2020

Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies

Ayesha Nadeem
University of Technology Sydney, ayesha.nadeem@student.uts.edu.au

Babak Abedin
Macquarie Business School, babak.abedin@mq.edu.au

Olivera Marjanovic
University of Technology Sydney, olivera.marjanovic@uts.edu.au

Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies

Completed research paper

Ayesha Nadeem
School of Professional Practice and Leadership
Faculty of Engineering & IT
University of Technology Sydney
Email: Ayesha.Nadeem@student.uts.edu.au

Babak Abedin
Macquarie Business School
Macquarie University
Email: Babak.abedin@mq.edu.au

Olivera Marjanovic
School of Professional Practice and Leadership
Faculty of Engineering & IT
University of Technology Sydney
Email: Olivera.Marjanovic@uts.edu.au

Abstract

The impact of artificial intelligence (AI) is significant in almost every industry. As many important decisions are now being automated by various AI applications, fairness is fast becoming a vital concern in AI. Moreover, the related literature and industry press suggest that AI systems are often biased towards gender. Thus, there is a need to better understand the contributing factors behind gender bias in AI, along with the current approaches taken to address it. Therefore in this paper, we aim to contribute to the emerging IS literature on AI by presenting a consolidated picture of the most often discussed contributing factors and approaches taken in relation to gender bias in AI in the multidisciplinary literature. Our findings indicates that the more frequently discussed contributing factors include lack of diversity in both data and developers, programmer bias, and the existing gender bias in society, now amplified through AI. Additionally, our findings indicate the most discussed approaches for addressing gender bias in AI include the implementation of diversity in society and data and fairness in AI development, as well as reducing bias in algorithms. Based on our findings, we indicate some future IS research for the better development of AI systems.

Keyword Bias in AI, Fairness and AI, Ethics in AI

1 Introduction

Artificial intelligence (AI) applications are penetrating all aspects of our daily lives, from shaping people's opinions to influencing their behaviours in daily routines (Leavy 2018). AI algorithms are starting to influence important decisions related to hiring employees, loan applications, the advertisements/ news people see, healthcare, etc. (Mehrabi et al. 2019). At the same time, AI is creating new ethical and research challenges requiring our urgent attention. According to Intel, the key areas of concern for AI policy are accountability, fairness, and human employment (Dwivedi et al. 2019).

Indeed, in the recent years there has been an outburst of research on fairness in AI algorithm (Bellamy et al. 2018; Zhong 2018; Leavy 2018; Jobin, Lenca & Vayena 2019). As Dwivedi et al. (2019) argue, it is imperative to study the fairness in AI algorithms as their use is not only limited to industry applications, but has entered our lives on daily basis. While the concept of fairness is very broad, gender-related fairness is considered an essential aspect of fairness (Leavy 2018).

To address the problem, AI ethical guidelines have been proposed and widely discussed by many government agencies and other organizations in many countries (Jobin, Lenca & Vayena 2019). Yet, the current solutions proposed to meet the challenges of ethical AI are significantly divergent (Jobin, Lenca & Vayena 2019).

However, there remain gaps and uncertainly as to which principle to be prioritized and the conflicts between the principles; moreover, these gaps could hardly be solved through technical expertise (Jobin, Lenca & Vayena 2019). However, research on AI bias is still emerging. In particular, there are still research gaps regarding our understanding of gender bias in AI applications, particularly what causes gender bias in AI, mitigation, and possibly prevention.

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). The biased and adverse outcomes of algorithm decisions reach beyond the individuals; harmful effects reach families, communities, and society at large (Altman, Wood & Vayena 2018). Gender bias has been observed at various levels in society – starting from within a family to education and employment. Women are generally considered underutilized resource with the abilities and skills needed for the IT industry – which limits their career choice as well (Ridley & Young 2012). For example, according to the UNESCO 2019 report, there is less than 1 % women applications for technical AI jobs, estimated by the recruiters of technology companies in silicon valley. Therefore gender bias which could be easily identified in the society is also slowing penetrating in the emerging technology i.e. AI. Thus it is important to scrutinize the underlying gender bias in AI for bringing fairness in the AI which is one of the crucial principles of AI ethical principles. The concern regarding the gender bias in AI has also been raised by the government and the research academia (Parikh, Teeple & Navathe 2019; Feast 2019; Parsheera 2018; Agarwal 2020).

Moreover, AI 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). As the data on which AI algorithm is being trained is from years of previous resumes, hence AI systems are expected to yield biased outcome (World Economic Forum 2019; Galleno et al. 2019). Moreover, as organizations are nowadays more and more relying on Al tools for talent recruitment, talent sourcing, and candidate screening and engagement. Therefore, 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).

Given the above, this paper aims to unpack what gender bias in AI is about, what are the contributing factors pertaining to gender bias in AI, and what approaches could be used to mitigate the gender bias in AI. Focusing on the multidisciplinary literature, in this paper, we seek to answer the following research question:

What gender bias in AI entails?

What are the contributing factors of gender bias in AI?

What approaches are used to mitigate gender bias in AI?

To answer these questions, we reviewed the multidisciplinary literature, following Webster & Watson's (2002) literature review method. Our findings show that gender bias in AI is manifested in various vital services (employment, medical health, mortgage lending), new applications of technology (autonomous vehicles), and justice systems (judicial trails).

These findings contribute to the emerging body of the literature on AI in IS and beyond by identifying and categorizing the insights about manifestations and contributing factors of gender bias in AI, as well as the approaches used to mitigate it. As such, this study paves a way for a more comprehensive study of gender bias in AI through, for example, case studies in a particular context. It could also offer insights to developers and managers of AI solutions about the issues reported in the literature that could be used to inform their practice.

2 Research approach

While the research in gender bias in AI is growing, still very few researchers have investigated the contributing factors behind gender bias in AI and the approaches used to address it. To bridge the gaps mentioned in the literature pertaining to gender bias in AI and to answer the research questions of this study, an in-depth investigation of the multidisciplinary literature review of the relevant research articles was conducted. In this paper, we have adopted the literature review research method by Webster & Watson (2002). Hence, the process of selection and identification of relevant articles was carried out through a rigorous method as shown below (Figure 1). This process enabled us to classify the relevant articles in a step-wise process.

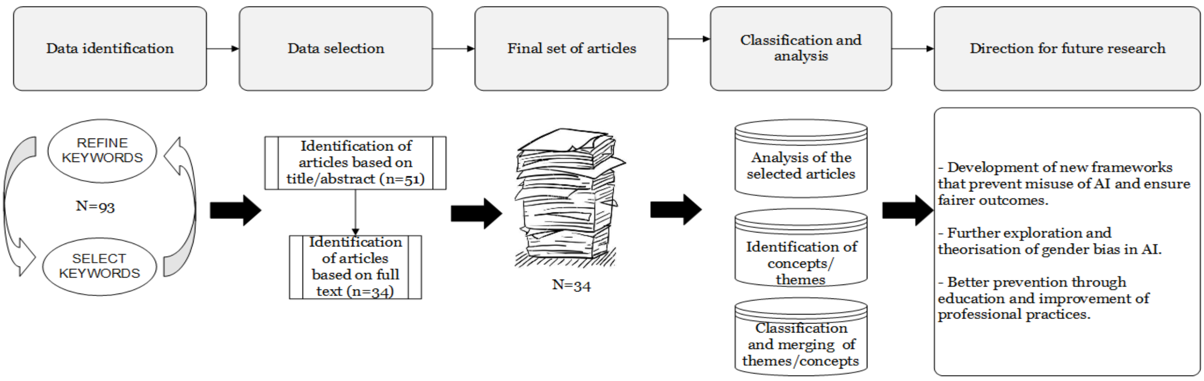


Figure 1. Research design and approach

2.1 Data Identification

The method for the selection of relevant articles started through the multidisciplinary literature review. The literature was thoroughly investigated by selecting and refining the appropriate keywords through the database search. The database used for this research was: Google scholar. The keywords that were used for this research were: Bias in AI, Fairness and AI, Ethics in AI. In this process of data identification, some papers were shown more than once in the result; therefore they were omitted from the final set of identified papers.

2.2 Data Selection and Classification

In this step, we considered only those papers that were directly dealing with gender bias in AI or the papers that discussed the procedures for mitigating gender bias in AI ranging from technical approaches that eliminate the bias from the algorithm or techniques that would be helpful in future in regards to addressing gender bias in AI. Therefore, we started by reading the titles and abstracts of the identified articles. After selecting the articles on the basics of their titles and abstract, we thoroughly read the full text of the articles. In this step, we excluded all the papers that were outside the scope of this research. Therefore only 34 papers were selected that were relevant to our research scope.

The data classification of the literature was based on the concept presented by Webster and Watson (2002). The identification of the concept and themes was carried out by reading the articles. 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.

2.3 Direction for Future Research

Through the in-depth analysis and synthesis of the relevant literature, the agenda for future and the gaps mentioned in the literature were also noted, as recommended by Webster and Watson (2002).

3 Findings and Discussion of Literature Analysis

In this section, we discuss the outcomes of the multidisciplinary literature review. Firstly, we identify what manifests and contribute to gender bias in AI, along with the approaches used to mitigate it. The term characteristics have been used in this paper to reflect the feature of gender bias in AI. The characteristics of gender bias in AI with their description and main themes are offered in Table 1.

Secondly, we present the contributing factors of gender bias in AI along with their description and main themes in Table 2, and thirdly the approaches that have been mentioned in past literature for mitigating the gender bias in AI have been compiled in Table 3. We grouped and categorized the characteristics, contributing factors, and approaches that appeared to be the same for a better understanding of this research phenomenon. Further, the graphical representation of the contributing factors themes and approaches themes for addressing gender bias in AI is displayed in Figure 2.

#	Characteristics of gender bias in AI	Source	Research Method/ Sector	Groups of characteristics	Description
1	Unethical practices	Wang 2020/ Australian Journal of Information Systems	Qualitative / case study /Vital services	Unethical practices	Gendered technology displays the lack of adherence to ethical values as it builds harm against a certain group, community, or individual (Wang 2020) and many applications not only replicates the bias but also amplifies it in the training data (Zhao et al. 2019)
2	Real with serious repercussions	Gallego et al. (2019)/ Boston Consulting group	Qualitative / Vital services		
3	Effect several applications	Zhao et al. 2019/ Proceedings of Association for Computational Linguistics	Qualitative /Natural language processing		
4	A complex issue which requires interdisciplinary collaboration	Sun et al. 2019/ Proceedings of computational linguistic	Qualitative /Natural language processing	Interdisciplinary concern	Gender bias in data arises from the level through which various languages express males and females of a referred human entity

5	Growing interdisciplinary concern	Bentivogli et al. 2020/ Proceedings of the association for computational linguistics	Qualitative / Speech translation		(Bentivogli et al. 2020)Interdisciplinary understanding of encoding and human biases as properties of data problem and biases are different across different fields (Sun et al. 2019).
6	Prejudice towards a certain gender	Sun et al. 2019/ Proceedings of computational linguistics	Qualitative analysis/ ELMO's	Gender stereotyping	Associating some characteristics to a certain gender because of the prejudice or preference towards one gender over the other (Sun et al. 2019).
7	Reinforces damaging stereotype	Sun et al. 2019/ Proceedings of computational linguistics	Qualitative /Natural language processing		

Table 1. Characteristics of gender bias in AI

We established 3 main characteristics of gender bias in AI that are: unethical practices, interdisciplinary concern, and gender stereotyping. Out of 34 articles, only 5 articles discussed the characteristics of gender bias in AI, as shown in the above table. Our findings further emphasize the fact already mentioned in the literature that gender bias in AI is embedding the already existent concept of gender stereotyping in the society by yielding biased outcomes against women (Sun et al. 2019) which is unethical (Wang 2020). Additionally, it is an interdisciplinary concern that is a sociological and engineering problem that requires an interdisciplinary understanding of encoding and human biases across different fields (Sun et al. 2019).

No	Contributing factors of gender bias in AI	Description	Grouping of themes
1	Lack of culturally diverse developers and designing of data	Training data if biased then the resulting outcome of AI will be affected as well (Gallego et al. 2019; Lambrecht & Tucker 2018; Parsheera 2018; Blodgett et al. 2020; Bentivogli et al. 2020; Font & Costa-Jussa 2019; Thelwall 2017; Arrieta et al. 2020; Edwards & Rodriguez 2019; Srivastava & Rossi 2018; Holstein et al. 2019; Brunet et al. 2019; Gonen & Goldberg 2019; Bolukbasi et al. 2016; Wang 2020; Parsheera 2018; Leavy 2018). Moreover, lack of culturally diverse developers and training data leads to gender bias in AI (Wang 2020; Parsheera 2018; Blodgett et al. 2020; Terrell et a. 2017; Dawson et al. 2019).	Lack of diversity in training data and developers
2	Gender stereotyping	Gender bias ingrained so much in society that even if a women beliefs herself as a feminist, she still unwittingly share prejudices against women (Wang 2020; Blodgett et al. 2020; Font & Costa-Jussa 2019; Leavy 2018; Brunet et al. 2019; Bolukbasi et al. 2016; Lambrecht & Tucker 2018; Parsheera 2018; Gibbs	Bias in society

		2016; Huang et al. 2020; Terrell et al. 2017; Bentivogli et al. 2020; Zhao et al. 2019; Thelwall 2017; Noriega 2020; Croeser & Eckersley 2019; Leavy 2018; Holstein et al. 2019; Brunet et al. 2019, Sun et al. 2020)	
3	Programmer bias conscious or unconscious	Programmer bias conscious or unconscious also seeps into the algorithm (Lambrecht & Tucker 2018; Parsheera 2018; Gibbs 2016; Beard & Longstaff 2018)	Bias in data due to the algorithm
4	Gender bias in everyday language	Blind adoption of word embedding techniques one of the reason causing gender bias (Parsheera 2018; Prates, Avelar & Lamp 2018; Huang et al. 2020; Zhao et al. 2019; Leavy 2018; Brunet et al. 2019; Gonen & Goldberg 2019; Bolukbasi et al. 2016).	
5	Language discrimination in data for gender	Language distinguish in data between females and males grammatically enforce a bias in the person's perception in the world (Prates, Avelar & Lamb 2018; Huang et al. 2020; Leavy 2018).	
6	Economic factors	Economic factors could be one of the reasons distorting AI decisions (Lambrecht & Tucker 2018; Sun et al. 2019)	Economic factors
7	Discriminatory actual behaviour	The AI algorithm learns the gender bias from the behaviour of the users (Lambrecht & Tucker 2018; Blodgett et al. 2020)	Biased behaviours and decisions
8	Historical discrimination	Historical discrimination against women is associated with bias as it is perpetuated in the algorithm and eventually amplifies the bias in society (Lambrecht & Tucker 2018)	

Table 2. Contributing factors of gender bias in AI

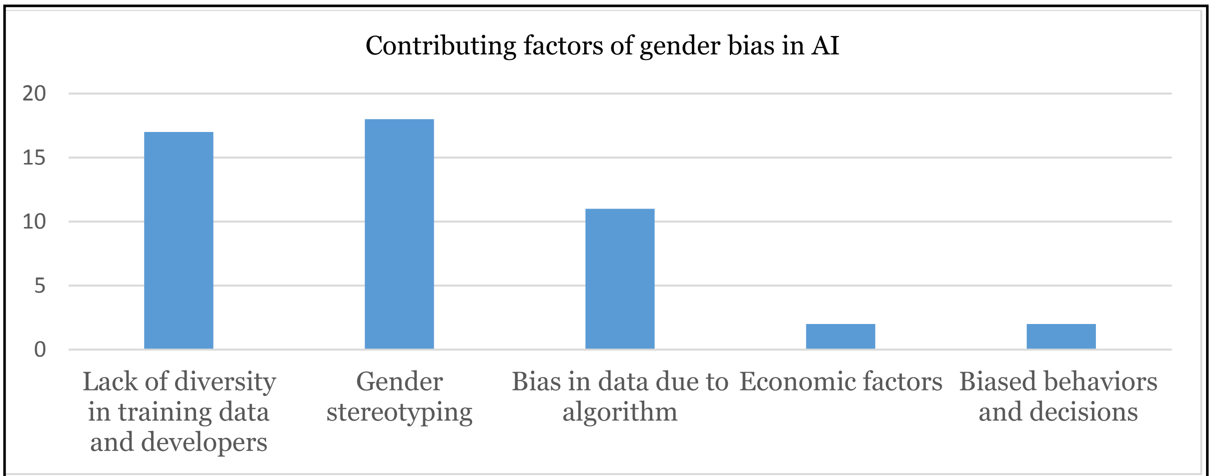
We established 5 main themes relating to contributing factors of gender bias in AI they are: Lack of diversity in training data and developers, Bias in society, the bias in data due to the algorithm and conscious or unconscious programmer bias, economic factors, and biased behaviours and decisions. Among them, the most discussed contributing factors are lack of diversity in data and developers, gender stereotyping and bias in data due to programmer conscious or unconscious bias as displayed in Figure 2.

No	**Approaches for addressing gender bias**	**Description**	**Grouping of themes**
1	Regulated policy in technological development	Taking into account the ethical consequences and regulated policy for the development of regulated AI systems i.e. Awareness, education, training, workshops, etc. (Wang 2020; Lambrecht & Tucker 2018; Parsheera 2018; Noriega 2020; Dawson et al. 2019).	Fair and ethical design and implementation of AI applications.
2	Designing ethical and fairness testing techniques	Investment in research and development of technical tools that can bring the concept of ethics and fairness principles to practice (Parsheera 2018; Croeser & Eckersley 2019; Dawson et al. 2019; Cho et al. 2019; Mehrabi et al. 2019; Leavy 2018; Holstein et al. 2019).	
3	Algorithmic transparency	The decisions that are being made by the algorithm should be visible and transparent to the users (Thelwall 2017; Arrieta et al. 2020).	

4	Trade-off	The trade-off between reducing the apparent bias & using economic mechanisms for allocating resources efficiently through an algorithm (Lambrecht & Tucker 2018; Sun et al. 2019)	Reducing bias from the algorithm
5	Post-processing word embedding/ gender tagging	Removing gender association from word embedding generated from a dataset (Parsheera 2018; Sun et al. 2019; Zhao et al. 2019; Font & Costa-Jussa 2019; Gonen & Goldberg 2019; Bentivogli et al. 2020)	
6	AI-enabled analysis	Analysis of training data and communication patterns about senders and receivers can be used to screen the bias patterns (Florentine 2016)	
7	Data augmentation	The dataset that is identical to the original datasets but biased towards the opposite gender of the original dataset will be useful in minimizing bias (Sun et al. 2019; Zhao et al. 2019; Font & Costa-Jussa 2019; Mehrabi et a. 2019)	
8	Bias control training	Removing biased data from the dataset (Terrell et al. 2017; Leavy 2018; Holstein et al. 2019, Brunet et al. 2019, Sun et al. 2019)	
9	Ensuring diversity in AI development	Creating cross-disciplinary teams of data scientists and social scientists that identify and address bias (Parsheera 2018; Gibbs 2020; Terrell et al. 2017; Edwards & Rodriguez 2019; Srivastava & Rossi 2018; Leavy 2018; Holstein et al. 2018; Beard & Longstaff 2018).	Ensuring diversity in AI development
10	Neutralization of language in data and algorithm	A move towards gender neutrality in the language in data and algorithm will promote and improve gender equity (Prates, Avelar & Lamb 2018; Cho et al. 2019; Parsheera 2018; Sun et al. 2019; Zhao et al. 2019; Mehrabi et al. 2019; Gonan & Goldberg 2019; Bolukbasi et al. 2016).	
11	Increase in the number of women in AI development	More representation of women in designing and developing of AI from the user interface, user experience, testing algorithm, and training can manage the gender bias (Gallego et al. 2019; Gibbs 2020)	

Table 3. Approaches for addressing gender bias in AI

The training data shapes the outcomes of machine learning systems and plays a novel role in the development of the algorithm and therefore, should be selected responsibly. One of the reasons for the training data to be biased against women could be because of historical bias against women. Gender ideologies are still present in today's digital and advanced world and therefore still embedded in the training data by the developers; hence result in the algorithm learning stereotypical concepts of gender (Leavy 2018). Experiments show in many articles, papers, and websites more female names being tagged as non-person than male names; amplifying gender stereotyping (Sun et al. 2020). Moreover, lack of culturally diverse developers and training data leads to gender bias in AI i.e. data of people with disability, data of minorities, data of specific group/race, etc. (Wang 2020; Parsheera 2018; Blodgett et al. 2020, Terrell et al. 2017; Dawson et al. 2019). Moreover, word embedding picks the gender bias displayed in the society and therefore perpetuate it in the training data and hence becomes one of the contributing factors of gender bias in AI.



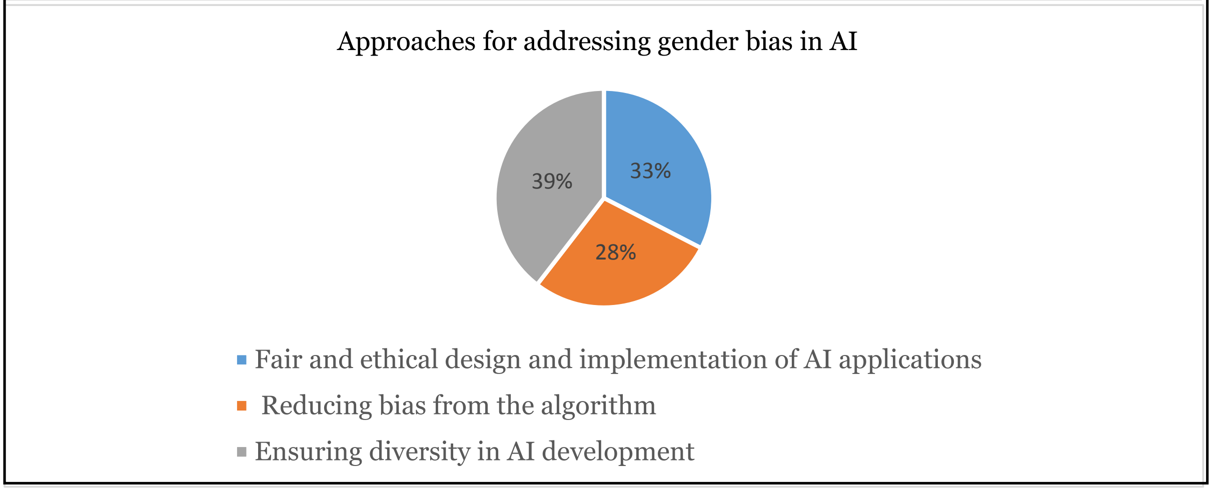


Figure 2. Graphical representation of the themes of contributing factors and approaches for addressing gender bias in AI.

Lastly, we found 3 main approaches for addressing gender bias in AI they are: Implementing fair and ethical AI development, reducing bias in the algorithm, and ensuring diversity in AI development. According to Leavy (2018), the designers of AI algorithms are majority males and that could be one of the contributing factors for the gender bias perpetuation in the resulting applications. Therefore bringing diversity in the developers' sector as an approach for addressing gender bias in AI would be effective. Techniques like gender tagging, bias control training, AI-enabled analysis, trade-offs, and data augmentation have also been discussed in the past as an approach for addressing gender bias. Fairness in AI is a significant and emerging concern for the users as well as for the developers; therefore fairness principle embedded in the AI algorithmic development would be highly significant for mitigating gender bias.

The fairness concept linked to justice could be introduced in AI algorithm (Jobin, Lenca & Vayena 2019). The idea of fairness embedded in AI algorithm development for bringing fairer AI outcomes will bring gender equity as well as fairness in AI procedures and outcomes. There have been few theories used in the literature to explain the phenomena of gender inequality in society, for instance using theoretical lens of Rawl's theory of social justice could be a significant step in future for investigating and analysing the fairness in AI relating to gender bias, as the theory of justice is about the fairness of procedures and outcomes (Ivaturi & Bhagwatwar 2020). As past literature on racial bias is evident that lack of fairness in outcomes leads to a trust deficit, therefore bringing fairness and diversity in AI algorithms through the lens of the Theory of Social justice will not only bring trust whilst will be efficient in mitigating the gender bias present in AI and therefore it will be investigated in our future research.

5 Conclusion, limitations, and future work

Gender bias has been observed at various levels in society. Women face barriers starting from an early level of education to barriers at various levels in various jobs, for instance STEM-related jobs have pronounced barriers at executive levels (Australian Academy of Science 2019). Therefore it is imperative to investigate if the upcoming emerging technologies i.e. AI systems are also perpetuating the bias in the society.

This research presents a deeper understanding of gender bias in AI and brings evidence from the multidisciplinary literature that gender bias does exist in AI systems. Hence the very first step towards addressing gender bias in AI is to explore the contributing factors of this bias that work as a catalyst in AI systems. After the identification of the contributing factors, it is imperative to further investigate and research the approaches that have been discussed in the literature that could mitigate the gender bias in AI. This research gives concise 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.

Our multidisciplinary literature review indicates that the most discussed contributing factor of gender bias includes lack of diversity in data and developers, the bias in society, and bias in data due to programmer conscious or unconscious bias. Therefore it is imperative to continue society-wide efforts in minimizing systemic gender bias as well as taking steps to minimize bias in the AI outcomes through improved design and implementation practices, guided by new ethical guidelines. Furthermore, solving the underline stereotyping in the society and ensuring diversity in AI development, reducing bias in the algorithm development, and fair and ethical implementation of AI applications would be the appropriate approaches for addressing gender bias in AI.

This research not only argues for the concept of justice in gender bias in the overall society, but it also paves the way for other researchers and users to better understand the notion of the gender bias in AI, its contributing factors, and the approaches for possible mitigation as reported in the current multidisciplinary literature.

Fair decisions are perceived to have a significant impact on an individual's satisfaction and loyalty towards a certain product, company, or even relationship. This paper aims to unpack the contributing factors of gender bias in AI and also the approaches that could be applied for mitigating gender bias in AI. Our future work includes further investigation and theorizing of gender bias in AI, using the social justice theory by Rawl (Ivaturi & Bhagwatwar 2020; Sen 1995). We expect this future research to set the foundations for new frameworks that, our literature review confirms, are still very much needed.

Our findings indicate some future IS research related to prevention, mitigation, and future theorization of gender bias in AI, in particular from the IS perspective. Firstly development of new frameworks focusing on gender bias in AI, that could prevent misuse of AI, ensure fairer outcomes (Dwivedi et al. 2019; Gummadi & Heidari 2019), and mitigate past injustices (Trewin 2018) is much needed. Secondly, further exploration and theorization of gender bias in AI, through the lens of Rawl's theory of justice (Kulik et al. 1996) will be an effective mitigating strategy. Lastly, better prevention through education and improvement of professional practices including design and development of AI applications (Leavy 2018; Dwivedi et al. 2019) is required.

The limitation of this study is that we only used Google scholar as a database for this research. Further, the keywords used were biased in AI, fairness, and AI, ethics in AI, alternative terms like analytics, and machine learning in addition to AI would bring more clarity and extension of this research in the future.

6 References

Australian Academy of Science. (2019) Women in STEM Decadal Plan (Australian Academy of Science)

Agarwal, P., 2020, "Gender bias in STEM: Women in Tech still facing discrimination", *Forbes*.

Altman. M, Wood, A., Vayena, E. 2018, "A harm-reduction framework for algorithmic fairness", *AI Ethics*.

Arrieta et al., 2020. "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI", *Information fusion*, vol. 58, pp. 82-115.

Bentivogli et al., 2020. “Gender in danger? Evaluating speech translation technology on the Must- She Corpus”, *58th Proceeding of association for computational linguistic*.

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.

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*.

Bellamy et al., 2018 “AL fairness 360: An extensible toolkit for detecting, understanding and mitigating unwanted algorithmic bias”, *Computer science*, 2018.

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.

Beard, M., Longstaff, S. 2018. “Ethics by design: Principles for good technology”, The ethics center.

Blodgett et al., 2020. “Language (Technology) is power: A critical survey of bias in NLP”, *Computational and language*.

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.

Croeser, S., Eckersley, P. 2019. “Theories of parenting and their application to artificial intelligence”, *Computers and society*.

Crawford, K. (2016, June 26). A.I.’s White Guy Problem. (Sunday Review Desk) (OPINION). *The New York Times*.

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*.

Daugherty, P., Wilson, H., Chowdhury, R. 2018, “Using artificial intelligence to promote diversity”, *MIT Sloan Management Review*.

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.

Edwards, J. S., Rodriguez, E. 2019. “Remedies against bias in analytics systems”, *Journal of Business Analytics*, vol. 2, issue. 1.

Feast, J., 2019. “4 ways to address gender bias in AI”, Harvard Business Review.

Font, J., Costa-Jussa, M.R. 2019. “Equalizing gender biases in neural machine translation with word embedding techniques”, *Computational and language*.

Florentine, S. 2016. “ How artificial intelligence can eliminate bias in hiring”, CIO.

Galleno, A., Krentz, M., Tsusaka, M., Yousif, N. 2019, “How AI could help or hinder women in the workforce”, *Boston Consulting Group*.

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.

Gummadi, K.P., Heidari, H. 2019, “Economic theories of distributive justice for fair machine learning”, *Companion proceedings of 2019 worldwide web conference*.

Holstein et al., 2019, “Improving fairness in machine learning systems: What do industry practitioners need?”, *ACM CHI Conference on human factors in computing sciences*.

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*.

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.

Jobin, A., Lenca, M., Vayena, E. 2019, "The global landscape of AI ethics guidelines", *Nature Machine Intelligence*, vol. 1, pp. 389-399.

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.

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

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.

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

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*

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.

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.

Parsheera, S. 2018, "A gendered perspective on artificial intelligence", *Machine learning for a 5G future*.

Parikh, R.B., Teeple, S., Navathe, A.M.2019, "Addressing bias in artificial intelligence in health care", *JAMA*.

Robnett, R. D., 2015, "Gender bias in STEM fields: Variation in prevalence and links to STEM self-concept", *Psychology of women quarterly*.

Ridley, G., & Young, J. (2012). "Theoretical approaches to gender and IT: examining some Australian evidence". *Information Systems Journal*, 22(5), 355–373.

Srivastava, B., Rossi, F. 2018, "Towards Compostable bias rating of AI service", *AAAI/ACM Conference on AI, Ethics and Society*, New Orleans, Louisiana.

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.

Sun et al., 2019. "Mitigating gender bias in Natural Language Processing: A literature review", 57th *proceeding of association for computational linguistics*, Italy.

Trewin, S. 2018, "AL fairness for people with disabilities: Point of view", *Computer science*.

Terrell et al., 2017. "Gender differences and bias in open source: pull request acceptance of women versus men", *Peer J computer science*.

The global gender gap report, World Economic Forum 2018.

UNESDOC Digital library, Artificial intelligence and gender equality: key finding of UNESCO's global dialogue. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000374174

Wang, L. 2020. "The three harms of gendered technology", *Australian Journal of Information Systems*, vol. 24.

Webster, J., Watson, R.T. 2002, "Analyzing the past to prepare for the future: Writing a literature review", *MIS Quarterly*, vol. 26, issue. 2, pp. 13-23

Zhao et al., 2019. "Gender bias in contextualization word embedding", *proceeding of association for computational linguistics*. Minnesota.

Zhong, Z. 2018, "A tutorial on fairness in machine learning", *Towards data science*.

Copyright