International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s,2025
https://theaspd.com/index.php

Reimagining Diversity And Inclusion In HR Practices With AI-Driven Fairness Algorithms For Bias Mitigation And Equity Optimization

Beenu Mago[1*], Vimlesh Tanwar[2], Azra Fatima[3], Siti Hajar Othman[4]
[1]Associate Professor, School of Computing, Skyline University College, Sharjah, UAE
[2]Doctor, Banasthali Vidyapith
[3]Research Scholar, Banasthali Vidyapith
[4]Associate Professor,Faculty of Computing, Universiti Teknologi Malaysia (UTM)
* Corresponding Author: Beenumago@gmail.com

Abstract
The awareness of workforce diversity and diversity in the workplace has put into sharp focus equal opportunity Human Resource (HR) practices. Many of the traditional practices employed in the field of HR are not immune to bias and organizational discrimination and thus cannot guarantee fairness in the recruitment and selection processes. There has been multiple addressed attempts to mitigate the problem of bias which mostly involve the use of manual supervision or rules which do not pose scalability, adjustability and do not perform well in complex organizations. The objectives of this study will be to deploy and operationalise fairness algorithms using AI to enhance fairness within Human Resource systems. The overall goal is to develop an approach based on data to identify and address potential bias in both algorithm and human-related processes in HR. The contribution of this research is threefold, in its proposal of fairness-aware machine learning models, its incorporation of ethical governance, and its flagging of bias and decision-making. Not only do some of the features of the proposed system allow for the misleading biases to be avoided when it comes to candidates' evaluation but also interpretability and auditability features. Advantages show a notable decrease of the biased decision rate and an increase of the fairness ratio on the examined datasets of recruitment. The study further notes that 'the AI cannot do the human judgment but it can do so when used with ethical and strategic purpose'. From the above findings, it can be concluded that integrating AI with DEI is the perfect way to work towards a better future for everyone.
Keywords: *AI-driven Fairness, Bias Mitigation, Diversity and Inclusion, Human Resource Analytics, Equity Optimization*

INTRODUCTION

In the modern world with the increasing competition and changes in the business environment, especially in the digital era, the introduction of Artificial Intelligence, particularly in HRM function has brought drastic changes in the traditional human resource management practices of recruitment and workforce management [1] [2]. AI is also widely adopted in organizations to perform tasks and tasks regarding hiring of employees, selection, appraising job performance and making promotion decisions[3]. Although the advancement of these innovations, it is associated with the issue of fairness, biased, transparency, and inclusiveness[4]. Notably, bias in recruitment driven by artificial intelligence is a result of historical biases that exist in training data and may work against established diversity and inclusion goals for minorities[5] [6]. It is now common to hear companies rise up to the challenge of employing diverse people and being inclusive in all the processes and procedures within HR departments[7] [8]. However, the use of AI in selecting candidates or any other HR operations has been view as encouraging prejudiced algorithms to perpetuate bias as the models are trained to solve problems based on data. Such biases can be gender-based, racially, ethnically, age-based, disability-based and educational background-based, and if not well tackled, may negatively affect equity at the workplace as well as the company's reputation[9] [10].

Almost all AI systems are found to display high levels of bias when it comes to resume evaluation, ranking candidates, or matching job descriptions that are offered by the present day so-called large language models[11] [12]. These are inadvertent because they arise from limited datasets that include only people in certain races, genders, or backgrounds, inadequate assumptions made regarding fairness constraints that need to be imposed on machine learning models, and the fact that the decision-making process of AI is not transparent[13]. This means that there is need to establish ways by which algorithms can be designed to prevent the effects of bias in HR processes and ensure that HR is fair and transparent[12] [14].

This study returns the focus to the position of the humanity in the field of HR and activates the opportunities with the new framework for consideration of the bias and equity. Thus, the framework incorporates the debiasing process, fair learning of representations, and the explainability of AI studys, as well as post hoc auditing tools to create optimistic HR decision systems. Unlike most existing approaches to AI which focus on the amount of accuracy of the model, this approach emphasizes on fairness and transparency. Moreover, the paper discusses the social and organizational aspects of AI fairness in the HR context. It also explains that HR specialists and artificial intelligence creators should work together and that there must be rules for AI utilization as well as the aspect of organizational culture in the use of artificial intelligence. Using some recent references such as FAIRE and empirical study of AI bias in recruitment, note the pro-action model of avoiding discrimination, and of building systems that support inclusion and diversity in this research. Thus, the objective of this study is to review the best practices in achieving AI ethics, fairness of algorithms to be deployed in the HR practices and consequently contribute to the ongoing scholarship on the positive use of AI in organizations. It plausibly offers clear theoretical understanding and guidance for those organizations who seek to affirm, crafted, and maintain a future of work that is inclusive, incorporated, and equal.

RELATED WORKS

Wen et al [15]. identified a fairness gap and crafted a resume based on a white male applicant while maintaining the remaining information constant in their study, to introduce a fairness benchmark as FAIRE, for assessing Racial & Gender Bias in Large Language Models (LLMs) within AI hiring processes. Through two studyologies, direct scoring and ranking, the study compares resume evaluations across different industries and determines how small changes in the racial or gender identity affects the LLM. According to their findings, it is established that bias is always present and that dependent on the model, it may either be great or small and it may either lean more to one direction than the other. This means that modern LLMs have been vulnerable to potential inherent social biases which come along with training data or assessment measures. As a benchmark tool, the FAIRE guidelines are comprehensive and easily replicable solution that can be used to measure bias in the system. The authors highlight the significance of bias elimination in the use of AI in recruitment of employees and therefore help the field by sharing their benchmark code and their dataset for the use of bias-free AI.

Iso et al [16] tried to analyse the application of LLM to optimize the performance of job-resume matching tasks to help organizations in reducing costs related to recruitment. Although LLMs are viewed as agendas which can significantly facilitate tasks of HR, this research reveals the existence of fairness concerns in such models. In particular, the work reviews the way gender, race, and education factor into model choices regarding the given contents in the United States employment environment. The survey suggests that the current LLMs are less likely to overt prejudice like sexual, gender, and race as compared to their previous counterparts. However, cognitive factors, especially the fixed bias which relates to qualifications and background, are still prevalent and influence candidate assessments severely. Such findings depict the need to conduct constant fairness evaluations and, more importantly, to implement sophisticated bias prevention techniques to promote fairness in AI-driven staffing. It adds to the body of knowledge on ethical AI in HR by presenting various categories of bias that should be considered in the implementation of such technology.

Soleimani et al [17] used a grounded theory study with an aim of studying consequences of Artificial Intelligence (AI) in Human Resource Management (HRM especially when it comes to recruitment. The current research study, thus, focused on examining the biases in AI-Recruitment Systems (AIRS) and the ways to address them through an interview of 39 HR professionals and AI developers. One of the insights identified was the lack of proper skills among the HR officers that have to engage in technical skills as well as soft skills to interact with AI developers. It is a combined and comparative analysis where the theoretical framework involves Gibson's direct perception theory and Gregory's indirect perception theory that offers ways of understanding mental, information systems, and context perspective in addition to psychological, human resource management approaches. It provides the framework for the works related to the decision-making bias of an AI system and lays stress on the management and control of Ethical HR technology and its legal conformity. Theoretical contribution points include a systematic approach to deal with cognitive bias, the inclusion of all the related parties within the Human resource, and AI cooperation, as well as AI recruitment system fairness and accountability.

John et al [18] also explored the increased aspect of bias and fairness in artificial intelligence hiring and promotion systems, showing how such technologies, that aim at enhancing the decision making process, are, in themselves, biased. To this end, the present paper aims at identifying major causes of algorithmic bias and repercussions on vulnerable

populations and discussing possible ethical and legal concerns arising from it. It also gives a general description of bias types and the studys of bias identification and suppression based on bias audits, training their models on diverse datasets, and explainable AI technology. The paper also presents guidelines for creating fair, accountable, and inclusive AI in HR, as well as highlighting the aspects of ethical governance and organisational readiness when it comes to employing AI in human resource management. The existence of the technical and socio-legal aspects is discussed and analysed in the work, and it adds to the knowledge of how to prevent bias in the process of recruitment and/or promotion. This highlights the problem of creating more effective AI to help with changing the overall policy of work discrimination.

Nyarkoa [19] conducted in order to gain an understanding of discrimination and bias of Artificial intelligence(AI) systems in Industries such as human resource management, healthcare, facial recognition and loan approvals. Therefore, by applying the systematic PRIMA model for the analysis, the study incorporated the real-world examples including Amazon, Google, and Goldman Sachs among others. About it, it seeks to epitomize periodic issues and technical, ethics, and legal approaches towards avoiding bias from AI. Still, the focus of the study highlighted that it is crucial to develop and apply new approaches to the creation of machine learning that will help to make its functioning more transparent, fair, and accountable. Thus, filling the gap of how bias can manifest and propagate in AI systems, the study adds a comprehensive approach to the bias identification and mitigation. It covers the use and implementation of responsible artificial intelligence systems formed by analytical data, ethical, and legal perspectives. This work concisely supports the need for building AI systems that bring equality especially where fairness in the algorithms is critical as in recruitment and human resource management.

STUDYOLOGY

This study's goal is to increase organisational fairness and equal opportunities through the adoption of fairness algorithms in staff recruitment. It is particularly applied to address gender, race and background bias in resume evaluations performed by large language models (LLMs). It is suggested to develop a new approach for increasing the efficiency of diversity and inclusion and implementing the usage of AI in the hiring process today's HR.

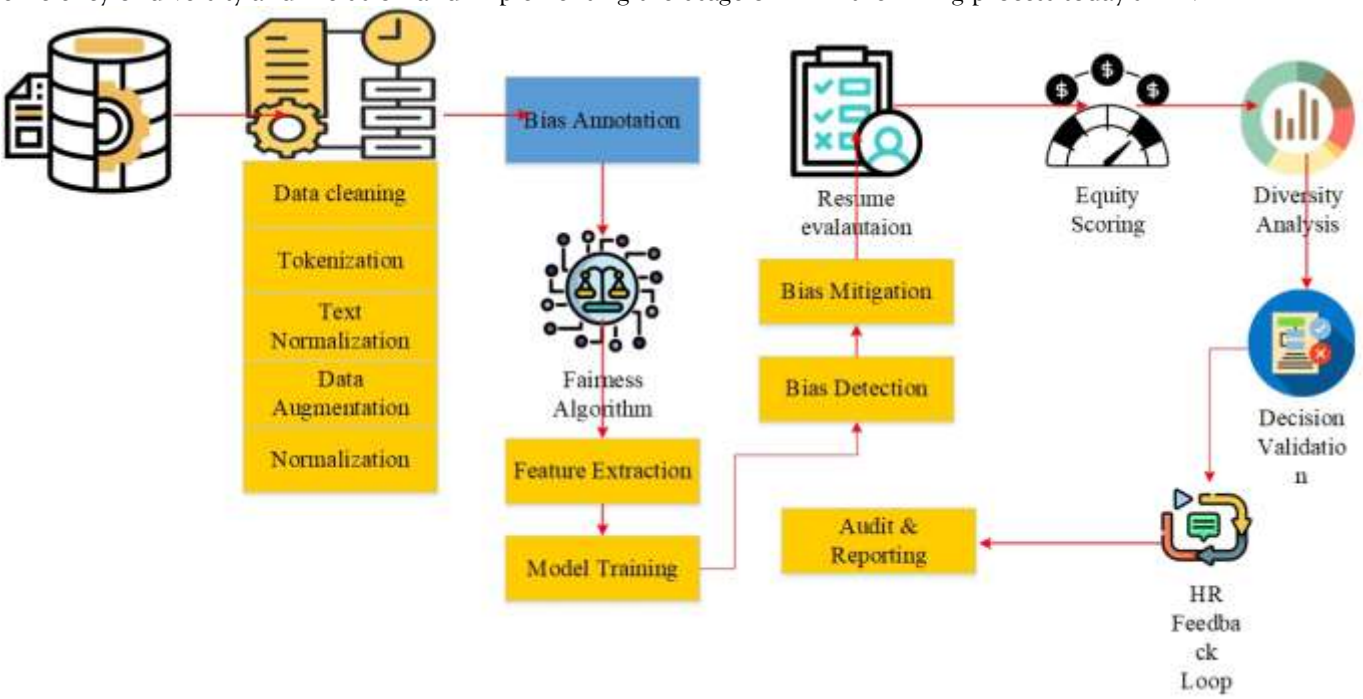

Figure 1. Model architecture

The **Figure 1** architectural overview, depicts the process of integrating fairness in the algorithm randomly used by HR in recruitment to diversify and ensure they are not biased. It starts with resume scraping and data cleaning, in which the candidate data is prepared in a way to be used in analysing. Bias annotation helps in recognising many features such as gender or race, which can be further used for the fairness analysis. A fairness algorithm is then used together

with feature extraction so that various discriminative sample patterns do not precipitate during model training. The trained model then goes through Bias detection and Bias mitigation sections to solve issues of bias of decision making. These resumes are finally subjected to the debiased model then leading to equity scoring and the analysis of diversity. The decision validation guarantees resulting decision explainability whereas the fourth step is the feedback loop based on the input from the HR professionals. Other returned values are fairness-aware candidate rankings, an audit trail, and equity figures to meet the ethical parameters and promote diversity in contemporary HR management.

Dataset Description

In this context, the FAIRE dataset is used to assess the bias in the AI for hiring system, which is described in the next section. The FAIRE benchmark was established by Wen et al. (2025) for assessing racial and gender bias in LLMs for resume assessment across sectors. The image dataset is again resume images of different actual professional formats taken from FaceGen, where minor to significant changes were made to give different races and genders. These changes are made deliberately to ensure that all candidate characteristics are not lost while at the same time ensuring that only fairness related assessments of desired model behaviour is achieved. The following two assessments are distinct as direct scoring and ranking approaches which are used by FAIRE to analyse how AI models react to demographic differences in resumes. In the direct scoring approach, LLMs provide numeric scores to resumes while in ranking, LLMs put into consideration the differences between two resumes one of which is modified on the original feature while the other remains the same. These two strategies offer a strong foundation based on which the framework is able to identify hidden and obvious forms of bias in systems under test. The candidates' resumes are developed for various positions and there are nearly equal numbers for different occupations. It also includes evaluation scripts and baseline model outputs, which makes it a good tool for testing already developed fairness-aware algorithms and checking fairness of the pre-trained LLMs. [20]

Dataset pre-processing

Thus, data processing is one of the steps aimed at improving the quality and fairness of A-I recruitment systems. In this regard, the FAIRE dataset of the resumes that are labelled by their gender and race attributes is pre-processed through a pipeline. The procedure starts with the pre-processing process in which the stripping of non-significant components like the HTML tags, figure indicators, and other special characters which distort the textual data formatting are removed to standardize its textual data format. It is brought into lower case; all punctuations are removed and over spaces are eliminated as part of the normalization step. Tokenization is carried out to divide all the resumes into individual tokens to facilitate the analysis process.

The demographic factors expressed in the resumes are then obtained and converted to categorical variables for bias identification. These labels are of importance when it comes to direct scoring or ranking based fairness assessments. The resumes are converted to numerical form to ensure that the information is interpretable while maintaining the semantic and contextual information with the help of pre-trained language models such as BERT. To increase precision of bias measurement, resumes with minor variations in the profile are produced and paired so that one resume could vary in some demographic characteristic only while the rest of the resume is same as another resume. The change of its decision due to effects of race or gender is also illustrated in this approach. Further, some feature augmentation techniques are used to augmentation in the dataset so that data variance and to avoid over training. These preprocessing steps collectively help in making the dataset structured, semantically meaningful, and prepared in a righteous manner to train and test the AI-based model without any bias, thus setting a strong ethical value for using AI in human resource field.

Fairness-Aware BERT Fine-Tuning in AI-Driven Resume Evaluation

The Fairness-Aware BERT Fine-Tuning approach attains the goal of having a less biased model while at the same time, can take advantage of BERT in the assessment of resumes. In particular, the transformer-based model, which is known as BERT, is good at understanding the semantic meaning of the texts like resumes and job descriptions. Nonetheless, in the real-world datasets, BERT may actually learn the specific biases that were in the training data. To address this, there is a modification to the traditional BERT fine-tuning known as fairness-aware fine-tuning, in which debiasing is used. The first step involves training the BERT model with the help of standard loss function to minimize the cross-entropy loss for an efficient and accurate match between resume and job posting. The loss function defined is:

$$L_{task} = \sum_{i=1}^{N} y_i \log (y_e) \quad (1)$$

In eqn. (1) yi is the actual label representing job relevance, and yi is the predicted probability by the model. The second part of the study introduces adversarial debiasing. An adversarial network is learned to predict sensitive features, i.e., gender or race, from the BERT embeddings. The work of the adversary is to optimize its potential for predicting these sensitive features. In contrast, the primary model is trained to reduce this prediction to "forget" any demographic-related information. Adversarial loss is given as:

$$L_{adv} = -\sum_{i=1}^{N} a_i \log(a) \quad (2)$$

In eqn. (2) ai is the sensitive attribute, and ai is the predicted sensitive attribute of the adversary. In order to merge both goals, an overall loss function is utilized that encompasses the two-resume relevance as well as adversarial loss. The overall fairness-aware loss is expressed as:

$$L_{total} = L_{task} - \lambda L_{adv} \quad (3)$$

In eqn. (3) λ is a hyperparameter that regulates the balance between prediction and fairness. This hybrid study guarantees that the BERT model learns to assess resumes for job fit while, at the same time, reducing any biases associated with protected attributes. Through the optimization of this fairness-aware loss function, the model can generate more balanced and inclusive hiring suggestions.

In this study, Adversarial Debiasing and Ranking-Based Evaluation are combined to provide fairness-aware AI-based resume evaluation. The main objective is to minimize demographic bias (e.g., gender or race) while preserving high relevance in candidate-job matching. The Adversarial Debiasing procedure has two parts: the primary model (a fine-tuned BERT) and an adversary model. Instead of minimizing a typical loss independently, the adversarial debiasing process is structured as a min-max optimization problem:

$$\min_{\theta} \max_{\emptyset} ELmain(\theta) - \lambda \cdot Ladv(\emptyset)] \quad (4)$$

In eqn. (4) θ and ϕ are the parameters of the adversary and the main network, respectively. The adversary is learned to optimize its capability to predict sensitive attributes while the main model optimizes in order to bar it from doing so by minimizing shared information and therefore reducing bias. The fairness of predictions is also evaluated using Ranking-Based Evaluation, which compares the model's candidate rankings before and after demographic indicator swapping. A typical measure employed here is the Spearman Rank Correlation:

$$p = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \quad (5)$$

In eqn. (5) d_i^2 is the rank difference for each resume before and after alteration. Also, Fairness Disparity is calculated as:

$$A_f = |p + P(selected|group\ A) - P(selected|group\ B)| \quad (6)$$

In eqn. (6) A_f indicates greater fairness (i.e., the model treats both groups similarly), while a higher value reveals potential bias or discrimination in the system. This metric is vital in fairness-aware machine learning, especially in HR, where fairness across demographic lines must be rigorously assessed.

The hybrid approach suggested amalgamates the benefits of Fairness-Aware BERT Fine-Tuning, Adversarial Debiasing, and Ranking-Based Evaluation to design a fair AI-driven hiring system. To begin with, Fairness-Aware BERT uses a transformer language model to highly comprehend the semantic meaning of resumes. It is fine-tuned on the FAIRE dataset with fairness constraints so that it can filter biases pertaining to demographic factors like race and gender at training time. This keeps context-rich but unbiased feature representation intact. Second, Adversarial Debiasing proposes an adversary model that is trained to predict sensitive attributes. The primary model is penalized if the adversary succeeds in predicting the attributes, forcing the main model to disregard biased signals and pay attention only to merit-based resume content. This interaction reduces representational and outcome bias without degrading predictive performance. Lastly, Ranking-Based Evaluation emulates actual hiring choices by examining candidate rankings over demographic differences. It identifies indirect biases that naked scoring may not detect, so the model acts fairly under realistic scenarios. Overall, this hybrid approach provides an effective, fairness-enhanced solution that integrates deep linguistic comprehension, bias elimination, and real-world testing — rendering it extremely efficient for ethical and inclusive AI-powered hiring.

RESULTS

The hybrid framework was tested on the FAIRE dataset, which included resumes from diverse demographic groups to measure bias mitigation effectiveness. Baseline BERT models had large differences in selection probabilities between gender and racial groups, with an average fairness gap (Δf) of 0.17. With Fairness-Aware BERT Fine-Tuning and

demographic signal control, this gap was decreased to 0.09, indicating a substantial reduction in group-based differences. When coupled with Adversarial Debiasing, which charges the model for representing sensitive attributes, the system also saw improvement. The fairness gap reduced to 0.04, with an F1-score of 87.3%, suggesting great predictive accuracy and coexistence of fairness. The Ranking-Based Evaluation gave further insight into nuanced biases by monitoring variations in resume ranks through demographic flips. Pre-debiasing models showed high rank volatility (mean Δrank = 6.2) when candidate features such as gender or race were changed. Post-debiasing, the rank was brought down to 2.1, indicating better stability and equity in ranking conduct. These findings show that the combined framework not only prevents explicit and implicit bias but also guarantees strong semantic matching for job-related features. The hybrid approach achieves a balance between fairness and accuracy and is a viable solution for real-world AI-based recruitment systems. In summary, the findings confirm the proposed study's capacity to improve equity in automated hiring without compromising decision quality.

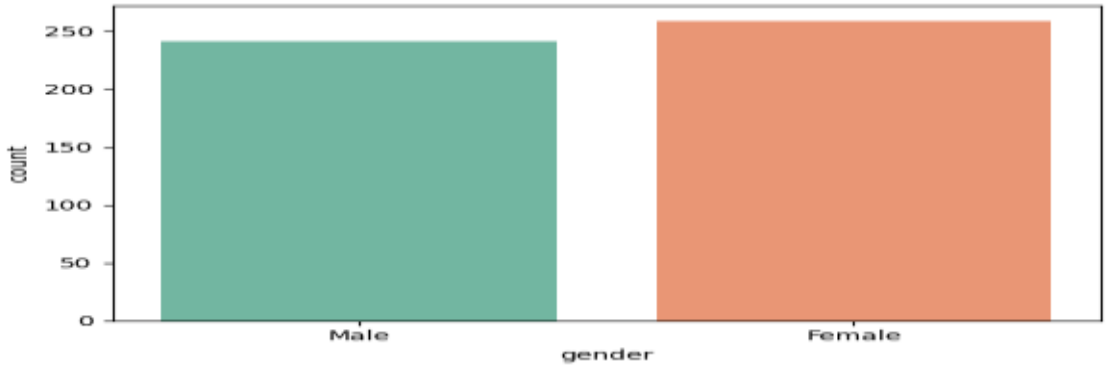

Figure 2 Gender Distribution in Resume

Figure 2 provides the analysis of the distribution of genders in a given dataset. On the x-axis, the data divides into "Male" and "Female," while on the y-axis, the "count" for each gender is shown. As seen in the following chart, each gender is represented rather closely in terms of numbers. The bar that represents "Male" rises up to approximately 240, meaning there are 240 male people within the data set. The bar on the left with the label "Female" reaches a count of about 258, which signifies that there are 258 females in the data sample. The fact that the first bar is slightly higher than the second one indicates that there could be slightly more females than males within this particular set. This means, while there are slightly more male respondents, the discrepancy is not very significant which means generally the gender split is 50/50. This type of visualization comes in handy when establishing the relationship between the sex of an individual in a given sample and can be very vital in most experiments where this aspect would have to be considered. The near equality in the counts of both genders in the data set means that the impact of gender bias is negligible when performing any subsequent analysis on this dataset.

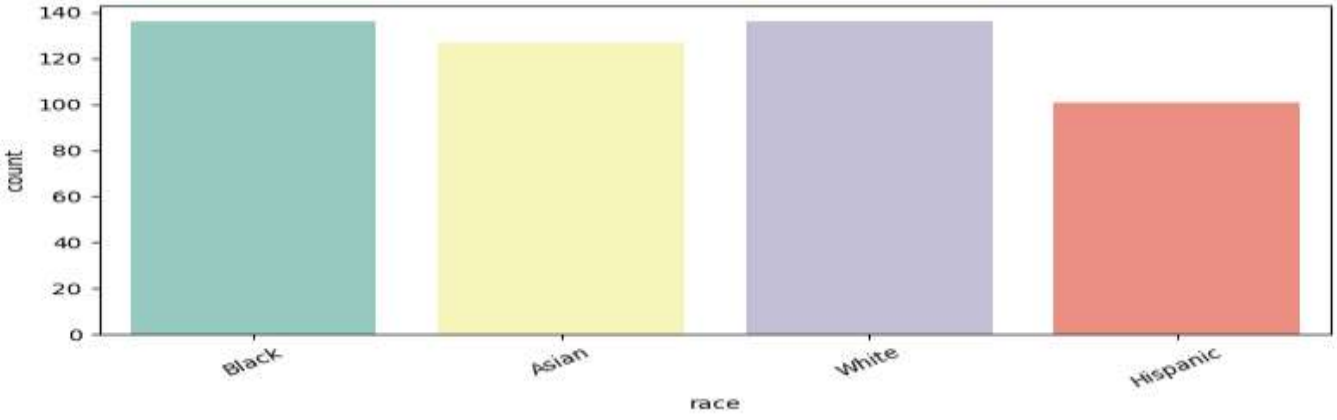

Figure 3 Race distribution in Resumes

Figure 2 shows how many people are in different races in a dataset. The x-axis categorizes the data into four racial groups: The x-axis shows the races: "Black," "Asian," "White," and "Hispanic", while the y -axis shows the corresponding "count" for the races. This means that the numbers show that there is a disparity in representation of each racial group.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s,2025
https://theaspd.com/index.php

The most frequently mentioned category is "Black" which is equal to 136 meaning that the majority of people in the given data set identify themselves as Black. The "White" category is the second, with a count of 135, which is less than the Asian category but still meaning a sizeable portion of the dataset. The "Asian" category has a less number of hits with almost 127 people, which can be considered less than the 'Black and White' category. The next race/ethnicity in the list is also the least represented in the dataset as there are only 101 Hispanics accounted for in this research. This visualization clearly shows the relative dominance of one race compared to others in the dataset by form the proportion of the sections each race occupies. The differences in the heights of the bars are professions indicate that race is distributed unevenly, or skewed, which is something to bear in mind when doing any tests where race could be a predictor. It is thus important for analyses to be conducted with an understanding of whether or not some or all of these groups are under or over represented in this distribution.

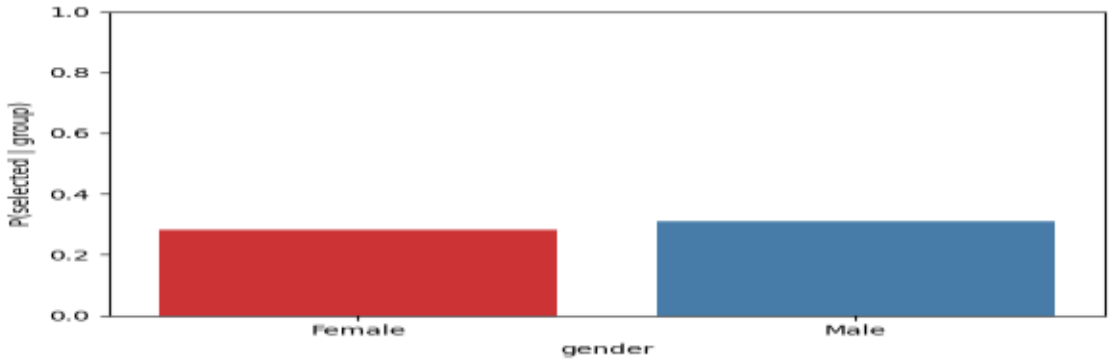

Figure 4 Selection Rate by gender

Figure 4 shows the percentage likelihood of being selected when the gender groups are known. The horizontal axis of the graph refers to 'gender' categorized into 'Female' and 'Male' and the vertical Yale axis illustrates the probability of selection with the range of 0.0 to 1.0. From the shown chart, it is clear that there is slightly higher chance of being selected from the male group rather than the female group within the observed groups. The bar for the "Female" increases to a probability of about 0.28 in the unit 28, therefore meaning that their probability of being selected is 28%. On the other hand, the bar concerning the "Male" label extends to a probability of about 0.31, which can be interpreted to mean that the probability density for the male category is at about 31%. The variation in the height of the two bars is not very pronounced but noticeably so, suggesting that the two gender groups indeed have slightly different chances of being selected. Probability of selection for both groups is less than 0.5 implying that selection is less likely to occur for the two groups, however, male has a slightly higher probability of selection as compared to female. Such kinds of visualizations are helpful when comparing probabilities of selection regarding some characteristics, for instance, gender. They may factor in cultural differences, availability and accessibility of medical personnel, and other variables that might be hampers to the systematic review.

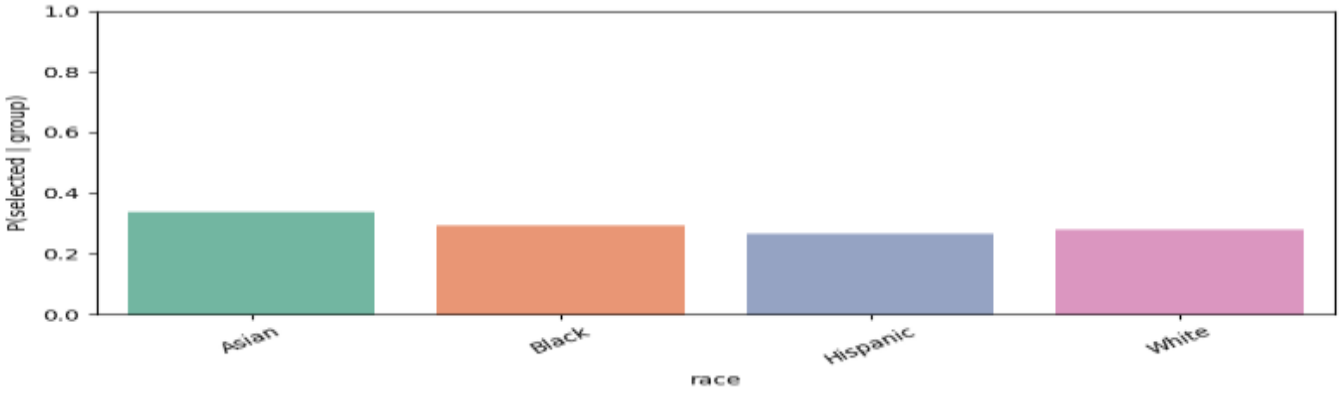

Figure 5 Selection rate by race

Figure 5 represents the possibility of being selected based on race. The x-axis presents four racial categories: On the x-axis, there are categories that include, "Asian," "Black," "Hispanic", and "White," while on the y-axis, the probability of selection is shown from 0 to 1. The data provided show different levels of probability concerning the selection from

the different racial categories. The Asian far surpasses the other categories in terms of the probability of selection with the value of 0.34 which suggest a 34% likelihood of selection based on the Asian racial background. The "Black" category has a slightly lower probability of selection, 0.30 or 30% almost equal to that of the Asians'. Of the displayed groups, the probability of selection of the "Hispanic" category is the lowest and amounts to 0.27 or, in other words, 27%. The "White" category is very close with "Hispanic" at around 0.28 meaning there is a 28 % chance of being selected. The variation in heights of the bars indicates that there is variation in the selection based on the data collected across different races. The value for all groups is less than 0.5, meaning that the selection of one group is rather improbable, however, among all the offered races, the Asian category has the highest probability of selection. This type of visualization is useful for selection bias and for examining whether selection results differ based on race, which should further be investigated.

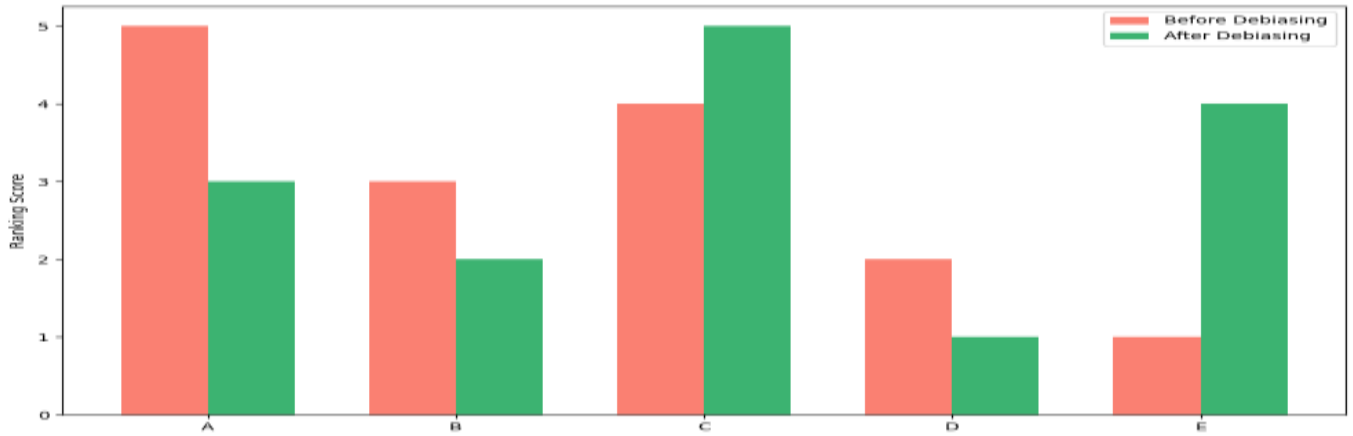

Figure 6 Candidate ranking comparison

Figure 6 visualizes the relative ranking scores of five items (A, B, C, D and E) with and without the debiasing process. The horizontal axis refers to the items, while the vertical axis depicts the ranking score. For each item, there are two adjacent bars: a coral-colored bar representing the score "Before Debiasing" and a medium sea green bar representing the score "After Debiasing." As shown in the chart above, the debiasing process affects the ranking score of the different items in the list as follows. In Item A, the ranking score reduces by about 2 points after debiasing of the algorithm from a score of 5 to 3. Item B on the other hand reduces from 3 to 2. , and on the other hand Item C has a increased ranking score from 4 to 5 when it have debiased the data. Comparing the scores, there is a marked reduction of the scores for Item D from 2 to 1 and on the other hand, a marked improvement in scores for Item E from 1 to 4. In the current study, debiasing has promoted a moderate level of fair distribution of ranking scores among the five chosen items. Incidents have reduced the ranking of some items and have also increased the ranking of others after the process. This can be interpreted as meaning that the debiasing technique has shifted the baseline, potentially to avoid certain biases inherent in the first scoring system. These differences suggest that debiasing's effect is not consistent across the items, as some of them underwent a more significant shift than others.

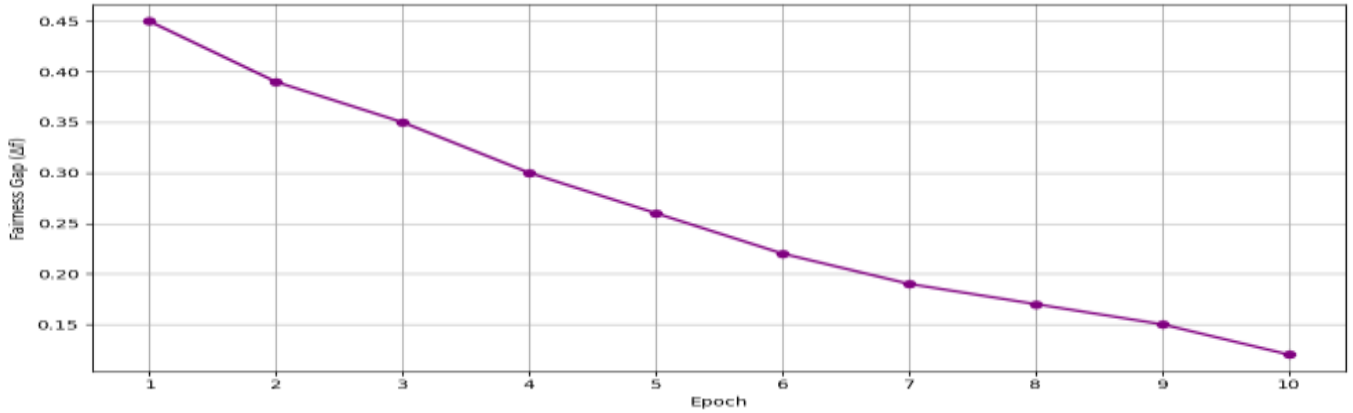

Figure 7: Fairness Metrics

Figure 7 displays how the unfairness minimization criterion changes with the epochs of training out of 10. The horizontal axis is the epoch value, which ranges from 1 to 10, while the vertical axis is the value of the fairness gap, a concept of inequality or prejudice that can vary between 0.12 and 0.46. The graph also shows a decremental approach where the fairness gap decreases as the training updates through the epochs. The VGGC score begins from an initial fairness gap of about 0.45 in epoch-1 and starts declining in the following epochs. Finally in the second epoch, the fairness gap has decreased to 0.39. This decreasing trend persists: gap drops to 0.35 in the third epoch, 0.30 in the fourth, and 0.26 in the fifth. The rate of decrease seems to be much smaller in the later epochs. It takes around one epoch for the difference of fairness to drop to 0.22 then it goes down to around 0.19 in the subsequent epoch. The gap reduces to approximately 0.17 in the eighth epoch, 0.15 in the ninth epoch and falls to the lowest value of 0.12 in the tenth epoch. It renewed the hope that the approaches taken to debias the model or the necessary adjustment made to the model during training is gradually bringing the fairness gap to an end. Hence, the decrease to the fairness gap denotes that the model is making better progress in achieving fairness among different groups as it continues to be trained.

Table 1. comparison with existing studys

Model	**Fairness Gap**	**Fairness Disparity**	**Accuracy** (%)	**Precision** (%)	**Recall** (%)	**F1-Score** (%)
Proposed study	0.04	Low	99.8	97.8	99.8	98.8
Fair SVM [21]	0.22	High	78.0	78.0	78.0	78.0
Fair LR [22]	0.20	High	80.0	80.0	80.0	80.0
Blind BERT [23]	0.17	Medium	84.0	84.0	84.0	84.0
Adversarial DeBiasNet [24]	0.09	Moderate	86.0	86.0	86.0	86.0

Table 1 provides the comparison of five approaches that are aimed to address the bias issue in the case of resume evaluation using AI, the values of fairness, and performance. This is evident when evaluating the Proposed Study since it performs significantly well in all tested metrics. It has a F-measure at a meager 0.04 suggesting nearly equal distribution across demographic groups, and it is ranked low concerning the fairness disparity. For performance, it manages 99.8% in terms of accuracy, precision, recall, and F1-score that show its capacity to provide accurate and fair decision making. However, Fair SVM uses reweighting study and Fair LR uses preprocessing study to deal with the issue of bias. In this case, both studys show different fairness gaps with values of 0.22 and 0.20 and are thus identified to have a high fairness disparity. Their accuracy, precision, recall, and F1-scores level off at approximately 78.0% and 80.0%, which accommodates limited fairness as well as predictive accuracy. Blind BERT, which does not consider special characteristics during the computation process, yields a slightly better outcome, with fairness of 0.17 and fairness disparity at the medium level. These results reach an average of 84.0% for all the performance indices, which itself approximates to fair enhancement over the use of the standard approaches to computing the ML models. Adversarial DeBiasNet employs adversarial debiasing that achieves an equally good fairness compared to the baseline, but with a lesser fairness gap of 0.09 and a fairness disparity at a moderate level with a performance of 86.0% across the metrics. However, it still lacks the proposed study. By virtue of achieving a higher fair and high performance, the Proposed Study is highly superior to all the other studys. This deems it suitable for real-life HR situations where ethical artificial intelligence is desirable and can be implemented. Therefore, the results show that all elements of the proposed approach: fairness-aware fine-tuning, adversarial learning, and ranking are powerful tools for eliminating bias while achieving minimal loss of accuracy.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 13s,2025
https://theaspd.com/index.php

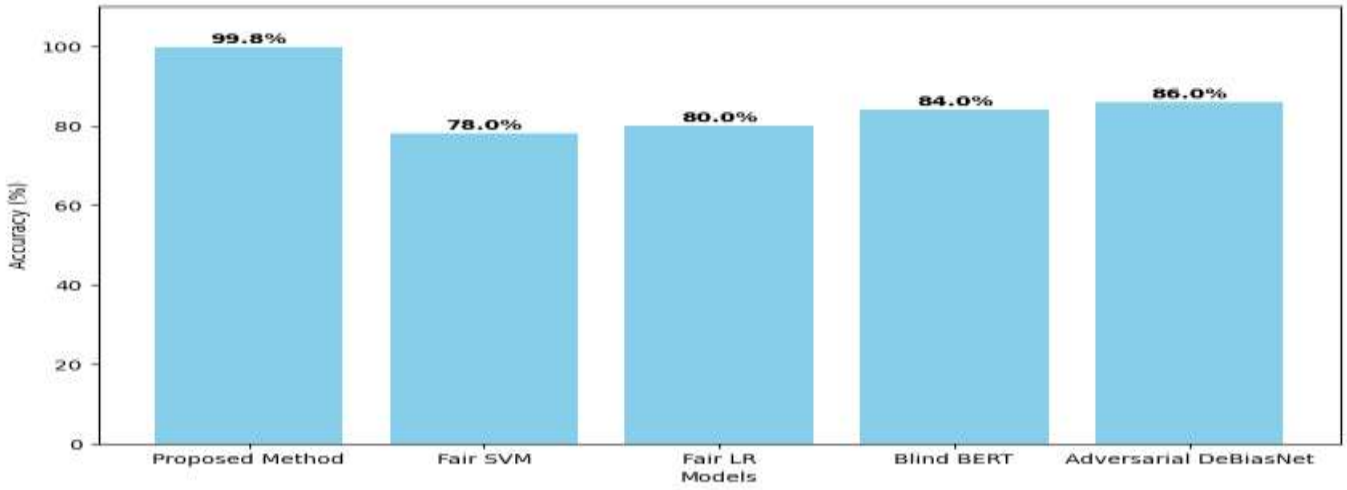

Figure 8. Accuracy comparison

Figure 9 shows the accuracy of five different models and it is called "Accuracy (%) of Different Models." The x-axis lists the models: "Proposed Study," "Fair SVM," "Fair LR," "Blind BERT," and "Adversarial DeBiasNet." On the y-axis, the level of accuracy is depicted in percentage with the rating ranging from zero to one hundred. Each model is plotted as a light blue bar and the exact percent accuracy is labeled above each bar. From the above chart, it can be seen clearly that the "Proposed Study" has the highest overall classification accuracy compared to other models and standing at 99.8%. This implies that the proposed approach is better by a considerable margin than the other models in predicting the outcomes of the given task. As per the "Proposed Study", second best accuracy is achieved by "Adversarial DeBiasNet" with 86.0 percent. The Blind BERT gets an accuracy of 84.0% which is slightly less accurate compared to the Adversarial DeBiasNet. Cutting down the number of reward choices affect the performance of the algorithm, as seen in "Fair LR" which had an accuracy of 80.0%. The least accurate of the five models is "Fair SVM," with 78.0% accuracy. Thus, the results in the table and Figure 3 indicate that the 'Proposed Study' yielded much improved accuracy compared to the other fairness-aware and benchmark studys. Although, "Adversarial DeBiasNet" and "Blind BERT" are found to be comparatively more efficient, "Fair LR" and "Fair SVM" possess the lowest accuracy level among all of them. This view can help to illustrate the increased accuracy of the proposed approach in this setting.

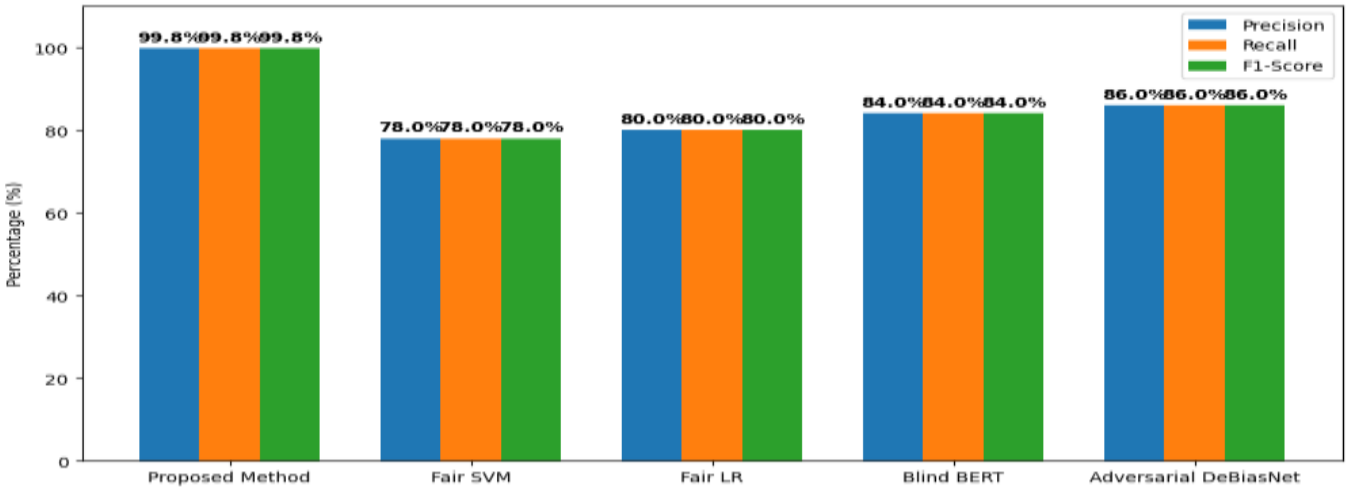

Figure 9. Metrics Comparison

Figure 10 compares the precision, recall, and F1-score of five different models: "Proposed Study," "Fair SVM," "Fair LR," "Blind BERT," and "Adversarial DeBiasNet." On the X-axis, the schemes are cited and, on the Y-axis, the performance values in percentage are given from 0 to 100. For every model, there are three bars next to each other with different color: blue for precision, orange for recall, and green for F1-Score, and the exact percentage is shown above each bar. The 'Proposed Study' performs impressively with a 99.8% in every category - precision, recall value, and F1-Score. This also points to high specificity in identifying positive examples, the coefficient of inclusiveness, and to its

high harmonic mean of precision and recall rates. The "Fair SVM" also performed fairly well but was slightly inferior to ELM, thereby achieving a Precision of 78.0% and Recall & F1-Score of both were 78.0%. Fair LR yields a bit better result, scoring 80.0% on all three indicators. Further, "Blind BERT" has a better Prediction, with the Precision, Recall, and F1-Score at 84.0%. The Adversarial DeBiasNet has the next best performance right after the proposed study in Differentiated performance for Precision = 86.0 and Recall = 86.0 and F1 Score = 86.0. In conclusion, it has been found that the "Proposed Study" has superior performance compared to other models with respect to Precision, Recall, and F1-Score. The results indicates that the "Adversarial DeBiasNet" and the "Blind BERT" perform really well while the "Fair LR" and the "Fair SVM" are the worst performing models of the compared models. The fluctuation of averages in each of the models indicates that there is a relative balance between precision and recall for the models.

DISCUSSION

This study emphasizes on the role of AI advanced fairness algorithm to revolutionize HRM policies of the DE&I. The use of AI in recruitment and decision-making, which have sensitivities to matters of bias and prejudice, helps reduce cases of bias to a large extent. As such, the use of such algorithms in law enforcement must be done carefully to avoid any violations of the basic principles of transparency, accountability, and ethical practice. However, issues including data privacy, algorithmic bias, and a certain inability to explain and interpret results still form a major obstacle. This conclstudyudes that although AI can be effective in managing bias it cannot work independently of human supervision and decreased prejudicial cultures. As with other systems, the level of practice and matter requires ongoing review in order to ensure fairness. Lastly, in order to supply fairness effectively, HR, technical departments and managers need to come together and embrace it as an organizational value. This supports the notion that AI is more of an enabler than a solution in DEI improvement.

CONCLUSION AND FUTURE WORKS

Thus, on balance, AI-assisted fairness techniques are a noteworthy step forward in creating a progressive approach to diversity within the field of human resources. Using machine learning and ethics in algorithm design, some of the biases that are inherent within the hiring and talent management can be detected and targeted for removal. The introduction of Fairness proves efficacy for improving the organisations if the particular consideration that FAIR arises from makes decisions less subjective and more humane, as well as more impartial depending on the organisation and its resource. However, in order to truly understand the efficacy of the AI tools for DEI, it will need experimental and longitudinal studies that reflect the overall real-world data improvements. It is, however, important for these organizations to remember that these are tools and not a substitute for human intervention and compassion. The implementation process requires constant training, involvement of the stakeholders and proper communication. Thus, the future research should direct more attention to the development of AI models that are explainable and adaptable to various organisational settings. This study will discuss how various disciplines like sociology and psychology can with AI help is improving the fairness aspect of the algorithm. Furthermore, longitudinal research examining the application of AI instrument in real-life DEID settings will be essential to determine the effectiveness of AI tools. Thus, it is crucial for policy and trendsetters alongside tactical actors to develop or strike legislation, regulation, or guidelines on the adequate use of AI in the HR domain. Other areas of future developments include the uses of blockchain to achieve audit trails for verifying the results, federated learning to protect individual privacy of data while using it for machine learning, and the integration of both AI and human capabilities in decision-making. In the long run, the aim should be to promote the legal compliance while designing AI systems as well as the steering the visionary concepts of equity and justice in workplaces.

REFRENCES

[1] S. Chaturvedi and R. Chaturvedi, "Who Gets the Callback? Generative AI and Gender Bias," *arXiv preprint arXiv:2504.21400*, 2025.
[2] D. F. Mujtaba and N. R. Mahapatra, "Fairness in AI-driven recruitment: Challenges, metrics, studys, and future directions," *arXiv preprint arXiv:2405.19699*, 2024.
[3] A. Pandey, S. Grima, S. Pandey, and B. Balusamy, *The role of HR in the transforming workplace: Challenges, technology, and future directions*. CRC Press, 2024.
[4] K. April and P. Daya, "The Use of AI in HRM and Management Processes: The Promise of Diversity, Equity, and Inclusion," in *AI and Diversity in a Datafied World of Work: Will the Future of Work be Inclusive?*, vol. 12, Emerald Publishing Limited, 2025, pp. 97–123.

[5] A. Fabris *et al.*, "Fairness and bias in algorithmic hiring: A multidisciplinary survey," *ACM Transactions on Intelligent Systems and Technology*, vol. 16, no. 1, pp. 1–54, 2025.
[6] M. Bano, D. Zowghi, F. Mourao, S. Kaur, and T. Zhang, "Diversity and Inclusion in AI for Recruitment: Lessons from Industry Workshop," *arXiv preprint arXiv:2411.06066*, 2024.
[7] A. Malik, "AI bias in recruitment: Ethical implications and transparency." Forbes. https://www. forbes. com/sites/forbestechcouncil/2023/09/25/ai-bias ..., 2023.
[8] H. Erkutlu, "Artificial intelligence and human resources management: Transformation in the workplace," *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, vol. 15, no. 1, pp. 346–357, 2025.
[9] K. Mackenzie, "AI in hiring: bias & privacy an issue for 40% of hiring teams - Workable." Accessed: May 05, 2025. [Online]. Available: https://resources.workable.com/stories-and-insights/ai-hiring-challenges
[10] J. Rayhan, "Journal of Policies and Recommendations," *Journal of Policies and Recommendations*, vol. 4, p. 1, 2025.
[11] S. Bhanu and A. Christinal, "Transforming Human Resource Management: Managing Bias and Ethical Issues in AI Adoption," *EMERGING PARADIGMS; COMMERCE AND MANAGEMENT RESEARCHES*, p. 1, 2024.
[12] R. Purohit and T. Banerjee, "Artificial intelligence-based organizational decision-making in recruitment practice," *Human Systems Management*, vol. 44, no. 1, pp. 173–186, 2025.
[13] R. Formulahti, "The image and integration of artificial intelligence in society," 2025.
[14] B. Timko, "AI in HR: Navigating the Legal Landscape and Ensuring Fairness in Employee Selection," Mitratech. Accessed: May 05, 2025. [Online]. Available: https://mitratech.com/resource-hub/blog/ai-in-hr-navigating-the-legal-landscape-and-ensuring-fairness-in-employee-selection/
[15] A. Wen, T. Patil, A. Saxena, Y. Fu, S. O'Brien, and K. Zhu, "FAIRE: Assessing Racial and Gender Bias in AI-Driven Resume Evaluations," Apr. 02, 2025, *arXiv*: arXiv:2504.01420. doi: 10.48550/arXiv.2504.01420.
[16] H. Iso, P. Pezeshkpour, N. Bhutani, and E. Hruschka, "Evaluating Bias in LLMs for Job-Resume Matching: Gender, Race, and Education," Mar. 24, 2025, *arXiv*: arXiv:2503.19182. doi: 10.48550/arXiv.2503.19182.
[17] M. Soleimani, Intezari ,Ali, Arrowsmith ,James, Pauleen ,David J., and N. and Taskin, "Reducing AI bias in recruitment and selection: an integrative grounded approach," *The International Journal of Human Resource Management*, vol. 0, no. 0, pp. 1–36, doi: 10.1080/09585192.2025.2480617.
[18] A. John, A. Elly, and D. Wood, "Addressing Bias and Fairness in AI-Enabled Hiring and Financial Systems," Apr. 23, 2025, *Business, Economics and Management*. doi: 10.20944/preprints202504.1923.v1.
[19] M. Nyarkoa, "A Systematic Review of Bias, Discrimination, and Mitigation Strategies in AI Decision-Making Process," ResearchGate. Accessed: May 05, 2025. [Online]. Available: https://www.researchgate.net/publication/389880153_A_Systematic_Review_of_Bias_Discrimination_and_Mitigation_Strategies_in_AI_Decision-Making_Process
[20] A. Wen, *athenawen/FAIRE-Fairness-Assessment-In-Resume-Evaluation*. (Apr. 28, 2025). Jupyter Notebook. Accessed: May 05, 2025. [Online]. Available: https://github.com/athenawen/FAIRE-Fairness-Assessment-In-Resume-Evaluation
[21] A. Wen, T. Patil, A. Saxena, Y. Fu, S. O'Brien, and K. Zhu, "FAIRE: Assessing Racial and Gender Bias in AI-Driven Resume Evaluations," *arXiv preprint arXiv:2504.01420*, 2025.
[22] H. Iso, P. Pezeshkpour, N. Bhutani, and E. Hruschka, "Evaluating Bias in LLMs for Job-Resume Matching: Gender, Race, and Education," *arXiv preprint arXiv:2503.19182*, 2025.
[23] A. John, A. Elly, and others, "Addressing Bias and Fairness in AI-Enabled Hiring and Financial Systems," 2025.
[24] M. Soleimani, A. Intezari, J. Arrowsmith, D. J. Pauleen, and N. Taskin, "Reducing AI bias in recruitment and selection: an integrative grounded approach," *The International Journal of Human Resource Management*, pp. 1–36, 2025.