Gender Bias in AI: Implications for Managerial Practices

Ayesha Nadeem, Olivera Marjanovic, Babak Abedin

To cite this version:

Ayesha Nadeem, Olivera Marjanovic, Babak Abedin. Gender Bias in AI: Implications for Managerial Practices. 20th Conference on e-Business, e-Services and e-Society (I3E), Sep 2021, Galway, Ireland. pp.259-270, 10.1007/978-3-030-85447-8_23 . hal-03648154

Gender Bias in AI: Implications for managerial practices

Ayesha Nadeem[1] and Olivera Marjanovic[2] and Babak Abedin[3]

1 University of Technology Sydney, NSW, Australia
Ayesha.Nadeem@student.uts.edu.au
[2] University of Technology Sydney, NSW, Australia
Olivera.Marjanovic@uts.edu.au
[3] Macquarie University, NSW, Australia
Babak.abedin@mq.edu.au

Abstract. Artificial intelligence (AI) applications are widely employed nowadays in almost every industry impacting individuals and society. As many important decisions are now being automated by various AI applications, fairness is fast becoming a vital concern in AI. Moreover, the organizational applications of AI-enabled decision systems have exacerbated this problem by amplifying the pre-existing societal bias and creating new types of biases. Interestingly, the related literature and industry press suggest that AI systems are often biased towards gender. Specifically, AI hiring tools are often biased towards women. Therefore, it is an increasing concern to reconsider the organizational managerial practices for AI-enabled decision systems to bring fairness in decision making. Additionally, organizations should develop fair, ethical internal structures and corporate strategies and governance to manage the gender imbalance in AI recruitment process. Thus, by systematically reviewing and synthesizing the literature, this paper presents a comprehensive overview of the managerial practices taken in relation to gender bias in AI. Our findings indicate that managerial practices include: better fairness governance practices, continuous training on fairness and ethics for all stakeholders, collaborative organizational learning on fairness & demographic characteristics, interdisciplinary approach & understanding of AI ethical principles, Workplace diversity in managerial roles, designing strategies for incorporating algorithmic transparency and accountability & ensuring human in the loop. In this paper, we aim to contribute to the emerging IS literature on AI by presenting a consolidated picture and understanding of this phenomenon. Based on our findings, we indicate direction for future research in IS for the better development and use of AI systems.

Keywords: Artificial Intelligence, Machine learning, Analytics, Gender, fairness

1 Introduction

It is true that AI applications offer solutions to various problems faced in different disciplines but simultaneously yield biased outcomes that could affect individuals or minorities of a certain race, gender, or color (Ntoutsi et al., 2019). The biased and adverse outcomes of algorithm decisions reach beyond the individuals and include harmful effects that reach families, communities, and society at large (Altman, Wood & Vayena 2018). The literature is evident that gender bias does exist in AI algorithms (Trewin 2018; Leavy 2018; Mehrabi et al. 2019; Dawson 2019; Kumar, Singh & Bhatanagar 2019; Canetti et al. 2019; Crawford 2016; Altman, Wood & Vayena 2018; Lambrecht & Tucker 2018; Galleno et al. 2019; Bolukbasi et al. 2016; Daugherty, Wilson & Chowdhury 2018, Dwivedi et al. 2019; Agarwal 2020; Robnett 2015; Nadeem, Abedin & Marjanovic 2020).

According to past literature bias is externalized and includes misguided conducts of "bad actors" which are either intentional or accidental and might not be easily traceable; therefore, such social and contextual issues are left beyond the law's reach (Hoffmann. A. L, 2019) and thus are deprived aspect of the society since long. AI algorithms are trained on datasets that are influenced by their creators' thinking, and as a result the pre-existing prejudices in the society "sneaks in" the AI systems thus amplifying the societal gender stereotyping and discrimination in society (Ntoutsi et al., 2019, Lee. N. T, 2018, Hoffmann. A. L, 2019).

It is noteworthy to mention here that lack of gender diversity and exceptionally homogenous and male domination in high tech industries and in the design & implementation of AI creating "blind spots" (Johnson. K. N, 2019, Lee. N. T, 2018, Wang. L, 2020, Martinez. C.F, Fernandez. A, 2020, Clifton. J, Glasmeier. A, Gray. M, 2020) that drives gender bias in AI.

Furthermore, AI enabled decision systems used in the recruitment software are found to be biased towards women according to a recent report of the Division of gender equality, UNESCO (2020) (Nadeem, Abedin, Marjanovic 2020). As AI algorithm are the reflections of the biased data (that comes from years of previous resumes) on which they are trained, hence AI systems are expected to yield biased outcome (World Economic Forum 2019; Galleno et al. 2019, Nadeem, Abedin, Marjanovic 2020). As organizations are relying on Al enabled decision tools for talent recruitment, talent sourcing, and candidate screening and engagement it is crucial to ensure that the decision taken by AI systems are not biased towards a certain group of people (Mehrabi et al. 2019; Jobin, Lenca & Vayena 2019).

Moreover, gender bias in AI is a complex and tricky matter, requiring attention from not just the technological aspect but also from the managerial aspects for dealing with data, people, and algorithms. Moreover, certain regulations, laws, and policies regarding fairness awareness/education if enhanced will improve the validation of the AI systems against gender bias and other discriminations. Given the above, this paper aims to answer the following research question:

What managerial practices are useful for organizations for mitigating gender bias in AI?

To answer this question, this study conducts a systematic literature review, following Webster & Watson's (2002) literature review method. The findings of systematic literature review (SLR) contribute to the emerging body of the literature on AI in Information Systems (IS) and beyond by identifying and categorizing the insights on managerial practices for mitigating gender bias in AI. As such, this study paves a way for a more comprehensive study of gender bias in AI through, for example, experts' interview in a particular context. It could also offer insights to data practices for developers and managerial practices for managers of AI to employ in order to avoid bias.

The remainder of this paper is organized as: section 2 presents the research design for this research along with the process of selection of related articles; section 3 presents the findings and discussion on managerial practices; section 4 & 5 presents the future work recommendations along with limitations and conclusion.

2 Research Methodology

According to Webster and Watson (2002), systematic literature reviews (SLR) thoroughly investigate research areas and opportunities for new research. As this area of research is relatively a new and emerging research field, therefore we conducted SLR by adopting Webster & Watson (2002) guidelines to acquire a better understanding of this phenomena by systematically analyzing the literature.

The process of selection and identification of relevant articles was carried out through a rigorous method. The first step of this research included a thorough investigation of the appropriate keyword selection. The keywords finally selected and used for this research were: Artificial Intelligence, Machine learning, Analytics, Gender, fairness.

For this research, we used Scopus as a source of search. We looked at various disciplines while selecting the articles in order to grasp a wider perspective on this area of research i.e. we selected computer science and business management (including Information Systems), social sciences and psychology to cover the social and behavioral aspects of gender bias in AI. Further, this search was limited to papers written in English.

There were 3817 articles that were captured through the selected keywords. The filtration of articles started by applying source type and inclusion criteria. A total number of 882 articles were recovered that met the inclusion criteria. In this step, we considered only those papers that were directly dealing with gender bias in AI or the papers that discussed the procedures or practices for mitigating gender bias in AI ranging from technical approaches to managerial approaches. Therefore, we started by reading the titles of the identified articles. The total number of articles that were recovered through titles were136. After selecting the articles on the basics of their titles, we recovered the articles on the basis of their abstract, which came to 65(this number included articles on fairness in AI, gender bias, AI ethics, discrimination in AI and AI in HR). We then thoroughly read the full text of the 65 articles. In this step we excluded all the articles that were outside the scope of this research. Therefore only 31 papers were selected that were relevant to our research scope. As this area of research is a fairly new and emerging topic, therefore 31 articles are a good number for analyzing the past literature systematically.

The analyses of the articles were carried out in a step wise process which included reading the articles line by line and highlighting the phrases/sentences (called excerpts) that were relevant to this research (Wolfswinkel, Furtmeuller and Wilderom 2013). The identification of the concepts and themes was carried out by organizing, analyzing and coding the final set of articles by following the guidelines by Wolfswinkel et al. (2013). Open coding, axial coding and comparative analysis was carried out as recommended by Wolifswinkel et al. (2013) for the development of the themes and concepts. Further, the themes that had almost the same meaning and were used in the same context and perspective were merged into high-level themes and concepts for better understanding and discussion.

3 Outcomes and discussion of systematic Review

In this section, we present the key findings from systematic literature review (SLR). The SLR analysis confirms that the level of publication activity in this field started to increase from 2017 and increased considerably in 2020, which shows that this is an emerging and fast-growing research area. Moreover, this trend highlights that although fairness in AI has been under discussion for the past few years, little has been published in IS journals so far. The results indicate that the research on gender bias is not yet well established, which highlights a great potential for future research in this field.

Prior research illustrates that there is a need to better understand and identify what manifests and contributes to gender bias in AI and what approaches should be undertaken for addressing this matter. Therefore, at this stage we will briefly discuss the contributing factors behind gender bias in AI. Our recent research shows there are eight main themes relating to contributing factors of gender bias in AI: Biased training datasets, gender stereotyping, biased behaviors and decisions, AI amplifying the bias, lack of gender diversity in training data and developers, lack of AI regulations, contextual/ socio-economic factors and other external factors.

It is noted in literature that the training datasets are often biased due to improper practices i.e. over, under, or misrepresentation of certain groups (Hayes. P, Poel. I.V.D, Steen. M, 2020), historical biases and gender stereotyping (Ntoutsi et al., 2019, Johnson. K. N, 2019). Moreover, unfair patterns in datasets (Veale. M, Binns. R, 2017) such as the correlation of data of sensitive variables/features i.e. proxy variable (salary serving as a proxy of gender, zip code serving as a proxy of background) make their way into AI algorithm and result in biased outcomes and contributing factors behind gender bias in AI.

Furthermore, the absence of gender disparity in developers, data miners, and datasets incorporate bias during the training phase of the algorithm (Martinez. C.F, Fernandez. A, 2020, Johnson. K. N, 2019, Lee. N. T, 2018, Wang. L, 2020, Clifton. J, Glasmeier. A, Gray. M, 2020), creates "blind spots" that emerge over time and are often difficult to predict (Hoffmann. A. L, 2019) that needs further attention.

3.1 Managerial practices for mitigating gender bias in AI

We established six main managerial practices for mitigating gender bias in AI and they are: better fairness governance practices, continuous training on fairness and ethics for all stakeholders, collaborative organizational learning on fairness & demographic characteristics, interdisciplinary approach & understanding of AI ethical principles, Workplace diversity in managerial roles, designing strategies for incorporating algorithmic transparency and accountability & Ensuring Human in the loop as shown in table 1 in appendices. We will now discuss these managerial practices and their implications accordingly.

Algorithms sift through datasets (Hoffmann. A. L, 2019) and discover the trends/ patterns and make predictions; it is thus important to rely on better, faster and more ubiquitous algorithms to make sense of the big datasets (Martin. K, 2018). Perhaps designing managerial strategies for fair AI compliance and audits and updated regulations to maintain certain minimum standards for the datasets (Johnson. K. N, 2019) could be adopted for neutralizing the gender bias in AI.

It is noted in literature that managerial practices, such as organizations investing in hiring, training/ workshops of expert programmers for regular maintenance and vetting of datasets is essential for mitigating gender bias in AI (Noriega. M, 2020). This would require for the organizations to develop improved and modernized fair and ethical internal structures and corporate strategies to govern and manage the gender imbalance (Johnson. K. N, 2019).

Moreover, organizational strategies that could bring awareness on the ethical and responsible AI is very much needed; including giving importance to work force diversity in an organization; including policies pertaining to gender diverse workplace that bring cultural diversity in data will be beneficial in neutralizing the gender bias in AI (Lee. N. T, 2018). Enhanced women representation/ gender inclusion in the technology sector especially STEM career domain (Lambrecht. A, Tucker. C, 2019), gender diversity among the member of boards, management and leadership roles (Johnson. K. N, 2019) plus focusing on the "blind spots" (Hoffmann. A. L, 2019) that are created by the lack of gender diversity in data and developers, will minimize the homogeneous and exceptionally male-dominated leaderships and decisions (Johnson. K. N, 2019) would offer a pathway towards mitigating gender bias in AI.

Additionally, emphasis on improved business model innovation addressing gender equity and fairness in AI should be designed (Arrieta et al., 2020, Feuerriegel. S, Dolata. M, Schwabe. G, 2020); including firms addressing fairness in the overall culture instead of focusing on filling the requirement of diversity quotas. Regular data audit practices for identifying datasets that correlate with protected characteristics and may therefore serve as a proxy for an attribute of protected classes (Johnson. K. N, 2019) would neutralize the misrepresentation of gender in the data and also in AI decisions.

It is noted in past literature that Algorithms for AI enabled decision systems should be designed to yield fair decisions. Developers should not only be responsible for ethical implications but they should shift algorithmic decision-making responsibility to the users as well (Martin. K, 2018). Practices and strategies that preclude individuals to take responsibility within a decision should be replaced with giving autonomy to users in decision making to bring fairness in the AI decisions (Martin. K, 2018, Hayes. P, Poel. I.V.D, Steen. M, 2020). Likewise, organizations should also deploy efficient quality control and assurance policies for better and improved algorithmic accountability and transparency for neutralizing gender bias in AI (Ntoutsi et al., 2019).

4 Directions for future research

Our findings indicate some future IS research related to prevention, mitigation, and future theorization of gender bias in AI. Following are a few of the suggested future directions in this area of research:

Research on AI policies and regulations to bring justice to society and in AI fair design should be enhanced. Including but not limited to exploring ways how organizations can ensure diversity in the workplace and how organizations can bring autonomy to users in AI enabled decision systems. Also, future work should focus on investigating and theorizing gender bias in AI (Ivaturi & Bhagwatwar 2020; Sen 1995).

It is noted in a recent survey, 38% of organizations are already using AI in their workplace with 62% expected to be using it near future (Martínez, Fernández 2020). Therefore, organizations need to deploy certain mechanisms to deal with gender bias in AI. Users of AI should use their own intuition while using AI tools. Therefore, organizations need to re-consider their managerial approaches including designing innovative business models and managerial strategies focusing on mitigating gender bias in AI.

Therefore, as a follow up of this paper, our next step in this research would be to collect empirical data by conducting experts' interviews from those subjects who are directly involved in managing AI in an organization and posits a broader overview of organizational AI-enabled decision systems. Empirical data will explore the current managerial practices and mechanism that are being carried out by organizations in this regard and also will present the guidelines on practices for organizations for better managing of gender bias in AI.

Experts' interviews. Interviews are considered as a primary data source (Tim et al. 2016, Myer & Newman 2007). Experts' interview would not only validate our SLR findings from the real-world perspective – it will present the insights from the experts' who have vast and extensive and precise knowledge and experience in this field and are able to manager AI within an organization (Merge, Edelmann, Haug 2019). An expert interview would provide this research with an in-depth insight from the user as well as from organizational perspective; also, would present the framework on strategies and practical approaches that would be useful for managing gender bias in AI.

5 Conclusion

Past data is mostly reflection of history and AI algorithms trained on past datasets could be discriminatory towards a certain group or individuals due to the various underlying drivers and factors. Hence people suffer from algorithms harm and there is no accountability for it.

Bias is externalized and includes misguided conducts of "bad actors" which are either intentional or accidental and might not be easily traceable; therefore, such social and contextual issues are left beyond the law's reach (Hoffmann. A. L, 2019) and thus are deprived aspect of the society since long.

Given the above, this paper aims to focus on unpacking the managerial approaches for mitigating gender bias in AI. In this research, we have presented a deeper understanding of gender bias in AI and have provided evidence from the systematic literature review that gender bias does exist in AI systems and the practices/ approaches required for addressing gender bias in AI. This research gives a concise overview of findings of gender bias in AI from past literature; in terms of what has been discussed and investigated and what needs to be researched in the future.

As the roots of gender bias in AI are not just technological, and as such technological solutions might not suffice; AI enabled decisions systems are being made on mathematical model that leads to a biased outcome. There has to be some checks for assurance against gender bias when making decisions through AI enabled decision systems. Hence, organizations need to follow some mechanisms and strategies to mitigate and address this matter timely and effectively. This research therefore presents a set of managerial practices, approaches and recommendations for organizations for better governance and management of gender bias in AI.

Due to time constraints we were not able to present empirical findings and their analyses with the lens of a theory, which is one of the limitations of this short paper. However, we aim to publish our empirical outcomes in near future.

References

1. Hoffmann. A.L. (2019). Where fairness fails: data, algorithms, and the limits of anti-discrimination discourse. *Information, communication & society*, 22 (7), 900-915.
2. Grari. V., Ruf. B., Lamprier. S., Detyniecki. M. (2020). Achieving fairness with decision tress: An adversarial approach. *Data science and engineering*, 5(2).
3. Martinez. C. F., Fernandez. A. (2020). AI and recruiting software: Ethical and legal implications. *Journal of behavioral robotics*, 11 (1).
4. Costa. P., Ribas. L. (2019). AI becomes her: Discussing gender and artificial intelligence. A journal of speculative research, 17 (1-2), 171-193.
5. Bellamy et al., (2018). AI fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. *Journal of research and development*.
6. Hayes. P., Poel. I.V.D., Steen. M. (2020). Algorithms and values in justice and security. *AI & Society*, 35 (3), 533- 555.
7. Schonberger. D. (2019). Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications. *International Journal of law and Information Technology*, 27 (2), 171-203.
8. Prates. M., Avelar. P., Lamb. L.C. (2019). Assessing gender bias in machine translation – A case study with google translate. *Neural computing and applications*.
9. Kyriazanos. D. M., Thanos. K.G., Thomopoulos. S.C.A. (2019). Automated decisions making in airports checkpoints: Bias detection toward smarter security and fairness. Automated security decision-making. *IEEE Security & applications*, 17 (2), 8-16.
10. Johnson. K.N. (2019). Automating the risk of bias. *George Washington law review*, 87 (6).
11. Lambrecht. A., Tucker. C. (2020). Algorithmic bias? An empirical study of apparent gender bias discrimination in the display of STEM career ads. *Management Science*, 65 (7), 2966-2981.
12. Ntoutsi et al., (2019). Bias in data-driven artificial intelligence systems – An introductory survey. *Data mining and knowledge discovery*, 10 (3).
13. Ibrahim. S.A., Charlson. M.E., Neill. D.B. (2020). Big data analytics and the structure for equity in healthcare: The promise and perils. *Health equity*, 4 (1), 99- 101.
14. Chen. I. Y., Szolovits. P., Ghassemi. M. (2019). Can AI help reduce disparities in general medical and mental health care? *AMA Journal of Ethics*, 22 (2), 167- 179.
15. Qureshi. B., Kamiran. F., Karim. A., Ruggieri. S., Pedreschi. D. (2020). Causal inference for social discrimination reasoning. *Journal of Intelligent Information Systems*, 54, 425-437.
16. Robert. L.P., Pierce. C., Marquis. L., Kim. S., Alahmad. R. (2020). Designing fair AI for managing employees in organizations: a review, critique, and design agenda. *Human-computer interaction*, 35 (5-6), 545- 575.
17. Lee. N. T. (2018). Detecting racial bias in algorithms and machine learning. *Journal of Information, communication, and ethics in society*, 16 (3), 252-260.
18. Martin. K. (2019). Ethical implications and accountability of algorithms. *Journal of business ethics*, 160, 835- 850.
19. Wu.W., Huang. T., Gong. K. (2019). Ethical principles and governance technology development of AI in China. *Engineering*, (6), 302-309.
20. Piano. S.L. (2020). Ethical principles in machine learning and artificial intelligence: cases from the field and possible ways forward. *Humanities and social sciences communications*, 9.
21. Miron. M., Tolan. S., Gomez. E., Castillo. C. (2020). Evaluating causes of algorithmic bias in juvenile criminal recidivism. *Artificial law and intelligence*.

22. Arrieta et al., (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities, and challenges towards responsible AI. *Information fusion*, 58, 82-115.
23. Feuerriegel. S., Dolata. M., Schwabe. G. (2020). Fair AI: Challenges & opportunities. *Business Information Systems Engineering,* 62 (4), 379-384.
24. Veale. M., Binns. R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big data and society*, 1-17.
25. Berk. R., Heidari. H., Jabbari. S., Kearns. M., Roth. A. (2018). Fairness in criminal justice risk assessments: The state of the art. *Sociological methods and research*.
26. Thelwall. M. (2017). Gender bias in machine learning for sentiment analysis. *Online information review*, 42 (3), 343- 354.
27. Paulus.J.K., Kent. D. M. (2020). Predictably unequal: Understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. *Digital medicine*, 99 (3).
28. Cirillo et al., (2020). Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. *Digital medicine,* 8 (3).
29. Noriega. M. (2020). The application of artificial intelligence in police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, cooperation, and false confessions. *Futures*, 117.
30. Wang. L. (2020). The three harms of gendered technology. *Australasian Journal of Information Systems*, 24.
31. Ahn. Y., Lin. Y.R. (2020). Fairsight: Visual analytics for fairness in decision making. *IEEE transactions on visualization and computer graphics*, 26 (1).
32. Clifton. J., Glasmeier. A., Gray. M. (2020). When machines think for us: the consequences for work and place. *Cambridge Journal of regions, economy, and society*, 13 (1), 3-23.
33. Webster, J., Watson, R.T. 2002, Analysing the past to prepare for the future: Writing a literature review, *MIS Quarterly*, 26(2), 3-23
34. UNESDOC Digital library, Artificial intelligence, and gender equality: key finding of UNESCO's global dialogue. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000374174
35. Australian Academy of Science. (2019) Women in STEM Decadal Plan (Australian Academy of Science)
36. Agarwal, P., 2020, "Gender bias in STEM: Women in Tech still facing discrimination", *Forbes*.
37. Altman. M, Wood, A., Vayena, E. 2018, "A harm-reduction framework for algorithmic fairness", *AI Ethics.*
38. Arrieta et al., 2020. "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information fusion*, vol. 58, pp. 82-115.
39. Bentivogli et al., 2020. "Gender in danger? Evaluating speech translation technology on the Must- She Corpus", *58th Proceeding of association for computational linguistic*.
40. Brunet, M.E., Houlihan. C.A., Anderson, C., Zemel, R. 2019, "Understanding the origins of bias in word embedding", *Proceedings of the 36 the International Conference on Machine Learning, Long Beach,* California.
41. Bolukbasi, T., Chang, K.W., Zou, J., Saligrama, V., Kalai, A. 2016, " Man is to computer programmer as women is to homemaker ? Debiasing word embeddings", *Cornell University computer science and artificial intelligence*.
42. Bellamy et al., 2018 "AL fairness 360: An extensible toolkit for detecting, understanding and mitigating unwanted algorithmic bias", *Computer science,* 2018.

43. Berger, K., Klier, J., Klier, M., Probst, F. 2014. "A review of information systems research on online social networks", *Communications of the Association for Information Systems*, vol. 35, no. 8, pp. 145-172, 2014.
44. Beard, M., Longstaff, S. 2018. "Ethics by design: Principles for good technology", The ethics center.
45. Blodgett et al., 2020. "Language (Technology) is power: A critical survey of bias in NLP", *Computational and language*.
46. Canetti et al., 2019, "From soft classifiers to hard decisions: How far can we be?" *Processing of the conference on fairness, accountability, and transparency*, p: 309-318.
47. Croeser, S., Eckersley, P. 2019. "Theories of parenting and their application to artificial intelligence", *Computers and society*.
48. Crawford, K. (2016, June 26). A.I.'s White Guy Problem. (Sunday Review Desk) (OPINION). *The New York Times*.
49. Dwivedi. Y.K et al., 2019, "Artificial intelligence (AI): Multidisciplinary perspective on emerging challenges, opportunities, and agenda for research, practice, and policy", *International Journal of Information Management*.
50. Daugherty, P., Wilson, H., Chowdhury, R. 2018, "Using artificial intelligence to promote diversity", *MIT Sloan Management Review*.
51. Dawson D and Schleiger E, Horton J, McLaughlin J, Robinson C∞, Quezada G, Scowcroft J, and Hajkowicz S (2019) Artificial Intelligence: Australia's Ethics [17] Framework. Data61 CSIRO, Australia.
52. Edwards, J. S., Rodriguez, E. 2019. "Remedies against bias in analytics systems", *Journal of Business Analytics*, vol. 2, issue. 1.
53. Feast, J., 2019. "4 ways to address gender bias in AI", Harvard Business Review.
54. Font, J., Costa-Jussa, M.R. 2019. "Equalizing gender biases in neural machine translation with word embedding techniques", *Computational and language*.
55. Florentine, S. 2016. "How artificial intelligence can eliminate bias in hiring", CIO.
56. Galleno, A., Krentz, M., Tsusaka, M., Yousif, N. 2019, "How AI could help or hinder women in the workforce", *Boston Consulting Group.*
57. Gonen, H., Goldberg, Y. 2019. "Lipstick on a pig: Debasing methods cover up systematic gender bias in words embeddings but do not remove them", *Proceeding of association for computational linguistics*, Minnesota.
58. Gummadi, K.P., Heidari, H. 2019, "Economic theories of distributive justice for fair machine learning", *Companion proceedings of 2019 worldwide web conference*.
59. Holstein et al., 2019, "Improving fairness in machine learning systems: What do industry practitioners need?", *ACM CHI Conference on human factors in computing sciences*.
60. Huang et al., 2020. "Historical comparison of gender inequality in scientific careers across countries and disciplines", *Proceedings of the National academy of Sciences of the U.S.A.*
61. Ivaturi, K., Bhagwatwar, A. 2020, "Mapping sentiments to themes of customer reactions on social media during a security hack: A justice theory perspective", *Information and Management*, vol. 57, issue. 4, pp.
62. Jobin, A., Lenca, M., Vayena, E. 2019, "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, vol. 1, pp. 389-399.
63. J.F Wolfswinkel, E. Furtmueller and C. P. M. Wilderom. 2013, "Using grounded theory as a method for rigorously reviewing literature", *European Journal of Information Systems*, 22(1), 45-55.
64. Kumar, G., Singh, G., Bhatanagar, V. 2019, "Scary side of artificial intelligence: A perilous contrivance to mankind", *Humanities and social sciences review*, vol. 7, issue. 5.

65. Kulik, C.T., Lind, E.A., Ambrose, M.L., Maccoun, R.J. 1996, "Understanding gender differences in distributive and procedural justice", *Social justice research,* vol. 9, issue. 4
66. Leavy, S. 2018, "Gender bias in artificial intelligence: The need for diversity and gender theory in Machine learning", *2018 IEEE/ACM First international workshop on gender equality in software engineering, Gothenburg,* Sweden.
67. Lambrecht, A., Tucker, C. 2018. "Algorithmic bias? An empirical study into apparent gender-based discrimination in the display of STEM career ads". Available at SSRN: https://ssrn.com/abstract=2852260 or http://dx.doi.org/10.2139/ssrn.2852260
68. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A. 2019, "A survey on bias and fairness in machine learning", *Computer science- Machine learning. Available at SSRN: https://arxiv.org/abs/1908.09635*
69. Mikolov et al., 2013, "Distributed representation of words and phrases and their compositionality", *Proceeding of the 26th International Conference on Neural Information Processing Systems,* vol. 2, pp. 3111-3119.
70. Noriega, M. 2020. "The application of police interrogations: An analysis addressing the proposed effect AI has on racial and gender bias, corporation and false confessions*", Futures,* vol. 117.
71. Parsheera, S. 2018, "A gendered perspective on artificial intelligence", *Machine learning for a 5G future.*
72. Parikh, R.B., Teeple, S., Navathe, A.M.2019, "Addressing bias in artificial intelligence in health care", *JAMA.*
73. Robnett, R. D., 2015, "Gender bias in STEM fields: Variation in prevalence and links to STEM self-concept", *Psychology of women quarterly*.
74. Ridley, G., & Young, J. (2012). "Theoretical approaches to gender and IT: examining some Australian evidence". *Information Systems Journal*, 22(5), 355–373.
75. Srivastava, B., Rossi, F. 2018, "Towards Compostable bias rating of AI service", *AAAI/ACM Conference on AI, Ethics and Society,* New Orleans, Louisiana.
76. Sen, A. 1995, 'Gender inequalities and theory of justice', in Nussbaum & Glover (eds), Women, culture, and development: A study of human capabilities", *Oxford university press,* New York.
77. Sun et al., 2019. "Mitigating gender bias in Natural Language Processing: A literature review", 57th *proceeding of association for computational linguistics,* Italy.
78. Trewin, S. 2018, "AL fairness for people with disabilities: Point of view", *Computer science*.
79. Terrell et al., 2017. "Gender differences and bias in open source: pull request acceptance of women versus men", *Peer J computer science*.
80. Myer, M, D., Newman, M. 2007, "The qualitative interview in IS research: Examining the craft", *Information and Management,* 7(10).
81. Mergel, I., Edelmann, N., Haug, N. 2019, "Defining digital transformation: Results from the expert interview", Government Information Quarterly, 36(4).
82. Nadeem, A., Abedin, B., & Marjanovic, O. 2020, "Gender bias in AI: A review of contributing factors and mitigating strategies", ACIS 2020 proceedings. 27. https://aisel.aisnet.org/acis2020/27/
83. Zhao et al., 2019. "Gender bias in contextualization word embedding", *proceeding of association for computational linguistics*. Minnesota.
84. Zhong, Z. 2018, "A tutorial on fairness in machine learning", *Towards data science*

Appendices

Table 1. Managerial practices for mitigating gender bias in AI

Grouping of concepts	**Concepts for mitigating gender bias in AI**	**Description**
Better fairness governance policies	Internal governance policies (Johnson. K. N, 2019) Internal structures and process-oriented corporate governance (Johnson. K. N, 2019, Martin. K, 2018)	Enhanced AI corporate governance for gender bias mitigation
Continues education/training on fairness and ethics for all stakeholders	Educational workshops and training on workplace fairness (Noriega. M, 2020) Certified professional required (Martin. K, 2018) Awareness of ethics and promoting responsible AI (Wu. W, Huang. T, Gong. K, 2019, Veale. M, Binns. R, 2017) Awareness of unintended bias in scientific community and technology industry (Cirillo et al et al., 2020)	Workshops/education that involves principles of ethics such as promoting ethical education for every stakeholder in AI research & development
Collaborative organizational learning on fairness & demographic characteristics	Business models and policy should be designed concerning fair AI (Feuerriegel. S, Dolata. M, Schwabe. G, 2020)	Design of business models and policies to consider AI principles.
Interdisciplinary approach & understanding of AI ethical principles	Interdisciplinary disciplines to work collaboratively to address ethical challenges (Wu. W, Huang. T, Gong. K, 2019, Ibrahim. S.A, Charlson. M.E, Neill. D.B, 2020)	Employment of a more diverse IT workforce to be included in the design and implementation of algorithms.
Workplace diversity in managerial roles	Gender diversity at managerial levels (Lee. N. T, 2018) Diversity in the development of AI systems (Costa. P, Ribas. L, 2019, Johnson. K. N, 2019, Ntoutsi et al., 2019, Arrieta et al., 2020, Clifton. J, Glasmeier. A, Gray. M, 2020) Gender diversity in the high-tech industry and STEM career (Lee. N. T, 2018, Johnson. K. N, 2019, Wang. L, 2020)	An increase in gender inclusion in the development of AI technologies will introduce diverse perspectives.
Designing strategies for incorporating algorithmic	Big data review board required (Martin. K, 2018)	AI audits to be conducted periodically to ensure AI compliance.

transparency and accountability	Incorporate regular audits of the data (Martinez. C.F, Fernandez. A, 2020, Johnson. K. N, 2019, Ibrahim. S.A, Charlson. M.E, Neill. D.B, 2020, Robert et al., 2020, Piano. S. L, 2020, Veale. M, Binns. R, 2017, Noriega. M, 2020) Designing strategies for fairness and ensure accountability (Hayes. P, Poel. I.V.D, Steen. M, 2020)	
Ensuring Human in the loop	Integrating human & AI decision making (Miron et al., 2020)	Design strategies like providing more autonomy to the users in decision-making would bring fairness to the AI decisions.