

Bias and Fairness in Automated Recruitment and Employee Selection Systems

Dr. Aida A. Kusairi

Associate Professor, University Malaya, Malaysia

Abstract

Automated recruitment and employee selection systems are increasingly used to identify applicants, screen résumés, administer assessments, rank candidates, and support hiring decisions. Their growing use has created expectations that data-driven decision-making can improve consistency, efficiency, and objectivity in personnel selection. At the same time, algorithmic systems may reproduce or intensify existing inequalities when they learn from historically biased employment data, rely on inappropriate proxy variables, use poorly validated outcome measures, or produce systematically different error rates across applicant groups. This study examines the sources, manifestations, and mitigation of bias in automated recruitment through a conceptual and simulation-based analytical framework. The analysis considers four stages of the recruitment process: candidate sourcing, résumé screening, pre-employment assessment, and final shortlisting.

A simulated selection-rate ratio is used to illustrate how an unmitigated automated system may produce unequal outcomes and how a fairness-governed system may reduce disparities through dataset auditing, job-related feature selection, subgroup validation, fairness testing, explainability, and structured human review. The simulated results indicate meaningful improvement across all four stages after governance mechanisms are introduced, although the study does not interpret mathematical parity as sufficient evidence of substantive fairness.

The paper argues that fair automated recruitment requires a sociotechnical governance approach in which technical validation is combined with job relevance, procedural transparency, accessibility, applicant safeguards, human accountability, and continuous post-deployment monitoring. Automated selection should therefore function as accountable decision support rather than an unquestionable substitute for professional judgment.

Keywords: algorithmic bias, automated recruitment, employee selection, artificial intelligence, hiring algorithms, algorithmic fairness, recruitment analytics, human oversight.

1. Introduction

Artificial intelligence and predictive analytics have altered the architecture of employee recruitment by allowing organizations to process applications at a scale and speed that would be difficult to achieve through conventional manual procedures. Automated systems can support job advertising, candidate sourcing, résumé classification, assessment administration, applicant scoring, interview scheduling, and shortlisting. These technologies are frequently promoted as mechanisms for improving efficiency and reducing subjective human judgment. Such claims are attractive because conventional employment

selection has itself been affected by stereotyping, inconsistent evaluation, affinity bias, and unequal access to occupational opportunities. Nevertheless, replacing or supplementing human judgment with an algorithm does not automatically remove the social and organizational conditions that produced those inequalities.

The principal concern is that automated selection systems learn from data generated within existing labor markets and organizations. Historical recruitment decisions, promotion patterns, performance ratings, educational opportunities, occupational segregation, and workforce composition may contain structural inequalities. Barocas and Selbst explain that data-driven systems can inherit discriminatory patterns even when protected characteristics are not explicitly included in the model. An algorithm trained to reproduce past hiring outcomes may therefore learn patterns associated with historical organizational preferences rather than genuine future job performance. Variables that appear technically neutral may also function as proxies for social or demographic characteristics.

Bias can enter automated recruitment at several points rather than appearing only in the final hiring decision. Candidate sourcing determines who encounters an employment opportunity. Résumé screening determines which applicants continue in the process. Automated assessments influence candidate rankings, while final shortlisting determines who receives meaningful consideration from decision-makers. Rieke and Bogen emphasize that recruitment should be viewed as a sequence of interconnected decision points. A disparity introduced during an early stage may substantially change the population available for subsequent stages and can therefore influence the final workforce even when later stages appear comparatively balanced.

Fairness is also more complex than achieving a single favorable statistical measure. Different fairness criteria may emphasize equal selection rates, comparable error rates, predictive validity, equal opportunity, procedural consistency, or protection against unjustified group disadvantage. These objectives may not always be simultaneously attainable. More importantly, an algorithm can satisfy a particular numerical fairness threshold while relying on measures that are poorly related to actual job requirements. Recruitment fairness must therefore involve both statistical evaluation and substantive examination of what the system measures, why those variables are relevant, how applicants are affected, and whether consequential decisions can be reviewed.

This study investigates bias and fairness in automated recruitment and employee selection systems through a conceptual analysis supported by an explicitly simulated comparison. It examines how bias may arise through training data, target variables, feature construction, automated assessments, ranking mechanisms, and human interaction with algorithmic recommendations. The paper further evaluates how fairness-oriented governance may improve selection processes through data auditing, validation, transparency, accessibility, subgroup testing, and structured human oversight. Rather than suggesting that automated hiring is inherently discriminatory or inherently objective, the study adopts a balanced position: recruitment algorithms can potentially reduce some forms of inconsistent human judgment, but only when their design and deployment are subjected to rigorous and continuing scrutiny.

2. Review of Literature

2.1 Historical Data, Proxy Variables, and Algorithmic Discrimination

Research on data-driven discrimination demonstrates why automation cannot be assumed to be neutral. Barocas and Selbst argue that machine-learning systems can reproduce social inequalities through biased

training examples, inadequately defined target variables, and correlations between apparently neutral features and protected characteristics. This insight is especially important in recruitment because historical organizational records often reflect decisions made under institutional, cultural, and labor-market conditions that may themselves have been unequal.

The problem persists even when variables such as sex, race, ethnicity, age, or disability status are removed from a model. Other variables can contain overlapping information and serve as proxies. Residential location, employment gaps, educational pathways, language patterns, previous employers, occupational history, or patterns of digital behavior may indirectly correlate with demographic characteristics. A technically sophisticated algorithm may therefore identify predictive relationships that reproduce existing social patterns without explicitly receiving sensitive attributes.

Köchling and Wehner's systematic review of algorithmic decision-making in HR recruitment and development confirms that discrimination and fairness concerns extend across the employment lifecycle. Their synthesis demonstrates that algorithmic outcomes depend heavily on the assumptions built into data, objectives, design choices, and organizational practices. Fairness must therefore be assessed within the complete recruitment system rather than reduced to the properties of an isolated machine-learning model.

2.2 Automated Assessments and Candidate Ranking

Automated employee selection may incorporate résumé parsing, psychometric tests, game-based assessments, predictive scoring, natural-language analysis, and other forms of computational evaluation. These tools promise consistency because applicants can theoretically be processed according to standardized criteria. However, standardization does not guarantee validity. A system can consistently measure something that has little connection with the actual knowledge, skills, abilities, or competencies required for successful job performance.

Raghavan, Barocas, Kleinberg, and Levy examined the development and validation claims associated with algorithmic pre-employment assessment tools. Their work highlights important questions concerning how prediction targets are selected, how models are validated, what data are collected, and how vendors attempt to evaluate or mitigate bias. A central implication is that bias assessment cannot be separated from questions of construct validity and job relevance. An apparently balanced model remains problematic when the underlying assessment does not measure a defensible occupational requirement.

Automated ranking introduces an additional concern. Precise numerical scores can create an impression of scientific certainty that exceeds the reliability of the underlying assessment. Recruiters may defer excessively to rankings because quantitative outputs appear more objective than qualitative judgment. The risk is especially significant when differences between candidates are small, measurement uncertainty is not communicated, or the model provides limited explanation for its recommendation.

2.3 Fairness Governance and Human Accountability

The literature increasingly treats automated recruitment as a sociotechnical system rather than a purely computational application. Tambe, Cappelli, and Yakubovich identify fairness, accountability, limited HR datasets, complexity of human behavior, and employee reactions as fundamental challenges for

artificial intelligence in human resource management. Their analysis demonstrates why technically accurate prediction is only one component of responsible implementation.

Human oversight is therefore essential, but simply placing a person after an algorithm is not sufficient. Reviewers need authority, relevant information, training, and clear criteria for challenging automated recommendations. If recruiters routinely accept algorithmic rankings without scrutiny, nominal human involvement offers little protection. Conversely, uncontrolled manual overrides can reintroduce the inconsistent judgment that automation was intended to reduce.

Fairness governance should consequently operate before, during, and after deployment. Before implementation, organizations should establish job relevance, assess dataset quality, identify sensitive variables and potential proxies, and conduct subgroup validation. During deployment, applicants should receive appropriate information and access to accommodation or review mechanisms. After deployment, organizations should monitor selection rates, errors, model drift, complaints, and changing job requirements. Fairness is therefore a continuing governance obligation rather than a one-time certification exercise.

3. Objectives of the Study

3.1 Analytical Objectives

The principal objective of the study is to examine how bias can emerge within automated recruitment and employee selection systems and to evaluate the mechanisms through which fairness-oriented governance can reduce unjustified disparities. The analysis considers recruitment as a sequential process because disparities generated at one stage can influence the composition of applicants available at later stages.

The study also distinguishes between statistical fairness and substantive employment fairness. Statistical indicators provide useful diagnostic evidence, but a fair selection system must additionally measure relevant occupational requirements, provide meaningful procedural safeguards, and remain accountable to human decision-makers.

The specific objectives are:

- To identify major sources of bias in automated candidate sourcing, screening, assessment, and shortlisting.
- To examine the relationship between historical data, proxy variables, model targets, and unequal applicant outcomes.
- To evaluate the contribution of subgroup validation, fairness auditing, explainability, and human review.
- To demonstrate through simulated data how governance interventions may influence selection-rate disparities.
- To develop a human-centered framework for responsible automated employee selection.

3.2 Fairness and Governance Objective

A further objective is to develop a governance interpretation of algorithmic fairness that extends beyond model accuracy. Recruitment technologies operate within organizational decision structures, and responsibility therefore remains with the employer and relevant decision-makers rather than being transferred to software or technology vendors.

The study accordingly evaluates fairness through interconnected dimensions involving statistical outcomes, job relevance, transparency, accessibility, reviewability, and accountability. This multidimensional approach avoids presenting any single fairness metric as a complete definition of equitable recruitment.

4. Research Methodology

4.1 Research Design

The study employs a conceptual-methodological research design supported by an illustrative simulation. It does not use actual applicant records, employer databases, biometric data, or recruitment outcomes. The numerical analysis is therefore intended to demonstrate how a fairness evaluation may be structured rather than to estimate the performance of a particular commercial hiring system.

The theoretical framework was developed from established literature concerning algorithmic discrimination, employment analytics, machine-learning fairness, automated hiring, and organizational algorithmic control. Four recruitment stages were selected because they represent distinct points at which automated decisions can influence access to employment opportunities.

4.2 Fairness Evaluation Framework

The analysis uses a simulated **selection-rate ratio** as a simplified diagnostic indicator. Conceptually, the ratio compares the selection rate of a monitored applicant group with the corresponding reference-group **selection rate**:

Selection-

$$\text{Rate Ratio} = \frac{\text{Selection Rate of Monitored Group}}{\text{Selection Rate of Reference Group}}$$

A value closer to 1.00 indicates increasingly similar selection rates in the simulated comparison. However, the ratio is not treated as definitive proof of fairness. It does not establish job-related validity, explain individual decisions, prove absence of discrimination, or account for all legally or ethically relevant circumstances.

Table 1: Framework for Fairness Assessment in Automated Recruitment

Evaluation Dimension	Principal Question	Proposed Control
Training Data	Does historical data reproduce previous inequalities?	Dataset audit and representativeness review
Prediction Target	Is the outcome genuinely related to job success?	Criterion and construct validation
Model Features	Could variables operate as unjustified proxies?	Feature audit and sensitivity analysis
Group Outcomes	Are materially different selection patterns emerging?	Subgroup fairness testing
Explainability	Can consequential recommendations be understood?	Decision documentation and interpretable outputs
Human Review	Can an automated result be questioned	Structured reviewer intervention

	or corrected?	
Post-Deployment Monitoring	Does fairness change over time?	Periodic audit and model-drift assessment

4.3 Simulation Procedure and Ethical Position

Two theoretical configurations were compared. The first represents an unmitigated automated recruitment system in which limited fairness controls are assumed. The second represents a fairness-governed system incorporating dataset review, job-related validation, subgroup testing, proxy-variable assessment, structured human oversight, and post-deployment monitoring.

Simulated selection-rate ratios were assigned across candidate sourcing, résumé screening, pre-employment assessment, and final shortlisting. These numbers were generated solely to demonstrate comparative analysis and should not be cited as observed outcomes from real organizations.

Because the study uses no human participants, personal information, identifiable applicant data, or actual employment decisions, participant-based research ethics approval was not required for the simulation itself. A future empirical study using recruitment records would require appropriate ethical review, privacy safeguards, lawful data processing, organizational authorization, and careful protection of sensitive demographic information.

5. Analysis

5.1 Simulated Recruitment Fairness Outcomes

The simulated analysis indicates substantial differences between the two theoretical configurations. In the unmitigated system, selection-rate ratios range between 0.69 and 0.76. Résumé screening records the lowest simulated value at 0.69, suggesting that automated screening may become an important point of concern when historical employment patterns, résumé characteristics, or proxy features influence model outputs.

After fairness-governance interventions are introduced, the simulated ratios rise across all four recruitment stages. Values range from 0.89 during résumé screening to 0.94 during final shortlisting.

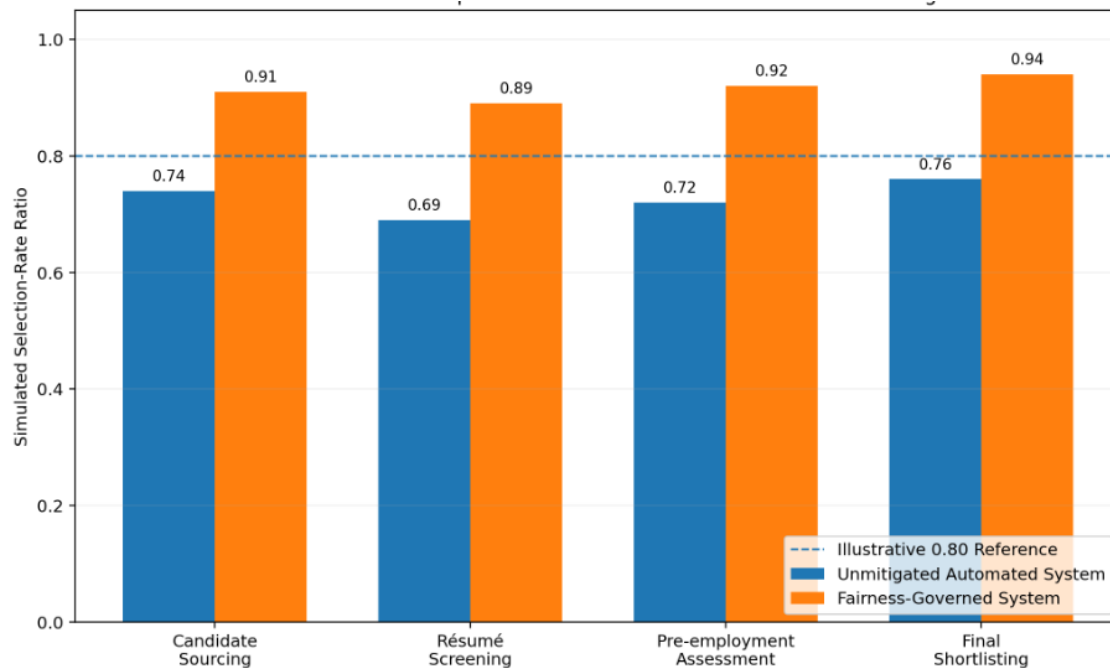
Table 2: Simulated Fairness Comparison Across Recruitment Stages

Recruitment Stage	Unmitigated Automated System	Fairness-Governed System	Simulated Improvement
Candidate Sourcing	0.74	0.91	+0.17
Résumé Screening	0.69	0.89	+0.20
Pre-employment Assessment	0.72	0.92	+0.20
Final Shortlisting	0.76	0.94	+0.18

Note: Values are illustrative and are not derived from real applicants, organizations, vendors, or employment decisions.

5.2 Graphical Fairness Comparison

Graph 1: Simulated Fairness Improvement Across Automated Recruitment Stages



The graph compares simulated selection-rate ratios before and after implementation of fairness-oriented governance. The dashed 0.80 line is included only as an illustrative analytical reference and should not be interpreted as a universal declaration that values above it establish legal or ethical fairness.

The largest simulated improvements occur during résumé screening and pre-employment assessment, each increasing by 0.20. These stages frequently depend on high-dimensional applicant information and model-generated scores, making them especially important locations for feature auditing, construct validation, and subgroup testing.

Candidate sourcing improves from 0.74 to 0.91, illustrating the importance of evaluating fairness before candidates formally enter a recruitment pipeline. Unequal targeting or recommendation can influence who learns about employment opportunities and therefore alter the applicant pool before conventional selection begins. Final shortlisting improves to 0.94, but even this comparatively balanced simulated outcome would still require substantive examination of validity and individual decision processes.

5.3 Interpretation of Fairness Improvement

The results suggest that algorithmic fairness is most effectively approached through multiple complementary interventions. Dataset balancing alone cannot correct a poorly defined prediction target. Removing demographic variables cannot eliminate proxies. Statistical parity cannot validate an irrelevant assessment. Explainability cannot compensate for systematically inaccurate predictions.

Human review cannot be effective when reviewers lack authority or blindly accept machine recommendations.

Fair recruitment consequently depends upon alignment among data quality, occupational relevance, model behavior, candidate safeguards, and organizational decision processes. This supports the broader view that automated recruitment should be evaluated as a complete decision system rather than as an isolated technical model.

A further implication concerns cumulative disadvantage. Minor differences at several sequential stages can combine to create a larger effect on final employment access. Organizations should therefore avoid limiting fairness audits to the final shortlist. Monitoring should examine sourcing, application completion, screening, testing, interview progression, and final decisions separately so that the origin of disparities can be identified more accurately.

6. Challenges

6.1 Hidden Bias and Proxy Discrimination

One of the most difficult challenges is identifying discrimination that does not arise from explicitly protected characteristics. Machine-learning systems may infer relationships among educational background, location, employment history, vocabulary, career interruptions, online behavior, or other attributes. These relationships can produce uneven outcomes even when developers have intentionally excluded demographic variables.

Historical data create another difficulty because past performance or hiring outcomes may not provide neutral measures of employee potential. Performance evaluations themselves can contain managerial subjectivity. Occupational opportunities may have been unequally distributed. Successful employees in historical records may therefore represent people whom previous organizational systems favored rather than the complete range of people capable of succeeding.

Fairness auditing should consequently examine how variables are generated, what they represent, and whether they are genuinely necessary for predicting relevant occupational outcomes.

6.2 Transparency, Explainability, and Applicant Rights

Automated selection can create substantial information asymmetry. Employers and technology providers may know how applicants are being scored while candidates receive little information concerning why they were rejected or which assessment influenced the outcome. Such opacity can undermine trust and make it difficult to identify errors or request appropriate review.

Complete publication of proprietary source code is not necessarily required for meaningful transparency. Organizations can instead communicate the purpose of an automated tool, the type of information being evaluated, the role of automation in the hiring decision, the availability of accommodations, and the mechanisms through which applicants may seek clarification or review.

Explainability should also be tailored to the decision context. Technical documentation may be appropriate for auditors, while recruiters require operational explanations and applicants require understandable information about consequential procedures. Transparency is therefore audience-dependent rather than a single disclosure obligation.

6.3 Human Oversight and Automation Bias

Human involvement is frequently proposed as a safeguard, but humans can themselves become influenced by algorithmic outputs. Numerical rankings, recommendation scores, and risk indicators may appear objective and can therefore receive excessive weight from decision-makers. This creates the possibility of automation bias, in which people rely on computational recommendations even when contextual evidence suggests that reconsideration is appropriate.

Effective human oversight should be structured. Reviewers should know when intervention is expected, which information should be checked, what uncertainty surrounds a model output, and how overrides should be documented. Organizations should additionally analyze patterns of human overrides because systematic differences may reveal either algorithmic weaknesses or reintroduced human bias.

7. Discussion

7.1 Fairness Is More Than Statistical Parity

The simulated analysis demonstrates that quantitative fairness indicators can help organizations identify potential disparities. However, a numerical ratio should function as a diagnostic signal rather than as a complete judgment of fairness. Two groups may display similar selection rates even when the assessment measures characteristics that are unrelated to occupational performance. Conversely, different selection rates require careful investigation of their causes rather than automatic assumptions about either fairness or discrimination.

Raghavan and colleagues' examination of algorithmic hiring practices supports this broader understanding. Model construction, validation, prediction targets, and bias-mitigation choices all shape the consequences of automated assessment. Fairness must therefore be considered simultaneously with validity and job relevance.

A defensible recruitment system should answer two fundamental questions: whether the procedure identifies characteristics that legitimately matter for the job and whether those characteristics are assessed in a manner that does not create unjustified disadvantage.

7.2 Human–Algorithm Decision Architecture

The analysis does not support either unrestricted automation or unrestricted human discretion. Conventional recruitment contains documented sources of subjective bias, while algorithmic systems can reproduce inequality at scale. The more productive design objective is a hybrid decision architecture in which each component is assigned tasks appropriate to its strengths.

Algorithms can support consistency, pattern identification, high-volume processing, and documentation. Human professionals contribute contextual judgment, interpretation of unusual circumstances, ethical reasoning, accommodation decisions, communication, and responsibility for employment consequences.

This division requires organizational accountability. Employers should not describe a discriminatory outcome as something that “the algorithm decided.” Organizations choose whether to deploy the tool, which data to provide, which outputs to use, what threshold to establish, and how recommendations influence recruiters. Responsibility therefore remains institutional and human.

7.3 Continuous Fairness Governance

Automated recruitment systems require continuing monitoring because employment environments change. Applicant populations evolve, occupational requirements change, educational pathways shift, economic conditions alter career histories, and new data may differ from those used during model development. These changes can reduce predictive validity or create new group disparities.

Organizations should consequently conduct periodic subgroup analysis, data-quality checks, job-validity reviews, model-drift testing, accessibility assessments, and examination of applicant complaints. Major model updates or substantial changes in job requirements should trigger renewed validation.

Fairness governance should additionally include procurement. Employers purchasing third-party systems need sufficient information to evaluate the vendor's data sources, validation procedures, model limitations, fairness testing, security arrangements, and update practices. Outsourcing the technology does not outsource responsibility for employment decisions.

8. Conclusion

Automated recruitment and employee selection systems have considerable potential to improve the scale, consistency, and analytical capacity of organizational hiring. Their value, however, depends on the quality of the decisions they support rather than the sophistication of the technology they employ. Algorithms trained on historical employment patterns can reproduce inequality, while proxy variables, inappropriate targets, poorly validated assessments, and excessive reliance on numerical rankings can create new forms of disadvantage.

The simulation presented in this study illustrates how fairness-oriented governance could reduce disparities across candidate sourcing, résumé screening, pre-employment assessment, and final shortlisting. These results are deliberately illustrative and should not be interpreted as observed evidence from an actual recruitment system. Their purpose is to demonstrate why fairness must be evaluated at multiple stages of the hiring pipeline rather than exclusively at the final selection decision.

A responsible automated recruitment framework requires rigorous job analysis, high-quality data, defensible prediction targets, subgroup validation, proxy-variable assessment, transparent documentation, accessibility, structured human review, and continuous monitoring. No single fairness metric can substitute for these safeguards. Statistical measures are valuable tools for detecting possible inequality but must be interpreted alongside occupational validity, procedural justice, and the consequences experienced by applicants.

The central conclusion is that automated recruitment should remain accountable to human and organizational governance. Artificial intelligence can help organizations make more consistent and evidence-informed employment decisions, but it should not be granted unquestioned authority over access to employment. Fairness is most likely to emerge when algorithms are treated as auditable decision-support systems operating within transparent, job-relevant, and continuously monitored recruitment processes.

References

1. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.15779/Z38BG31>
2. Mann, G., & O'Neil, C. (2016). Hiring algorithms are not neutral. *Harvard Business Review*.
3. Bogen, M., & Rieke, A. (2018). *Help wanted: An examination of hiring algorithms, equity, and bias*. Upturn.
4. Ajunwa, I., Crawford, K., & Schultz, J. (2016). Limitless worker surveillance. *California Law Review*, 105, 735–776.
5. Ajunwa, I., Friedler, S. A., Scheidegger, C., & Venkatasubramanian, S. (2016). Hiring by algorithm: Predicting and preventing disparate impact. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2746071>
6. Barach, M. A., Leung, M. D., & Lu, S. (2017). Man versus machine: The effect of algorithms on employer search and hiring behavior. *Academy of Management Proceedings*, 2017(1). <https://doi.org/10.5465/AMBPP.2017.16697>
7. Kleinberg, J., Ludwig, J., Mullainathan, S., & Rambachan, A. (2018). Algorithmic fairness. *AEA Papers and Proceedings*, 108, 22–27. <https://doi.org/10.1257/pandp.20181018>
8. Kallus, N., & Zhou, A. (2018). Residual unfairness in fair machine learning from prejudiced data. *Proceedings of the 35th International Conference on Machine Learning*, 80, 2439–2448.
9. Zafar, M. B., Valera, I., Rodriguez, M. G., & Gummadi, K. P. (2017). Fairness constraints: Mechanisms for fair classification. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 54, 962–970.
10. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. *Advances in Neural Information Processing Systems*, 29, 3315–3323.
11. Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (pp. 214–226). ACM. <https://doi.org/10.1145/2090236.2090255>
12. Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). Learning fair representations. In *Proceedings of the 30th International Conference on Machine Learning* (Vol. 28, pp. 325–333).
13. Calders, T., Kamiran, F., & Pechenizkiy, M. (2009). Building classifiers with independency constraints. In *2009 IEEE International Conference on Data Mining Workshops* (pp. 13–18). IEEE. <https://doi.org/10.1109/ICDMW.2009.83>
14. Pedreschi, D., Ruggieri, S., & Turini, F. (2008). Discrimination-aware data mining. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 560–568). ACM. <https://doi.org/10.1145/1401890.1401959>
15. Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, 29, 4349–4357.
16. Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186. <https://doi.org/10.1126/science.aal4230>
17. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77–91.



18. Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.
19. O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
20. Leavy, S. (2018). Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In *Proceedings of the 1st International Workshop on Gender Equality in Software Engineering* (pp. 14–16). ACM. <https://doi.org/10.1145/3195570.3195580>