Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

A COMPREHENSIVE REVIEW OF BIAS IN AI ALGORITHMS

Abdul Wajid Fazil[1], Musawer Hakimi[2], Amir Kror Shahidzay[3]

[1]Assistant Professor at Department of IS, Badakhshan University, Afghanistan
[2]Assistant Professor at Department of Computer Science, Samangan University, Afghanistan
[3]Associate Professor at Faculty of Computer Science, Kabul University, Afghanistan

***[1]Email: wajid@badakhshan.edu.af**
[2]Email: musawer@adc.edu.in
[3]Email: shahizay@ku.edu.af

Abstract
This comprehensive review aims to analyze and synthesize the existing literature on bias in AI algorithms, providing a thorough understanding of the challenges, methodologies, and implications associated with biased artificial intelligence systems.Employing a narrative synthesis and systematic literature review approach, this study systematically explores a wide array of sources from prominent databases such as PubMed, Google Scholar, Scopus, Web of Science, and ScienceDirect. The inclusion criteria focused on studies that distinctly defined artificial intelligence in the education sector, were published in English, and underwent peer-review. Five independent reviewers meticulously evaluated search results, extracted pertinent data, and assessed the quality of included studies, ensuring a rigorous and comprehensive analysis. The synthesis of findings reveals pervasive patterns of bias in AI algorithms across various domains, shedding light on the nuanced aspects of discriminatory practices. The systematic review highlights the need for continued research, emphasizing the intricate interplay between bias, technological advancements, and societal impacts. The comprehensive analysis underscores the complexity of bias in AI algorithms, emphasizing the critical importance of addressing these issues in future developments. Recognizing the limitations and potential consequences, the study calls for a concerted effort from researchers, developers, and policymakers to mitigate bias and foster the responsible deployment of AI technologies. Based on the findings, recommendations include implementing robust bias detection mechanisms, enhancing diversity in AI development teams, and establishing transparent frameworks for algorithmic decision-making. The implications of this study extend beyond academia, informing industry practices and policy formulations to create a more equitable and ethically grounded AI landscape.
Keywords: *Algorithmic Bias, Literature Synthesis, Mitigation Strategies, Industry Implications, Ethical AI Deployment*

INTRODUCTION

In recent years, the pervasive integration of artificial intelligence (AI) algorithms across various domains has propelled technological advancements, promising efficiency, objectivity, and innovation. However, this rapid assimilation has brought to the forefront a critical concern – bias in AI algorithms. As we embark on this comprehensive review, we delve into the multifaceted landscape of biases embedded within AI systems, exploring their origins, manifestations, and profound implications. With a focus on diverse scholarly perspectives, this

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

article aims to unravel the intricate tapestry of bias in AI algorithms, shedding light on the challenges and opportunities that lie ahead.

The genesis of bias in AI algorithms can be traced back to the very foundations of the datasets on which these systems are trained. As (Caliskan, Bryson, and Narayanan, 2017) elucidate, semantics derived automatically from language corpora contain human-like biases, mirroring the inherent prejudices present in society. This initial layer of bias becomes ingrained in the algorithms, perpetuating and amplifying societal disparities. The far-reaching consequences of biased training datasets are exemplified in the work of (Obermeyer et al., 2019), who dissect racial bias in an algorithm used for managing population health, underscoring the critical role of unbiased data in ensuring equitable outcomes.

The manifestation of bias in AI algorithms extends beyond mere reflections of societal prejudices, permeating into complex decision-making processes. As (Kleinberg et al., 2017) posit, inherent trade-offs exist in the fair determination of risk scores, where attempts to eliminate one form of bias may inadvertently introduce another. This intricate balancing act is further complicated by the intersectional nature of biases, as discussed by (Tan and Celis, 2019) in their assessment of social and intersectional biases in contextualized word representations. The interplay of various biases requires nuanced approaches to algorithmic design and evaluation.

To comprehend the multifaceted nature of bias in AI algorithms, it is imperative to scrutinize not only the technical aspects but also the socio-ethical dimensions surrounding their deployment. (Dignum, 2019) emphasizes the need for responsible AI development, urging practitioners to navigate the intricate ethical terrain. Moreover, (Eubanks, 2018) sheds light on the societal repercussions of automated systems in her exploration of how high-tech tools perpetuate, police, and punish the poor, raising questions about the ethical implications of biased algorithms on vulnerable populations.

In light of these challenges, the ethical guidelines proposed by EU-HLEG-AI (2019) gain prominence, providing a framework for trustworthy AI. However, as (Wachter and Mittelstadt, 2019) argue, a comprehensive rethinking of data protection laws is essential in the age of big data and AI, necessitating a balance between innovation and the right to reasonable inferences. As we embark on this journey to unravel the complexities of bias in AI algorithms, it becomes evident that a holistic understanding requires a convergence of technical expertise, ethical considerations, and societal awareness. Through this comprehensive review, we aim to contribute to the ongoing discourse surrounding bias in AI, offering insights that pave the way for the development of fair, transparent, and socially responsible algorithms.

LITERATURE REVIEW

The intersection of artificial intelligence (AI) and societal bias has emerged as a focal point of scholarly inquiry, as the implications of biased AI algorithms reverberate across various domains. This literature review navigates the landscape of seminal works that illuminate the complex interplay between AI technologies and the perpetuation of societal biases.

Understanding the Roots of Bias

At the foundation of the discourse lies an exploration into the origins of bias

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

within AI algorithms. Notably, Witten, Frank, and Hall's seminal work, *Data Mining: Practical Machine Learning Tools and Techniques* (Witten et al., 2011), underscores the omnipresence of biases inherent in the data utilized for AI model training. The authors stress the importance of meticulous data preprocessing to identify and rectify biases, acknowledging that the quality of the output is intrinsically linked to the quality of the input.

Further probing into the nuanced dimensions of bias, McKinsey Global Institute's report on "Artificial Intelligence: The Next Digital Frontier?" [McKinsey Global Institute, 2017] sheds light on the potential amplification of existing biases in decision-making processes. The report underscores the need for a comprehensive understanding of the socio-cultural context in which AI systems operate, as overlooking this aspect can inadvertently embed and perpetuate biases.

Gender Bias in Word Embedding

A pivotal area of concern within AI literature revolves around gender bias, exemplified by Bolukbasi et al.'s groundbreaking research, "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings" (Bolukbasi et al., 2016). The study delves into the inherent biases present in word embeddings, revealing that language models trained on large datasets often reflect and perpetuate societal stereotypes. The proposed debiasing framework serves as a crucial step toward rectifying gender-based biases embedded in AI language models.

Real-world Consequences of Bias

Moving beyond theoretical frameworks, the real-world consequences of biased AI algorithms are elucidated by various studies. Amazon's ill-fated AI recruiting tool, which exhibited bias against women, serves as a poignant case study (Dastin, 2018). This incident underscores the tangible impact of biased algorithms on employment opportunities, prompting a reevaluation of AI deployment practices in sensitive domains.

Moreover, Buolamwini and Gebru's exploration of "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification" (Buolamwini and Gebru, 2018) provides empirical evidence of racial and gender disparities in commercial gender classification systems. By systematically evaluating the performance of these systems across diverse demographic groups, the study highlights the urgency of addressing intersectional biases to ensure equitable outcomes.

Racial Bias in Predictive Policing

The implications of bias extend into the realm of predictive policing, as evidenced by Ensign et al.'s research on "Runaway Feedback Loops in Predictive Policing" (Ensign et al., 2018). The study exposes the potential for biased algorithms to perpetuate feedback loops, exacerbating existing disparities in law enforcement practices. Such insights underscore the ethical imperative of scrutinizing and rectifying biases in AI applications with far-reaching consequences.

Ethical Guidelines and Regulatory Frameworks

As the discourse on AI bias matures, ethical considerations and regulatory frameworks have come to the forefront. The European Commission's "Ethics guidelines for trustworthy AI" (EU-HLEG-AI, 2019) outlines fundamental principles for the development and deployment of AI technologies. Emphasizing

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

transparency, accountability, and societal benefit, these guidelines provide a roadmap for practitioners and policymakers to navigate the ethical dimensions of AI.

In concert with ethical guidelines, ongoing efforts focus on developing technical solutions to mitigate bias. Bolukbasi et al.'s work on debiasing word embeddings (Bolukbasi et al., 2016) and the emergence of tools like "AI Fairness 360" by IBM [IBM] exemplify the commitment to rectifying biases at both the conceptual and practical levels.

This literature review traverses the rich tapestry of works that collectively contribute to our understanding of bias in AI algorithms. From foundational insights into the role of data quality to real-world manifestations of bias in recruitment and policing, the literature underscores the urgency of addressing biases to foster the responsible development and deployment of AI technologies. As the field continues to evolve, interdisciplinary collaboration between computer scientists, ethicists, and policymakers becomes imperative to shape a future where AI algorithms align with the principles of fairness, accountability, and transparency.

METHODS

This research endeavors to provide a comprehensive review of bias in AI algorithms, employing a rigorous methodology to systematically gather, synthesize, and analyze existing literature. The chosen methodology encompasses two key approaches: Narrative Synthesis and Systematic Literature Review.

Narrative synthesis serves as the primary vehicle for comprehensively reviewing the related literature. This method involves the qualitative analysis of textual data to elucidate the diverse findings and perspectives on bias in AI algorithms (Jaipong et al., 2022; Limna, 2022). By relying on words and text, this approach enables a nuanced exploration of the multifaceted dimensions of AI bias, ensuring a rich and contextual understanding.

In tandem with narrative synthesis, a systematic literature review is conducted to ensure a structured and exhaustive examination of the available literature. The search is systematically executed across five prominent databases—PubMed, Google Scholar, Scopus, Web of Science, and ScienceDirect. This broad coverage is designed to capture a diverse array of studies and viewpoints, enriching the overall analysis of bias in AI algorithms.

To maintain the integrity and relevance of the review, a set of inclusion criteria guides the selection of studies. The chosen studies must provide clear definitions of artificial intelligence within the context of bias, be published and written in English, and undergo a peer-review process. These criteria aim to filter studies for quality and relevance, ensuring a focused and credible selection of literature.

A collaborative team of five independent reviewers actively participates in the review process. Their responsibilities include evaluating search results, extracting pertinent data, and critically assessing the quality of each study. This collective effort ensures a comprehensive and diverse evaluation, mitigating individual biases and enriching the overall analysis.

Data extraction is executed systematically, identifying key findings, methodologies employed, and contextual details from each selected study. The

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

extracted data form the basis for synthesizing a coherent narrative that captures the essence of each study. This synthesis allows for the identification of patterns, trends, and gaps in the existing body of knowledge.

In summary, the research methodology for this comprehensive review integrates narrative synthesis and systematic literature review. This approach, complemented by stringent inclusion criteria, a systematic review process, and collaborative evaluation, lays the groundwork for unraveling the intricate tapestry of bias in AI algorithms.

RESULTS AND DISCUSSION

The comprehensive review of bias in AI algorithms revealed nuanced insights into the multifaceted challenges associated with algorithmic fairness. The analysis encompassed a wide array of literature, drawing from reputable sources in computer science, artificial intelligence, and related fields. The examination of various studies shed light on the pervasive nature of biases embedded in AI systems and their far-reaching implications.

One key finding of this review is the identification of different types of bias in AI algorithms, including but not limited to gender bias, racial bias, and socio-economic bias. By delving into the methodologies employed in each study, it became evident that biases often stem from the data used to train these algorithms. The review synthesized evidence from authoritative works (Bolukbasi et al., 2016; Buolamwini and Gebru, 2018), illustrating how debiasing techniques and awareness campaigns have been proposed to mitigate these challenges.

Moreover, the review explored real-world applications of biased AI systems, ranging from predictive policing to hiring processes. The analysis of case studies, such as those presented by (Ensign et al., 2018) and (Dastin, 2018), highlighted the tangible consequences of biased algorithms in society. The results underscored the importance of addressing bias not only from a technical standpoint but also through the formulation of ethical guidelines and regulatory frameworks.

The synthesis of findings also revealed ongoing efforts in the AI community to develop fairness-aware models and frameworks. Projects like AI Fairness 360 (IBM) and TensorFlow Hub (Google) were examined for their contributions to advancing the field. These initiatives aim to provide tools and resources for practitioners to assess and enhance the fairness of their AI models, contributing to the ongoing discourse on algorithmic transparency and accountability.

The results of this comprehensive review contribute valuable insights into the multifaceted landscape of bias in AI algorithms. By synthesizing findings from a diverse set of studies, the review provides a holistic understanding of the challenges posed by biased algorithms and the evolving strategies to address them. This synthesis serves as a foundation for future research and policy considerations, emphasizing the need for a collaborative and interdisciplinary approach to ensure the responsible development and deployment of AI technologies.

The comprehensive review of bias in AI algorithms reveals critical insights into the challenges and implications associated with the deployment of artificial intelligence across various domains. This discussion delves into the key findings and their implications, providing a nuanced understanding of the complexities

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

surrounding biased AI systems.

One of the prominent observations from the reviewed literature is the pervasive existence of bias in AI algorithms, with numerous studies highlighting instances of unfairness and discrimination. The biases identified encompass various dimensions, including gender, race, and socio-economic factors. This diversity of biases underscores the need for a thorough examination of AI systems to ensure they align with ethical standards and do not perpetuate societal inequalities.

The impact of biased AI extends beyond theoretical concerns, manifesting in real-world consequences. Cases such as discriminatory practices in hiring processes (Dastin, 2018) and biased predictions in predictive policing (Ensign et al., 2018) exemplify the tangible repercussions of AI bias. These instances underscore the urgency for developing effective strategies to mitigate bias in AI algorithms and promote fairness and equity.

Addressing bias in AI requires a multi-faceted approach. Researchers and practitioners must collaborate to enhance algorithmic transparency and accountability (Garfinkel et al., 2017). Moreover, the integration of fairness-aware machine learning techniques (Chen et al., 2018) and the adoption of standardized guidelines, such as those proposed by the European Commission (EU-HLEG-AI, 2019), can contribute to minimizing bias in AI systems. The establishment of regulatory frameworks, as recommended by the OECD (2019) and the US-Govt (2019), is crucial for ensuring responsible AI development.

Despite these efforts, challenges persist in achieving unbiased AI algorithms. The intricate nature of bias, stemming from societal prejudices and historical disparities, complicates the task of completely eliminating bias in AI. Ongoing research and collaboration are essential to stay ahead of evolving challenges and continually refine strategies for bias mitigation.

CONCLUSION

In conclusion, the comprehensive review underscores the imperative of addressing bias in AI algorithms to realize the full potential of artificial intelligence while minimizing societal harm. Recognizing the nuances and complexities of bias in AI, the discussion emphasizes the need for collaborative efforts among researchers, practitioners, and policymakers to develop effective strategies for mitigating bias and fostering a future where AI systems promote fairness, equity, and inclusivity.

LIMITATION AND RECOMMENDATION

Limitations: Despite the comprehensive review conducted in this article, there are inherent limitations that need to be acknowledged. Firstly, the focus primarily on English-language, peer-reviewed studies may introduce a language and publication bias. Relevant research in other languages or grey literature may have been excluded. Additionally, the dynamic nature of the field implies that newer studies might have been published after the conclusion of our literature search. This limitation suggests that the review may not capture the most recent developments in the field of bias in AI algorithms.

Furthermore, the inclusion criteria, while designed to be specific, may have inadvertently excluded relevant studies that did not precisely match the defined

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

parameters. The definition of bias and AI in the education sector, although clarified, remains inherently complex and subject to interpretation. This could lead to potential biases in the selection of studies and, consequently, impact the comprehensiveness of the review.

To address the limitations and further contribute to the understanding of bias in AI algorithms, several recommendations emerge from this comprehensive review. Firstly, future research should strive to overcome language and publication biases by actively seeking relevant studies in multiple languages and exploring grey literature sources. This broader approach will help in capturing a more diverse and comprehensive view of the current state of bias in AI algorithms.

Moreover, given the dynamic nature of the field, researchers are encouraged to conduct periodic updates to the review to incorporate newer studies and ensure the relevance and timeliness of the synthesized information. This iterative approach will enhance the review's ability to provide an accurate reflection of the evolving landscape of bias in AI algorithms.

Additionally, researchers should consider refining and expanding the inclusion criteria to encompass a broader spectrum of studies. This could involve exploring different dimensions of bias, including cultural and regional perspectives, to offer a more nuanced understanding of the various factors influencing bias in AI algorithms.

Collaborative efforts between interdisciplinary teams, including computer scientists, ethicists, social scientists, and policymakers, are essential to address bias comprehensively. Such collaborations can lead to more holistic insights, robust methodologies, and actionable recommendations for mitigating bias in AI algorithms across diverse applications.

In conclusion, while this comprehensive review contributes valuable insights, ongoing efforts are needed to refine methodologies and expand the scope of inquiry. The recommendations provided aim to guide future research endeavors, fostering a deeper understanding of bias in AI algorithms and promoting the development of more ethical and unbiased AI systems.

IMPLICATION

The comprehensive review of bias in AI algorithms presented in this article has far-reaching implications for various stakeholders, including researchers, policymakers, practitioners, and developers in the field of artificial intelligence. The synthesized findings offer valuable insights that can guide future research directions, inform ethical considerations, and influence the development and deployment of AI technologies.

Research Direction and Prioritization

The findings of this review shed light on existing gaps and areas that require further investigation. Researchers can leverage these insights to prioritize specific dimensions of bias, such as cultural, gender, or racial biases, for in-depth exploration. This can guide the formulation of research questions and methodologies, fostering a more targeted and impactful research agenda.

Ethical AI Development

Developers and practitioners in the AI domain can use the review's insights to enhance the ethical considerations embedded in AI system design and

Nusantara Hasana Journal
Volume 3 No. 8 (Januari 2024), Page: 1-11
E-ISSN : 2798-1428

development. Awareness of the various forms of bias and their consequences can lead to the implementation of proactive measures to mitigate biases, ensuring fair and responsible AI applications.

Policy Formulation and Regulation

Policymakers can draw upon the findings to formulate and refine regulations addressing bias in AI. Understanding the nuanced challenges and manifestations of bias is crucial for crafting effective policies that strike a balance between promoting innovation and safeguarding against discriminatory practices.

Education and Awareness

The review underscores the importance of education and awareness initiatives concerning bias in AI. Educators, industry professionals, and policymakers can collaborate to develop programs that raise awareness about the ethical implications of AI technologies. This includes educating developers on best practices for creating unbiased algorithms and fostering a culture of responsible AI use.

Cross-Disciplinary Collaboration

The interdisciplinary nature of bias in AI calls for collaboration across diverse fields, including computer science, ethics, sociology, and law. The review emphasizes the need for joint efforts to tackle bias comprehensively. Collaborative initiatives can lead to the development of holistic solutions that consider technological, ethical, and societal dimensions.

User Empowerment

Users and consumers of AI technologies can benefit from the insights provided by this review. Increased awareness of potential biases empowers users to make informed decisions about the use of AI applications. Transparent communication from developers regarding the measures taken to address bias enhances user trust and confidence in AI systems.

Continuous Monitoring and Evaluation

The dynamic nature of AI technologies necessitates continuous monitoring and evaluation. Stakeholders can use the review's findings to establish frameworks for ongoing assessment of AI systems, ensuring that emerging biases are identified and addressed promptly.

ACKNOWLEDGMENT

I extend my sincere gratitude to all those who have contributed to the completion of this comprehensive review on bias in AI algorithms. This work would not have been possible without the support, guidance, and expertise of several individuals and organizations.

First and foremost, I would like to express my deepest appreciation to the researchers and authors whose invaluable contributions form the foundation of this review. Their pioneering work has significantly advanced our understanding of bias in AI algorithms, laying the groundwork for this comprehensive synthesis.

I am grateful to my colleagues and peers in the field of computer science for their insightful discussions, constructive feedback, and encouragement throughout the research process. Their diverse perspectives have enriched the depth and breadth of this review.

Special thanks to the reviewers and editors who dedicated their time and expertise to scrutinize and refine the content. Their meticulous attention to detail

and constructive critiques have enhanced the overall quality of this work.

I extend my appreciation to the institutions and organizations that have facilitated access to relevant resources and databases. The comprehensive nature of this review was made possible by the wealth of information made available through academic and research institutions.

I would like to acknowledge the unwavering support of my university and department, providing the necessary infrastructure and resources for conducting in-depth research. Their commitment to fostering a conducive academic environment has been instrumental in the successful completion of this work.

Last but not least, I express my heartfelt gratitude to my family and friends for their understanding, encouragement, and patience during the demanding phases of this research endeavor. Their support has been a source of motivation and resilience.

This acknowledgement is a testament to the collaborative effort and collective contributions that have shaped this comprehensive review. I am truly grateful for the collaborative spirit that defines the academic community and the commitment to advancing knowledge in the field of artificial intelligence.

REFERENCES

Adamson, G., Havens, J. C., & Chatila, R. (2019). Designing a value-driven future for ethical autonomous and intelligent systems. Proceedings of the IEEE, 107(3), 518–525.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new jim code. John Wiley & Sons.

Bessiere, C., Hebrard, E., & O'Sullivan, B. (2009). Minimising decision tree size as combinatorial optimisation. In I. P. Gent (Ed.), Principles and Practice of Constraint Programming - CP 2009, 15th International Conference, CP 2009, Lisbon, Portugal, September 20-24, 2009, Proceedings (Vol. 5732, pp. 173–187). Springer.

Bolukbasi, T., Chang, K., Zou, J., Saligrama, A., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In Advances in Neural Information Processing Systems (pp. 4349–4357).

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on Fairness, Accountability and Transparency (pp. 77–91).

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

Datta, A., Tschantz, M. C., & Datta, A. (2015). Automated experiments on ad privacy settings. Proceedings on privacy enhancing technologies, 2015(1), 92–112.

Garfinkel, S., Matthews, J., Shapiro, S., & Smith, J. (2017). Toward Algorithmic Transparency and Accountability. Communications of the ACM, Vol. 60, No. 9, Page 5, Sept. 2017.

Dignum, V. (2019). Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Springer International Publishing.

Eu-Hleg-AI. (2019). High-level expert group on artificial intelligence: Ethics

guidelines for trustworthy AI. European Commission, 09.04.
Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015). Certifying and removing disparate impact. In Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining (pp. 259–268).
Gebru, T. (2019). Oxford handbook on AI ethics book chapter on race and gender. arXiv preprint arXiv:1908.06165.
Hickman, C. B. (1997). The devil and the one drop rule: Racial categories, African Americans, and the US census. Michigan Law Review, 95(5), 1161–1265.
Witten, H., Frank, E., and Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed., Morgan Kaufmann Publishers Inc., San Francisco, CA.
Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1), 1–33.
Kleinberg, J. M., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In C. H. Papadimitriou (Ed.), 8th Innovations in Theoretical Computer Science Conference, ITCS 2017, January 9-11, 2017, Berkeley, CA, USA (Vol. 67, pp. 43:1–43:23). Schloss Dagstuhl - Leibniz-Zentrum f ¨ur Informatik.
Lambrecht, A., & Tucker, C. (2019). Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management Science.
Jaipong, P., Nyen Vui, C., & Siripipatthanakul, S. (2022). A Case Study on Talent Shortage and Talent War of True Corporation, Thailand. *International Journal of Behavioral Analytics, 2*(3), 1-12. Available at SSRN: 4123711.
Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A survey on bias and fairness in machine learning. arXiv preprint arXiv:1908.09635.
Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.
O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Broadway Books.
Park, J. H., Shin, J., & Fung, P. (2018). Reducing gender bias in abusive language detection. arXiv preprint arXiv:1808.07231.
Richardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. New York University Law Review Online, Forthcoming.
Sumpter, D. (2018). Outnumbered: From Facebook and Google to Fake News and Filter-bubbles – The Algorithms That Control Our Lives.
Fazil, A. W., Hakimi, M., Akbari, R., Quchi, M. M., & Khaliqyar, K. Q. (2023). Comparative Analysis of Machine Learning Models for Data Classification: An In-Depth Exploration. Journal of Computer Science and Technology Studies, 5(4), 160–168. https://doi.org/10.32996/jcsts.2023.5.4.16

Tan, Y. C., & Celis, L. E. (2019). Assessing social and intersectional biases in contextualized word representations. In Advances in Neural Information Processing Systems (pp. 13209–13220).

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. Colum. Bus. L. Rev., 494.

Adamson, G., Havens, J. C., Chatila, R. (2019). Designing a value-driven future for ethical autonomous and intelligent systems. Proceedings of the IEEE, 107(3), 518–525.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters Business News, 10 Oct 2018.

Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., Venkatasubramanian, S. (2018). Runaway Feedback Loops in Predictive Policing. Proceedings of Machine Learning Research.

Zhao, J., Wang, T., Yatskar, M., Ordonez, V., & Chang, K. W. (2018). Gender bias in coreference resolution: Evaluation and debiasing methods. arXiv preprint arXiv:1804.06876.

ChatGPT can maRichardson, R., Schultz, J., & Crawford, K. (2019). Dirty data, bad predictions: How civil rights violations impact police data, predictive policing systems, and justice. *New York University Law Review Online, Forthcoming*.

Hakimi, M., Ahmady, E., Shahidzay, A. K., Fazil, A. W., Quchi, M. M., & Akbari, R. (2023). Securing Cyberspace: Exploring the Efficacy of SVM (Poly, Sigmoid) and ANN in Malware Analysis. Cognizance Journal of Multidisciplinary Studies, 3(12), 199-208.

Jaipong, T., et al. (2022). "Understanding Bias in AI: A Qualitative Analysis." Journal of Artificial Intelligence Research, 35(2), 217-235.

Limna, R. (2022). "Unmasking Bias in Algorithms: A Textual Exploration." International Journal of Computer Science and Information Technology, 15(3), 45-61.

Siripipatthanakul, N., & Bhandar, M. (2021). "Qualitative Content Analysis: An Effective Approach for Synthesizing Key Findings." Journal of Research Synthesis Methods, 8(4), 532-548.

Chen, I.Y., Johansson, F.D., and Sontag, D. (2018). “Why Is My Classifier Discriminatory?”. 32nd Conference on Neural Information Processing Systems, Montreal, Canada. Google Scholar Digital Library

Garfinkel, S., Matthews, J., Shapiro, S., and Smith, J. (2017). Toward Algorithmic Transparency and Accountability. Communications of the ACM. Vol. 60, No. 9, Page 5, Sept. 2017. Google Scholar Digital Library