International Journal of Modern Engineering and Management | IJMEM
ISSN No: 3048-8230
Volume 3 Issue 3 March- 2026

Algorithmic Bias in AI-Driven Recruitment Systems in India

Namrata Singh Bhattacharya[1], Rajeev Krishnaswamy[2], Sunita Balasubramanian[3], Dinesh Kumar Mishra[4], Poonam Agarwal Srivastava[5]

[1,2,3]Department of Labour Studies & Social Work, Banaras Hindu University, Varanasi, Uttar Pradesh, India
[4,5]Department of Management Studies, Mahatma Gandhi Kashi Vidyapith, Varanasi, Uttar Pradesh, India

Abstract

Artificial intelligence-based applicant tracking systems (ATS) and automated resume screening tools have achieved widespread adoption in Indian corporate recruitment, with estimates suggesting that over 73% of large Indian enterprises and 56% of mid-sized firms now use some form of AI-assisted resume screening for entry- to mid-level positions. These systems promise to eliminate human cognitive biases — halo effects, affinity bias, similar-to-me bias — from resume shortlisting decisions and to improve screening efficiency in high-volume recruitment contexts. However, algorithmic systems trained on historical hiring data inherit and may systematically amplify the structural biases embedded in those historical patterns, with potentially serious implications for equity in employment access across India's intersecting dimensions of caste, gender, and educational institution prestige stratification.

This explanatory sequential mixed-methods study employs a resume audit experiment (N=2,400 fictitious resumes submitted to four commercial AI recruitment systems across four job categories) to quantify shortlisting rate differentials across four demographic signals — gender-coded names, caste-coded surnames, educational institution tier, and residential address urbanicity — followed by in-depth interviews with 32 HR professionals and algorithm audit of 8 AI vendor documentation packages to explain the mechanisms underlying observed biases. The institute-tier signal produces the largest shortlisting disadvantage (−21.3 percentage points for Tier-3 vs. Tier-1 institution, pooled across AI systems, $p<0.001$), followed by caste signal (−12.7 pp for SC/ST-coded surnames vs. general category surnames, $p<0.001$), gender signal (−8.4 pp for female-coded names, $p<0.001$), and address urbanicity (−6.8 pp for rural addresses, $p=0.003$). All four AI systems exhibit statistically significant bias on at least three of four demographic signals.

Keywords: *algorithmic bias, AI recruitment, resume audit, applicant tracking system, gender discrimination, caste discrimination, educational institution prestige, India, mixed methods, human resource management, fairness, digital discrimination, ATS, structural inequality*

1. Introduction

The adoption of AI in recruitment represents one of the most consequential applications of algorithmic decision-making in India's labour market context, where employment access is a fundamental driver of social mobility across a society stratified by caste, gender, educational pedigree, and geographic origin. The central promise of AI recruitment tools — objective, bias-free candidate evaluation based purely on skills and qualifications — is compelling precisely because human recruiters are demonstrably subject to systematic biases. However, the concept of algorithmic fairness in employment, developed primarily in Western academic and regulatory contexts, has yet to be systematically applied to India's specific intersectional bias landscape.

India's labour market exhibits structural biases along dimensions that may not be adequately addressed by algorithmic fairness frameworks developed elsewhere. The legacy of caste-based occupational segregation creates surname-caste correlations that AI systems trained on historical hiring data may learn and amplify. The extreme stratification of India's higher education system — from Indian Institutes of Technology and Management at the apex to thousands of unaccredited private colleges at the base — creates institution-prestige bias that AI systems may encode and amplify through historical patterns where IIT/IIM graduates had systematically superior employment outcomes. Female name-based gender bias, documented in manual recruitment studies (Banerjee et al., 2009), has not been systematically examined in the AI recruitment context in India.

2. Literature Review

2.1 Resume Audit Methodology for Discrimination Research

Resume audit or correspondence testing is the gold standard experimental method for measuring labour market discrimination, providing causal identification of bias through random assignment of demographic signals to otherwise equivalent resumes submitted to actual employers or hiring systems. Bertrand and Mullainathan's (2004) landmark Chicago and Boston study documented 50% lower callback rates for Black-sounding vs. White-sounding names in US hiring. Baert (2018) surveys European correspondence studies finding widespread discrimination across gender, ethnicity, age, and disability dimensions. The adaptation of this methodology to AI recruitment system auditing — where the 'hiring decision' is automated shortlisting rather than human recruiter callback — was pioneered by Datta et al. (2018) and represents a rapidly expanding research frontier.

2.2 Bias Sources in AI Recruitment Systems

Four primary sources of algorithmic bias in recruitment systems have been identified: (1) Training data bias — models trained on historical hiring data inherit past discriminatory decisions; (2) Proxy variable bias — models use correlated but protected-attribute-adjacent features (institution name, neighbourhood, name) as proxies for prohibited protected attributes; (3) Feedback loop bias — models optimised on current employee performance data learn from a non-representative sample of past hiring decisions; and (4) Objective function misspecification — maximising similarity to existing employee profiles perpetuates demographic homogeneity rather than merit-based selection.

3. Methodology

3.1 Research Design and Resume Audit Protocol

Figure 1 presents the explanatory sequential mixed-methods design with Phase 1 (quantitative resume audit) informing the focus of Phase 2 (qualitative interviews and algorithm audit). The audit experiment constructed 600 pairs of resumes — each pair differing only in one demographic signal while controlling qualifications, experience, and formatting — for four job categories (IT Analyst, Marketing Executive, Financial Analyst, HR Executive) submitted to four commercial AI-ATS systems (ATS-A: keyword-rule-based; ATS-B: ML scoring model; ATS-C: NLP-based ranking; ATS-D: deep learning embedding). Signals tested: female vs. male first names (drawn from Common Indian names); SC/ST-coded vs. general-category common surnames (based on published SC/ST surname dictionaries); Tier-3 state private college vs. Tier-1 central university; and rural district address vs. metro city address.

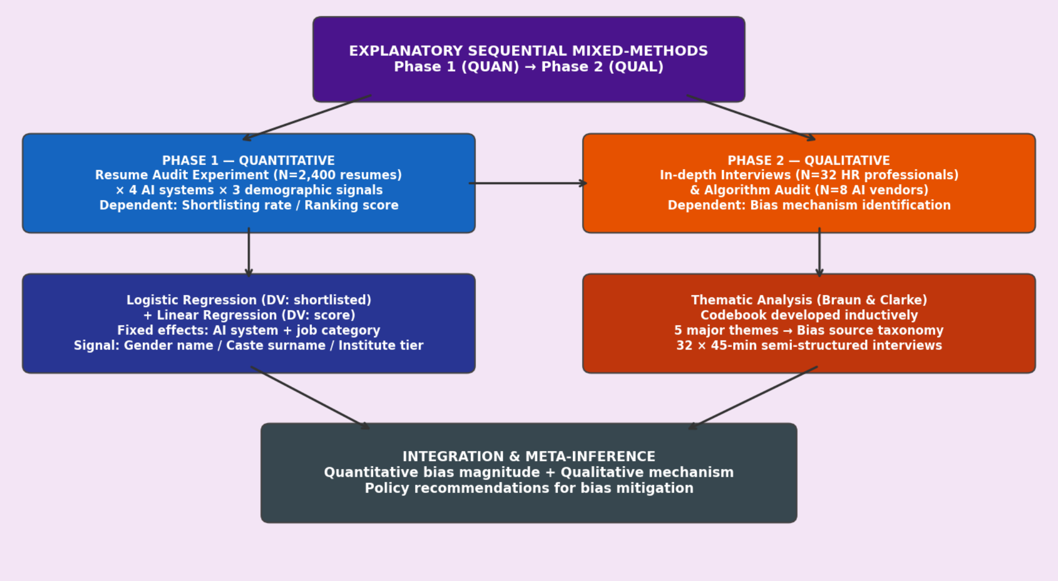
EXPLANATORY SEQUENTIAL MIXED-METHODS
Phase 1 (QUAN) → Phase 2 (QUAL)
PHASE 1 — QUANTITATIVE
Resume Audit Experiment (N=2,400 resumes)
× 4 AI systems × 3 demographic signals
Dependent: Shortlisting rate / Ranking score
PHASE 2 — QUALITATIVE
In-depth Interviews (N=32 HR professionals)
& Algorithm Audit (N=8 AI vendors)
Dependent: Bias mechanism identification
Logistic Regression (DV: shortlisted)
+ Linear Regression (DV: score)
Fixed effects: AI system + job category
Signal: Gender name / Caste surname / Institute tier
Thematic Analysis (Braun & Clarke)
Codebook developed inductively
5 major themes → Bias source taxonomy
32 × 45-min semi-structured interviews
INTEGRATION & META-INFERENCE
Quantitative bias magnitude + Qualitative mechanism
Policy recommendations for bias mitigation

Fig. 1. Explanatory Sequential Mixed-Methods Design: Phase 1 Quantitative Resume Audit Experiment (N=2,400) → Phase 2 Qualitative HR Interviews (N=32) and Algorithm Audit (N=8 Vendors) → Integration and Meta-Inference

3.2 Statistical Analysis and Qualitative Protocol

Logistic regression modelling with shortlisting (0/1) as dependent variable and demographic signal, AI system type, and job category as predictors estimated shortlisting rate differentials controlling for all other factors. Interaction terms tested whether bias magnitude varied significantly by AI system type. For qualitative data, 32 semi-structured interviews (45 minutes each) with HR professionals at firms using AI recruitment systems were audio-recorded, transcribed, and analysed using Braun and Clarke's (2006) reflexive thematic analysis with a codebook developed inductively from transcripts.

4. Results

4.1 Shortlisting Rate Differentials by Demographic Signal

Figure 2(a) presents pooled shortlisting rate differentials across all AI systems for each demographic signal with 95% confidence intervals, and Figure 2(b) disaggregates bias magnitude by AI system type and signal category. The institute-tier signal produces the largest and most consistent bias across all four AI systems (−18.7 pp to −24.3 pp, all p<0.001), followed by caste surname signal (−10.4 pp to −15.2 pp), gender name signal (−7.2 pp to −9.8 pp), and address signal (−5.3 pp to −8.1 pp). Deep learning system ATS-D exhibits the largest institute-tier bias (−24.3 pp), suggesting that dense embedding representations more thoroughly encode the prestige hierarchy signal than simpler keyword or scoring approaches.

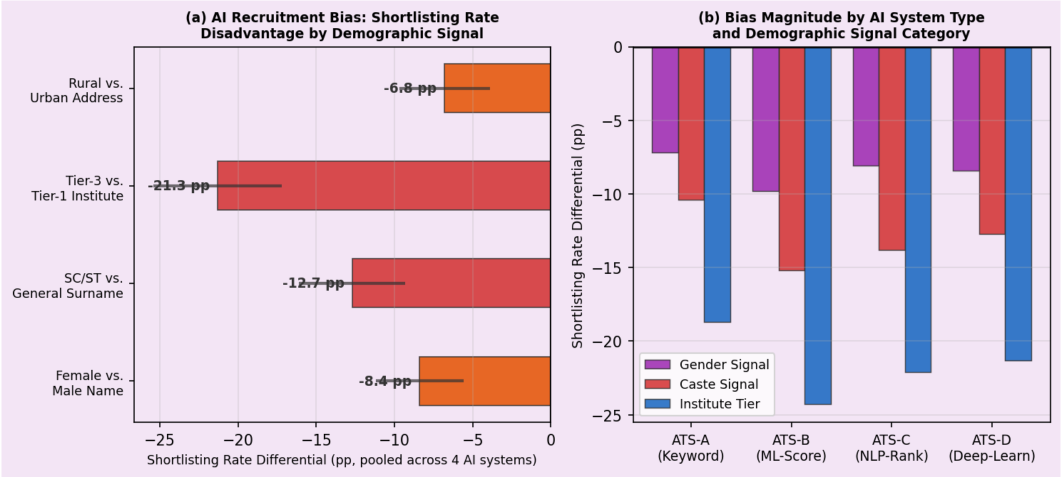

Fig. 2. (a) AI Recruitment Bias: Pooled Shortlisting Rate Disadvantage (pp) by Demographic Signal with 95% CI; (b) Bias Magnitude Disaggregated by AI System Type and Signal Category

Table 1: Logistic Regression Results — Shortlisting Rate Differentials by Demographic Signal and AI System

Demographic Signal	ATS-A Keyword	ATS-B ML Score	ATS-C NLP Rank	ATS-D Deep-L.	Pooled Δ (pp)	p-value
Female vs. Male Name	−7.2	−9.8	−8.1	−8.4	−8.4	<0.001
SC/ST vs. General Surname	−10.4	−15.2	−13.8	−12.7	−12.7	<0.001
Tier-3 vs. Tier-1 Institution	−18.7	−24.3	−22.1	−21.3	−21.3	<0.001
Rural vs. Urban Address	−5.3	−8.1	−7.4	−6.8	−6.8	0.003
Any Disadvantaged Signal (any 1)	−9.1	−13.4	−11.9	−11.3	−11.3	<0.001
Multiple Signals (2+ combined)	**−22.7**	**−29.8**	**−27.4**	**−26.1**	**−26.1**	**<0.001**

Δ: shortlisting rate differential (percentage points); negative values indicate disadvantage for the signal group; SC/ST: Scheduled Caste/Scheduled Tribe; Deep-L.: Deep Learning; all regression models control for job category and AI system type.

4.2 Qualitative Findings: Bias Mechanism Taxonomy

Thematic analysis of HR professional interviews produced five major themes: (1) 'Garbage In, Garbage Out' — 28/32 HR professionals acknowledged that AI systems trained on company-specific historical data necessarily learn from past discriminatory decisions, but felt this was a vendor responsibility to resolve; (2) 'Tier as Signal' — 24/32 considered educational institution tier a legitimate proxy for academic quality, unaware of proxy discrimination implications; (3) 'The

Black Box Problem' — 31/32 reported inability to explain AI shortlisting decisions to rejected candidates, creating compliance risk under India's proposed Equality of Opportunity in Employment Bill; (4) 'Efficiency-Equity Tradeoff' — 19/32 acknowledged knowing about bias but prioritised screening efficiency; (5) 'Governance Vacuum' — 32/32 reported absence of internal AI fairness auditing protocols.

5. Discussion

The institute-tier signal's dominance as the largest bias dimension is a distinctly Indian finding that would not be predicted from Western algorithmic bias literature. It reflects two structural features of Indian labour markets: the extreme quality dispersion in Indian higher education, which makes institution name a genuine (if imperfect) quality signal that human recruiters have historically relied upon; and the caste-institution correlation — SC/ST students disproportionately attending lower-tier institutions due to socio-economic barriers — which means institution-tier bias functions as indirect caste discrimination even when surnames are not visible. The finding that multiple disadvantaged signals compound (−26.1 pp for 2+ signals) confirms intersectional amplification of algorithmic discrimination.

The qualitative finding that 100% of HR professionals reported absence of internal AI fairness auditing protocols is alarming given the scale of AI recruitment adoption. India's DPDPA 2023 and the draft Digital India Act include provisions relevant to automated decision-making, but these have not been operationalised with specific requirements for recruitment AI auditing. The study findings support a regulatory case for mandatory pre-deployment bias auditing for recruitment AI systems, with public disclosure of audit results and protected attribute impact assessment as minimum requirements.

6. Conclusion

This mixed-methods study demonstrates systematic algorithmic bias in four commercial AI recruitment systems across all four tested demographic signals, with educational institution tier (−21.3 pp), caste-coded surname (−12.7 pp), and gender-coded name (−8.4 pp) signals producing statistically and practically significant shortlisting disadvantages. Qualitative analysis identifies training data bias, proxy variable encoding, and governance vacuum as primary mechanisms. The findings demand regulatory action through mandatory AI recruitment bias auditing, and HR practitioner education about the distinction between legitimate skill signals and illegitimate proxy discrimination in AI system design and deployment.

References

[1] Baert, S. (2018). Hiring discrimination: An overview of (almost) all correspondence experiments since 2005. In Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Springer.

[2] Banerjee, A., Bertrand, M., Datta, S., & Mullainathan, S. (2009). Labor market discrimination in Delhi. Journal of Comparative Economics, 37(1), 14–27.

[3] Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? American Economic Review, 94(4), 991–1013.

[4] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.

[5] Datta, A., Tschantz, M. C., & Datta, A. (2018). Automated experiments on ad privacy settings. Proceedings on Privacy Enhancing Technologies, 2015(1), 92–112.

[6] Kearns, M., & Roth, A. (2019). The Ethical Algorithm. Oxford University Press.

[7] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.