+18], thus reinforcing known stereotypes about the role of women and men in specific professional fields and in society more generally.

Relational bias is less well understood in this context, but it is clearly relevant. Who a person is connected with can reveal a lot of information not otherwise provided by candidates, and AI is routinely used to discover these connections for purposes such as targeted advertising [RWIS]. It is not outside the realm of possibility that these are currently being used within labor market AI. If this information is used to supplement the materials directly provided by the candidate as part of determining job fit, it can disadvantage people on the basis of who they are associated with. In particular, if men are more likely to be connected through social networks to individuals in particular professions or with particular job titles, this may bias hiring algorithms toward favorably rating male candidates for roles.

To eliminate all these negative consequences, a great deal of care should be taken to ensure that gender-biased behaviors in jobs advertising are eliminated. It is important to develop methods for removing bias, but these should also carefully attend the ways in which a model may make indirect use of a proscribed characteristic, e.g. though network effects as well as through emerging use of facial recognition technology.

[21] https://www.independent.co.uk/life-style/gadgets-and-tech/news/googles-algorithm-shows-prestigious-job-ads-to-men-but-not-to-women-10372166.html

[22] https://marketingland.com/carnegie-mellon-study-finds-gender-discrimination-in-ads-shown-on-google-134479

3.3. Gender Bias in Software as an obstacle to career progression

Moreover, a lot is still to be done for gender inclusiveness in software [VZH+19, MSS+18, BFI+10, BBW+05]. Since the labor market of software engineers is male-dominated, current software supports styles mostly preferable to men. Examples of such software include, 1) visualization systems (where evidence has been provided [BYB+13, TCR03] that there are gender-specific navigation benefits [TCR03], but also gender bias, in visualization systems that have been employed, e.g., for the quantitative analysis of research scientific data that can be used for further research [BYB+13]), 2) educational software authoring systems [BBG+06], 3) web automation and media authoring systems [BBG+06, RP18, RSE10], 4) intelligent agents [KSW+11, SG18], etc. Such bias usually reflects gender differences in problem-solving or cognitive processes: it is acknowledged [BBG+06] that women have different ways to process information and solve problems than men.

These forms of bias could complicate the career progression of women researchers and professionals.

4. Recommendations

4.1. Technical

Theoretical Framework. Several research questions are still to be answered to fully understand the discrimination behavior of AI and come up with the appropriate computational means to prevent them from being discriminatory.

Non-discrimination regulations in the EU (as defined at national, European, or international level) usually determine sensitive data characteristics and groups of people that are subject to discrimination [Z17, BS16]. Such legislation includes (among others) (1) Articles 8 and 19 of the Treaty of the Functioning of European Union[23], (2) the Council Directive 2004/113/EC [24], 3) the Employment Equality Framework Directive 2007/78/EC, 4) the Equal Treatment Directive 2006/54/EC, 5) the Gender Goods and Services directive 2006/113/EC, and generally the European employment discrimination law [25]. Additionally, the General Data Protection Regulation (GDPR) [26] disallows the use of algorithmic profiling based on sensitive personal data, which could be translated [G16-II] as a prohibition against processing data revealing membership in special categories.

Such regulations should be translated to formal non-discrimination constraints, which should be taken into consideration when developing predictive modeling algorithms [Z17].

These constraints should formally capture the degree of "explainable" differences between groups, allowing people who belong to the same group to be treated fairly. Moreover, they should be open and

[23] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12012E/TXT:EN:PDF

[24] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:373:0037:0043:en:PDF

[25] https://fra.europa.eu/sites/default/files/fra_uploads/1510-fra-case-law-handbook_en.pdf

[26] https://gdpr-info.eu/

up for discussion. This will contribute to developing explainable AI techniques and supporting the "right to explanation," i.e., the right for a person to ask why and how an algorithmic decision was made about her/him [DHP+12, G16]. There is also a need to come up with the appropriate research techniques to be able to computationally explain the roots of gender bias.

Moreover, no consensus has been reached yet on how to measure the performance of discrimination-aware data mining techniques [PRT12, Z17]. Producing a unifying view of performance criteria when developing techniques for non-discriminatory predictive modeling poses additional challenges.

Reducing gender bias in the data training process. Scientists from different disciplines need to collaborate to fully understand the data sets that are used in the training process of data-driven AI systems that may exhibit gender-discrimination behaviors, as well as improve their quality and gender-neutrality. The training data sets must be rich and diverse. However, diversity in features does not necessarily imply fairness in sensitive attributes [BS16, CDK+16].

Openness in datasets is helpful as it enables the understanding of sources of bias and may contribute to the production of gender-fair datasets. More generally, mechanisms for identifying and measuring gender discrimination in datasets, and gender-sensitive processes for data collection are necessary to guarantee that non-discriminatory models are produced.

Data-Driven AI technologies should undergo meticulous assessment for testing the quantity and the quality of the training and test data, as well as the fairness ensured by the computing process and models and by all other components they are comprised of throughout their entire lifecycle. Frequent re-assessments should also be included in the testing cycle[27]. The goal should be to ensure that the entire system complies with non-discrimination and fairness policies based on gender and other sensitive attributes.

Recommendation 1: Establish means, measures and standards to define, measure, and address gender bias in data-driven AI systems.

This will require a multi-stakeholder, multi-disciplinary collaboration to develop a theoretical framework, standards and practices to measure and prevent discrimination resulting from the use of data-driven AI technologies. The work will draw on expertise from computer science, law and ethics amongst others.

The framework should support the development of solutions to problems such as 1) the translation of current and future anti-discriminatory and gender-equality regulations to formal constraints without sacrificing precision, 2) the lack of consensus in defining gender fairness, and 3) measuring gender bias in a systematic and accountable way. It should also address issues of negative bias in the creation of AI algorithms.

[27] https://www.womenatthetable.net/blog/affirmative-action-for-algorithms

Recommendation 2: Ensure that the design of data driven AI technology takes into consideration principles and practices for ensuring gender fairness and avoidance of bias.

In particular, AI technology should be designed on the basis of principles such as openness and algorithmic transparency[1]. Best practices and policies for avoiding discriminative behavior should be taken into consideration. AI technology should undergo rigorous validation and frequent re-assessment.

4.2. Ethical

When it comes to algorithmic assessments, a number of ethical issues are discussed in [NRC13, L91] and are summarized in Table 1. Note that many of these issues may have implications in achieving gender-equality in the AI ecosystem. For instance, in a male-dominated field, as is ICT, assessments based on profiles' inferences (C1) or taking into consideration whether the applicant is a good fit for a particular team (C3) may have discrimination implications. Regarding C3, the influence of time/age in abilities may be different for men and women. In the context of C4, the provided feedback should be independent of gender or other sensitive characteristics. It thus becomes apparent that the ethical approaches to some of the issues raised in Table 1 may be important for achieving a gender-neutral ecosystem in labor market and research, and other aspects of the economy and the society.

Ethical Implications related to Automated Hiring Technologies [NRC13, L91]	
C1	Are there ethical implications when assessing individuals based on collective profiles' inferences versus individual attributes?
C2	How often should an individual be assessed? Should it be at the basis of evaluating her/him once or many times over her/his career?
C3	Is it ethical to take into consideration in the assessment process whether an applicant is a good fit for the team s/he is supposed to work with?
C4	What is the feedback to be returned to the applicant?

TABLE 1

Ethical questions that require debate may be also raised when exploiting the potential of AI to be used to redress gender inequality. For instance, if developers so choose, they can fix certain model parameters to bias in favor of women. Ethical issues may also come up [P17] when different algorithmic fairness constraints are impossible to be satisfied simultaneously. It is therefore critical that humans remain in the loop to assist in determining whether AI is contributing to gender inequality. To this end, the logic that drives AI should be transparent to both the employer and the potential candidate, to promote confidence that resulting decisions are based on criterion constructs that are ethically sound.

Recommendation 3: Ensure that the focus in the development of data-driven AI systems and processes is ethics-, rights-, and values-based.

In particular:

1. Ensure that data-driven algorithmic assessments and automated screening and recruitment recommendation processes are objective and gender fair.
2. Provide a development framework (e.g., through appropriate standardization efforts) to avoid unforeseen and unsubstantiated differential outcomes for applicants.
3. Ensure that a human remains in the loop to assist in determining whether AI is contributing to gender inequality.

4.3. Legal

The risk that AI may disadvantage women and other groups that are protected by anti-discrimination laws has already been touched upon in previous sections. In addition, many AI tools rely on extensive data gathering that often target individuals (e.g., applicants), and therefore their operation should comply to GDPR and other privacy-related regulations.

Recruiting and hiring algorithms frequently use observed correlations to predict how individuals will behave in the future. However, because these predictions are not always based on causal factors, they can result in significant biases [K18].

Several surveys and reports (see e.g. [BS16, K18, SDE20]) discuss how data-driven AI may result in discrimination which is caused by algorithmic effects regardless of jurisdiction [RBK+19]. Another report [A20] discusses legal frameworks to consider issues raised by algorithmic tools. Additional research is necessary [CD19] to analyze existing legislation related to AI from a gender fairness perspective, as well as to figure out how to appropriately exploit such legislation for boosting gender equality.

Recommendation 4: Develop a clear legal framework for ensuring gender fairness in data-driven AI systems and for promoting accountability in whatever concerns their use and impact.

Recommendation 5: Ensure that people have the right to know the impact that AI systems may have in their lives.

4.4. Economic

Several studies [CED16, EIG18, EP15] reveal that Europe faces a noticeable labor and skill shortage in the ICT and more generally the STEM sector. Within the EU, the requirement for STEM-skilled personnel has increased by 12% between the years 2000 and 2013 [EP15, CED16], whereas based on European Parliament estimations [EP15], over 7 million job openings are forecast until 2025 in the STEM field. Thus, enhancing the labor force and increasing the recruitment capacity in the field is considered one of the big challenges for the EU economy. A contributing factor to this shortage is women's enduring under-representation in the field [FPP19].

It therefore becomes apparent that achieving gender inclusiveness in STEM will positively influence the economic growth of Europe. As estimated by EIGE[28], the effect of closing the gender gap in STEM on employment will be significant [EIG18] (see Figure 1).

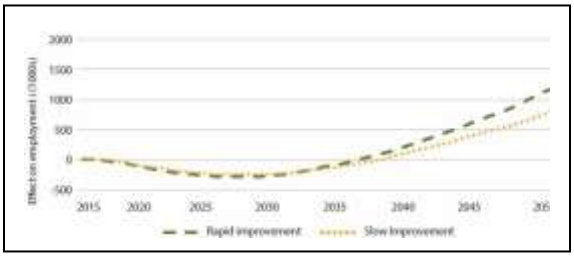

FIGURE 1. THE EFFECT OF CLOSING THE GENDER GAP IN STEM ON EMPLOYMENT (SOURCE: [EIG18, FIGURE 1]).

In addition, Figure 2 illustrates the increase on GDP per capita by 2030 that will result from the female participation in STEM related fields. Other studies [EIG18] also confirm that this increase will range between 2.2% and 3.0% by 2050. In monetary terms, closing the STEM gender gap leads to an improvement in GDP by 610-820 billion euros by 2050. Another study [EP15] foresees an increase in women's employment, productivity and wages, under the assumption that the EC measures will be effectively implemented. These will result in better long-term competitiveness of the EU economy and an improved balance of trade [EIG18]. Since achieving gender-fairness in AI may seriously influence the percentage of women in STEM (as discussed in previous sections), it follows that it will also contribute to the economic growth of Europe.

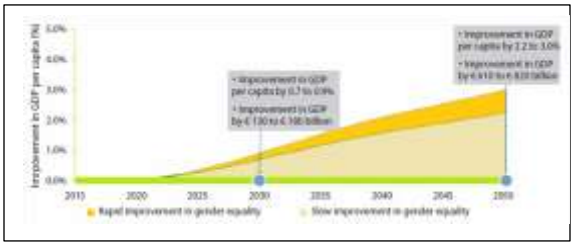

FIGURE 2. THE EFFECT OF CLOSING THE GENDER GAP IN STEM ON GDP PER CAPITA (SOURCE: [EIG18, FIGURE 2]).

Recommendation 6: Ensure adequate consideration of the consequences of data-driven AI technologies on job markets with particular emphasis on the effects on the gender gap in STEM-related professions.

Several studies have pointed out that AI and more generally smart automation systems, have the potential to boost productivity and result in the creation of new products and better services, thus bringing benefits to the economy. In a study by PwC[29], it is estimated that the contribution of these technologies to global GDP by 2030 can be up to 14%. Yet, another analysis by the same source [PWC18] shows that many occupations will be affected unevenly over time by such technologies. Specifically, the analysis reveals that highly educated employees performing clerical tasks and analytical jobs could, for example, be relatively exposed in the short term. So, these technologies may cause disruption to the job markets with unclear consequences to the gender aspects of their compositions.

[28] https://eige.europa.eu/

[29] PwC, "Sizing the prize, " 2017. Available at https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

4.5. Societal

Reducing the gender imbalance in STEM-related job markets will have obvious consequences for women's lives. Women will have better opportunities to find jobs and enjoy better salaries. Additionally, the risk of social exclusion of women will be reduced with pronounced benefits to the society.

Several articles (see e.g. [FPP19]) refer to the positive societal effects of the better inclusion of women in STEM studies, research, and related professions. These include the enrichment of skills in the labor market, the enhancement of the research process and its outcomes, the intensification of the innovation potential, and the boosting of major sectors of the economy. Moreover, gender equality is understood as a social justice and fairness issue in Europe and it is one of the EC's priorities.

To eliminate the negative consequences that AI may have in social cohesion by discriminating against women, it is therefore essential to take into consideration gender fairness when developing data-driven AI systems, as well as in all phases of the production cycle of AI technology.

Evidence has been provided [DKT18, H88, H91] that the production of knowledge pertains to the status of researchers, including their historical, conceptual, cultural and social backgrounds. Given that STEM-related sciences and professions have been and still are so demographically skewed, with women being constantly underrepresented in them, these could imply that conclusions reached about algorithmic discrimination and decisions made on algorithmic fairness are gender-insensitive and therefore, they do not always reflect the needs of the entire society [L13].

Ensuring gender-balanced development teams for AI software is therefore of crucial importance. It is also important to understand the interaction of human decision makers with automated decision making tools.

Ensuring gender-sensitive research and improving gender inclusion in software may have significant societal implications in many domains: healthcare, finance, education, and others. Women account for 50 percent of people on payroll [EHH10, W14]. They also account for a large portion of the consumers' landscape. They have a strong presence in social media and online services, and often use modern technology in their everyday life. They comprise a significant part of the society and economy that cannot be ignored [W14, VZH+19].

However, studies show that assuming gender-neutrality when developing software is not the norm. Increasing gender-awareness is, therefore, in many cases necessary [VZH+19]. This awareness should address all forms of potential gender-related bias in software development, including focusing only on a male perspective, underestimating gender differences, stereotyping gender features, showcasing non-significant differences between genders, etc. [DKT18].

Recommendation 7: Strive for gender-sensitive research and gender inclusion in the development of software.

Development teams of software (and in particular of data-driven AI systems and tools) should follow best practice in diversity and inclusion. This may require the involvement of non-profit organizations to formulate gender-inclusiveness guidelines and assign roles and responsibilities to appropriate bodies for their application. Gender-awareness should be considered necessary and basic knowledge for computer scientists and engineers.

As discussed in Section 3.2, linguistic bias reinforces gender norms and fosters gender stereotyping [CBN+16, HB+18, HTS17, PHL18, FCL12]. To achieve the use of a gender-fair language in machine-automated translation, as well as on information found on the web, techniques such as language neutralization (or even proper language feminization in order to enhance visibility of women in male-dominated job markets and fields) is recommended.

Recommendation 8: Ensure that gender-sensitive language is used in advertisements, news, and all other Internet material. Ensure also that linguistic bias is avoided in machine translation systems and other Natural Language Processing (NLP) software.

Data-driven AI systems may have significant impact in people's lives. This spans from making hiring decisions that can be life-transformative, to enforcing gender stereotypes and determining how people get informed by online news and advertisements systems (through their ranking algorithms) with severe consequences in the development of critical thinking, intelligence, and other necessary skills for a cohesive and healthy society.

Recommendation 9: Ensure the provision of ways and means that will make data-driven AI systems and their consequences explainable to society.

People should be given effective ways and means to get informed about the impact that data-driven AI systems may have in their lives. It should also be easily possible to obtain information about privacy, data acquisition and processing practices, fairness and anti-discrimination policies governing such systems. Finally, people should be informed about the accountability implications of the operation of such systems.

4.6. Educational

According to UNESCO [UN03], removing any racial, gender and cultural prejudice in education contributes to the development of solid foundations for a civil, open and global society. Thus, integrating the gender dimension in research and education is of crucial important. Indeed, it is one of the six European Research Area (ERA) priorities.

Recommendation 10: Ensure that the gender dimension is included in all levels of education.

Studies in STEM-related programs should not provide only technical knowledge. It should be built upon a complete framework, which should equip students with disciplines, ethics, attitudes, practices, and

all other necessary skills and attributes that when combined together could maximize benefit for society. Moreover, multi-disciplinarity and inter-disciplinarity must be promoted so that STEM studies contribute more widely to the economic and social progress.

Recommendation 11: Develop a framework for enhancing awareness of ethics and social responsibilities in all educational levels.

The wide spread of data-driven AI technologies and its presence in several aspects of everyday life leads to the necessity of appropriately educating people in order to understand its use and its implications. This encompasses appropriate adjustments of the schools' and universities' curricula, as well as the development of a stronger educational framework in terms of ethics, critical judgment, and digital knowledge and intelligence [EUACM18].

Recommendation 12: Stimulate AI-related technical education in university programs curricula.

5. Conclusion

We have demonstrated that there is a substantial gender imbalance in the STEM-based job market. We have also presented evidence that data-driven AI technologies, if not carefully designed and implemented, may well exacerbate this situation. As examples, we have discussed automated recruitment assistant ML software, usually designed with the goal of improving the quality of hiring process, which however, if developed (and trained) without care in terms of gender fairness, may end up to have negative consequences to the careers of women in STEM, decisive for their professional lives. We have also examined gender bias in recommendation systems that govern the operation of social media, and argued that it might influence the kind of job advertisements and news that are forwarded to women through the internet, as well as the creation of their professional network. Finally, we have touched upon the impact that the disregard of the gender dimension in software may have on the career progression of women. Apparently, such forms of gender bias reinforce gender stereotypes, have undesirable influence on the number of women that choose to follow careers in STEM, and empower existing obstacles and glass ceilings to their career development.

It is clear that the world economy would be considerably boosted, if we take care to ensure that women are given equal opportunity to progress through STEM-based careers. In addition, that would result in significant advancements on research and innovation, it will enhance women's employment and productivity, thus diminishing the risk of women's social exclusion, with obvious benefits to the society.

In particular, as every stage of the employment lifecycle becomes more technology mediated, steps should be taken to ensure that these technologies are gender fair. This implies that changes are needed to the ways that the technology is developed, to the ethical and legal frameworks applied, to the societal understanding of the issues and to the educational systems that train the future STEM workforce. Our recommendations are designed to assist policy makers to navigate their way towards a fairer job market where all citizens, regardless of gender, are given the opportunity to excel.

6. Bibliography

[A16]	H. L. Alford, "Gender Bias in IT Hiring Practices: An Ethical Analysis," 2016. Available at https://digitalcommons.olivet.edu/cgi/viewcontent.cgi?article=1000&context=csis_stsc
[A20]	I. Ajunwa, "The paradox of automation as anti-bias intervention," Cardozo Law Review,41, 2020. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746078
[AJ17]	D. Alvarez-Melis and T. S. Jaakkola, "A causal framework for explaining the predictions of black-box sequence-to-sequence models," Conference on Empirical Methods in Natural Language Processing, 2017. Available at https://people.csail.mit.edu/tommi/papers/ AlvJaa_EMNLP2017.pdf
[BBG+06]	L. Beckwith, M. M. Burnett, V. Grigoreanu, and S. Wiedenbeck, "Gender HCI: What about the software?," IEEE Computer, 39(11), 97–101, 2006.
[BBW+05]	L. Beckwith, M. Burnett, S. Wiedenbeck, C. Cook, S. Sorte, and M. Hastings, "Effectiveness of end-user debugging software features: Are there gender issues?," SIGCHI Conference on Human Factors in Computing Systems (CHI '05), 869-878, 2005. http://doi.acm.org/10.1145/1054972.1055094
[BCZ+16]	T. Bolukbasi, K.-W. Chang, J. Zou, V. Saligrama, A. Kalai, "Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings," Arxiv, 2016. Available at https://arxiv.org/abs/1607.06520
[BFI+10]	M. Burnett, S. D. Fleming, S. Iqbal, G. Venolia, V. Rajaram, U. Farooq, V. Grigoreanu, and M. Czerwinski, "Gender differences and programming environments: across programming populations," ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '10), Article No. 28, pp. 1–10, 2010, https://doi.org/10.1145.
[BH+14]	B. Burchell, V. Handry, J.,Rubery, and M. Smith, "A New Method to Understand Occupational Gender Segregation in European Labour Markets," Publications Office of the European Union, https://ec.europa.eu/info/sites/info/ files/150119_ segregation_report_web_en.pdf
[BLG+20]	M. Babaeianjelodar, S. Lorenz, J. Gordon, J. Matthews, and E. Freitag, "Quantifying Gender Bias in Different Corpora,'' Companion Proceedings of the Web Conference (WWW),pp. 752–759, 2020. Available at https://dl.acm.org/doi/10.1145/3366424.3383559
[BM13]	J. Burn-Murdoch, "The problem with algorithms: magnifying misbehavior," The Guardian, 2013, http://www.theguardian.com/news/datablog/2013/aug/14/problem-with-algorithms-magnifying-misbehavior
[BMJ88]	British Medical Journal, London, Saturday 5 March 1988, http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC2545288&blobtype=pdf
[BS16]	S. Barocas and A.D. Selbst, "Big Data's Disparate Impact," California Law Review, 2016. Available at http://www.courts.ca.gov/documents/BTB24-2L-2.pdf
[BYB+13]	M. A. Borkin, C. S. Yeh, M. Boyd, P. Macko, K. Z. Gajos, M. Seltzer, and H. Pfister, "Evaluation of file-system provenance visualization tools," IEEE Transactions on Visualization and Computer Graphics 19, 12 (December 2013), 2476-2485, 2013. Available at http://dx.doi.org/10.1109/TVCG.2013.155
[CBN+16]	A. Caliskan Islam, J. J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora necessarily contain human biases," Arxiv, 2016. Available at https://arxiv.org/abs/1608.07187

[CD19]	C. Collett and S. Dillon, “AI and Gender: Four Proposals for Future Research,” Cambridge: The Leverhulme Centre for the Future of Intelligence, 2019. Available at http://lcfi.ac.uk/media/uploads/files/AI_and_Gender___4_Proposals_for_Future_Research.pdf
[CDK+16]	L. Elisa Celis, A. Deshpande, T. Kathuria, and N. K. Vishnoi, “How to be Fair and Diverse?,” 2016. Available at https://arxiv.org/abs/1610.07183
[CED16]	European Centre for Development of Vocational Training, “Skill shortage and surplus occupations in Europe,’’ Briefing Note, October 2016. Avaialble at http://www.cedefop.europa.eu/files/9115_en.pdf
[CP14]	D.K. Citron, F.A. Pasqualle III, “The scored society: Due process for automated predictions,” Washington Law Reviews, 89:1–33, Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209
[CPF+16]	S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, “A computer program used for bail and sentencing decisions was labeled biased against blacks. Its actually not that clear,” The Washington Post. https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/canan-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?utm_term=.4ded14bf289e. Available also at http://www.cs.yale.edu/homes/jf/Feller.pdf
[CSV11]	S. Cheryan, J. O. Siy, M. Vichayapai, B. J. Drury, and S. Kim, “Do female and male role models who embody STEM stereotypes hinder women's anticipated success in STEM?,” Social Psychological and Personality Science 2 (6), 656-664, 2011. Available at https://www.researchgate.net/publication/258189201_Do_Female_and_Male_Role_Models_Who_Embody_STEM_Stereotypes_Hinder_Women%27s_Anticipated_Success_in_STEM
[CTY18]	P. Cappelli, P. Tambe, and V. Yakubovich, “Artificial intelligence in human resources management: Challenges and a path forward,” 2019. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3263878
[CZ13]	T. Calders, I. Zliobaite, “Why unbiased computational processes can lead to discriminative decision procedures,” In B. Custers, T. Zarsky, B. Schermer, T. Calders (eds): Discrimination and privacy in the information society—Data mining and profiling in large databases, Springer, pp 43–57, 2013. https://link.springer.com/chapter/10.1007/978-3-642-30487-3_3
[DB+10]	A. B. Diekman, E. R. Brown, A. M. Johnston, and E. K. Clark, “Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers,” Psychological Science 21(8), 1051-1057, 2010. Available at https://pdfs.semanticscholar.org/804f/0e7c4a33ec8e7033e41bb23e71a35d988034.pdf?_ga=2.183143086.920150581.1585234569-1877060260.1560988157
[DHP+12]	C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. Zemel, “Fairness through awareness,” 3rd Innovations in Theoretical Computer Science Conference, ITCS ’12, pages 214–226, 2012. Availabe at https://dl.acm.org/doi/10.1145/2090236.2090255
[DKT18]	C. Draude, G. Klumbyte, P. Treusch, “Re-Considering Bias: What Could Bringing Gender Studies and Computing Together Teach Us About Bias in Information Systems?,” 2018, Available at http://ceur-ws.org/Vol-2103/paper_3.pdf
[DTD15]	A. Datta, M. C. Tschantz, and A. Datta, “Automated experiments on ad privacy settings,” Privacy Enhancing Technologies 2015 (1), 92-112, 2015.
[EHH10]	A. English, H. Hartmann, and J. Hayes, “Are women now half the labor force? The truth about women and equal participation in the labor force,” Institute for Women's Policy Research, 2010. Available at http://www.iwpr.org/publications/pubs/are-women-now-half-the-labor-force-the-truth-about-women-and-equal-participation-in-the-labor-force

[EIG18]	EIGE, "Economic benefits of gender equality in the EU. How gender equality in STEM education leads to economic growth," 2018. Available at https://eige.europa.eu/rdc/eige-publications/economic-benefits-gender-equality-eu-how-gender-equality-stem-education-leads-economic-growth.
[EP15]	M. Caprile. R.,Palmen, P. Sanz. G. Dente. Encouraging STEM studies – Labour Market Situation and Comparison of Practices Targeted at Young People in Different Member States, European Parliament, DG for Internal Policies. Policy Department A, 2015.
[ES+18]	I. Erel, L. H. Stern, C. Tan, and M. S. Weisbach, "Research: Could Machine Learning Help Companies Select Better Board Directors?," *Harvard Business Review*, April 9, 2018. Available at https://hbr.org/2018/04/research-could-machine-learning-help-companies-select-better-board-directors
[EUACM18]	Informatics Europe & EUACM, "When Computers Decide: European Recommendations on Machine-Learned Automated Decision Making," 2018. Available at https://www.acm.org/binaries/content/assets/public-policy/ie-euacm-adm-report-2018.pdf
[FC19]	J. Escude Font and M. R. Costa-Jussa, "Equalizing gender biases in neural machine translation with word embeddings techniques," Arxiv, 2019. Available at https://www.arxiv-vanity.com/papers/1901.03116/
[FCL12]	J. L. Prewitt-Freilino, T. A. Caswell, E. K. Laakso, "The Gendering of Language: A Comparison of Gender Equality in Countries with Gendered, Natural Gender, and Genderless Languages," Sex Roles, 66(3):268–281, 2012. Available at https://link.springer.com/article/10.1007/s11199-011-0083-5.
[FPP19]	P. Fatourou, Y. Papageorgiou, and V. Petousi, "Women are needed in STEM: European Policies and Incentives, Communications of the ACM, Vol. 62 No. 4, 2019. Available at https://cacm.acm.org/magazines/2019/4/235600-women-are-needed-in-stem/abstract
[G16]	M. Garcia, "Racist in the Machine," World Policy Journal, Vol. 33, No. 4, pp. 111-117, 2016. https://muse.jhu.edu/article/645268/pdf
[G16-II]	B. W. Goodman, "A Step Towards Accountable Algorithms? Algorithmic Discrimination and the European Union General Data Protection," 29th Conference on Neural Information Processing Systems (NIPS), Barcelona, Spain, 2016. Available at http://www.mlandthelaw.org/papers/goodman1.pdf
[GG18]	R. Greenfield and R. Griffin, "Can Artificial Intelligence Take the Bias Out of Hiring?," *Boston Globe*, August 12, 2018. Available at https://www.bostonglobe.com/business/2018/08/12/can-artificial-intelligence-take-bias-out-hiring/SXVO2h7eYeQArnnjp0kPtL/story.html
[H18]	L. Hardesty, "Study Finds Gender and Skin-Type Bias in Commercial Artificial-Intelligence Systems," *MIT News,* February 11, 2018. Available at http://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212
[H88]	D. Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective," Feminist Studies 14(3), pp. 575–599, 1988. Available at https://msu.edu/~kg/874/Haraway_1988__Situated_Knowledges.pdf
[H91]	S. Harding, "Whose Science? Whose Knowledge? Thinking from Women's Lives," Cornell University Press, Ithaca, 1991.
[HB+18]	C. Harris, N. Biencowe, D. Telem, "What's in a pronoun? Why gender-fair language matters?," Annals of Surgery, 266(6): pp. 932-933, 2017. Available at https://www.ncbi.nlm.nih.gov/pmc/articles/ PMC5774006/

[HTS17]	D.M. Hechavarria, S.A. Terjesen, P. Stenholm, M. Brännback, S. Lang, "More Than Words: Do Gendered Linguistic Structures Widen the Gender Gap in Entrepreneurial Activity?," Entrepreneurship Theory and Practice, 2017. Available at https://www.researchgate.net/publication/326906637_More_than_Words_Do_Gendered_Linguistic_Structures_Widen_the_Gender_Gap_in_Entrepreneurial_Activity
[EC14]	European Commission, "Study on Women Active in the ICT Sector,'' Luxembourg, Publications Office of the European Union, 2013. Available at https://www.itu.int/en/ITU-D/Digital-Inclusion/Women-and-Girls/Girls-in-ICT-Portal/Documents/women_active_in_ict.pdf
[ISCO08]	http://www.ilo.org/public/english/bureau/stat/isco/isco08/
[K18]	P. Kim, "Big Data and Artificial Intelligence: New Challenges for Workplace Equality," University of Louisville Law Review, 2018. Available at https://ssrn.com/abstract=3296521
[KSW+11]	T. Kulesza, S. Stumpf, W.-K. Wong, M. Burnett, S. Perona, A. Ko, and I. Oberst, "Why-oriented end-user debugging of naïve Bayes text classification," ACM Transactions on Interact. Intell. Syst. 1, 1, Article 2, October 2011. Available at http://doi.acm.org/10.1145/2030365.2030367
[L13]	L. C. Landivar, "Disparities in STEM employment by sex, race, and Hispanic origin," Education Review, 2013. Available at https://www.sigmaxi.org/docs/default-source/News-Documents/keyed-in/disparities_in_science.pdf?sfvrsn=2
[L91]	R.L. Lowman, "The Clinical Practice of Career Assessment: Interests, Abilities, and Personality," Washington, DC: American Psychological Association, 1991. Available at
[LT18]	A. Lambrecht and C. E. Tucker, "Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads," 2018. Available at https://ssrn.com/abstract=2852260.
[M18]	D. Meyer, "Amazon Reportedly Killed an AI Recruitment System Because It Couldn't Stop the Tool From Discriminating Against Women," Fortune, October 10, 2018. Available at http://fortune.com/2018/10/10/amazon-ai-recruitment-bias-women-sexist/
[M18-II]	A. P. Miller, "Want Less-Biased Decisions? Use Algorithms," Harvard Business Review, July 26, 2018. Available at https://hbr.org/2018/07/want-less-biased-decisions-use-algorithms
[MSS+18]	C. Mendez, H. Susmita Padala, Z. Steine-Hanson, C. Hilderbrand, A. Horvath, C. Hill, L. Simpson, N. Patil, A. Sarma, and M. Burnett, "Open Source barriers to entry, revisited: A sociotechnical perspective," ACM/IEEE International Conference on Software Engineering, pp. 1004-1015, 2018. Available at https://dl.acm.org/doi/10.1145/3180155.3180241
[NRC13]	National Research Council, "New Directions in Assessing Performance Potential of Individuals and Groups: Workshop Summary," Washington, DC: The National Academies Press, 2013. Available at https://doi.org/10.17226/18427
[P17]	E. Pierson, "Gender differences in beliefs about algorithmic fairness," ArXiv, 2017. Available at https://arxiv.org/pdf/1712.09124.pdf
[P19]	L. Pasquarella Daley, "AI and Gender Bias: Trend Brief," Catalyst, 2019. Available at https://www.catalyst.org/research/trend-brief-gender-bias-in-ai/
[PAL19]	M. Prates, P. H. Avelar, and L. C. Lamb, "Assessing gender bias in machine translation: a case study with Google Translate," Neural Computing and Applications, Springer, 2019. Available at https://doi.org/10.1007/s00521-019-04144-6
[PHL18]	M. Prates, P. Henrique da Costa Avelar, L. C. Lamb, "Assessing Gender Bias in Machine Translation - A Case Study with Google Translate," September 2018. Available at

	https://www.researchgate.net/publication/327549663_Assessing_Gender_Bias_in_Machine_Translation_--_A_Case_Study_with_Google_Translate
[PRT12]	D. Pedreschi, S. Ruggieri, F. Turini, "A study of top-k measures for discrimination discovery," 27th annual ACM Symposium on Applied Computing (SAC), pp 126–131, 2012. Available at https://dl.acm.org/doi/10.1145/2245276.2245303
[PWC18]	PwC, "Will robots really steal our jobs? An international analysis of the potential long term impact of automation," 2018. Available at https://www.pwc.co.uk/economic-services/assets/international-impact-of-automation-feb-2018.pdf
[RBK+19]	M. Raghavan, S. Barocas, J. Kleinberg, K. Levy, "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices," 2019. Available at https://arxiv.org/abs/1906.09208
[RN+18]	R. Rudinger, J. Naradowsky, B. Leonard, and B. Van Durme, "Gender bias in coreference resolution," Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2, 2018. Available at https://arxiv.org/abs/1804.09301
[RP18]	J. A. Rode and E. Shehan Poole, "Putting the gender back in digital housekeeping," 4th Gender and IT Conference, 2018. Available at https://doi.org/10.1145/3196839.3196845
[RSE10]	M. Beth Rosson, H. Sinha, and T. Edor, "Design planning in end-user web development: gender, feature exploration, and feelings of success," IEEE Symposium on Visual Languages and Human-Centric Computing (VLHCC '10), 141-148, 2010. Available at http://dx.doi.org/10.1109/VLHCC.2010.28
[RWIS]	Robert Walters White Paper, "Using Social Media in the Recruitment Process," 5th White Paper, Robert Walters Insight Series. Available at https://www.robertwalters.us/content/dam/robert-walters/country/united-kingdom/files/whitepapers/rw-social-media-whitepaper.pdf
[S13]	L. Sweeney, "Discrimination in online ad delivery," ACM Queue 11 (3), 10, 2013. Available at https://queue.acm.org/detail.cfm?id=2460278
[S18]	N. Scheiber, "Facebook Accused of Allowing Bias Against Women in Job Ads," *The New York Times*, September 18, 2018. Available at https://www.nytimes.com/2018/09/18/business/economy/facebook-job-ads.html
[SDE20]	J. Sanchez-Monedero, L. Dencik, and L. Edwards, "What does it mean to solve the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems," Conference on Fairness, Accountability, and Transparency, ACM, pp. 458-468, 2020. Available at https://arxiv.org/abs/1910.06144
[SG18]	D. Showkat and C. Grimm, "Identifying gender differences in information processing style, self-efficacy, and tinkering for robot tele-operation," International Conference on Ubiquitous Robots (UR), 2018. Available at https://www.researchgate.net/publication/327196388_Identifying_Gender_Differences_in_Information_Processing_Style_Self-efficacy_and_Tinkering_for_Robot_Tele-operation
[SSZ19]	G. Stanovsky, N. A. Smith, and L. Zettlemoyer, "Evaluating Gender Bias in Machine Translation," Arxiv, 2019. Available at https://arxiv.org/pdf/1906.00591.pdf
[SW12]	J. R. Shapiro and A. M. Williams, "The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields" Sex Roles 66 (3-4), 175-183, 2012. Available at https://link.springer.com/article/10.1007/s11199-011-0051-0
[T18]	S. Tolan "Fair and Unbiased Algorithmic Decision Making: Current State and Future Challenges," JRC Technical Reports, Digital Economy Working Paper 2018-10, 2018. Available at https://arxiv.org/ftp/arxiv/papers/1901/1901.04730.pdf

[TCR03]	D. S. Tan, M. Czerwinski, and G. Robertson, “Women go with the (optical) flow,” ACM SIGCHI Conference on Human Factors in Computing Systems (CHI ’03), pp. 209-215, 2013. Available at http://doi.acm.org/10.1145/642611.642649
[UN03]	UNESCO, “Gender and Education for All: The Leap to Equality,” Global Monitoring Report 2003-2004. Available at https://unesdoc.unesco.org/ark:/48223/pf0000132513_eng
[VZH+19]	M. Vorvoreanu, L. Zhang, Y. H. Huang, C. Hilderbrand, Z. Steine-Hanson, and M. Burnett, “From Gender Biases to Gender-Inclusive Design: An Empirical Investigation,” Conference on Human Factors in Computing Systems, Paper No.: 53 Pages 1–14, 2019. Available at https://dl.acm.org/doi/10.1145/3290605.3300283
[W14]	G. Williams, “Are you sure your software is gender-neutral?,” Interactions 21(1), pp. 36–39, 2014. Available at https://interactions.acm.org/archive/view/january-february-2014/are-you-sure-your-software-is-gender-neutral
[WC15]	W.M. Williams and S. J. Ceci, “National hiring experiments reveal 2:1 faculty preference for women on STEM tenure track,” National Academy of Sciences 112 (17), pp. 5360-5365, 2015. Available at https://gap.hks.harvard.edu/national-hiring-experiments-reveal-21-faculty-preference-women-stem-tenure-track
[WH16]	The White House, “Big data: a report on algorithmic systems, opportunity, and civil rights,” Executive office of the president, 2016. Available at https://www.academia.edu/14047984/National_hiring_experiments_reveal_2_1_faculty_preference_for_women_on_STEM_tenure_track
[WWC19]	S.M. West, M. Whittaker, and K. Crawford, “Discriminating Systems: Gender, Race and Power in AI,” AI Now Institute, 2019. Available at https://ainowinstitute.org/discriminatingsystems.html.
[Z17]	I. Žliobaitė, “Measuring discrimination in algorithmic decision making,” Data Mining and Knowledge Discovery, 31, pp. 1060–1089, 2017. Available at https://link.springer.com/article/10.1007%2Fs10618-017-0506-1
[ZWY+18]	J. Zhao, T. Wang, M. Yatskar, V. Or-donez, and K.-W. Chang, “Gender bias in coreference resolution: Evaluation and debiasing methods,” NAACL-HLT, 2018. Available at https://arxiv.org/abs/1804.06876