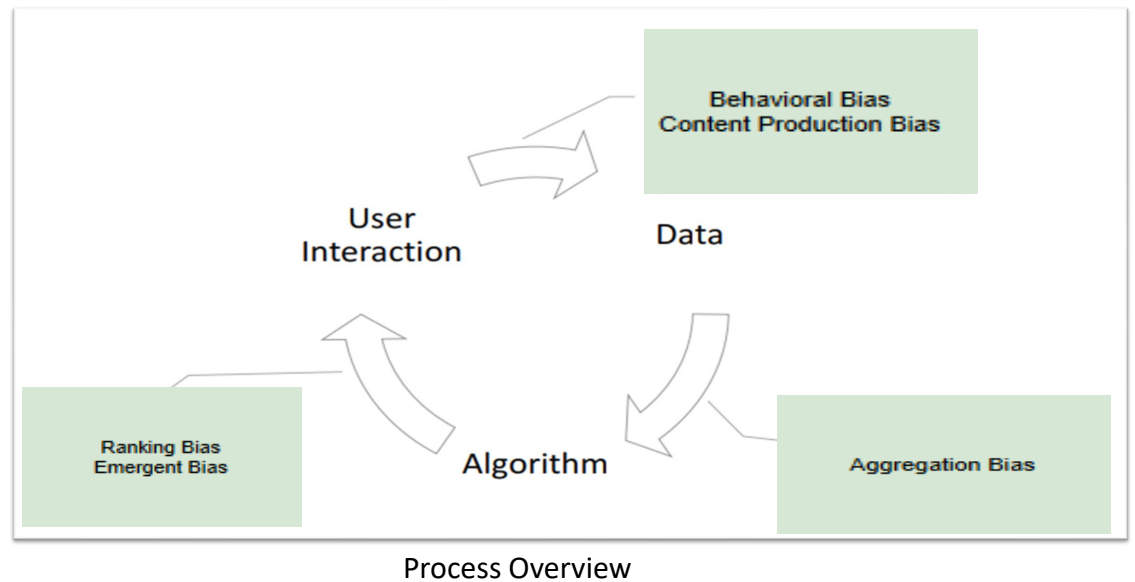


Promoting Fairness in Human-Centered AI Applications in Recruitment Systems

Oussama Dhouioui | Budapest University of technology and economics



Abstract: This research focuses on promoting fairness in AI-driven recruitment systems by identifying and mitigating demographic and social biases in the hiring process. It explores the application of deep learning techniques to address biases in candidate selection and ensure equitable outcomes for all applicants. The study aims to develop fairness frameworks and evaluate AI models using fairness metrics to promote diversity and inclusivity. Additionally, it proposes regulatory adjustments to improve transparency and data protection in AI recruitment systems. Ultimately, the research seeks to enhance trust in AI while fostering a more diverse workforce.



Process Overview

Introduction

The rapid adoption of AI in recruitment processes has brought significant efficiencies but also raised concerns about fairness and bias. AI systems, which analyze large datasets to make hiring decisions, can inadvertently perpetuate societal biases embedded in historical hiring patterns. These biases, whether related to gender, race, age, or other socio-demographic factors, can lead to unfair outcomes that disadvantage certain groups of candidates. As these systems are increasingly used across industries, it is crucial to ensure that they promote diversity, equity, and inclusion by identifying and mitigating biases in their algorithms. This research focuses on examining the biases present in AI recruitment systems and developing methods to ensure that these systems are fair, transparent, and trustworthy for all candidates. By leveraging deep learning techniques and fairness frameworks, the study aims to foster more inclusive hiring practices and guide the development of regulations that ensure the ethical use of AI in recruitment.



Ethical Considerations

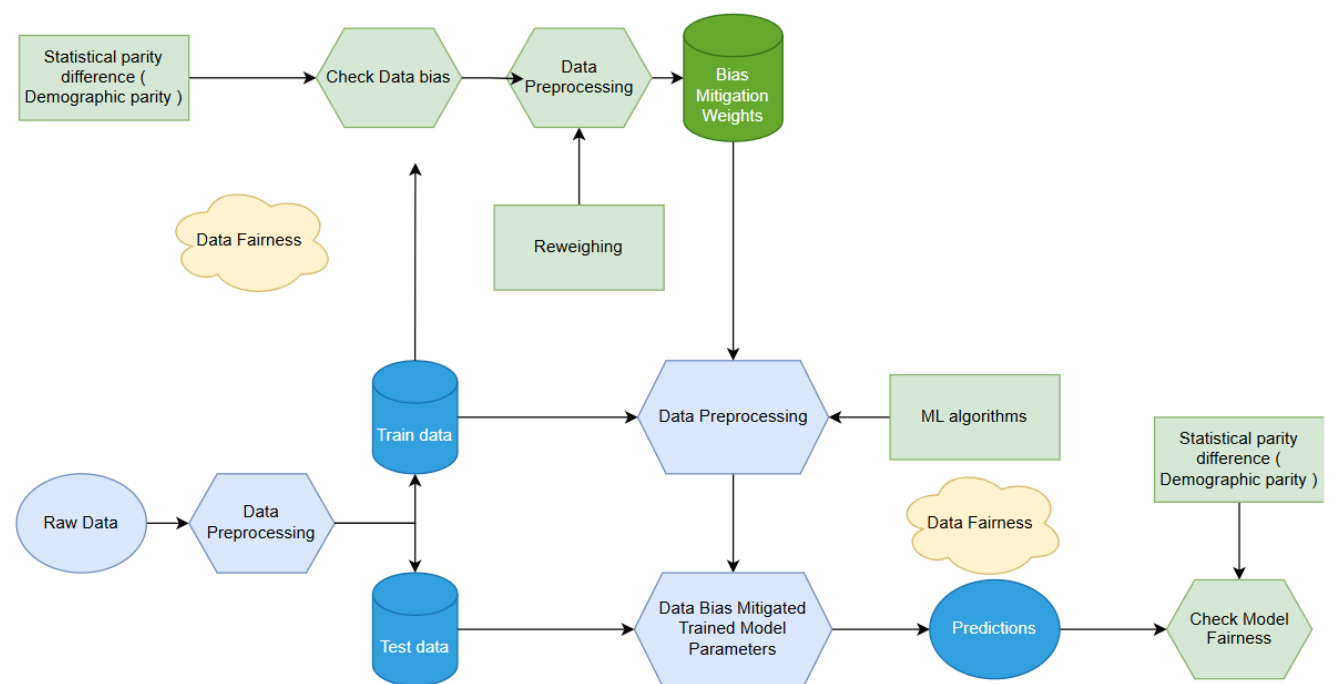
Research Methodology

The research methodology involves a mixed approach combining data analysis, deep learning techniques, and fairness evaluation. First, recruitment datasets will be analyzed to identify and assess demographic biases. Then, deep learning models will be trained using fairness-aware algorithms to mitigate bias in AI recruitment systems. Fairness metrics, such as demographic parity and equal opportunity, will be applied to evaluate the outcomes. Additionally, the study will propose regulatory adjustments and ethical guidelines to ensure transparency and fairness.

Literature Review

Artificial intelligence (AI) has increasingly been integrated into recruitment processes, where machine learning algorithms are used to screen resumes, evaluate candidates, and even conduct preliminary interviews. AI promises increased efficiency and reduced human bias in hiring. However, studies have raised concerns that the data-driven nature of these systems could inadvertently perpetuate or even amplify existing biases in hiring practices (Binns, 2018; Dastin, 2018).

For example, in a well-known study, Amazon's AI recruitment tool was found to have a bias against female candidates because it was trained on resumes predominantly submitted by men, reflecting historical gender imbalances in the tech industry (Dastin, 2018). This highlights the potential for recruitment systems to unintentionally reinforce societal biases and gender disparities.



Proposed Pipeline for Bias checking

Conclusion

This research aims to address the critical issue of bias and fairness in AI-driven recruitment systems. By identifying and mitigating demographic biases through deep learning techniques and fairness-aware algorithms, the study seeks to ensure that AI recruitment models offer equitable outcomes for all candidates. Additionally, the research emphasizes the importance of transparency, human involvement, and robust regulatory frameworks to uphold ethical standards and protect candidate data. Ultimately, the findings aim to foster trust in AI systems, promote diversity in hiring, and guide the development of more inclusive, fair, and transparent recruitment practices.

Acknowledgements.

The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant NoCEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This poster was created as part of the Blended Intensive Programme organised under the Erasmus + Programme of the European Union.

References:

- [1] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, Aram Galstyan. A Survey on Bias and Fairness in Machine Learning. arXiv:1908.09635
- [2] Kamiran F, Calders T (2009a). Classifying without discriminating. In: Proceedings of IEEE IC4 international conference on computer, Control & Communication. IEEE press
- [3] Kamiran, Faisal and Calders, Toon. Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1):1-33, 2012
- [4] Hardt, M., Price, E., and Srebro, N. (2016). "Equality of opportunity in supervised learning," in Advances in Neural Information Processing Systems, eds D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett (Barcelona: Curran Associates, Inc.), 3315-3323.
- [5] Chouldechova, A. (2017). Fair prediction with disparate impact: a study of bias in recidivism prediction instruments. Big Data 5, 153-163. doi: 10.1089/big.2016.0047