Fairness of AI-powered Recruitment Processes

Group C5

Adam Eljasiak [13974084]
Andesh Haribhajan [1008454]
Antonis Klironomos [13827731]
Roopa Sudhakar [13523775]
Sehra Elahi [13636804]

Fundamentals of Data Science (5294FUDS6Y) - Assignment 2
MSc. Information Science: Data Science Track
University of Amsterdam
October 18th, 2021

2093 words

1 Introduction

Organizations are leaning more towards **Artificial Intelligence (AI) hiring tools** to handle the magnitude of filtering applications and also trying to elude human bias in doing so [1] [2]. The **Applicant Tracking Systems (ATS)** designed for recruitment processes are advertised as unprejudiced tools, competent of avoiding human subjectivity from the employee hiring process. The use of AI in ATS for candidate assessment is still in its early stages; with not much being known about the practice and implications of this method. Wrong decisions made by algorithms can have harmful repercussions, resulting in unfair biases [3] [4] [5]. This paper revolves around the central question: *What are the ethical values of fairness in recruitment by using AI for candidate selection?*

Around 50% of millennials prefer working for a business with good ethical practices [6] [7] and a survey by Nielsen [8] highlighted that 67% of respondents wanted to work for a socially responsible company. Due to this, the biases in this field are of much interest to the authors of the paper, who themselves are Millenials and quite active in the job market as applicants - with one of the authors having experience as a recruiting management consultant. This is, therefore, a relatively interesting topic of concern to inspect further.

The foundation of our paper is built on the use case of an AI application, used at a large multinational company 'Multico' (pseudonym). This AI application was built by a third-party vendor 'NeuroYou' (pseudonym), and outsourced to 'Multico', with a purpose to remove bias from the workplace, by eliminating human decisions involved in recruitment [9]. The motivation behind this was to increase diversity - not only in terms of race or gender, but also with respect to thought and abilities.

This paper centralises around the context of stakeholders, the harmful impacts and the law surrounding the use case of 'NeuroYou'. Issues concerning fairness within the design of the system are identified and analysed, through conducting a technology ethics analysis using the Networked Systems Ethics (NSE) guidelines [10]. The 'Analysis' section of the paper is divided into two main halves corresponding to the summary questions - (A) context and (B) ethical aspects; from the NSE guidelines. The results of these analyses are discussed in the 'Discussion' and the 'Conclusion' sections at end of the paper.

2 Analysis

2.1 Context

The recruitment software by NeuroYou was implemented to tackle biases in recruitment, caused by human decisions for MultiCo. A pilot stage was released in the EU to suppress company biases and promote fairness. The ATS was able to achieve this by providing higher levels of accuracy and consistency for candidate evaluation. Nonetheless, the outcome of the NeuroYou algorithm was eventually deemed not fair, as it only selected a certain group of candidates that met distinct characteristics, such as; extroverted, leadership quality, etc. [9]. Hence, the tool was not assessing the right capabilities of the applicants and therefore suggested bias. The harms and benefits of NeuroYou are analysed and summarised in table 1.

Table 1: NeuroYou Context Analysis

Characteristic	Benefit	Harm
1. Consistency	Comparisons of applicants across different programs, locations, time and the threshold number for selecting a specific number of candidates, for each program [9]	No substantial consideration of threshold for different programs and locations. HR professionals deviated away from fixed threshold by making changes to it.
2. Accuracy	Developing hiring algorithms to predict future performance and extracting desired qualities from the dataset concerning top-performing employees.	Problems with accuracy and the rationality between desired skills and the candidate's ability for assessments. Certain applicants 'gamed the system', through learning the behaviour of the decision-making process of the algorithm. Managers stressed the algorithm was judging candidates unfairly, on the wrong abilities, by accepting candidates of similar profiles, erasing the aim of striving for diversity and fairness.
3. Reliability	Efforts for prediction successful on the training dataset.	Unreliability in the use of the application, as not all recruiters used the algorithm in all the steps for candidate selection, and some used the wrong algorithm. This also resulted in inconsistency within the data.

2.1.1 Stakeholder Benefits

Each stakeholder group can benefit from the presence of AI recruiting software. First of all, the HR department has fewer responsibilities as its members do not need to ensure fairness during the recruitment process [9]. Moreover, a robust AI tool might assess the abilities of candidates and match only the qualified ones to the job openings, which saves time for recruiters and in effect company's money. Additionally, the candidates are given the chance to experience a simple and fun testing process that can include gamification features and limit the impact of stress on less confident applicants. Furthermore, a candidate can expect a fair selection process. This would be beneficial for both the candidates and the company as it would lead to a diversity of thought among the company's personnel [9]. The last group of stakeholders are software developers and data scientists, who develop such solutions. They find employment in the field and improve existing solutions to be more fair and accurate.

2.1.2 Risk of Harm and Burdens

Analysis of the subject of AI-based recruitment presents certain risks which will be addressed in this section. Firstly, applicants can be encouraged by this paper's references to past cases against discrimination in the hiring process. Consequently, they may try to sue companies by which they suspect that they have been mistreated. From the companies' perspective, this can lead to economic problems but most importantly their reputation could diminish. Moreover, candidates may get skeptical about the AI-based recruitment process and apply for companies that are not utilizing similar solutions. As a result, employers who choose to use it might experience a decrease in the number of applications. Last, but not the least, some candidates might even start taking advantage of the flaws in the software to increase their chances of getting an interview. This situation is harmful because it introduces another aspect of unfairness.

2.1.3 Law

In 2020, there has been a growing interest in AI and algorithmic decision making, by the EU Commission and the Council of EU[11]. Many algorithmic software systems originate from the US, however, the legal framework for the US and EU are not compatible. This paper will focus EU laws.

Employers must follow legal obligations relating to decision-making algorithms, under the rules of equality law and the **General Data Protection Regulation (GDPR)**. Equality law, concerning direct and indirect discrimination, is domestic and imposed by each EU country's agencies. Data protection law on the other hand is stable across EU nations, as an EU law[11]. GDPR provides a coherent and stable framework of data protection laws across the EU. It controls all proceedings where automated algorithms are utilised to make human decisions. According to GDPR article 22[12], employers can use automated systems if they can show their necessity for their organisation. In the case of a fully automated algorithm, however, the candidates must be informed of the type of activity they are participating in and the logic behind the decision-making and prediction. Furthermore, Recital 71 of the GDPR[13] emphasises the importance of explanations of the outcomes reached, as a suitable safeguard. Under the GDPR, if an employer uses AI for hiring decisions, a Data Protection Impact Assessment (DPIA) must be conducted[14]. The DPIA considers elements of bias in algorithms, and the steps taken to reduce the bias.
The **enforcement of these laws** in practice can be performed by **individuals**, **public bodies** and the **DPIA**[11]. For individuals, it is almost impossible to find evidence of discrimination even if they suspect it. While GDPR rights allow the candidates to know if they have been assessed by an algorithm and getting the data they submitted, GDPR does not state the right for them to be informed about the existence of bias in the algorithm.

2.2 Technology Ethics

2.2.1 Function Creep/Technology Assessment

Systems used in ATS that utilize data science methods can be divided into two categories - those used during and those used outside an interview. Bias can be found in any of these systems.

CV parsing is the first stage of the ATS and its goal is to automatically map information entered in a CV to the categories of interest of a company, which will appear in future searches to match candidates to the job offerings. Unfortunately, this technology is far from perfect and does not guarantee to retrieve all relevant data correctly from an arbitrary CV, despite, parsing models utilizing advanced techniques, like Hidden Markov Models with SVN or deep natural language processing[15]. These models deal with simply structured CVs that consist only of plain text and bullet points but might fail with tables, graphical elements or freely placed text[16]. Moreover, even if a CV has been parsed correctly, there still exists a risk of missing a qualified candidate if he/she used diacritic characters or synonyms of keywords that are searched for a job opening. This system presents a degree of unfairness towards candidates that are unaware of its presence since candidates often try to impress the future employer with the appearance of the CV, which might be filtered out or parsed incorrectly. This also presents an opportunity for unqualified candidates to know how to work the system to at least get invited for an interview by abusing keywords and the structure of the document.

Data science is also used during the interview to extract additional metrics that might support the recruiter's decision. NeuroYou and other solutions [17] enhance this stage of recruitment with video-based emotion recognition tools to build a psychological profile of the candidate. Advanced and expensive machine learning models based on high-quality datasets achieve the performance of recognizing emotions from facial expressions with circa 70% accuracy [18]. So, it is reasonable to say that the performance might be lower while using real-time models based on images captured by mediocre-quality webcams. The severity of misclassification depends on the actions of a human interviewer based on the labels produced by ATS, however, it adds a weak link in the process. Unfortunately, research proves that non-white candidates' emotions are misclassified more often [19].

2.2.2 Data Governance

Analysis of the data governance is an important part of the Technical Ethics Analysis (TEA). The data governance analysis focuses on the following 5 aspects: (i) Dissemination, (ii) De-identification, (iii) Unforeseen Risk Management, (iv) Security, and (v) Data Retention. The EU released "Ethics Guidelines for Trustworthy AI" [20] [21] which enforces these principles legally. This has a global impact which obliges worldwide organizations to have their governance frameworks scrutinized. The companies should not only set these principles but also adopt adaptive-governance [22] for data and its processing.

Dissemination, De-identification, Unforeseen Risk Management & Security

In the case of NeuroYou, the risks related to data dissemination are present. The data has managed access. Only the recruiters have access to the applicant data that is necessary. Most of the personal information is not revealed for analysis purposes (de-identification). Only for contacting the applicant, some basic personal information will be accessible. However, the data can be shared with the 'Multico' and other relevant parties for evaluation. The data compliance policies were not made clear in the source of this case. The poor description of data compliance policies coupled with the sharing of data can result in an unwanted risk of data breaches. It is, therefore, recommended that most sensitive information is only accessible to a few albeit the extent of this is unknown. In the currently examined case, there were no clear security measures mentioned that were in place except for managing the access.

Data Retention

According to EU data retention policies "data should be held no longer than necessary" [23]. However, there is no particular data retention timeline specified because these time periods are context-specific. Determining the length of the data retention period is not always clear. Some companies delete the data after the conclusion of the hiring process within three months, while some keep it up to a year. Furthermore, one of the most difficult issues involved in the retention of data, according to Kiansing & Huan [24], is not 'knowing how long' the data must be retained but 'starting from when' the data is to be retained. In the case of NeuroYou the data retention period and policies are unknown.

3 Discussion

Although AI allows companies to save time and money with regards to the hiring process, each introduced system that carries a risk of *misclassification bias* diminishes the fairness of the process. HR stands for human resources and each data science technique introduced in the hiring process removes the human factor from it, increasing the possibility of making unfair decisions. The 'NeuroYou' AI implementation tries to address several relevant key aspects of fairness. However, we believe that there is room for further improvement. For example, the data retention policies and the security aspects are not sufficiently taken care of. Furthermore, GDPR should be expanded to include more requirements that will guard human rights in the recruitment process. For instance, the addition of obligatory publication of GPIA would be beneficial for both the individual and the organizational cases. Finally, the partially automated recruitment process ought to be supervised to ensure that the decision of HR team does not rely completely on the output of an algorithm.

4 Conclusion

Machine learning has started to play a crucial role in the hiring procedure as various employers have tried to use relevant algorithms to automate parts of the process. As mentioned above, introducing such techniques of decision-making can provide benefits to the stakeholders. On the other hand, these kinds of methods can lead to conflicts among the company's personnel and unfairness towards candidates due to (sometimes unwanted) discrimination [9]. Nowadays, there are laws (e.g. GDPR) which, if enforced properly, can contribute to the minimization of bias in algorithms. Nonetheless, if such bias is suspected by an individual or an organization, the process of proving it can be hard and complicated to ensure ethically sound and fair implementation of AI in the hiring process.

References

[1] Kimberly A Houser. Can ai solve the diversity problem in the tech industry: Mitigating noise and bias in employment decision-making. *Stan. Tech. L. Rev.*, 22:290, 2019.

[2] Manish Raghavan, Solon Barocas, Jon M Kleinberg, and Karen Levy. Mitigating bias in algorithmic employment screening: Evaluating claims and practices. 2019.

[3] Akhil Alfons Kodiyan. An overview of ethical issues in using ai systems in hiring with a case study of amazon's ai based hiring tool. *Researchgate Preprint*, 2019.

[4] Joo-Wha Hong, Sukyoung Choi, and Dmitri Williams. Sexist ai: An experiment integrating casa and elm. *International Journal of Human–Computer Interaction*, 36(20):1928–1941, 2020.

[5] Akhil Alfons Kodiyan. An overview of ethical issues in using ai systems in hiring with a case study of amazon's ai based hiring tool, 11 2019.

[6] T Deloitte. Big demands and high expectations: The deloitte millennial survey, 2014.

[7] Jessica Pruessmeyer and Andreas Ludwig. The good, the bad and the ugly substrates for adam10 and adam17 in brain pathology, inflammation and cancer. In *Seminars in cell & developmental biology*, volume 20, pages 164–174. Elsevier, 2009.

[8] Thomas E Patterson. Doing well and doing good. *Available at SSRN 257395*, 2000.

[9] Elmira van den Broek, Anastasia Sergeeva, and Marleen Huysman. Hiring algorithms: An ethnography of fairness in practice. 2019.

[10] Bendert Zevenbergen, Brent Mittelstadt, Carissa Véliz, Christian Detweiler, Corinne Cath, Julian Savulescu, and Meredith Whittaker. Philosophy meets internet engineering: Ethics in networked systems research.(gtc workshop outcomes paper). In *GTC Workshop Outcomes Paper)(September 29, 2015)*, 2015.

[11] Aislinn Kelly-Lyth. Challenging biased hiring algorithms. *Oxford Journal of Legal Studies [Forthcoming]*, 2020.

[12] Tal Z Zarsky. Incompatible: the gdpr in the age of big data. *Seton Hall L. Rev.*, 47:995, 2016.

[13] Bryce W Goodman. A step towards accountable algorithms?: Algorithmic discrimination and the european union general data protection. In *29th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona. NIPS Foundation*, 2016.

[14] Felix Bieker, Michael Friedewald, Marit Hansen, Hannah Obersteller, and Martin Rost. A process for data protection impact assessment under the european general data protection regulation. In *Annual Privacy Forum*, pages 21–37. Springer, 2016.

[15] Aakankshu Rawat, Siddharth Malik, Seema Rawat, Deepak Kumar, and Praveen Kumar. A systematic literature review (slr) on the beginning of resume parsing in hr recruitment process & smart advancements in chronological order. 2021.

[16] Katarina Drucker. Avoiding discrimination and filtering of qualified candidates by ats software. 2016.

[17] Hilke Schellmann and Jason Bellini. Artificial intelligence: The robots are now hiring. *The Wall Street Journal*, 20, 2018.

[18] Yousif Khaireddin and Zhuofa Chen. Facial emotion recognition: State of the art performance on fer2013. *arXiv preprint arXiv:2105.03588*, 2021.

[19] Alicia A Grandey, Lawrence Houston III, and Derek R Avery. Fake it to make it? emotional labor reduces the racial disparity in service performance judgments. *Journal of Management*, 45(5):2163–2192, 2019.

[20] Nicol Turner Lee, Paul Resnick, and Genie Barton. Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings Institute: Washington, DC, USA*, 2019.

[21] Luciano Floridi. Establishing the rules for building trustworthy ai. *Nature Machine Intelligence*, 1(6):261–262, 2019.

[22] Friedrich Holotiuk and Daniel Beimborn. Critical success factors of digital business strategy. 2017.

[23] Ian Brown. Communications data retention in an evolving internet. *International Journal of Law and Information Technology*, 19(2):95–109, 2011.

[24] Kiansing Ng and Huan Liu. Customer retention via data mining. *Artificial Intelligence Review*, 14(6):569–590, 2000.