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

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 compared 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 as well as 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 currently being 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). This form of bias is known as representation bias, where associations between concepts and gender are embedded within AI systems due to material it has been trained on. The other form of bias is known as allocation bias, where algorithms reward the majority gender within documents (Crawford, 2017). This form of gender bias has been observed with the Amazon company resumé rating system, it 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). Currently, countries such as the UK are focusing on creating ethical frameworks for the development of AI (such as the UK Data Ethics Framework) and enforceable law (such as the EU General Data Protection and Regulation Act). As these become more widely accepted companies will have clearer instruction and guidance on how to best avoid not only gender bias but also consider ethical considerations and address them within their products (UK Data Ethics, 2020).

The removal of gender bias within AI systems is still a work in progress. However, as seen, a number of issues must be addressed. These being the material within older corpora containing inbuilt bias and male-concentric viewpoints, as well as current AI training techniques rewarding majority classes. Another being the male dominated field of AI development which can have an unconscious yet negative impact on female representation within AI products. Finally, to implement these changes, governments and businesses need to be involved and make an active effort to combat gender bias.

# Section B

According to the UK Information Commissioners Office (ICO), a personal data breach is the “breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data.” There are a number of reasons why a data breach may occur, with the majority of them not containing explicit malicious intent by bad actors. A number of examples from the ICO itself contain scenarios such as:

* Sending an email with sensitive information to the wrong recipient
* Misplacing a laptop or thumb drive with sensitive data
* Deletion or alteration of data
* An unauthorised third-party gaining access to data
* An inadequate response or data protection by the data controller

The Marriott case is connected to the last 2 of these examples, due to the fact that the data was breached by a third party and the company had inadequate protection in place. In terms of Marriott’s fault in relation to its protection policy, the intrusion was actually put into the Starwood Resorts system in 2014. Marriott subsequently purchased Starwood in 2016 and did not notice the intrusion until 2018, giving the attackers another two years to potentially steal data. Due to Marriott being the data controller under the definition of section 6 of DPA, they have been found at fault in their ability to adequately protect personal data as required by Article 32 of GDPR as well as Article 5(1)(f).

The most critical issue of this data breach was the amount of data that the attackers had access to, with an estimated 339 million users potentially at risk from the breach. Looking at the Marriott hotels current data collection standards, we can see the vast amount of information that was vulnerable during the breach period:

* Name
* Gender
* Postal address
* Telephone number
* Email address
* Financial information (such as credit and debit card number or other payment data)
* Language preference
* Date and place of birth
* Nationality, passport, visa, or other government-issued identification data
* Important dates: birthdays, anniversaries, and special occasions
* Membership or loyalty program data (including co-branded payment cards, travel partner program affiliations)
* Employer details (for business-related bookings)
* Travel itinerary, tour group, or activity data
* Prior guest stays or interactions, goods and services purchased, special service and amenity requests
* Social media account ID, profile photo and other data publicly available, or data made available by linking your social media and loyalty accounts

Data such as financial, passport, and visa information is extremely valuable and can cause devastating harm if put into the wrong hands. Not only this but clearly identifiable information such as names, telephone numbers and addresses that are classified as personal data are also collected every time you make a booking. Due to the sensitivity of the data that Marriott holds, it is therefore an even greater example of negligence or mismanagement of data protection that allowed access to this amount and level of data. In response to this, the ICO was able to administer a fine of £18.4million under articles 83(1) and (2) of GDPR, under the belief that it was a proportional response and would be effective in making Marriott increase its data security.

In relation to the UK Data Ethics Framework, the principle that relates most to this case would be that of accountability. This requires there to be effective governance and oversight for any data project, which Marriott hotels either were not able to or declined to provide. To further simplify this, Marriott were required to keep the data that they processed (under article 4(2) of UK GDPR) in a secure manner to prevent unauthorised actors from gaining access to it. Due to a data breach occurring, Marriott failed in this capacity and thus should be held accountable for its inaction in upholding data security protocol.

From an interview with a person knowledgeable about this subject (referred to as the Expert), they bring up a number of points regarding data breaches, the nature of the data that Marriott was taking from customers as well as the limitations of UK GDPR and the future that UK GDPR is currently moving towards.

In regards to the data breach, the Expert mentions the fact that this exploit had been in the system for 4 years until it was discovered. You would expect during this time period that actions taken to both upgrade existing security and look for potentially infected systems would have been taken. Also, the fact that Marriott took over the original Starwood system and did not notice the malware suggests that they were not thorough in their takeover process and may not have been willing to make sure all systems were up to standard. The Expert attributes this seemingly lackadaisical approach to either ignorance of data standards and appropriate security measures, or possible negligence by Marriott hotels due to the fact that the Data Protection Act was updated in 2018 and they did not understand how it could affect them. However, when the exploit was discovered, Marriott hotels did act promptly in informing the ICO and performing damage control with its customers, including installing preventative security measures. It is unfortunate that it took a breach of this scale for a company of this size to revisit its security protocol and patch up holes in their system. Another point the Expert makes is how this was a 4-year long attack with the breach of 339 million users and Marriott hotels was fined £18.4million. On the surface this sounds like a large sum of money, however the Marriott Group (that owns Marriott hotels) operates 30 brands internationally with a revenue of $20billion in 2018. To a corporate entity such as Marriott group this fine would barely create any lasting impact and may have been seen as a minor nuisance. It is certainly possible that Marriott were happy to not upgrade their system until forced to due to their knowledge that any fine enacted on them would not have any significant impact. Another point to take into consideration is that Marriott likely profited from this data during the period. From the Marriott privacy policy, they state that they share your data not only with franchises within the Marriott group, but also third-party services such as airlines, car rental services, restaurants and promotional marketing companies that will target you based on your data. With this in mind, £18.4 million is likely dwarfed by the profit they made from selling their customers data during the time period.

Another point mentioned by the Expert was on the type of data declared vulnerable by Marriott hotels, and the data that Marriott collects and has access to as a whole. The official ICO report states that users may have had their “names, email addresses, phone numbers, unencrypted passport numbers, arrival/departure information, guests’ VIP status and loyalty programme membership number” compromised. However, as seen above it is known that Marriott also collects information such as gender, employer, place of birth, any data that is linked to social media accounts.

# 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)

UK, (2020), UK Data Ethics Framework. Available at:

[Data Ethics Framework - GOV.UK (www.gov.uk)](https://www.gov.uk/government/publications/data-ethics-framework/data-ethics-framework-2020)

Crawford, Kate. (2017). The Trouble with Bias. Keynote at Neural Information Processing Systems (NIPS‘17). Available at: [The Trouble with Bias, by Kate Crawford (Revolutions) (revolutionanalytics.com)](https://blog.revolutionanalytics.com/2017/12/the-trouble-with-bias-by-kate-crawford.html) [Accessed 26/03/2022]

# Appendices