ETHICS, REGULATION AND LAW IN ADVANCED DIGITAL INFORMATION PROCESSING AND DECISION MAKING

ECS7025P

MSc FT Data Science and Artificial Intelligence

Report title here

Elliot Linsey

Submission Date: HERE

Word Count: HERE

Table of Contents

[List of Abbreviations and Glossary 2](#_Toc53045814)

[List of Figures 3](#_Toc53045815)

[Abstract 5](#_Toc53045817)

[Introduction 6](#_Toc53045818)

[Literature Review 7](#_Toc53045819)

Methodology Research  [8](#_Toc53045820)

Data Analysis [& discussion 9](#_Toc53045821)

[Conclusion 10](#_Toc53045822)

[Bibliography 11](#_Toc53045823)

[Appendices 12](#_Toc53045824)

# List of Abbreviations and Glossary

# List of Figures

No table of figures entries found.

# Acknowledgements

# Executive Summery

# Introduction

# Literature review

Paragraph on what the focus of the literature review is on and what view it is supporting (that there is bias in machine learning algorithms). What focus within algorithmic bias am I looking at (gender, racial etc)? The following question has been asked: How has gender bias infiltrated machine learning and what are the effects of this bias etc?

This literature review shall be exploring the notion of gender bias within machine learning and AI algorithms. At the present day, there is clear evidence of gender bias infiltrating these systems from a multitude of angles. These include biased training data, a lack of diversity within AI development, economic factors, and potential inbuilt misogyny within the social setting. This bias causes negative effects to the female sex by potentially penalising them within applications or work-related roles as well as categorising them in different and demeaning manners to males. A large amount of research done in this subject has been led by female researchers who are best able to identify and relate to the issues that gender bias causes within AI but also the STEM field in general.

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.” (Tom Mitchell, 1994). This famous quote is used to summarise the process of a machine learning algorithm and although advancements have been made with new techniques and methods, the general formula has stayed the same. The issue that is being currently faced by machine learning scientists and AI practitioners is to do with the experience section of the above quote, where imbalanced datasets that do not contain the full range of experience to learn from can lead to “bias in machine learning algorithms that have troubling implications and deleterious consequences” (Weiss et al, 2018)

Due to the material within older corpora that machine learning algorithms are being trained on, gender bias has been observed within current algorithms due to the outdated ways of referring to men and women (Leavy, 2018). This is due to the heavily male-centric way of thinking and writing within periods such as the 50s and 60s, with women being referred to with far more appearance-based descriptions and metaphors compared to men, who are described according to accomplishment (Hines, 1999). Within algorithms trained on more recent data, gender bias has also been observed as the Amazon company resumé rating system started to penalise those that contained vocabulary such as “women’s chess captain,” and those that attended all-women colleges (Dastin, 2018). This is due to the vast majority of previous successful applicants that the algorithm was trained on being male, therefore it would reward those resumés that were similar to the ones that it had been told were previously successful and discount the female resumés.

Whilst the data these algorithms are being trained on may be problematic, another influencing factor on the gender bias seen within machine learning could also be related to the programmers and developers themselves (Nadeem et al, 2020). Within AI development the male perspective has been dominant through sheer force of numbers. An example of this type of gender bias can be observed with the Github platform, where women’s acceptance rates are higher for open-source projects only when they are not openly identifiable as women (Terrell et al, 2017). There are more initiatives being developed to encourage and support female entrants into the field, however this is only a first step into eliminating gender bias (Parsheera, 2018). Successive measures such as cross-disciplinary teams, bias identification methods and fairness measures being built into evaluation metrics could be the next stepping stone for reducing the impact of gender bias during AI development. Whilst these are excellent goals to strive for, the fact of the matter is that a lot of the data being used to train AI models still uses heavily biased terms, such as ‘chairman’ instead of ‘chairperson’. The most difficult step would be correcting this training data itself. Education could be implemented to work with the current generation of digital content producers but also to influence the next generations to use more gender-neutral terms in their creations.

An issue with implementing these changes to the development of AI is that of economic factors, as well as research capability. Technology is designed and created by engineers, but the actual reason for its creation is down to socio-political factors or profit motivation (Wang, 2020). Due to this, there would need to be an active involvement by businesses and governments to tackle the effects of gender bias. Some governments have already stated their desire to be involved more prominently in specific areas of AI development, such as the UK wishing to play a greater role in the ethics element of AI creation (House of Lords, 2018).

Hidden biases such as gender can infiltrate machine learning without the creators or users knowing about it, but even when there are active efforts to remove protected attributes such as race or gender from training data, issues still arise (Kilbertus et al, 2017). For example, algorithms such as PredPol are based on the geographical area that crime occurs in with no inbuilt racial knowledge. However, due to certain neighbourhoods being dominated by a particular racial group this leads to the potentiality for racial bias. If an algorithm predicts crime happening regularly in a particular location, this could lead to an unfair response by law enforcement (ACLU, 2016).

# Section B

# Recommendations

# Conclusion

# Bibliography

Yapo, Adrienne., Weiss, Joseph., 2018/01/01 “Ethical Implications of Bias in Machine Learning” 10.24251/HICSS.2018.668

Mitchell, Tom M., et al. 1997, “Machine learning” pp. 870–877.

Leavy, S., "Gender Bias in Artificial Intelligence: The Need for Diversity and Gender Theory in Machine Learning," 2018 IEEE/ACM 1st International Workshop on Gender Equality in Software Engineering (GE), 2018, pp. 14-16.

Dastin, J,. “Amazon scraps secret AI recruiting tool that showed bias against women”, 2018, Reuters. Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> [Accessed 24th February 2022]

Hines, Caitlin. 1999. Rebaking the pie: The woman as dessert metaphor. Reinventing identities: The gendered self in discourse, ed. by M. Bucholtz, A.C. Liang & L.A. Sutton, 145-62. New York, Oxford: Oxford University Press.

Kilbertus, N., Rojas Carulla, M., Parascandolo, G., Hardt, M., Janzing, D. and Schölkopf, B., 2017. Avoiding discrimination through causal reasoning. *Advances in neural information processing systems*, *30*.

ACLU, 2016. Predictive Policing Today: A Shared Statement of Civil Rights Concerns. Available at: <https://www.aclu.org/sites/default/files/field_document/jointstatementpredictivepolicing17orgs.pdf> [Accessed 24th February 2022]

Parsheera, Smriti, (2018). A Gendered Perspective on Artificial Intelligence. Proceedings of ITU Kaleidoscope 2018 -- Machine Learning for a 5G Future, Available at SSRN: <https://ssrn.com/abstract=3374955> or [http://dx.doi.org/10.2139/ssrn.3374955](https://dx.doi.org/10.2139/ssrn.3374955)

Terrell J, Kofink A, Middleton J, Rainear C, Murphy-Hill E, Parnin C, Stallings J. (2017). Gender differences and bias in open source: pull request acceptance of women versus men. *PeerJ Computer Science* 3:e111 <https://doi.org/10.7717/peerj-cs.111>

Nadeem, Ayesha; Abedin, Babak; and Marjanovic, Olivera, (2020) "Gender Bias in AI: A Review of Contributing Factors and Mitigating Strategies". ACIS 2020 Proceedings. 27.  
<https://aisel.aisnet.org/acis2020/27>

Wang, L. (2020). The Three Harms of Gendered Technology. *Australasian Journal of Information Systems*, *24*. <https://doi.org/10.3127/ajis.v24i0.2799>

House of Lords, 2018: House of Lords, Select Committee on Artificial Intelligence, AI in the UK: ready, willing and able? available at: [https://publications.parliament.uk/pa/ld201719/ldselect/ldai /100/100.pdf](https://publications.parliament.uk/pa/ld201719/ldselect/ldai%20/100/100.pdf)

# Appendices