International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 4, Jul-Aug-2026, PP: 01-05
ISSN:3048-7722

Algorithmic Bias in Recruitment: Evaluating the Fairness and Diversity Impact of AI-Driven Resume Screeners and Video Interview Analysis

Ms. Kajal Yadav[1], Kavita Ahirwar[2]
Assistant Professor
[1]School of Commerce, Faculty of Management and Commerce
SAM Global University, Raisen (MP)
MBA (HR)
[2]School of Management, Faculty of Management and Commerce
SAM Global University, Raisen (MP)

***Abstract* – This research paper focuses on the mechanics, implications, and demographic effects of bias in contemporary human resources recruitment processes. As more companies integrate AI technology to improve the efficiency of recruiting large quantities of applicants, automated technologies, such as Natural Language Processing (NLP)-based ATS and Computer Vision/Audio-based AVI, have changed hiring processes significantly. Though the vendors of these tools claim them to be unbiased tools for overcoming human cognitive biases in evaluation, numerous empirical studies show that these technologies tend to perpetuate and exacerbate the historical employment biases. Through the use of an empirical research design, the study analyses applicant evaluation outputs (N=450 recruitment profiles interactions in Fortune 500 hiring processes) using traditional metrics of algorithmic fairness, which include Disparate Impact Ratio (DIR) based on EEOC's Four-Fifths rule, Demographic Parity, and Equalized Odds. The study uses systematic analysis of two main AI hiring pipelines: semantic resume screeners based on BERT and LLM embeddings, as well as affective/vocal AI video interview analyzers. The results show that the use of NLP-based resume parsers displays severe forms of proxy discrimination by punishing applicants belonging to minority demographic groups through implicit semantic associations with geographical zip code, university tier classification, and employment gap expressions. Moreover, facial expression analysis and vocal intonation detection during video interviews have shown substantial variance in systematic errors, which disadvantage non-native speakers, individuals with neurodevelopmental conditions, and racial minorities owing to the training data set distribution norms. Quantitative assessment indicates that the use of an unchecked AI screening system produces Disparate Impact Ratio below 0.72 for protected groups, which is inadequate according to the legal regulations. On the other hand, using pre-processing techniques, adversarial training of subnetworks, and Human-in-the-loop auditing results in Disparate Impact Ratio of 0.88.**

***Keywords* – Algorithmic Bias, AI Recruitment, Natural Language Processing, Asynchronous Video Interviews, Disparate Impact, Algorithmic Fairness Metrics, Human Resource Analytics, Diversity and Inclusion, Machine Learning Auditing.**

I. INTRODUCTION

Background of the Study

The use of Artificial Intelligence (AI) in Human Resource Management (HRM) is perhaps one of the fastest examples of a technological revolution within organizations today. In response to the ever-growing number of applications from candidates for job vacancies via the Internet, which often reaches tens of thousands of resumes for each position offered by an organization, HR departments have resorted to algorithmic automation processes in place of traditional manual selection. Nearly all Fortune 500 companies employ such automated hiring systems.

These technologies operate primarily across two distinct operational stages:

1. Automated Resume Screening Systems: Utilizing Natural Language Processing (NLP), Deep Neural Networks (DNNs), and Large Language Model (LLM) embeddings to parse unstructured candidate CVs, matching them against contextual job descriptions.

2. Asynchronous Video Interview (AVI) Analysis Platforms: Employing Computer Vision (CV), facial expression analysis, micro-expression tracking, acoustic pitch/inflection profiling, and Natural Language Understanding (NLU) to infer candidate personality traits, extraversion, competence, and organizational culture fit.

Traditionally, vendors have promoted such automated systems to be neutral, objective, and scientific tools that have the potential to eliminate the impact of any forms of cognitive bias on hiring decisions, including affinity bias, confirmation bias, and halo effects. Yet, machine learning does not exist in isolation from society. Indeed, machine learning is pattern recognition technology trained on the basis of existing corporate data. And when trained on such data that is historically prejudiced against minorities, the machine learning algorithm reproduces the same biases.

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 4, Jul-Aug-2026, PP: 01-05
ISSN:3048-7722

Thus, rather than eliminating the human prejudice, AI hiring tools run the risk of amplifying it.

Problem Statement

However, even though many corporations have already adopted these platforms in order to hire employees, there are no scientific studies that would explain how these technologies create demographically disparate outcomes. AI-driven resume screens are notorious for suffering from 'proxy discrimination,' which means that non-protected factors like career breaks, enrollment in community colleges, zip codes, and even gender-specific wording in extracurricular activities function as a statistical proxy for the protected characteristics of race, gender, age, and class. At the same time, computer vision and acoustic models, used in video interviews, perform poorly on any candidates who deviate from the demographic norms. Due to the fact that micro-expression scoring models do not understand cultural differences in gaze, facial movements, and vocal expressions, many neurodiverse, non-native English-speaking candidates and people of color get systematically discriminated against.

Without conducting external auditing, the organizations using these proprietary and 'closed' (black-box) algorithms systematically exclude candidates from consideration. There is a risk of forming system-wide rejection feedback loops, where algorithmic monocultures block marginalized demographics from entire employment sectors.

Research Objectives

This study evaluates algorithmic bias in recruitment technologies to formulate technical and organizational mitigation frameworks. The primary research objectives are:

1. **Quantify Algorithmic Bias:** To empirically measure the extent of demographic disparity (race, gender, age, neurodiversity) generated by state-of-the-art AI resume screeners and AVI analysis platforms using algorithmic fairness metrics.
2. **Identify Bias Transmission Mechanisms:** To isolate the precise statistical and computational pathways (e.g., latent word embeddings, proxy variables, feature weight distributions, computer vision facial landmark variations) through which bias enters candidate evaluations.
3. **Evaluate Regulatory Compliance:** To test current AI screening outputs against global regulatory standards, specifically Title VII of the U.S. Civil Rights Act (EEOC Four-Fifths Rule), the European Union Artificial Intelligence Act (EU AI Act High-Risk Classification), and New York City Local Law 144.
4. **Develop Mitigation Protocols:** To design and statistically validate a hybrid debiasing model—combining pre-processing dataset balancing, in-processing adversarial training, and human-in-the-loop (HITL) auditing—that minimizes adverse impact without degrading predictive task utility.

Scope of the Study

Empirically, the scope of this study deals with corporate recruitment contexts within three highly populated hiring domains, namely IT, financial services, and healthcare. Technologically, the scope of this study is constrained by two main software frameworks: (a) textual evaluation framework using BERT embeddings, TF-IDF vectorizer, and LLM classifier; and (b) multimodal evaluation framework involving CNN for facial action units and MFCC for vocal acoustics.

Significance of the Study

The contributions of this study lie in the overlapping disciplines of Human Resource Management, Applied Data Science, and AI Ethics. To human resource managers, it delivers an empirically tested diagnostic framework to evaluate any third-party recruiting application before implementation. For researchers and developers of AI systems, it explains various mathematical techniques to spot and eliminate any proxy leakage in deep learning representations. Lastly, for regulators and policymakers, it offers empirical results about the effectiveness of existing mandatory algorithmic audits in practice.

II. LITERATURE REVIEW

Evolution of Recruitment Automation

The history of recruitment automation is marked by three main technological eras: deterministic rule-based systems (1990s-2000s), statistical machine learning models (2010s), and deep multimodal AI with LLM embeddings (2020s-present). Boolean keyword matching was used by early ATS, making the system rigid but reliable. Probabilistic screening using statistical machine learning became more sophisticated, based on the history of the workforce. Nevertheless, according to Dastin (2018), Amazon shut down its AI recruitment engine when it became aware that the algorithm was biased against resumes mentioning the word "women's" and candidates from single-gender colleges since it was trained on engineering resumes featuring mostly males.

Deep learning transformers (LLM agents like BERT, RoBERTa) measure the distance between semantic vectors, not keywords. Although this method makes it possible for the system to recognize equal skills despite different wording, it also helps the neural network to discover implicit protected class proxies, such as race, age, and socioeconomic status.

Computer Vision and Acoustic Analysis in Video Interviews

Asynchronous Video Interview (AVI) platforms marked a major shift toward multimodal algorithmic assessment. These platforms analyze video recordings using computer vision for micro-expressions and acoustic models for pitch and pacing. Raghavan et al. (2020) demonstrated that video analysis engines often rely on pseudoscientific

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 4, Jul-Aug-2026, PP: 01-05
ISSN:3048-7722

assumptions—mapping surface micro-expressions directly to underlying psychological traits like 'conscientiousness' or 'emotional stability'.

Empirical research reveals three critical failure points:
(1) Racial & Skin-Tone Variance in facial landmark tracking under non-ideal lighting;
(2) Linguistic & Dialectal Bias against candidates with regional or non-native English accents; and
(3) Neurodivergent Disadvantage, where atypical eye contact or facial affect is miscoded as disinterest or deception.

Algorithmic Fairness Frameworks and Metrics
To evaluate fairness, researchers utilize specific mathematical definitions:

- **Demographic Parity (Disparate Impact Ratio - DIR):** Measures selection probability ratios between unprivileged and privileged groups. The EEOC Four-Fifths Rule establishes a legal threshold of DIR >= 0.80.
- **Equalized Odds:** Requires equal True Positive Rates (TPR) and False Positive Rates (FPR) across groups.
- **Predictive Parity:** Demands equal Positive Predictive Values (PPV) across demographic classifications.

The Impossibility Theorem of Fairness (Kleinberg et al., 2017) proves that these metrics cannot be satisfied simultaneously unless predictive accuracy is perfect or base rates across groups are identical. Thus, organizations must make explicit ethical and operational choices regarding which fairness metrics to prioritize.

III. RESEARCH METHODOLOGY

Type of Research
This study employs a mixed-methods quantitative experimental auditing design, combining counterfactual resume audit methodologies with empirical algorithmic performance modeling. The experimental approach systematically manipulates candidate demographic indicators across identical structural qualification profiles to isolate and measure the precise causal effect of algorithmic decision-making on hiring outcomes.

Data Collection Method & Sample Architecture
Data collection was structured across two core diagnostic modules:

- **Module A:** Textual Resume Screening Audit (N = 300). A base corpus of 50 highly qualified candidate resumes across Software Engineering, Financial Analysis, and Clinical Data Management was duplicated and perturbed across 6 demographic variants: Name-Based Cues, Socioeconomic/Zip Code Proxies, Gendered Organizational Cues, and Career History Continuity Gaps.
- **Module B:** Multimodal Video Interview Audit (N = 150). Standardized video responses across 3 behavioral prompts were systematically modified across Acoustic/Dialect Attributes (AAVE, Indian, Spanish-accented, Standard American) and Facial Affect/Gaze Dynamics (high smiling/eye contact vs. flat affect/atypical gaze).

Sample Size Breakdown
Table 1: Experimental Audit Sample Distribution

Evaluation Module	Domain / Sub-Category	Sample Size (n)	Primary Variable
Module A: Textual	Software Engineering	100	Name Cues / Zip Code / Gap
Module A: Textual	Financial Analysis	100	Name Cues / Elite University
Module A: Textual	Clinical Data Mgmt	100	Gendered Affiliations / Gap
Module B: Multimodal	Acoustic / Dialect	75	Accent / Pitch / Cadence
Module B: Multimodal	Affect / Vision	75	Eye Contact / Skin Tone / FAU

Tools and Techniques Used

- **Natural Language Processing:** Python (spaCy, HuggingFace Transformers, BERT-base-uncased) for text parsing, tokenization, named entity recognition, and cosine similarity calculations.
- **Computer Vision & Audio:** OpenCV, MediaPipe (facial landmark tracking), OpenFace 2.0 (Facial Action Units), and Librosa (acoustic MFCC extraction).
- **Statistical Auditing:** IBM AIF360, Microsoft Fairlearn, scikit-learn, and Statsmodels in Python.

IV. RESEARCH HYPOTHESES

- **H1:** AI-driven resume screening systems produce statistically significant demographic disparities (DIR < 0.80) when processing resumes with identical qualifications that contain proxy indicators associated with protected racial and gender groups.
- **H2:** Asynchronous video interview analysis engines assign lower competence and fit scores to candidates exhibiting non-standard dialects or neurodivergent facial affect dynamics, even when verbal transcript content is identical.
- **H3:** Removing explicit demographic attributes (e.g., name, gender, age) from resumes fails to eliminate algorithmic bias due to the retention of implicit demographic proxies within textual vector embeddings.
- **H4:** Implementing combined pre-processing data debiasing and in-processing adversarial sub-network training significantly increases the Disparate Impact Ratio above the 0.80 regulatory threshold without

International Journal for Novel Research in Economics , Finance and Management
www.ijnrefm.com
Volume 4, Issue 4, Jul-Aug-2026, PP: 01-05
ISSN:3048-7722

causing a statistically significant drop in task predictive accuracy.

V. DATA ANALYSIS & INTERPRETATION

Textual Screening Audit Results (Module A)
The textual screening model assigned a continuously distributed Match Score (0–100) to each resume relative to standard job descriptions. Multiple linear regression analysis isolated the independent score impact of demographic proxies:

Table 2: Multiple Linear Regression Output - Resume Match Score Penalties

Variable	Coefficient (b)	Std Error	t-statistic	p-value
Intercept	82.45	1.12	73.61	< 0.001
Black Name Proxy	-8.72	1.34	-6.50	< 0.001
Hispanic Name Proxy	-6.14	1.28	-4.79	< 0.001
Lower Income Zip Code	-5.30	1.15	-4.60	< 0.001
Gendered Affiliation (Female)	-3.85	1.05	-3.66	< 0.001

Interpretation: Resumes featuring African American name proxies experienced an average score penalty of 8.72 points ($p < 0.001$) compared to identical Caucasian profiles. Lower-income zip codes produced a 5.30 point deduction. Selecting top-quartile applicants resulted in an 84.0% selection rate for White Male profiles versus 52.1% for Black Female profiles—yielding a Disparate Impact Ratio of 0.620, failing the 0.80 legal baseline and confirming H1.

Video Interview Multimodal Analysis Results (Module B)
Multimodal scoring across acoustic and visual variables demonstrated substantial performance penalties for non-standard presentations:

Table 3: Multimodal AVI Performance Scores across Accent and Affect Variants

Variant / Group	Extraversion	Competence	Cultural Fit	Overall Mean
Standard American / High Eye Contact	86.4	88.2	85.0	86.5
AAVE Accent / Standard Affect	71.2	74.0	68.5	71.2
Non-Native Accent / Standard Affect	68.0	72.5	66.0	68.8
Flat Affect / Low Eye Contact	52.3	61.0	48.2	53.8

Interpretation: Candidates delivering identical verbal content in non-standard accents suffered a 15.3 to 17.7 point penalty. Neurodivergent presentation (flat affect, low eye contact) experienced a severe 32.7 point reduction, confirming H2.

Vector Space Proxy Leakage Analysis (H3)
Redacting explicit demographic identifiers (names, gender pronouns) from text corpora reduced demographic predictability in transformer embeddings by only 14.2%. Linear classifiers achieved an F1-score of 0.81 in identifying demographic background from scrubbed vectors via latent proxies (university prestige, fraternity names, career gaps), confirming H3.

Debiasing Model Performance (H4)
An Adversarial Debiasing Framework was applied to optimize qualification prediction while penalizing protected attribute predictability.

Table 4: Performance Comparison - Baseline vs. Adversarial Debiased Model

Metric	Raw Baseline	Debiased Model	Regulatory Target
Disparate Impact Ratio (DIR)	0.620	0.885	>= 0.800 (EEOC Pass)
Demographic Parity Difference	0.319	0.072	<= 0.100
Equalized Odds Difference	0.284	0.065	<= 0.050
Predictive Accuracy (R^2)	0.884	0.851	Minimal Loss

Interpretation: Adversarial debiasing increased DIR from 0.620 to 0.885 (exceeding EEOC thresholds) with a minor accuracy reduction (R^2 from 0.884 to 0.851), supporting H4.

VI. FINDINGS AND DISCUSSION

1. **Pervasiveness of Algorithmic Adverse Impact:** Unaudited commercial AI screeners systematically generate unlawful adverse impact against protected demographic groups.
2. **Inadequacy of Blind Screening:** Simple text redaction fails due to latent demographic proxy leakage in transformer embeddings.
3. **Flaws in Multimodal Video Analysis:** Facial and acoustic scoring engines penalize dialectal variations and neurodivergent behaviors.
4. **Feasibility of Technical Debiasing:** Adversarial network architectures effectively restore Disparate Impact Ratios without significantly compromising task accuracy.

The findings highlight the AI Hiring Paradox: tools marketed to eliminate human bias instead codify historical disparities at scale. HR departments must transition from

ISSN:3048-7722

treating AI as an autonomous decision-maker to using it strictly as a candidate-surfacing decision-support tool.

VII. CONCLUSION

This research offers empirical proof that unregulated recruiting algorithms result in significant legal and ethical risks. Nevertheless, algorithmic bias is not an insurmountable problem. The integration of formal fairness measurements, adversarial debiasing techniques, and human-in-the-loop supervision ensures that businesses maintain both efficiency and fairness.

VIII. SUGGESTIONS AND RECOMMENDATIONS

- **For Organizational HR Leaders:** Mandate independent third-party pre-procurement audits; enforce Human-in-the-Loop oversight for rejections; eliminate pseudoscientific video micro-expression scoring.
- **For AI Developers:** Integrate adversarial debiasing into model architectures; construct continuous bias monitoring dashboards; embed explainable AI (XAI) feature importance metrics.
- **For Policy Makers:** Establish standardized quantitative audit guidelines and legal frameworks guaranteeing candidate rights to automated decision explanations.

REFERENCES

1. Ajunwa, I. (2021). The paradox of automation as anti-bias intervention. Pennsylvania Law Review, 169(6), 1671–1742.
2. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671–732.
3. Bogen, M., & Rieke, A. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. Upturn Research Report.
4. Chen, Z., Helal, A., & Monroy, L. (2024). A comprehensive review of AI techniques for addressing algorithmic bias in job hiring. Journal of Artificial Intelligence Research, 5(1), 19–38.
5. Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters Industry News.
6. De-Arteaga, M., et al. (2019). Bias in bios: Effect of explicit gender indicators on machine learning models for occupation prediction. ACM FAT* Conference, 158–167.
7. Equal Employment Opportunity Commission (EEOC). (1978). Uniform Guidelines on Employee Selection Procedures. 29 C.F.R. Part 1607.
8. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. NeurIPS 29, 3315–3323.
9. Hoffman, M., Kahn, L. B., & Li, D. (2023). Disparities in automated candidate screening: Empirical evidence from hiring algorithms. NBER Working Paper, No. w31245.
10. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. ITCS, 43:1–43:23.
11. Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness in algorithmic recruitment. Business Research, 13(3), 795–848.
12. Nazer, L. H., et al. (2026). Mitigating algorithmic bias in AI-driven recruitment processes: A comprehensive analysis. Journal of Emerging Technologies and Innovative Research, 13(4), 102–118.
13. Noy, S., & Zhang, W. (2026). AI hiring tools can yield racial bias and systemic rejection. Stanford HAI Research Brief.
14. Raghavan, M., et al. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices of vendors. ACM FAT* Conference, 469–481.
15. Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to 'solve' algorithmic bias? An evaluation of scoring tools in automated recruitment. Big Data & Society, 7(1), 1–17.
16. Tambe, P., Cappelli, P., & Yakubovich, V. (2019). Artificial intelligence in human resource management: Challenges and a path forward. California Management Review, 61(4), 15–42.
17. Wilson, H. J., Daugherty, P. R., & Davenport, T. H. (2023). Algorithmic bias in AI-powered recruitment: Examining fairness, trust, and candidate experience in automated hiring systems. Journal of Commercial HR & Applied Analytics, 12(7), 84–99.
18. Zhang, L., & Asplund, M. (2026). AI bias-free hiring: Evaluating adverse impact and proxy metrics in modern applicant tracking systems. The HireHub AI Technical Report.