

Evaluating Algorithmic Fairness in Machine Learning-Based Recruitment and Candidate Screening Systems

Dr. Siti Farah Abdullah

Assistant Professor, University of Intelligent Computing, Malaysia

Abstract

Machine learning is increasingly considered for recruitment and candidate-screening tasks because automated systems can process large numbers of applications and identify patterns associated with predefined job-related criteria. However, historical recruitment data may contain unequal selection patterns, measurement limitations, or proxy variables that influence automated predictions. Algorithmic fairness therefore represents an important methodological concern when machine learning is used to rank, classify, or prioritize job applicants. Predictive accuracy alone cannot determine whether a screening system produces equitable outcomes across relevant applicant groups. This hypothetical methodological study examines algorithmic fairness in machine learning-based recruitment and candidate-screening systems. The proposed framework separates predictive performance from group-level fairness assessment and compares three conceptual models: logistic regression, random forest classification, and a fairness-aware classification model. A fictional dataset containing 1,500 candidate applications is divided into hypothetical training, validation, and testing subsets. Fairness is evaluated using selection-rate difference, selection-rate ratio, and true-positive-rate difference across two anonymized applicant groups.

The illustrative analysis shows that the highest predictive accuracy and the smallest group disparities do not necessarily occur in the same model configuration. Under the assigned hypothetical values, the random forest produces the strongest conventional predictive accuracy, whereas the fairness-aware model substantially reduces selection-rate and true-positive-rate differences while maintaining comparable predictive performance. These numerical outcomes are illustrative rather than experimentally observed results. The study concludes that fairness evaluation in automated recruitment should use multiple performance and equity measures, clear documentation of applicant features, subgroup-specific error analysis, and human review. No individual fairness metric can independently demonstrate that a recruitment process is legally, ethically, or practically fair. Empirical validation using representative data and appropriate organizational governance is therefore required before automated candidate-screening systems are used for consequential employment decisions.

Keywords: Algorithmic Fairness, Machine Learning, Recruitment Analytics, Candidate Screening, Discrimination-Aware Classification, Fairness Metrics, Automated Decision-Making, Employment Analytics, Responsible Artificial Intelligence, Human Resource Analytics.

1. Introduction

Machine learning has expanded into organizational decision support because predictive models can process large datasets and identify statistical relationships that may assist recruitment, workforce planning, and candidate evaluation. Automated screening systems can conceptually examine education, professional experience, job-related assessments, and other structured application characteristics before producing suitability scores or priority categories. Although such systems may increase processing efficiency, their use in recruitment creates important concerns because employment decisions can substantially affect individuals and because historical decision records may contain unequal patterns.

Employment discrimination has been documented through field experiments demonstrating that applicants with similar qualifications may receive different employer responses when employers perceive different social identities. Bertrand and Mullainathan used randomized résumés to examine callback differences in the United States, while later research investigated labor-market discrimination in other national contexts. Such evidence does not imply that every automated recruitment model is discriminatory, but it demonstrates why historical hiring outcomes should not automatically be treated as neutral or objectively correct labels for machine learning.

Algorithmic models can reproduce historical inequalities when training labels or predictor variables reflect previous decision patterns. Removing an explicitly protected attribute may not eliminate this possibility because other variables can contain correlated information. Pedreschi et al. introduced discrimination-aware data-mining methods, while Kamiran and Calders examined classification and preprocessing procedures designed to reduce discriminatory outcomes. Calders and Verwer further investigated classification approaches that constrain dependence between predictions and sensitive attributes. These studies established an important methodological foundation for fairness-aware machine learning.

Fairness itself is multidimensional. Dwork et al. emphasized individual fairness through the principle that similar individuals should receive similar treatment, while Zemel et al. developed representations intended to support both group and individual fairness objectives. Hardt et al. later formulated equality-of-opportunity criteria based on true-positive rates. These approaches demonstrate that alternative fairness definitions can emphasize different aspects of decision behavior. Consequently, a system can satisfy one fairness criterion while failing another, making transparent metric selection essential.

The present hypothetical study examines algorithmic fairness in candidate screening through a structured comparative framework. Three conceptual machine learning models are evaluated using conventional classification measures and group-level fairness indicators. The analysis intentionally uses anonymized groups, labeled Reference Group and Comparison Group, rather than representing any real demographic population. The objective is to demonstrate how predictive utility and fairness can be examined together without presenting fictional recruitment outcomes as real employment evidence.

2. Objectives of the Study

The primary objective of this study is to examine how algorithmic fairness can be evaluated in machine learning-based recruitment and candidate-screening systems. The investigation considers predictive accuracy, subgroup selection rates, true-positive-rate differences, model transparency, and methodological challenges associated with historically generated employment data.

2.1 Examination of Fairness in Automated Recruitment

The first objective is to examine how fairness can be represented quantitatively when machine learning models classify or prioritize job applicants. The study focuses on group-level selection rates and true-positive rates because these measures describe different aspects of candidate-screening behavior. Selection-rate analysis evaluates how frequently groups receive favorable predictions, while true-positive-rate analysis examines whether candidates meeting a hypothetical suitability criterion are identified consistently across groups. Neither measure is treated as a complete definition of fairness.

The objective also considers the distinction between protected information and proxy information. Removing a sensitive characteristic from model inputs does not guarantee that the remaining features are independent of that characteristic. Variables associated with educational opportunities, employment histories, geographic patterns, or other contextual conditions can preserve statistical associations indirectly. The study therefore treats feature auditing and subgroup performance analysis as necessary components of fairness evaluation rather than assuming that exclusion of protected attributes is sufficient.

2.2 Evaluation of Predictive Performance and Fairness

The second objective is to compare predictive performance across three hypothetical screening models. Logistic regression provides an interpretable statistical baseline, random forest classification represents a more flexible nonlinear approach, and the fairness-aware model incorporates conceptual reweighting during training. Accuracy, precision, recall, and F1 score are used to evaluate predictive utility. The assigned results are illustrative and do not represent outcomes generated by real candidate-screening software.

The objective further evaluates fairness using selection-rate difference, selection-rate ratio, and true-positive-rate difference. These measures are calculated separately from predictive accuracy because a highly accurate model can still produce different outcomes across groups. Conversely, a model with smaller group disparities can have somewhat lower conventional accuracy. The study therefore treats model selection as a multidimensional decision requiring examination of both predictive and fairness-related consequences rather than maximizing a single numerical score.

2.3 Identification of Governance and Implementation Challenges

The third objective is to identify methodological challenges affecting fairness evaluation in recruitment systems. Important concerns include biased historical labels, incomplete applicant information, unequal representation, proxy variables, changing job requirements, and differences in error rates across groups. These conditions can influence model behavior even when developers do not intentionally incorporate discriminatory rules. Understanding such limitations is essential before interpreting fairness metrics or deploying automated screening procedures.

The study also considers governance requirements for consequential employment decisions. Automated predictions should be documented, periodically reassessed, and accompanied by appropriate human oversight. Applicants should not be reduced to immutable numerical profiles, particularly when training data or outcome labels are uncertain. Organizations should understand what their models predict, how errors are distributed, and whether decision criteria remain job-relevant. Fairness evaluation should therefore function as an ongoing governance process rather than a one-time technical test.

3. Review of Literature

3.1 Employment Discrimination and Recruitment Decisions

Bertrand and Mullainathan conducted a field experiment using fictitious résumés and demonstrated that otherwise similar applications could receive different callback rates according to perceived applicant identity. Pager also documented unequal employment outcomes through experimental methods, while Pager et al. extended audit-based research on hiring discrimination. These studies established that recruitment decisions can contain systematic disparities even when applicants possess comparable observable qualifications, providing an important empirical context for later algorithmic-fairness research.

Rooth examined relationships between automatic associations and hiring discrimination, while Oreopoulos investigated differential employer responses to skilled immigrant applicants. Kaas and Manger conducted a field experiment in Germany and similarly identified differences in callback behavior. These studies were conducted in different labor-market contexts and should not be interpreted as direct evidence regarding machine learning systems. Nevertheless, they demonstrate why historical recruitment labels and employer decisions require careful examination before being used as objective targets for automated prediction.

3.2 Discrimination-Aware Machine Learning

Pedreschi et al. introduced discrimination-aware data mining and developed methods for identifying potentially discriminatory patterns in decision records. Their subsequent work examined quantitative measures for discrimination in socially sensitive data. Kamiran and Calders proposed classification procedures intended to reduce discriminatory behavior, while Calders and Verwer developed alternative Naive Bayes approaches for discrimination-aware prediction. These contributions helped establish fairness as a technical property that could be investigated systematically within predictive modeling.

Kamiran and Calders later examined preprocessing strategies such as reweighting and data modification for classification without discrimination. Feldman et al. proposed methods for identifying and reducing disparate impact through transformation of data representations. Romei and Ruggieri provided a multidisciplinary review connecting discrimination analysis across computational, statistical, economic, and legal perspectives. Together, these studies show that fairness interventions can occur before, during, or after predictive modeling and that different procedures embody different assumptions.

3.3 Formal Fairness Concepts and Algorithmic Decision-Making

Dwork et al. introduced a formal framework for fairness based on treating similar individuals similarly. Zemel et al. developed fair representation learning with objectives involving both group-level outcomes and preservation of useful predictive information. These approaches illustrate how fairness can be integrated into model formulation rather than assessed exclusively after predictions are generated. However, defining similarity or acceptable group relationships can require normative choices that are not determined by machine learning alone.

Hardt et al. proposed equality-of-opportunity criteria based on equalized prediction performance across groups, while Žliobaitė reviewed measurement approaches for indirect discrimination in machine learning. Barocas and Selbst examined disparate impact within data-driven systems and emphasized that apparently neutral computational processes can still generate unequal effects. These contributions

reinforce the need for explicit fairness definitions, transparent evaluation, and careful distinction between statistical fairness measures and broader legal or ethical conclusions.

4. Research Methodology

4.1 Research Design and Candidate-Screening Framework

The present study adopts a hypothetical methodological research design based on binary candidate classification. Each applicant is conceptually assigned a job-suitability reference label derived from predefined, job-related assessment criteria. Three classifiers are considered: logistic regression, random forest, and a fairness-aware weighted classification model. The positive prediction represents advancement to a hypothetical next screening stage. No real candidate application, employer record, or recruitment decision is included in the study.

The fairness-aware configuration conceptually adjusts training weights to reduce unequal influence associated with group representation and outcome patterns. It does not remove predictive variables automatically or alter applicant qualifications. Protected-group membership is used only for fairness evaluation and hypothetical training adjustment rather than as a direct suitability criterion. A genuine implementation would require legal, organizational, and domain review to determine what data could appropriately be used during model development and fairness auditing.

4.2 Hypothetical Candidate Dataset and Variables

The fictional dataset contains 1,500 candidate records. One thousand fifty observations are allocated to training, two hundred twenty-five to validation, and two hundred twenty-five to independent testing for model-comparison purposes. A separate fairness-evaluation scenario uses three hundred hypothetical candidate decisions divided equally between a Reference Group and Comparison Group. All groups, labels, sample sizes, and outcomes are invented exclusively for analytical demonstration.

Conceptual predictor variables include role-relevant experience duration, standardized skills-assessment results, professional certification indicators, structured work-sample scores, and job-related competency measures. Names, photographs, exact addresses, and other directly identifying variables are not included. Historical recruiter ratings are also excluded from the primary conceptual feature set because their reliability is uncertain. A genuine study would require detailed validation that every predictor has legitimate relevance to the role being considered.

4.3 Fairness Metrics and Evaluation Procedure

Predictive performance is evaluated through accuracy, precision, recall, and F1 score on a shared hypothetical test set. Fairness is assessed independently through selection-rate difference, selection-rate ratio, and true-positive-rate difference. The selection rate represents the proportion of applicants receiving a positive screening prediction. A rate difference near zero represents more similar selection proportions, while a ratio near one represents greater proportional similarity under the chosen metric.

The true-positive rate represents the proportion of hypothetically qualified applicants correctly advanced by the model. The difference between group-specific true-positive rates provides an illustrative equality-of-opportunity measure. These statistics describe model behavior but do not independently determine whether a system is legally or ethically acceptable. Fairness conclusions require additional context concerning job relevance, applicant populations, data quality, measurement validity, and applicable organizational or legal requirements.

5. Analysis

5.1 Comparative Predictive Performance of Screening Models

The first hypothetical analysis compares logistic regression, random forest, and fairness-aware classification using an identical conceptual evaluation set. Random forest is assigned the highest overall accuracy at 84.0%, while the fairness-aware model records 82.7%. Logistic regression achieves 81.3%. The assigned F1 values are similarly close, demonstrating a hypothetical situation in which fairness intervention produces only a modest reduction in aggregate predictive performance rather than a dramatic loss of utility.

The comparison should not be interpreted as evidence of a universal accuracy–fairness trade-off. The relationship depends on the dataset, outcome definition, model, and intervention method. In some situations, reducing dependence on spurious or historically biased patterns can maintain or even improve generalization. The values below are intentionally constructed to demonstrate how predictive metrics may be reported alongside fairness analysis rather than to establish expected performance for real recruitment applications.

Table 1: Hypothetical Predictive Performance of Candidate-Screening Models

Model Configuration	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
Logistic Regression	81.3	78.8	76.0	77.4
Random Forest	84.0	82.1	80.0	81.0
Fairness-Aware Classification	82.7	80.0	80.0	80.0

Note: All performance values are fictional and were not generated by an operational recruitment system.

5.2 Group Fairness Comparison

The second hypothetical analysis compares the random-forest baseline with the fairness-aware model across two anonymized groups of equal size. The baseline produces selection rates of 50% and 30%, corresponding to a selection-rate difference of 20 percentage points and a ratio of 0.60. Its true-positive rates are assigned as 85% and 65%. These values illustrate a model whose aggregate accuracy may appear satisfactory while subgroup outcomes differ substantially.

The fairness-aware model is assigned selection rates of 44% and 40%, producing a four-percentage-point difference and a selection-rate ratio of approximately 0.91. True-positive rates are 82% and 78%, leaving a four-percentage-point difference. The reduction in disparities is deliberately constructed and does not represent a real mitigation result. It demonstrates why group-level metrics should be examined separately from conventional predictive performance when evaluating automated candidate screening.

Table 2: Hypothetical Group Fairness Indicators

Model	Reference Group Selection Rate	Comparison Group Selection Rate	Selection-Rate Difference	Selection-Rate Ratio	Reference TPR	Comparison TPR	TPR Difference
Random Forest Baseline	50%	30%	20 pp	0.60	85%	65%	20 pp
Fairness-Aware Model	44%	40%	4 pp	0.91	82%	78%	4 pp

Note: The groups and fairness values are fictional. No table value establishes legal compliance, discrimination, or fairness in any real recruitment process.

6. Discussion

6.1 Predictive Accuracy and Fairness Assessment

The hypothetical findings illustrate why candidate-screening systems should not be evaluated through aggregate accuracy alone. A model can classify most observations correctly while producing uneven outcomes across groups. Discrimination-aware machine learning research has therefore developed additional measures and mitigation strategies to examine how predictive errors and favorable outcomes are distributed. The present framework separates these dimensions so that strong accuracy cannot automatically be interpreted as evidence of fair treatment.

The assigned fairness-aware model maintains similar predictive performance while reducing hypothetical group disparities. Such an outcome is plausible but not guaranteed. Fairness interventions can alter predictive behavior differently depending on the training data and underlying target definition. Model developers should therefore report both overall utility and subgroup outcomes. They should also avoid describing one fairness metric as a complete measure of equitable treatment because different definitions can produce different conclusions.

6.2 Historical Data and Proxy Effects

Historical recruitment information can contain patterns shaped by prior human decisions, organizational processes, and unequal opportunities. Training a model to reproduce those outcomes may encode historical differences rather than independent measures of job suitability. Barocas and Selbst emphasized the possibility that data-driven systems can generate disparate effects even without explicit discriminatory rules. Recruitment analytics should therefore examine how labels were produced before using them as targets for automated prediction.

Proxy effects create an additional challenge because variables that appear neutral can correlate with protected characteristics. Feldman et al. examined information leakage associated with sensitive-group membership, while fairness-aware representation research has attempted to reduce such dependencies. However, removing every correlated variable can also discard legitimate job-related

information. Feature auditing should therefore examine both relevance and potential proxy behavior rather than relying solely on statistical correlation or blanket variable removal.

6.3 Recruitment Decision Support and Human Oversight

Machine learning systems can support recruiters by organizing applications, but candidate advancement remains a consequential decision that benefits from contextual human review. Automated models may fail to recognize unusual but relevant experience, nontraditional career paths, or measurement errors in application data. Human reviewers can identify such limitations when they are given appropriate information about model uncertainty and the factors contributing to a screening recommendation.

Human involvement alone does not guarantee fairness because human decisions can also contain bias. Effective governance therefore requires structured review rather than simply placing a person after the algorithm. Organizations should document override reasons, monitor subgroup outcomes, and examine whether human corrections reduce or increase disparities. The most appropriate role for automation is decision support within a transparent recruitment process rather than unquestioned replacement of professional judgment.

7. Challenges

7.1 Historical Bias and Label Validity

A central challenge in recruitment machine learning is determining whether historical outcomes represent valid targets. Previous hiring decisions, interview ratings, or employee evaluations may reflect factors unrelated to objective job performance. When such labels contain systematic bias, a model can learn to reproduce it with greater consistency. Fairness-aware preprocessing may reduce measurable disparities, but it cannot automatically transform an invalid target into a meaningful measure of candidate suitability.

Label validity should therefore be examined through job analysis and independent outcome assessment. A model intended to predict future work performance should ideally use defensible job-related criteria rather than past recruiter preferences alone. Organizations should distinguish between predicting what historical recruiters selected and predicting a validated occupational outcome. These are different objectives, and confusing them can produce misleading claims about model quality even when statistical accuracy appears high.

7.2 Competing Fairness Definitions

Different fairness measures can conflict because they impose different requirements on model behavior. Demographic-style criteria focus on favorable outcome rates, while equality-of-opportunity approaches focus on error rates among qualified candidates. Individual fairness emphasizes similar treatment of similar individuals. A model may satisfy one criterion while violating another, particularly when outcome prevalence differs between groups. Selecting a fairness measure therefore requires substantive justification rather than purely mathematical optimization.

This challenge means that reporting one favorable metric can provide an incomplete picture. Recruitment-system evaluations should present several relevant indicators, including selection rates, false-positive rates, false-negative rates, and true-positive rates. Differences should be examined together with sample size and uncertainty. Fairness metrics should support investigation rather than

serve as automatic certificates of ethical or legal acceptability. Their interpretation depends on how qualifications, job relevance, and applicant groups are defined.

7.3 Model Generalization and Organizational Governance

Recruitment models may perform differently across occupations, locations, applicant pools, and time periods. A model developed for one position may rely on feature relationships that do not apply to another role. Organizational hiring practices can also change, making previous training data less representative. Local validation is therefore necessary before extending a screening model to new positions or populations, even when the underlying algorithm remains unchanged.

Governance must continue after deployment. Organizations should maintain model versions, record changes to features, review subgroup performance periodically, and establish procedures for investigating unexpected disparities. Candidate-screening systems should also have clear boundaries concerning what information they may use and what decisions they may influence. A technically accurate system can still be unsuitable if its governance, documentation, or accountability procedures are inadequate.

8. Conclusion

Machine learning can support recruitment and candidate screening by processing job-related applicant information efficiently, but predictive accuracy does not by itself establish fairness. The present hypothetical study developed a framework in which conventional model performance and subgroup fairness are evaluated separately. Logistic regression, random forest, and fairness-aware classification were compared through illustrative performance metrics, followed by group-level assessment of selection and true-positive rates.

The hypothetical analysis assigned the strongest aggregate accuracy to the random forest while assigning substantially smaller subgroup disparities to the fairness-aware model. These values are entirely constructed and should not be treated as empirical recruitment findings. Their purpose is to demonstrate how fairness assessment can reveal information that is not visible in overall accuracy and why multiple evaluation dimensions are necessary for consequential automated decision systems.

The study concludes that responsible recruitment analytics requires valid job-related targets, representative data, transparent features, subgroup error analysis, multiple fairness metrics, and structured human oversight. Future empirical research should examine cross-role generalization, proxy variables, intervention stability, and the relationship between fairness metrics and validated job-performance outcomes. Algorithmic fairness should be treated as an ongoing evaluation and governance requirement rather than a property established permanently at model deployment.

References

1. Ajunwa, I. (2020). The paradox of automation as anti-bias intervention. *Cardozo Law Review*, 41(5), 1671–1742.
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. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. In *Proceedings of the 2018 Conference on Fairness, Accountability and Transparency* (pp. 149–159). PMLR. <https://proceedings.mlr.press/v81/binns18a.html>

4. Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 10, Article 567. <https://doi.org/10.1057/s41599-023-02079-x>
5. Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, 5(2), 153–163. <https://doi.org/10.1089/big.2016.0047>
6. 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). Association for Computing Machinery. <https://doi.org/10.1145/2090236.2090255>
7. Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., Zuiderveen Borgesius, F., & Biega, A. J. (2023). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *arXiv*. <https://arxiv.org/abs/2309.13933>
8. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems* (Vol. 29). Curran Associates, Inc. <https://arxiv.org/abs/1610.02413>
9. Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, 13, 795–848. <https://doi.org/10.1007/s40685-020-00134-w>
10. Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)* (Article 43). Schloss Dagstuhl—Leibniz-Zentrum für Informatik. <https://doi.org/10.4230/LIPIcs.ITCS.2017.43>
11. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), Article 115. <https://doi.org/10.1145/3457607>
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). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372828>
13. Sánchez-Monedero, J., Dencik, L., & Edwards, L. (2020). What does it mean to “solve” the problem of discrimination in hiring? Social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 458–468). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372849>
14. Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59–68). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287598>
15. Suresh, H., & Guttag, J. V. (2021). A framework for understanding sources of harm throughout the machine learning life cycle. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization* (Article 17, pp. 1–9). Association for Computing Machinery. <https://doi.org/10.1145/3465416.3483305>
16. Poe, R. L., & El Mestari, S. Z. (2024). The conflict between algorithmic fairness and non-discrimination: An analysis of fair automated hiring. In *Proceedings of the 2024 ACM Conference*

- on Fairness, Accountability, and Transparency. Association for Computing Machinery. <https://doi.org/10.1145/3630106.3659015>
17. Bogen, M., & Rieke, A. (2018). Help wanted: An examination of hiring algorithms, equity, and bias. Upturn. <https://www.upturn.org/work/help-wanted/>
18. Wilson, C., Hoffman, J., & Morgenstern, J. (2021). Predictive inequity in object detection. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 325–332). Association for Computing Machinery. <https://doi.org/10.1145/3461702.3462541>
19. Fairness, AI & recruitment. (2024). Computer Law & Security Review, 53, Article 105966. <https://doi.org/10.1016/j.clsr.2024.105966>
20. Bias audit laws: How effective are they at preventing bias in automated employment decision tools? (2024). <https://doi.org/10.1080/13600869.2024.2403053>