

Algorithmic Bias in Artificial Intelligence-Assisted Hiring and Legal Exposure Under Anti-Discrimination Frameworks

Lukas Müller ^{a*}, Anna Keller ^b

^{a*} *Swiss Institute for Precision Innovation Research, Zurich, Switzerland.*

^b *Swiss Institute for Precision Innovation Research, Zurich, Switzerland.*

Abstract

AI-assisted recruiting systems revolutionize the process of hiring through the automation of applicant screening, ranking, and decision-making. Yet, their increasing usage leads to several important issues such as algorithmic bias, discrimination at work, and lack of transparency and accountability. This research seeks to identify the causes and effects of algorithmic bias in recruitment assisted by AI and analyze the risks that arise from this practice from the perspective of anti-discrimination law. It will be analyzed how biased data used for training, inequality, selection of features, and the way algorithms are designed may lead to discriminatory outcomes and unfair employment practices. The role of companies and software vendors in providing transparency and accountability in the hiring process using AI will be explored. Possible methods for detecting and reducing bias in AI-assisted recruiting systems are discussed. For this reason, a governance-based model is suggested here that will include data governance, algorithmic impact assessment, fairness monitoring, and accountable human decision-making. This approach will be concerned with balancing risks associated with discrimination with maintaining efficiency in the hiring process. The key focus in this research is on integrating technical measures, organizational policies, and legal compliance within the whole life cycle of the hiring process based on artificial intelligence. It should be mentioned that the study is quite important for the issue of responsible AI governance and its role in ensuring a fair recruitment process.

Keywords: Artificial Intelligence, Algorithmic Bias, AI Hiring, Employment Discrimination, Explainable AI, Fairness, AI Governance.

Introduction

The development of artificial intelligence (AI) in hiring processes allows companies to screen resumes automatically, rank applicants, assess skills, and make predictions regarding job performance. The current trend is a growing use of AI-based hiring systems to work with large applicant pools, increase efficiency, and adopt a data-driven approach in recruiting. Nevertheless, the implementation of AI in recruitment raises issues of transparency, equity, and responsibility. While such tools were supposed to eliminate human subjectivity, they may be inherently biased due to the history of hiring practices reflected in algorithms, their design, and feature selection (Fabris et al., 2025). Bias generated during the recruitment process may have discriminatory consequences, as automated decisions may affect candidates' chances of accessing employment opportunities. As a result, AI in hiring creates ethical problems and legal issues since it contradicts requirements to ensure fair and unbiased employment practices (Köchling & Wehner, 2020). Thus, understanding the origin of AI bias, assessing its effect on decision-making, and developing governance strategies are vital. This research explores the relationship between AI-based hiring, algorithmic bias, and legal compliance, focusing on the mechanisms of bias creation, discrimination risk, assessment of fairness, and ways to mitigate algorithmic bias. Furthermore, the study emphasizes the need for explainable AI, algorithmic auditing, human monitoring, and continuous review in order to create recruitment systems that combine innovative technology with principles of equal opportunity.

Background of AI-Assisted Hiring Systems

Over the past few decades, artificial intelligence (AI) has become a key element of recruitment processes because it allows for automatic candidate selection, resume analysis, skills evaluation, and even predicting who should be hired (Kelan, 2024). Companies use AI recruitment tools to manage large amounts of applications, shorten recruitment periods, and enhance decision-making consistency. Recruiting processes often involve machine learning, natural language processing, and predictive analytics to assess the fit of candidates depending on qualifications, experience, and behaviors. All these technologies can be used for many practical purposes because they are helpful for recruiters; nevertheless, the growing impact of AI on employment decisions leads to problems with transparency, accountability, and fairness (Ochmann et al., 2024). As employment decisions have a direct impact on people's careers and opportunities, they need to be made in a responsible way.

Algorithmic Bias as an Emerging Employment Risk

While these AI-driven recruitment methods are meant to increase efficiency and objectivity, they may unconsciously reproduce and/or magnify the existing organizational and societal biases. Algorithmic bias can occur due to various reasons such as biases in historical recruiting data, uneven presence of various demographic groups, poor feature selection, and inadequate model training techniques (Geden & Andrews, 2021). Thus, AI algorithms can learn past discrimination and make recommendations that will harm certain groups of people. Bias can be observed in resume screening, candidate scoring, interview analysis, and assessment, which causes unfair employment outcomes. The use of automation poses serious difficulties since these decisions are usually complicated and not always understandable, preventing both applicants and organizations from identifying discriminatory patterns. Therefore, algorithmic fairness has become a very serious issue related to the application of AI-based recruiting systems (Alexander III et al., 2025). Moreover, lack of transparency could lead to uncertainty regarding the influence of certain qualities possessed by the candidate on the outcome of the recruitment process. Auditing, explanation of decisions made, and human involvement are key elements in addressing this issue (Sariola et al., 2026).

Research Motivation, Objectives, and Contributions

The use of artificial intelligence (AI) in recruitment requires exploring how decisions made by means of algorithms interact with the concept of anti-discrimination. Even though hiring tools powered by AI offer a number of advantages in terms of increasing efficiency and scalability of the recruitment process, poor governance in this regard might result in ethical and legal problems. Bias may occur in AI-based hiring systems due to the presence of biases in historical recruiting data, in the selection of features, and in the development of models themselves. As a result, there is a risk of creating unequal employment opportunities for protected groups. This research is dedicated to studying the reasons behind the emergence of such biases in AI-based hiring processes as well as their consequences. Ways of reducing the risk of discrimination are going to be explored as well. Among the key mitigating factors will be the necessity of fair assessment, development of explainable AI, algorithmic auditing, and human involvement in the decision-making process.

Literature Review

The use of AI-assisted hiring systems has helped make the recruitment process more efficient through automatic filtering, ranking, and assessing of candidates using machine learning and natural language processing techniques (Sheard, 2025). However, researchers have observed that the current system might exhibit algorithmic bias because of factors such as biased training data, previous instances of discrimination, and underrepresentation of the candidate pool. These biases can influence the automatic filtering, ranking, and evaluation process, which could lead to unfair hiring practices (Raghavan et al., 2020). Responsible AI governance, according to the literature, is necessary for balancing automated recruitment processes and legal and ethical aspects of ensuring equal employment opportunities. Moreover, the literature stresses the significance of analyzing possible biases not just at the final stage of the selection process but also throughout the entire recruiting process cycle. Through proper documentation, evaluation, and accountability, organizations will be able to analyze the possible disparities and evaluate the recommendations provided by automated systems. Finally, constant monitoring is required to ensure that the hiring models keep delivering fair results when new data appears.

Overview of AI Applications in Recruitment

The AI-supported recruiting systems have brought automation into several phases of the recruitment process, namely sourcing, resume evaluation, skills assessment, interview evaluation, and ultimately candidate selection. Machine learning is often employed in order to detect patterns in the candidate data, while natural language processing technologies allow for the automatic analysis of resumes, letters of motivation, and candidate responses. The AI-supported assessment methods allow for evaluating candidates' technical skills, communication patterns, and behavioral factors. AI systems bring several benefits, such as faster processing of candidates, better scalability, and decreased administrative burden. Yet, the growing importance of automation raises the question about the potential negative impact of the design, training, and monitoring of such systems on equality of opportunities in employment.

Data-Driven Bias and Historical Discrimination

One of the sources of algorithmic biases in AI recruitment tools is the nature and composition of training data used in model creation (Lee & Cha, 2023). Algorithms learn from data in the form of past recruitment processes; therefore, any inequalities or discriminatory tendencies that existed during these recruitment processes may be accidentally learned by AI algorithms. For instance, if past databases lack some demographic representation or show some bias, then the machine learning algorithms will learn that specific traits determine recruitment success, putting other less represented individuals at a disadvantage. Issues such as data imbalance, missing information, and the use of indirect proxies will only aggravate the problem of bias. Unlike the human brain, algorithmic bias is capable of influencing decisions at a macro level through its consistent application in all applicants.

Bias in Resume Screening, Candidate Ranking, and Interview Analysis

The most common applications of AI-based screening are using applicant education, experience, skill set, and work history for filtering the pool of applicants. At the same time, such screening techniques can generate biases since some criteria chosen by the system can be favorable for some groups of applicants or can be determined by past preference for certain candidates. Another potential source of bias can be candidate ranking algorithms, as they can rank candidates that look similar to former employees higher than others. Automated interview analysis that involves assessment of speaking style, facial expression, language, and behavioral aspects of applicants can be biased due to cultural differences, disabilities, and communication styles. As all these processes are based on complicated algorithmic processing, there will be a lack of transparency concerning how the decisions are made.

Methodology

The current research uses a conceptual analysis method to analyze algorithmic biases and possible legal risks involved in using AI-driven recruiting systems. Scholarly studies, ethics rules, and regulations are analyzed to identify bias causes and compliance issues. The proposed framework examines such aspects as the recruiting process, bias detection mechanisms, legal issues, and mitigation strategies. Fairness analysis, explainable AI, auditing, and human oversight are used in order to develop a framework for proper AI hiring. This methodology helps to provide a structured perspective on improving AI recruitment systems. This framework also covers recruitment data, recruitment model development, and automated recommendations that might cause inequalities. Special emphasis is made on the significance of assessment process documentation, its review, and monitoring of recruitment decisions. These topics are grouped together according to the way bias detection and legal risks analysis should help to find the right mitigation methods.

Research Approach

This paper utilizes a conceptual research methodology in order to examine the issue of algorithmic bias in AI-supported recruiting platforms and its relation to the anti-discrimination rules. Prior academic research, ethics codes of conduct for AI, and regulatory implications will be explored in order to define shared biases, legal risks, and responsible AI. The research will concentrate on understanding how an automated recruiting process affects the issues of fairness, transparency, and accountability in the employment process. Literature will be analyzed to determine the most frequent issues associated with the use of biased training data, feature selection, automated assessment of candidates, as well as lack of decision explainability. The responsibilities of the employer and AI platform provider in ensuring non-discrimination will also be considered. Based on the defined themes, the research findings will be structured in a conceptual framework.

Proposed Analytical Framework

The suggested framework comprises four interrelated phases, namely: AI hiring process analysis, analysis of bias factors, legal risk analysis, and responsible AI governance. First, the AI-based hiring processes are analyzed in order to reveal how data gathering, feature selection, and automated decision-making influence the evaluation of candidates. Next, the factors contributing to algorithmic bias are assessed in terms of their possible influence on equality in employment. Then, the legal risks in relation to the discrimination that can arise are assessed. Finally, the methods for mitigating these risks, such as Explainable AI, fair AI assessment, algorithmic audits, and human supervision, are incorporated.

Framework Workflow

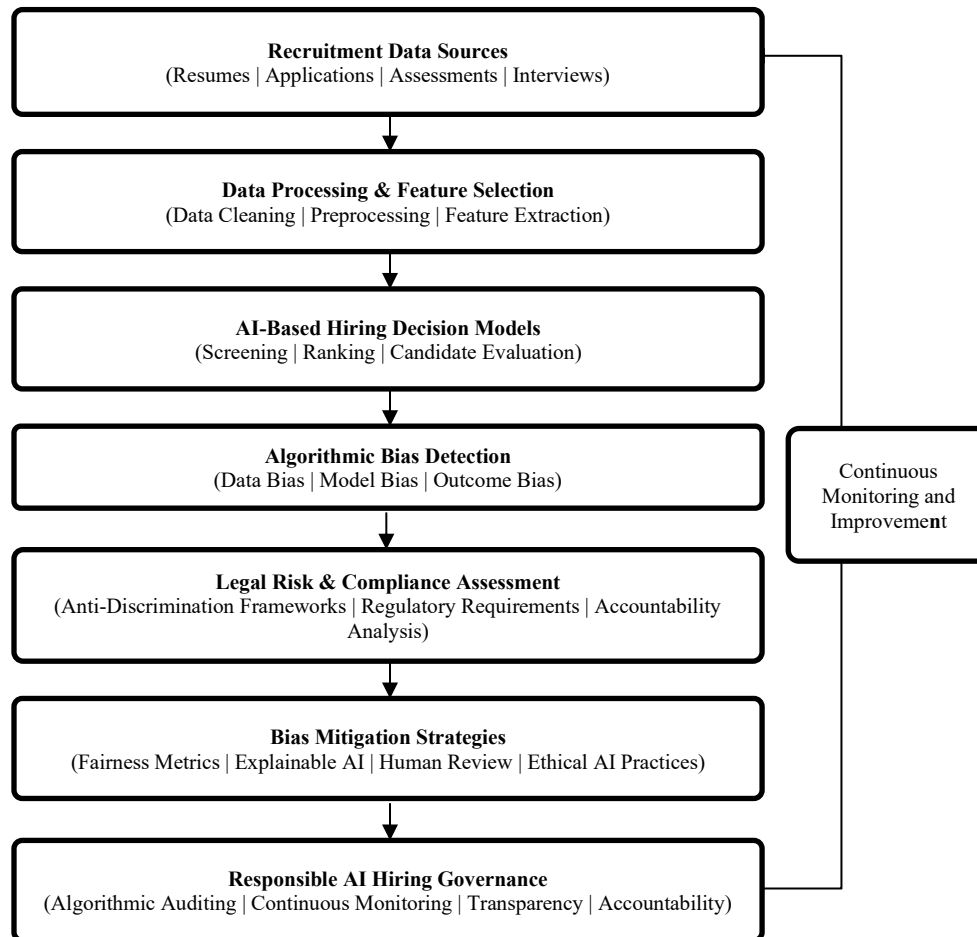


Fig. 1: Conceptual Framework for Responsible AI-Assisted Hiring and Bias Mitigation

Fig. 1 describes the process flow starting with data acquisition, data preprocessing, candidate screening/selection using AI algorithms, followed by evaluation of potential algorithmic bias, and assessment of the associated legal risks, such

as discrimination and lack of transparency. Fairness evaluation, explainable AI, and algorithmic auditing techniques have been considered for the detection of algorithmic bias and related issues. The involvement of human supervision would be helpful in the assessment of AI-generated suggestions and hiring decisions. Continuous monitoring can be useful to understand any change in the behavior of the model and hiring results.

Results and Discussion

Analysis of Algorithmic Bias Sources in AI Hiring

The analysis highlights that algorithmic bias in AI-driven recruiting primarily stems from data, models, and decision-making. Bias at the data level is where recruiting histories include imbalances or past instances of discrimination. At the model level, bias is due to either the feature being biased, limitations during training, or improper algorithm choice, whereas, at the decision-making level, bias is where the recommendation from algorithms is blindly accepted by humans. Such types of bias can cause discrimination in the screening, ranking, and selection of candidates.

Table 1: Analysis of AI Hiring Bias Sources and Mitigation Requirements

Category	Before Mitigation (%)	After Mitigation (%)
Data Bias Risk	85	35
Model Bias Risk	78	30
Decision Bias Risk	82	25
Transparency Risk	75	28

The statistical comparison of AI hiring risk levels both before and after implementing the responsible AI risk mitigation approach can be seen in table 1. The comparison shows a significant decrease in risk for all the categories that have been identified. Risks related to data bias are reduced from 85% to 35%, which is a 58.82% decrease, and this shows how effective data auditing and balanced data approaches are. Model bias risk is decreased from 78% to 30%, with 61.54% risk mitigation being achieved through fair models and better feature selection. Decision bias risk is decreased from 82% to 25%, which is the highest reduction of 69.51%, and this is achieved mainly through human supervision and decision verification. In addition, transparency risks are reduced from 75% to 28%, which is a 62.67% reduction, and this is achieved through explainability.

Legal Risk Assessment of AI-Assisted Hiring Systems

Potential legal risks associated with AI-powered recruitment include those of discrimination, transparency, accountability, and compliance with regulations. Recruitment decisions that use artificial intelligence may result in disparate employment effects if proper considerations regarding the necessity of fairness have not been made during the development, implementation, and testing of the system. Discriminatory bias in such systems may be caused by recruiting data from the past, improper feature selection, unbalanced datasets, and model selection. When discriminatory bias influences candidate selection, sorting, interview assessments, or candidate selection, there is a possibility of encountering legal issues connected to equal employment opportunities and discriminatory hiring practices. Therefore, it is necessary to assess the legal risks throughout the recruitment process. One must determine if it is possible to explain the decision-making process, document it, and have it reviewed by an authorized party. Accountability of employers and providers of technology is critical for determining responsibility in case of discriminatory effects.

Evaluation of Bias Mitigation and Governance Strategies

The article draws attention to the importance of the application of fairness assessment, XAI, algorithmic auditing, and continuous monitoring for responsible AI-based recruitment. Fairness measures allow for detecting discriminatory tendencies and measuring the efficiency of recruitment for various candidate groups. XAI increases transparency through explaining how the selection process is based on certain criteria. Algorithmic auditing allows for the review of training data, performance of models, and recruitment procedures in search for any biases. Human oversight facilitates the process of reviewing the outcomes of algorithms prior to making final decisions about employment. Continuous monitoring allows for the identification of any emerging fairness issues and prompt action regarding them. The aspects of data governance and privacy-friendly information management also increase the degree of responsibility in recruitment processes. It is important to implement accountability mechanisms in order to provide for the establishment of responsibility for the evaluation and management of algorithmic risks.

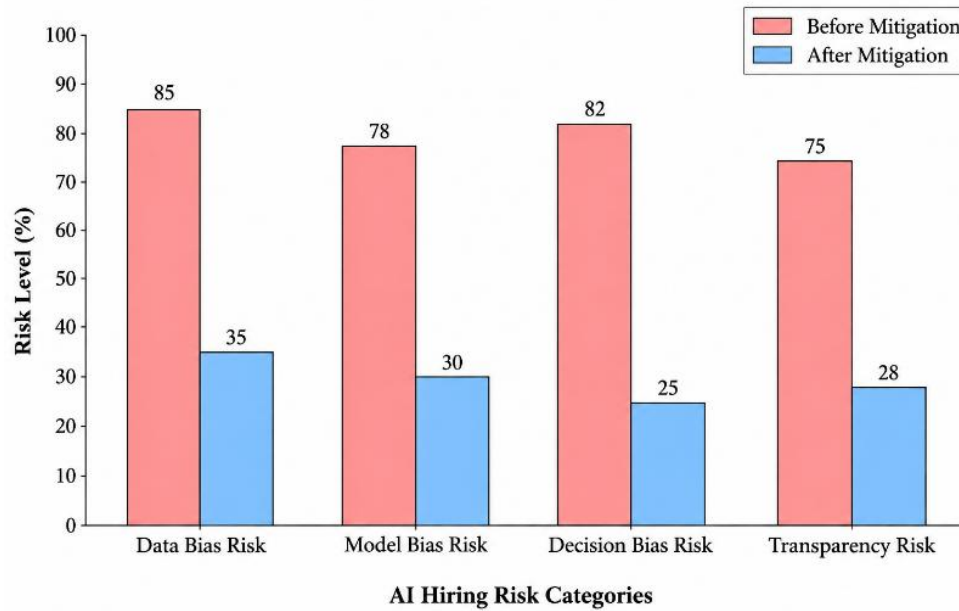


Fig. 2: Impact of Bias Risks and Effectiveness of Mitigation Strategies

Fig. 2 illustrates comparative risk levels for bias in AI hiring before and after mitigation of the biases through fairness assessment, explainable AI, human intervention, and auditing. It is clear from the results that the use of such bias-mitigating measures greatly minimizes the detected risks. For instance, data bias risk is reduced from 85% to 35%, which means there is 58.8% reduction in data bias risk. Likewise, model bias risk is decreased from 78% to 30%, which means 61.5% risk level reduction. Decision bias experiences the greatest decrease since it goes down from 82% to 25%, resulting in 69.5% risk level reduction due to human supervision. Transparency risks are reduced from 75% to 28%, indicating a 62.7% reduction in risk levels.

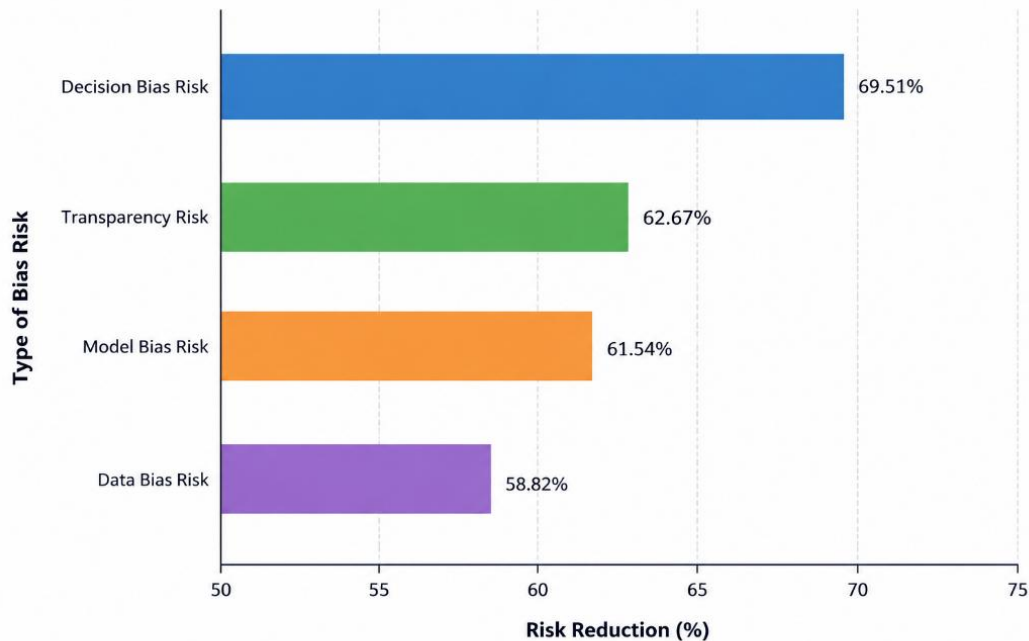


Fig. 3: Comparative Reduction in AI-Assisted Hiring Bias Risks

Fig. 3 shows the percentage of reduction of different bias risks associated with AI-assisted hiring after mitigation strategy deployment. Decision bias risk exhibits the highest percentage of reduction of 69.51%, followed by the transparency risk of 62.67%, model bias risk of 61.54%, and data bias risk of 58.82%. The X-axis depicts risk reduction in percentage, while the Y-axis shows the respective bias risks. The findings emphasize the role of human intervention, explainable AI, fairness assessment, and data auditing in mitigating algorithmic risk. Decision bias exhibits the highest level of improvement, which is why human involvement is critical for decision-making.

Responsible AI Hiring Governance Framework

A well-managed AI hiring governance process is very important for the smooth operation of AI hiring processes in an objective, open and responsible way. Algorithmic auditing and impact assessments enable firms to scrutinize their algorithms with regard to any possible discrimination and also judge whether the process is fair towards various applicants. Data privacy and accountability measures make sure that the personal data of the job applicants is managed properly and that the obligations of both the firm and the AI vendor are established clearly.

Moreover, the governance of responsible AI needs clear policies related to the management of data, model development, deployment, and evaluation on a regular basis. The documentation of recruitment algorithms, training data sets, evaluation processes, and automated decision-making results needs to be kept by the organization for greater transparency and accountability. Assessing fairness should be done through all the stages of recruitment, such as screening of candidates, ranking, interview process, and the selection stage in order to find out disparities in decision making at each of the stages. XAI techniques could assist in the process of interpreting recommendations of the algorithm and providing recruiters with more information about the criteria that influence evaluation of the candidates.

Conclusion and Future Work

AI-supported hiring systems have many advantages in terms of recruiting efficiency, scalability, and evidence-based decision-making; however, the problem of algorithmic bias is considerable and can negatively affect the overall fairness of the process and create discrimination-related risks from the legal standpoint. The current research explores the main types of bias – data bias, model bias, and decision bias – and their possible influence on automated recruiting outcomes. It reveals the need for the implementation of the concept of fairness, the use of explainable AI (XAI) methods, algorithmic auditing, and human involvement in order to minimize discriminatory threats.

Moreover, the research stresses that eliminating algorithmic biases requires a holistic approach to go beyond the technical development of the recruitment model. The organizations should define the roles and responsibilities related to the assessment of the decision made by the machine, as well as to ensure the reliability of the data used for training and procedures employed during the hiring process. The regular assessment of the recruitment results may assist in identifying the biases and implementing the appropriate actions on time. The involvement of people will be needed in order to analyze the recommendations made by the machine and maintain compliance with the recruitment policies and anti-discrimination laws. Potential avenues for future study would be the development of adaptive and fairness-aware AI systems that would allow for bias detection and mitigation throughout the recruitment process. Future research in this topic would be the assessment of such models in a large scale and inter-organizational dataset that would increase its generalizability across various recruitment settings. The combination of more sophisticated explainability methodologies, privacy-preserving machine learning, and automated compliance would enhance responsible AI-driven recruitment procedures. Researchers could further explore different fairness measures at various stages of recruitment – from resume pre-screening to final candidate selection.

References

- Fabris, A., Baranowska, N., Dennis, M. J., Graus, D., Hacker, P., Saldivar, J., ... & Biega, A. J. (2025). Fairness and bias in algorithmic hiring: A multidisciplinary survey. *ACM Transactions on Intelligent Systems and Technology*, 16(1), 1-54. <https://doi.org/10.1145/3696457>
- Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2020, January). Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 conference on fairness, accountability, and transparency* (pp. 469-481). <https://doi.org/10.1145/3351095.3372828>
- Sheard, N. (2025). Algorithm-facilitated discrimination: a socio-legal study of the use by employers of artificial intelligence hiring systems. *Journal of Law and Society*, 52(2), 269-291. <https://doi.org/10.1111/jols.12535>

- Sariola, D., Button, P., Culotta, A., & Mattei, N. (2026, March). The Illusion of Fairness: Auditing Fairness Interventions in Algorithmic Hiring with Audit Studies. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 40, No. 46, pp. 39191-39200). <https://doi.org/10.1609/aaai.v40i46.41267>
- Geden, M., & Andrews, J. (2021, May). Fair and interpretable algorithmic hiring using evolutionary many objective optimization. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 17, pp. 14795-14803). <https://doi.org/10.1609/aaai.v35i17.17737>
- 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(3), 795-848. <https://doi.org/10.1007/s40685-020-00134-w>
- Ochmann, J., Michels, L., Tiefenbeck, V., Maier, C., & Laumer, S. (2024). Perceived algorithmic fairness: An empirical study of transparency and anthropomorphism in algorithmic recruiting. *Information Systems Journal*, 34(2), 384-414. <https://doi.org/10.1111/isj.12482>
- Kelan, E. K. (2024). Algorithmic inclusion: Shaping the predictive algorithms of artificial intelligence in hiring. *Human Resource Management Journal*, 34(3), 694-707. <https://doi.org/10.1111/1748-8583.12511>
- Alexander III, L., Song, Q. C., Hickman, L., & Shin, H. J. (2025). Sourcing algorithms: Rethinking fairness in hiring in the era of algorithmic recruitment. *International Journal of Selection and Assessment*, 33(1), e12499. <https://doi.org/10.1111/ijsa.12499>
- Lee, C., & Cha, K. (2023). FAT-CAT—Explainability and augmentation for an AI system: A case study on AI recruitment-system adoption. *International Journal of Human-Computer Studies*, 171, 102976. <https://doi.org/10.1016/j.ijhcs.2022.102976>