**Coursework Assignment: Data Science Project Proposal**

**AI Effect on Workforce Diversity**

**Abstract**

Artificial intelligence (AI), more than ever before, is being used in the recruitment industry. With the ability of machine learning algorithms to automate and speed up the hiring process, more and more employers are turning to AI-based recruitment. As with any technology, AI-based recruitment not only has many pros, but it also presents with several cons. Bias and discriminatory hiring has been identified as one problem that need to be addressed. The concept of diversity, equity, and inclusion (DEI) is a goal being prioritized to benefit employers and employees. The concept of explainable AI goes hand in hand with AI-based recruitment as a mitigation method to alleviate this bias. Humans have biases, and there is concern that these biases are transferred to the AI recruitment software through legacy data, programmer code and unconscious bias. Biased algorithms such as PredPol and Amazon’s recruiting platform, amongst others, have already been identified. Mitigation of these biases through statistical measures, improvements in input data, together with the regulation of AI recruitment software have been proposed. The need for transparency in the recruitment pipeline is evident.

**Keywords:** Artificial intelligence · Algorithmic hiring · Employee selection · Ethical recruitment · Ethics of AI · Bias of AI · Explainable AI

**Introduction:**

AI is now commonly used in the recruitment industry. From around 2017, discussions about AI in recruitment became the norm and since then is the latest trend in the talent industry, with the COVID-19 pandemic accelerating automation in algorithmic hiring (Anon., 2018). AI-based recruiting is used in job postings, sourcing and pre-selecting candidates, and screening as well as digitized interviews. AI is also utilized to perform background and reference checks on candidates. Discriminatory hiring, based on race, gender, age or location for example is an issue associated with AI recruitment, it mostly results from a lack of transparency in the system (Oliver, 2022). Discrimination is a big risk when training AI models. IBM believe that trusted AI can be explained by five pillars of trust, performance, fairness, explainability, robustness and transparency. The notion of explainable AI (XAI) is a system that can be understood by humans, it is now at the forefront of AI as one method of mitigating the bias. Explainable AI promotes fairness, model clarity and accountability (Krishnan, 2020).

**AI-based recruitment benefits**

The four stages in the recruitment process are sourcing, screening, interviewing and selection. The potential for AI automation is apparent in the first three stages of the process (Raghavan, 2019). Some of the benefits of using AI in recruitment is its ability to save companies a lot of time by sifting through ever increasing numbers of resumes per job (Marr, 2022). Identifying the candidates with the right skills for the job from a vast pool of applicants is a major efficiency of AI systems. Although a benefit of using AI in recruitment is its potential ability to reduce human bias, there is growing concern that AI systems are susceptible to bias (Jake Silberg, 2019).

**Bias and recruitment algorithms**

AI in recruitment has the potential to rectify bias or it could also magnify it (Chin, 2019). Bias, with marginalized communities can result from historical data. Bias could also be introduced by programmer code during the development of an algorithm or from the model’s parameters (Dilmegani, 2020). Subconscious bias or the lack of diversity among developers could have an influence on how AI is trained and can disseminate the bias further (Deloitte, 2022).

**Biased algorithms**

Several biased algorithms have been identified, these include Amazon’s recruiting engine, which was found to be biased against women (Robinson, 2019). Google photos algorithm discriminated against black people, while IDEMIA’s facial recognition algorithm discriminated against black women are notable examples (Dilmegani, 2020). Correctional Offender Management Profiling for Alternative Sanctions or COMPAS is an algorithm used for scoring a criminal defendant’s likelihood of reoffending, recidivism; the algorithm is used by judges and parole officers was found to be biased in favour of white defendants (Anon., 2017). TaskRabbit and Fiverr were found to have racial, and gender biased against black and female workers (Anikó Hannák, n.d.). PredPol, an algorithm that predicts crime was biased against minority groups such as those with low-income communities as well as black communities (Reynolds, 2018).

**The need for regulation**

The rise of inequalities in the employment sectors in the US and other prominent economies is believed to be because of AI automation technologies Unregulated AI markets will lead to discrimination, monopolization of data (Acemoğlu, 2021). In the 2017 National Governors Association meeting, Elon Musk remarked that unregulated AI could be one of the biggest risks to human civilization. He called for regulations to be implemented in AI (Musk, 2017). The mostly unregulated use of AI in the recruitment industry has resulted in a white paper presented by the EU in 2020. The report proposes legislation for high-risk AI applications in the recruitment industry. The report outlines five major components for fair AI. These relate to the use of training data, keeping tracking of data sets, informing the end user of AI usage, technical robustness and accuracy of AI, and human oversight over the AI system. The white paper identified explainability as a key component in increasing trust in AI (EU, 2020).

**Gaps in research**

In a 2020 Master’s thesis on fairness in AI-based recruitment, research gaps in AI-based recruitment were identified as the need to develop models for job-specific requirements; these tools would vary by occupations. Mujtaba also proposes that fairness in job postings and hiring decision transparency should also be considered in future research (Mujtaba, 2020). In a 2022 white paper by Enspira, the unity of AI with human intelligence as a means of combatting hiring bias is discussed (Anon., 2022). The paper identifies a five-step recruitment process for using AI as a supplemental tool to reduce bias and improve diversity in hiring. The steps proposed include reviewing job descriptions, the formulation of diversity, equity, inclusion and belonging (DEIB) initiative teams, casting a wider net in the application process, assembling a diverse list of candidates, standardization of the interview process, being skilled in the AI technology, and being proactive with best practices.

**Proposed project:**

Determine how a dataset can be biased and consider mitigation strategies.

**Motivation:**

The motivation for the project stems from the need to identify, manage and reduce the potentially harmful effects of bias in artificial intelligence (AI) systems. Bias could cause harm and discrimination to certain members of society. The mitigation of the sources of this bias is therefore important, not only for the public perception, trust, and acceptance of AI but also for the reputation of the companies providing such AI algorithms.

**Aims and objectives:**

* Demonstrate how bias aggravates already occurring inequalities.
* Show how bias intensifies gender, racial and other stereotypes in the labour market.
* Show how improving existing methods in employment can affect people’s ability to participate in the economy and society.
* Assess the fairness and social benefit of AI-based systems by understanding its transparency and interpretation.

**Stakeholders:**

Ethnic minority groups and the LGBTIQ community as well as women, all at risk of discrimination will benefit from this type of research.

**Related works:**

**AI Fairness 360:**

IBM has an open-source library to detect and mitigate bias in unsupervised learning algorithms called The *AI Fairness 360* toolkit. The toolkit, developed by around 34 contributors has been developed to help detect and mitigate bias in machine learning models. It is available for Python and R and is designed to test for biases. The following library of metrics and debiasing algorithms are included in the API.

Currently it only supports binary classification and will need to be extended to multiclass and regression problems for more complex problems.

## Supported bias mitigation algorithms

* Optimized Preprocessing
* Disparate Impact Remover
* Equalized Odds Postprocessing
* Reweighing
* Reject Option Classification
* Prejudice Remover Regularizer
* Calibrated Equalized Odds Postprocessing
* Learning Fair Representations
* Adversarial Debiasing
* Meta-Algorithm for Fair Classification
* Rich Subgroup Fairness
* Exponentiated Gradient Reduction
* Grid Search Reduction
* Fair Data Adaptation
* Sensitive Set Invariance/Sensitive Subspace Robustness

## Supported fairness metrics

* Comprehensive set of group fairness metrics derived from selection rates and error rates including rich subgroup fairness
* Comprehensive set of sample distortion metrics
* Generalized Entropy Index
* Differential Fairness and Bias Amplification
* Bias Scan with Multi-Dimensional Subset Scan

**IBM Watson OpenScale:**

The Watson OpenScale checks for bias in real time. The application is an enterprise-grade environment that information into how an AI is built, used and delivers return on investment. It allows businesses to have transparent, explainable output without bias. It can be used in a private cloud or be hosted on-premises.

**Google’s What-if Tool:**

Google’s tool will test performance in hypothetical situations, analyse data features and visualize model behavior across multiple models and subsets of input data for different machine learning fairness metrics.

**IBM TakeTwo API:**

The TakeTwo project uses natural language understanding to detect and eliminate racial bias by flagging and classifying phrases and words that are perceived as racially biased, particularly in the United States. As text is composed, the TakeTwo API scans the content for potentially racially based language. The API is built using Python and FastAPI. The Chrome extension uses JavaScript. It can also be run on IBM Cloud using Docker and run locally with CouchDB or IBM Cloudant database.

**EU Horizon project:**

Mitigating diversity biases of AI in the labor market. The EU Horizon project is a consortium of nine institutions to detect and mitigate unfairness in AI-driven recruitment tools. The aim is that both AI-tools and employers use relevant information for hiring, rather than basing decisions on existing bias and irrelevant sensitive features such as race, gender, sexual orientation amongst others.

**IBM Diversity in Faces Dataset:**

The aim of the dataset is to advance the study of fairness in facial recognition systems.

**Data:**

As there were no databases with recruitment data suitable to this case study, an experimental dataset would need to be used to show bias in the machine learning pipeline.

**Methods:**

What methods will be employed to achieve the aims of the proposed project? How will you analyse the data? Explain why those methods are suitable for the project? [15 marks]

Conduct an exploratory data analysis of the experimental dataset to find the patterns of gender and ethnicity bias throughout the recruitment phases.

Investigate whether bias exists in all the stages of recruitment or just the shortlisting stage.

Use some of the tools mentioned above to alleviate bias in the dataset.

Apply machine learning to the dataset to predict who will be shortlisted and who is more likely to obtain the job in question.

Provide recommendations on the findings.

During the Machine learning process, bias could be introduced in any of the stages:

The stages of the Machine Learning:

* Data collection
* Data cleaning
* Feature engineering
* Data split
* Model training
* Model testing

some features that are excluded during the data cleaning stage can lead to exclusion bias. Exclusion bias can also occur when removing duplicates.

sampling bias can occur when the data collected doesn’t represent the environment that a program is expected to implement. The data needs to be carefully chosen to suit the algorithm.

Observer bias can occur during data collection; the observer may only record certain instances of the data and skip other parts which could also be beneficial.

Analysis of the model performance

The following machine learning techniques could be employed to achieve the project aims:

KNN classifier

Support Vector classifier - with grid search algorithm

Decision Trees – Gini and Entropy criterion, also with grid search algorithm

Random Forest classifier - with grid search algorithm

Logistic Regression classifier - with grid search algorithm

Naïve Bayes classifier

**Conclusion**

Although AI-based recruitment is not yet ready to take the place of recruiters entirely yet, it is a valuable tool in the recruitment process, making it easier and more efficient. Fair, transparent AI will encourage trust in AI and promote a more diverse workplace, employee retention is then more likely.

# References

Acemoğlu, D., 2021. *cepr.org.* [Online]   
Available at: https://cepr.org/voxeu/columns/dangers-unregulated-artificial-intelligence

Anikó Hannák, C. W. D. G. A. M. M. S. C. W., n.d. *Bias in online freelance marketplaces: evidence from TaskRabbit and Fiverr,* s.l.: s.n.

Anon., 2017. *kaggle.com.* [Online]   
Available at: https://www.kaggle.com/datasets/danofer/compass  
[Accessed November 2022].

Anon., 2018. *talentlyft.com.* [Online]   
Available at: https://www.talentlyft.com/en/blog/article/207/the-role-of-artificial-intelligence-ai-in-recruitment

Anon., 2022. *enspirahr.com.* [Online]   
Available at: https://www.enspirahr.com/news/white-paper-ai-hiring-bias

Blazek, P. J., 2022. *Why we will never open deep learning's black box.* [Online]   
Available at: https://towardsdatascience.com/why-we-will-never-open-deep-learnings-black-box-4c27cd335118

Chin, C., 2019. *Assessing employer intent when AI hiring tools are biased,* s.l.: Center for Technology Innovation.

Deloitte, 2022. *Deloitte.com.* [Online]   
Available at: https://www2.deloitte.com/uk/en/pages/financial-services/articles/banking-on-the-bots-unintended-bias-in-ai.html

Dilmegani, C., 2020. *AI Multiple.* [Online]   
Available at: https://research.aimultiple.com/ai-bias/  
[Accessed 21 November 2022].

Dilmegani, C., 2022. *Explainable AI (XAI) in 2022: Guide to enterprise-ready AI.* [Online]   
Available at: https://research.aimultiple.com/xai/

EU, 2020. *White Paper: On Artificial Intelligence - A European approach to excellence and trust,* s.l.: s.n.

Jake Silberg, J. M., 2019. *mckinsey.com.* [Online]   
Available at: https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

Krishnan, P., 2020. *ibm.com.* [Online]   
Available at: https://www.ibm.com/blogs/watson/2020/10/how-ibm-makes-ai-based-on-trust-fairness-and-explainability/

leVine, S., 2017. *Axios.com.* [Online]   
Available at: https://www.axios.com/2017/12/15/artificial-intelligence-pioneer-says-we-need-to-start-over-1513305524

Marr, B., 2022. *Forbes.com.* [Online]   
Available at: https://www.forbes.com/sites/bernardmarr/2022/05/06/the-benefits-and-dangers-of-using-ai-in-recruitment/?sh=789a8afd292f

Mujtaba, D., 2020. *Michigan State University.* [Online]   
Available at: https://d.lib.msu.edu/etd/48478/datastream/OBJ/View/

Musk, E., 2017. *Ahead of the curve* [Interview] (15 July 2017).

Oliver, C., 2022. *didlaw.com.* [Online]   
Available at: https://didlaw.com/discrimination-in-recruitment

Raghavan, M., 2019. *brookings.edu.* [Online]   
Available at: https://www.brookings.edu/research/challenges-for-mitigating-bias-in-algorithmic-hiring/

Reynolds, M., 2018. *newscientist.com.* [Online]   
Available at: https://www.newscientist.com/article/mg23631464-300-biased-policing-is-made-worse-by-errors-in-pre-crime-algorithms/

Robinson, D., 2019. *ns-businesshub.com.* [Online]   
Available at: https://www.ns-businesshub.com/technology/ai-recruitment-bias/