AI Recruitment Bias Governance through Multi-Case Comparative Study: From the Infeasibility of "Zero Bias" to Auditable Compliance and Engineering Practices

Zihe Qi

Laber and Industrial Relations, University of Illinois Urbana-Champaign Champaign, The US

ziheqi2@illinois.edu

Abstract. Artificial intelligence is revolutionizing recruitment, with 83% of employers using automated screening systems [14]. While boosting efficiency, AI introduces algorithmic bias and transparency issues, as seen in cases involving Amazon and HireVue [14]. Through case studies of Harver, Eightfold AI, HireVue, and LinkedIn, this study finds bias stems from the interaction of data, algorithms, and human interpretation [2;12]. Bias in AI recruitment systems does not arise from isolated technical flaws but from a structural coupling between social inequality and computational optimization. Historical labor market inequalities shape training data distributions; these distributions are then formalized through algorithmic objective functions (e.g., predictive accuracy or retention likelihood), which systematically privilege historically dominant groups. I propose the Auditable Fairness Framework (AFF)—based on Auditability, Engineering, Control, and Remediation—shifting the goal from unachievable "zero bias" to establishing detectable, explainable, and correctable governance.

Keywords: AI recruitment; algorithmic bias; FATE framework; fairness governance; third-party audit; human-AI collaboration; ethics and compliance

1 Introduction

Existing research has identified key components of this process. Barocas and Selbst (2016) demonstrate how ostensibly neutral features act as proxies for protected attributes [2]. Raghavan et al. (2020) highlight organizational overreliance on automated tools, while Crawford (2021) situates algorithmic bias within broader political and economic structures [12;3]. However, prior studies predominantly examine bias either at the technical level (model design, data imbalance) or the normative level (ethics, regulation), leaving insufficient attention to how organizational governance practices operationalize fairness over time.

This study addresses three research questions: (1) How does AI recruitment bias arise technically and socially? (2) How do leading companies manage bias through

T. G. Guan et al. (eds.), *Proceedings of the 2026 4th International Conference on Digital Economy and Management Science (CDEMS 2026)*, Advances in Economics, Business and Management Research 392,
https://doi.org/10.2991/978-94-6239-699-9_3

technical and institutional methods? (3) How can we construct a sustainable, auditable governance framework for AI fairness?

This study addresses this gap by shifting the analytical focus from identifying bias to governing bias. Rather than treating fairness as an ex-ante design property or an abstract ethical ideal, this paper examines how organizations embed auditability, engineering controls, and remediation mechanisms into AI recruitment systems. By doing so, it reconceptualizes fairness as an ongoing organizational capability rather than an unattainable state of "zero bias."

2 Research Methodology

This study employs a multiple-case design and systematic literature review to analyze AI recruitment bias. It compares four exemplars (Harver, Eightfold AI, HireVue, LinkedIn) and three failure cases (Amazon, iTutorGroup, Money Bank) using Yin's (2009) pattern matching, examining distinct AI approaches. The literature review covered key publications (2016-2025) on bias and mitigation [1;2;4;6; 9;13; 8]. Data analysis integrated Yin's (2009) and Eisenhardt's (1989) methods via document screening, FATE pattern matching, and cross-case synthesis, forming a framework (Fig. 1) to examine how organizational maturity shapes fairness outcomes [15;5].

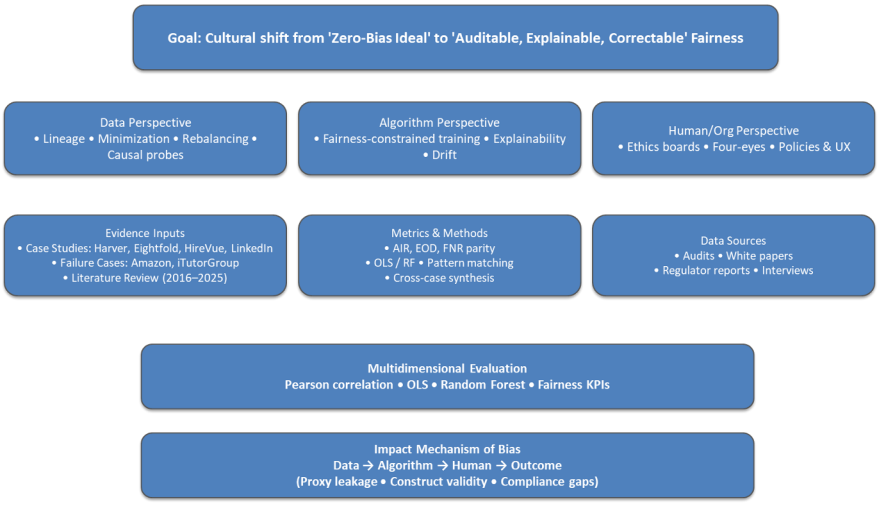

Fig. 1. Framework of this study

Case selection followed a theoretical sampling logic rather than statistical representativeness. The selected cases represent recurring governance challenges in AI recruitment: (1) historical data bias (Amazon), (2) explicit rule-based discrimination (iTutorGroup), (3) opaque vendor-controlled systems (Money Bank), and (4) mature governance-oriented implementations (Harver, Eightfold AI, HireVue, LinkedIn). Together,

these cases capture the most common failure modes and mitigation strategies documented in AI hiring practice.

These cases are representative not because they exhaust all AI recruitment systems, but because they exemplify structural problems shared across the field, including proxy discrimination, lack of explainability, over-automation, and insufficient accountability mechanisms.

This study adopts Yin's (2009) pattern-matching logic and Eisenhardt's (1989) cross-case synthesis to identify recurring causal mechanisms. However, these methods have limitations when applied to complex AI systems [15;5]. First, causal attribution is constrained by limited access to proprietary model architecture. Second, AI systems evolve continuously, challenging the assumption of stable case boundaries. Accordingly, findings should be interpreted as mechanism-oriented explanations, not deterministic causal proofs.

3 Case Study Analysis

3.1 Case Company Profiles

Four AI recruitment firms exemplify progressive governance maturity through their approaches: Harver with audited gamified psychometrics, Eightfold AI with certified deep-learning models, HireVue with its ORCAA-certified system, and LinkedIn with continuous monitoring via the LiFT toolkit [1;9;11].

Case selection followed a theoretical sampling logic rather than statistical representativeness. The selected cases represent recurring governance challenges in AI recruitment: (1) historical data bias (Amazon), (2) explicit rule-based discrimination (iTutorGroup), (3) opaque vendor-controlled systems (Money Bank), and (4) mature governance-oriented implementations (Harver, Eightfold AI, HireVue, LinkedIn). Together, these cases capture the most common failure modes and mitigation strategies documented in AI hiring practice.

These cases are representative not because they exhaust all AI recruitment systems, but because they exemplify structural problems shared across the field, including proxy discrimination, lack of explainability, over-automation, and insufficient accountability mechanisms.

This study adopts Yin's (2009) pattern-matching logic and Eisenhardt's (1989) cross-case synthesis to identify recurring causal mechanisms. However, these methods have limitations when applied to complex AI systems [15;5]. First, causal attribution is constrained by limited access to proprietary model architectures. Second, AI systems evolve continuously, challenging the assumption of stable case boundaries. Accordingly, findings should be interpreted as mechanism-oriented explanations, not deterministic causal proofs.

3.2 Applications Across Industries and Misjudgment Examples

AI recruitment systems exhibit persistent discrimination, from Amazon's gender bias and iTutorGroup's age discrimination to a UK case overemphasizing physical cues and

Money Bank's opaque algorithms, highlighting systemic flaws in data, transparency, and feature selection [1;8;14].

3.3 In-depth Analysis of Misjudgments (Amazon, Money Bank, iTutorGroup)

The failure cases analyzed here demonstrate that algorithmic bias arises from joint causation across technical, organizational, and cultural dimensions. In Amazon's case, a resume-ranking model trained on historical hiring outcomes implicitly learned gendered correlations embedded in feature representations (e.g., term frequency–inverse document frequency vectors penalizing female-associated terms). Technically, the absence of fairness constraints during model optimization allowed these correlations to dominate predictions. Organizationally, rapid deployment prioritized efficiency over model interrogation, while culturally, engineering teams treated past hiring outcomes as objective ground truth.

In contrast, iTutorGroup's case reflects explicitly encoded bias. The system implemented deterministic age-based rejection rules at the data-processing layer, bypassing statistical learning entirely. This failure was not due to model opacity but to organizational intent embedded in code, underscoring the necessity of rule-level audits and governance oversight.

The Money Bank case illustrates a different causal pathway: reliance on a third-party black-box model without access to feature importance, training data composition, or decision thresholds. Here, technical opacity combined with organizational outsourcing practices and a compliance-driven culture resulted in systematic but undetectable bias. These cases collectively demonstrate that bias cannot be mitigated through technical fixes alone; it is produced through interlocking socio-technical mechanisms.

3.4 Low Probability and Non-Causal Associations in AI Analysis

AI systems frequently confuse correlation with causation, amplifying statistically correlated but causally irrelevant signals [12]. The fundamental difference can be formalized through causal inference:

$$\boldsymbol{\tau} = \mathbf{E}[\mathbf{Y}(\mathbf{1}) - \mathbf{Y}(\mathbf{0})]$$

where $\mathbf{Y}(\mathbf{1})$ and $\mathbf{Y}(\mathbf{0})$ represent potential outcomes under treatment and control conditions, respectively. To address this, mitigation requires implementing causal constraints, multimodal validation, and human-machine collaboration [3].

3.5 Expanded Comparative Synthesis: Beyond Data–Algorithm–Human

Achieving AI recruitment fairness requires integrating technical, organizational, and cultural dimensions. Exemplary firms employ causal inference (e.g., Eightfold AI, LinkedIn) and continuous audits (e.g., Harver, HireVue), complemented by oversight tools like LinkedIn's fairness dashboards and institutionalized ethics committees. This synergy makes fairness a core organizational capability [1;11;14].

4 Comparative Findings

Cross-case analysis reveals distinct bias patterns across the data-algorithm-human triad. Five key themes emerge from comparing governance exemplars (Harver, Eightfold AI, HireVue, LinkedIn) with failure cases (Amazon, iTutorGroup, UK makeup-artist incident), highlighting critical intervention points for bias mitigation (Fig. 2).

Dimension	Harver (governance)	Eightfold AI (governance)	HireVue (governance)	LinkedIn (governance)	Amazon (failure)	iTutorGroup (failure)
Data governance	Anonymized inputs; data minimization; third-party audit under NYC LL144	Data lineage, debiased embeddings; ISO/IEC 42001; client audits	Reduced reliance on visual data; dataset scope documented	LiFT monitoring; fairness dashboards; dataset documentation	Historical male-skewed data taught proxy gender bias	Rule-based exclusion by birth year; explicit age thresholds
Algorithm design	Psychometric games + fairness constraints; continuous testing	Deep skill matching + fairness constraints; causal probes	Removed facial analysis; risk controls before release	Recommendation fairness with LiFT; explainability APIs	Resume-term penalization (e.g., women-coded terms)	Deterministic rule in code (ADEA violation)
Human oversight	Behavioral scientists + data scientists; evidence-based reviews	Cross-functional fairness board (engineers, legal, ethics)	Ethics board; four-eyes principle to ship	CDO-led Responsible AI governance program	Reactive shutdown after discovery; limited pre-checks	Insufficient legal pre-clearance; lack of human review
Transparency/Auditability	Independent audit summary disclosed	Client & third-party audits; public summaries	ORCAA safety/quality certification	Public paper & ongoing fairness toolkit	Internal prototype; limited disclosure	EEOC enforcement exposed rules
Primary risk addressed	Proxy leakage; construct validity	Spurious correlations; representation bias	Measurement bias in video analytics	Ranking fairness; exposure fairness	Statistical discrimination replicated from history	Direct discrimination by rule
Outcome	Passed bias audit; governance exemplar	Widely adopted by enterprises; maturing governance	Pivoted product; improved compliance posture	Industry reference for fairness ops	Project sunset; reputational cautionary tale	Settlement; legal penalty and remediation

Fig. 2. Multi-Case Comparative Analysis of AI Recruitment Bias Governance

4.1 Dataset Bias and Algorithmic Decision Impact

Amazon's algorithm, trained on male-dominated data, perpetuated discrimination by penalizing female-associated terms [8]. This bias can be quantified using the demographic parity difference:

$$\Delta_{\mathbf{DP}} = \mathbf{P}(\hat{\mathbf{Y}} = \mathbf{1} | \mathbf{G} = \mathbf{male}) - \mathbf{P}(\hat{\mathbf{Y}} = \mathbf{1} | \mathbf{G} = \mathbf{female})$$

where $\hat{\mathbf{Y}} = \mathbf{1}$ denotes a positive screening decision and $\mathbf{G}$ represents gender groups. A nonzero $\Delta_{\mathbf{DP}}$ indicates systematic discrimination, demonstrating how historical biases become embedded in algorithmic outputs [2]. Pre-training interventions like data rebalancing thus offer a fairer alternative [8; 10].

4.2 Programming Rule–Based Discrimination

The iTutorGroup incident constitutes explicit algorithmic discrimination, with age-based rejection rules directly encoded into the system [14]. Unlike statistical biases, this deterministic rule encoding underscores the critical need for human oversight and version-controlled rule auditing as essential safeguards [14].

4.3 Integrated Analysis

Multiple cases reveal interconnected challenges, from overreliance on irrelevant features to the limited efficacy of anonymization against structural biases [1;7]. Effective mitigation thus necessitates a combination of technical measures, continuous diagnostics, and candidate empowerment—exemplified by tools like LinkedIn's LiFT which enhances transparency and literacy [1;8;11;10].

5 Systematic Literature Review

AI recruitment bias stems from interconnected technical and social factors. Training data embeds societal prejudices, as demonstrated by differential callback rates for identical resumes with racially distinct names, which algorithms then replicate [2]. Bias also arises from algorithmic designs prioritizing accuracy over fairness and human factors like automation bias [9;12]. The black-box problem exacerbates accountability gaps, requiring interpretable models [3;9;10]. Regulatory immaturity necessitates mandatory audits, while sociolinguistic biases demand semantic debiasing [1;3;11]. Mitigation requires multi-objective optimization, adversarial debiasing, and cross-cultural validation throughout the AI lifecycle.

6 Discussion

Comparative analysis reveals that success cases do not merely apply more fairness tools; they institutionalize causal feedback loops that continuously detect, explain, and correct bias. For example, LinkedIn's LiFT system integrates real-time disparity monitoring with organizational decision rights, ensuring that detected bias triggers human review and model adjustment. Eightfold AI's use of causal inference techniques limits reliance on spurious correlations, while Harver's third-party audits introduce external accountability that constrains internal optimization pressures.

In contrast, failure cases lack mechanisms for error visibility and organizational learning. Bias persists not because it is unknown, but because systems lack structured pathways for contestation, diagnosis, and remediation. The critical distinction, therefore, lies not between biased and unbiased algorithms, but between auditable and unauditable systems.

6.1 Root Causes of Misjudgment Risks

Algorithmic bias in recruitment stems from systemic interactions among historically biased data, value-laden algorithm design, and human interpretation. Historical inequalities embed in training data, optimization processes amplify these patterns, and cognitive biases cement them in decisions [2;8]. Resilient organizations treat fairness as an integrated lifecycle constraint, requiring continuous monitoring and multi-level interventions.

6.2 Algorithm Transparency and Explainability

Industry evolution is evidenced by HireVue abandoning facial analysis and LinkedIn deploying LiFT dashboards [9]. We propose a practical three-layer framework: input traceability (signals used), model interpretability (feature effects), and output intelligibility (decisions and appeals) [3;11]. This enables actionable explanations through templates, threshold rationales, and competency-based feedback.

7 Recommendations and Implementation Strategies

Effective AI recruitment bias mitigation requires integrated technical and governance measures. Technically, implement data cleansing with proxy removal, regular fairness audits using Adverse Impact Ratio metrics, and third-party verification like LinkedIn's LiFT system [1;8;11]. Procedurally, adopt context-aware keyword screening and hybrid human-AI evaluation using structured interviews and work samples [12]. Governance should institutionalize Candidate Data Rights Charters, mandate bias impact assessments, and align with regulations like the EU AI Act [8]. These interconnected layers form a comprehensive framework for continuous bias monitoring and control throughout the recruitment lifecycle.

8 Conclusion

This study establishes algorithmic bias in AI recruitment as systemic, rooted in biased data, value-laden design, and human interpretation. We propose the Auditable Fairness Framework (AFF)—built on Auditability, Engineering, Control, and Remediation—to shift the goal from unattainable “zero bias” to achieving detectable, explainable, and correctable fairness. Practical implementation requires collaboration across enterprises, policymakers, and researchers. While limited by its reliance on public cases, this study posits that treating fairness as an organizational capability, beyond mere compliance, can transform hiring into a model of ethical innovation.

References

1. BABL AI. (2023). Summary of bias audit results. BABL AI Inc. https://www.paramount.com/sites/g/files/dxjhpe226/files/2023-07/Harver_Pymetrics-Final_Audit_Summary-2023-06-29.pdf.
2. Barocas, S., & Selbst, A. D. (2016). Big data’s disparate impact. California Law Review, 104(3), 671–732. https://doi.org/10.15779/Z38BG31.
3. Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press. https://yalebooks.yale.edu/book/9780300264630/atlas-of-ai/.
4. Eightfold AI. (2025, March 21). Summary of bias audit results. https://eightfold.ai/wp-content/uploads/eightfold-summary-of-bias-audit-results.pdf.

5. Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550. https://doi.org/10.5465/amr.1989.4308385.
6. Electronic Privacy Information Center [EPIC]. (2021, January 12). *HireVue*, facing an FTC complaint from EPIC, halts use of facial recognition. https://epic.org/hirevue-facing-ftc-complaint-from-epic-halts-use-of-facial-recognition/.
7. Jain, P. (2023). Impact of artificial intelligence on recruitment: A case study of Flipkart. *International Journal of Research Publication and Reviews*, 6(6), 3127–3134. https://ijrpr.com/uploads/V6ISSUE6/IJRPR48610.pdf.
8. Larsson, S., White, J., & Ingram Bogusz, C. (2024). The artificial recruiter: Risks of discrimination in employers' use of AI and automated decision-making. *Social Inclusion*, 12, Article 7471. https://doi.org/10.17645/si.7471.
9. Morgan Stanley. (2025). Bias audit of Morgan Stanley's use of the Eightfold model for scoring applicants. https://www.morganstanley.com/content/dam/msdotcom/en/disclosures/Bias_Audit_of_Morgan_Stanleys_Use_of_Eightfold_Model_For_Scoring_Applicants.pdf.
10. O'Neil, C., & Mann, G. (2016). Hiring algorithms are not neutral. Harvard Business Review. https://hbr.org/2016/12/hiring-algorithms-are-not-neutral.
11. Quinonero-Candela, J., Wu, Y., & Hsu, B. (2023). Disentangling and operationalizing AI fairness at LinkedIn. arXiv preprint arXiv:2306.00025. https://arxiv.org/abs/2306.00025.
12. Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020). Mitigating bias in algorithmic hiring: Evaluating claims and practices. *In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (pp. 469–481).* https://dl.acm.org/doi/10.1145/3351095.3372828.
13. Soleimani, M., Intezari, A., Arrowsmith, J., & Pauleen, D. (2025). Reducing AI bias in recruitment and selection: An integrative grounded approach. *International Journal of Human Resource Management.* https://www.researchgate.net/publication/390032264_Reducing_AI_bias_in_recruitment_and_selection_an_integrative_grounded_approach.
14. U.S. Equal Employment Opportunity Commission [EEOC]. (2023, September 11). iTutorGroup to pay $365,000 to settle EEOC discriminatory hiring suit. https://www.eeoc.gov/newsroom/itutorgroup-pay-365000-settle-eeoc-discriminatory-hiring-suit.
15. Yin, R. K. (2009). Case study research: Design and methods (5th ed.). *Sage*.