Interdisciplinary Research Term Paper

IDS 300W - Interdisciplinary Theory and Concepts (Spring 2024)

Nicholas Gray

Old Dominion University

IDS300W

Pete Baker

04/27/2025

Algorithmic Bias in Automated Hiring Systems: An Interdisciplinary Approach to Mitigating Inequity

Abstract

As artificial intelligence (AI) continues to be incorporated into human resource management, most notably in automated recruiting systems, the problem of algorithmic bias has become a pressing concern. In using Repko and Szostak's (2016) ten-step interdisciplinary research process to analyze how psychological, economic, and technological disciplines explain, intersect, and ultimately enhance a more holistic view of algorithmic discrimination in recruiting, the empirical literature of these fields is drawn from to assess the systemic causes of algorithmic bias, the psychological effect of such bias on the applicant, and the technological architecture that fosters inequity. The paper integrates these findings into an interdisciplinary solution based on transparent algorithmic design, behavioral testing, and regulatory frameworks.

Step 1: Define the Problem or State the Focus Question

Algorithmic bias in the hiring system describes the process by which AI-driven decision-making machines inadvertently reproduce or exaggerate historical biases embedded in data, often to the disadvantage of groups protected by law. Although the systems are designed to be more efficient and objective, they tend to be based on training data sets that reflect the existing inequalities in the institution (Raji & Buolamwini, 2022). Subsequently, automated application evaluation tools can reinforce discrimination based on race, gender, and socioeconomic status instead of doing away with discrimination. An example is that job applicants who are given the identifiable race-based names of "Lakisha" or "Jamal" tend to get fewer callback offers compared to those who are given conventionally white-sounding names such as "Emily" or "Greg" (Bertrand & Mullainathan, 2004; Pedulla et al., 2022). This reveals how systemic bias can be embedded in human and machine-based evaluations.

The central research inquiry is: **How can the psychological, economic, and technological dimensions of algorithmic bias in hiring be integrated to develop more equitable AI systems?** To answer this inquiry necessitates a systematic interdisciplinary approach, one that takes into account the AI system's technical architecture in addition to discrimination's socio-psychological mechanisms, as well as the labor markets in which such tools are implemented (Binns, 2018; Emerson et al., 2024; Finocchiaro et al., 2021).

Step 2: Justify Using an Interdisciplinary Approach

The algorithmic employment discrimination issue is multidisciplinary in that it can be fully appreciated and solved by neither a single discipline alone. To address the issue, technologists identify the machine learning system architecture, such as preprocessing, model training, and auditing (Kroll, 2015; Raji & Buolamwini, 2019). The technical orientation, however, underemphasizes the human and behavioral effects of algorithmic choice. Economists evaluate the dynamics of the labor market and study the effect of discrimination on the efficiency of the market, relying most often upon experimental designs to detect employer response bias (Bertrand & Mullainathan, 2004; Jibuti, 2024). Economic theory, in turn, too often relies upon rational agents, potentially under-theorizing ethical trade-offs or subjective perceptions of fairness. Psychology contributes to a deeper understanding of the mental health, self-esteem, and stereotype-activated responses of job hunters who face algorithmic discrimination (Steele, 1997; Emerson et al., 2024). Yet, psychologists do not deal with the logic of incentives or the computation embedded in computational systems.

Therefore, informed development demands the integration of these three fields: technology (for determining designs' constraints and audibility), economics (for analysis of systemic effects), and psychology (for explanation of experience and mental processing). Only by such integration can we be able to develop interventions that are both technologically feasible and socially responsive (Finocchiaro et al., 2021; Binns, 2018, 2024).

Step 3: Identify Relevant Disciplines

In order to discuss the unequal impacts of automation on hiring, three disciplines were chosen for their unique but complementary lens. Computer Science and AI Ethics study the technical origins of bias in algorithms and discuss interventions such as fairness-aware machine learning as well as auditing mechanisms (Finocchiaro et al., 2021; Raji & Buolamwini, 2022). This area also discusses normative questions regarding justice as well as transparency in automated systems (Binns, 2024; Kroll, 2015). Behavioral and Social Psychology explores ways that rejection by algorithms might activate stereotype threat, decrease trust in institutions, and inform identity formation in marginalized populations (Steele, 1997; Emerson et al., 2024).

These psychological impacts have a lasting influence on workforce participation and social justice. Labor Economics supplies empirical evidence on discrimination, illustrating ways that automated tools for hiring might replicate past biases and misallocate labor market efficiencies (Bertrand & Mullainathan, 2004; Jibuti, 2024; Mengel, 2024). Together, these areas allow for an in-depth interdisciplinary examination of discrimination by algorithms.

Step 4: Conduct a Literature Search

An exhaustive interdisciplinary search from academic databases like JSTOR, Google Scholar, IEEE Xplore, and PsychINFO was carried out. The search terms applied included algorithmic discrimination, AI hiring, automated recruiting bias, stereotype threat hiring, and labor market fairness. This generated a set of peer-reviewed empirical studies, theoretical models, and interdisciplinary critiques related to AI in hiring as well as labor discrimination.

One early empirical study is Bertrand and Mullainathan's (2004) audit study, which showed evidence of racial bias in resume callbacks by releasing fake resumes with racially identifiable names. Their finding remains commonly referenced in fairness in algorithmic recommender systems, particularly since machine learning systems tend to reproduce

historical employment data that have these same racial imbalances. Building on these observations, Pedulla et al. (2022) investigate to what extent the design and origin of job advertisements affect observed rates of racial discrimination, proposing that mediation by technology may combine with prevailing systemic inequalities in unexpected ways.

Drawing on AI law and ethics, which include auditing commercial facial analysis tools, Raji and Buolamwini in 2019 and 2022 revealed huge racial and gender discrepancies. By their "actionable auditing" approach, their work has been instrumental in pointing to empirical failure in automated system use in contexts where high stakes exist, including employment.

To frame these biases in terms of larger philosophical and political theory, Binns (2018, 2024) extends Rawlsian ideas to rule-based decision-making, questioning whether leading measures of fairness effectively confront structural inequality. Parallel ideas from psychology, such as Steele's (1997) theory about stereotype threat, explain ways in which identity-related stress degrades candidate performance on tests used in hiring situations, perhaps magnifying harms from algorithms.

Lastly, work by Finocchiaro et al. (2021) and Kroll (2015) bridges design at the technical level with accountability, considering particular ways in which algorithmic transparency and auditability could support institutional reform. This body of work as a whole represents an active and rich discourse across computer science, economics, psychology, and philosophy, all of which are needed to achieve rich interdisciplinary knowledge about algorithmic discrimination in employment.

Step 5: Develop Adequacy in Each Relevant Discipline

Acquiring disciplinary adequacy demanded an in-depth study of foundational concepts, empirical research, and case studies in all three chosen disciplines—Computer Science, Psychology, and Economics.

In Computer Science, adequacy was also developed from close reading of technical reports and ethics guidelines from prominent professional bodies like IEEE and ACM, as well as recent developments in fairness-aware machine learning. Publications such as Raji and Buolamwini's (2019, 2022) audits on commercial facial recognition systems demonstrate that technical bias in design in algorithms can have disparate impacts on marginalized populations. Kroll's (2015) dissertation on algorithms that account for their behavior and Finocchiaro et al. (2021) bridge mechanism design via machine learning to analyze structure limitations in fairness in algorithms.

Theories in Psychology, namely, stereotype threat theory (Steele, 1997) and empirical social identity threat analyses (Emerson et al., 2024), were used to study how automated recruitment systems influence self-concept and behavior among applicants. Theories in organizational justice helped to elucidate perceived process fairness in algorithmic tests.

In economics, models of labor marketplace discrimination in terms of statistical discrimination, as well as segmentation theory, were examined by employing experimental data from Bertrand & Mullainathan (2004), Pedulla et al. (2022), as well as Jibuti (2024), in order to posit a structural focus on AI in reproducing economic inequality.

Step 6: Analyze the Problem and Evaluate Each Insight

Computer Science Insight: Technical Bias is Embedded in Data and Models

Algorithmic bias in hiring tools is usually not due to malicious intent but to bias in training data. As machine learning algorithms are trained on historical data about being hired, they replicate discriminatory habits in training data. Raji and Buolamwini (2019) notoriously revealed that commercial facial recognition systems produced dramatically higher rates of error for dark-skinned women compared to light-skinned men, unveiling race and gender bias in training sets. One glaring example is Amazon's now-discontinued AI recruiter tool that

stigmatized résumés with the word "women's"—for example, in "women's chess club captain"—because such mentions were associated in history with underrepresentation among successful candidates (Raji & Buolamwini, 2022). Even if not designed to discriminate, the lack of mechanisms for being held to account, as well as the lack of explainability, in these systems makes technical bias run unmitigated (Kroll, 2015).

Psychological Insight: Bias Diminishes Candidate Self-Perception

Psychologically, algorithmic recruitment can evoke stereotype threats and harm the self-efficacy of marginalized candidates. As per Steele (1997), stereotype threat is evoked when an individual has fear about confirming stereotypes related to their social group, triggering performance decline and disengagement. This threat is aggravated when prospects are spurned by impenetrable systems that do not even render feedback. Current research has found that perceived fairness is a potent predictor of applicant reactions and motivation (Emerson et al., 2024). That there is little human contact in algorithmic recruitment exacerbates alienation, especially for individuals who already feel uncertain about their place in the professional environment.

Economic Insight: Bias Undermines Labor Market Efficiency

Economically, built-in discrimination in algorithmic decision-making distorts efficiency in labor markets. Bertrand and Mullainathan (2004) showed that résumés with Black-sounding names were called back substantially less often compared to résumés with white-sounding names, even if qualifications were the same. If algorithmic systems replicate such biases, qualified individuals will be excluded on a systematic basis, compromising allocative efficiency as well as causing wage stagnation and underemployment for minorities (Mengel, 2024; Jibuti, 2024). Not only do such inefficiencies hurt individuals, but the productivity and flexibility in aggregate of the labor force also decrease.

Step 7: Identify Conflicts Between Disciplinary Insights

Evident tensions exist between the disciplinary knowledge of technology, psychology, and economics when it comes to tackling algorithmic discrimination. From a tech angle, most computer scientists prefer "fairness through unawareness," withholding protected features such as race or gender from data to maintain fairness. Psychologists, on the other hand, point to how bias still bleeds through by means of proxy variables like names, addresses, or ZIP codes that strongly align with demographic identity (Bertrand & Mullainathan, 2004; Emerson et al., 2024). This leaves omissions of overt variables as insufficient in eliminating discriminative outcomes.

Economists and psychologists also differ in their priorities. Economists tend to emphasize efficiency and cost-effectiveness, using algorithms that speed up recruitment and augment productivity (Jibuti, 2024; Mengel, 2024). Psychologists, on their part, warn that such efficiency is done at the expense of fairness and human decency, particularly if it is likely to increase stereotype threat or marginalization (Steele, 1997; Emerson et al., 2024).

In addition, computer scientists and economists tend to have differing views on regulation. Economists tend to favor state intervention to ensure competitive fairness in markets, but technologists tend to oppose such regulations on the grounds of innovation costs and regulatory lag concerns (Finocchiaro et al., 2021; Kroll, 2015). Such unresolved conflicts make uniform policy or tech-based response more challenging.

Step 8: Create or Discover Common Ground

In order to settle disciplinary appeals, this project builds on a common conceptual architecture for algorithmic fairness, defined as: *A set of technical, psychological, and economic criteria that ensures AI hiring systems neither reproduce nor exacerbate existing labor inequalities and are perceived as legitimate by users.*

This definition weaves together multiple disciplinary concerns. From both technical and economic considerations, fairness in algorithms is about maximizing performance while reducing bias in data inputs, decision rules, and results (Raji & Buolamwini, 2019; Finocchiaro et al., 2021). From a psychological consideration, there is a focus on perceived fairness, trustworthiness, and identity safety for job applicants, particularly historically underrepresented group members (Steele, 1997; Emerson et al., 2024). Finally, from an economic perspective, there is an appreciation for how recruitment algorithms can drive macro-level labor market inequities (Bertrand & Mullainathan, 2004; Mengel, 2024; Jibuti, 2024).

The approach does not overlook disciplinary objectives but instead synthesizes them into an integrated construct that facilitates empirical measurement as well as normative criticism. The approach allows for alignment with stakeholder values but anchors fairness in both procedural seriousness as well as subjective legitimacy (Binns, 2018; Kroll, 2015). This common ground is used to offer some basis for interdisciplinarity in future analysis.

Step 9: Integrate Insights Through Comprehensive Understanding

An inter-disciplinary approach drawing from various fields offers an advanced model for equitable hiring through algorithms. This is an integration of views from both technology, psychology, economics, and sociology to promote fairness and inclusivity in hiring practices.

1. ***Bias Auditing Frameworks (Tech)***

Technological solutions such as routine third-party audits play an indispensable role in detecting and mitigating bias in AI systems. As an example, IBM's AI Fairness 360 is equipped with tools that measure fairness metrics such as demographic parity and equal opportunity to help identify and confront discrimination patterns before AI systems are implemented (Raji & Buolamwini, 2022). This process assists in preventing algorithms from

reinforcing historical imbalances, making audits an integral part of responsible algorithmic design.

2. User-Centric Design (Psychology)

Psychological considerations highlight user experience as central to preventing dehumanizing impacts from algorithms in employment. Providing features to offer feedback to applicants, as well as having an appeals process in place, can help to minimize alienation and exasperation (Emerson et al., 2024). In addition, affective computing techniques that measure emotional indicators can recognize expressions of psychological distress in users to ensure that harm is avoided in a timely manner while enhancing the hiring experience (Steele, 1997).

3. Regulatory and Market Incentives (Economics)

From an economic perspective, regulations such as the proposed U.S. Algorithmic Accountability Act can make companies responsible for making their algorithmic decision-making processes transparent to ensure that their hiring processes are fair (Kroll, 2015). Market incentives, for instance, in terms of tax relief or Environmental, Social, and Governance (ESG) labels, could also incentivize companies to incorporate fair hiring that is valued by society at large and that is diverse in composition (Binns, 2018).

4. Intersectional Datasets (All Disciplines)

Lastly, by incorporating intersectionality in data collection as well as in designing algorithms, training data sets prove to encompass the multidimensionality of human identity. Since this sociological approach involves data collection that is heterogeneous in reflecting on the multidisciplinary aspect of identity, homogenizing individuals in terms that prove to be unrepresentative is avoided. Through intersectionality recognition, algorithms have fewer chances to misrepresent marginalized populations, hence resulting in more equitable results (Bertrand & Mullainathan, 2004; Finocchiaro et al., 2021).

By integrating these disciplinary understandings, we can create a stronger, more equitable design for algorithmic employment, one that tackles bias, user experience, regulation, and data representation in their complexity. These tactics are essential for creating inclusive recruitment processes that treat all candidates equitably.

Step 10: Communicate and Test It

Pilot programs must be implemented in large companies like Google, IBM, and Accenture in order to measure the efficacy of this interdisciplinary approach. Such programs can monitor diversity in hiring, rates of algorithmic errors, recruiter satisfaction, and speed over time. Measures must be compared to historical performance to ascertain if the infusion of ethics, psychology, and economics results in quantifiable improvement (Pedulla et al., 2022; Raji & Buolamwuni, 2019). Respected academic research institutions like MIT's Media Lab or Stanford's Institute for Human-Centered Artificial Intelligence might collaborate with these companies to perform serious, longitudinal studies (Mengel, 2024; Emerson et al., 2024).

Interdisciplinary communication calls for adapting insights for an array of stakeholders. Technologists need to have certain design constraints in place, for example, not using ZIP codes as surrogates for race, that can recreate redlining using machine learning models (Bertrand & Mullainathan, 2004; Finocchiaro et al., 2021). Economists need to have cost-benefit estimates presented to them to confirm the efficiency and productivity of fairness interventions (Jibuti, 2024). Psychologists and HR professionals, in turn, need to be trained on mitigating bias systems and offering human-focused feedback cycles (Steele, 1997; Emerson et al., 2024). This testing process needs to conform to the standard of actionable auditing, empirically measuring if fairness mechanisms realize real-world gains (Raji & Buolamwini, 2022).

Conclusion

Algorithmic bias in computerized recruitment systems is not just an issue in terms of design; it is not just an engineering problem. Rather, it is an interdisciplinary challenge located at the intersection between technology, psychology, and economics. This paper showed that an interdisciplinary approach, predicated on Repko and Szostak's (2016) typology, makes possible a more robust understanding of both why such bias occurs as well as with what effects. In addition, it outlined an integrative approach to AI recruitment systems that are both efficient and lawful as well as psychologically compassionate and economically equitable. Only by unifying multiple perspectives in this way can we navigate towards fair employment in the digital era.

References

Bertrand, M., & Mullainathan, S. (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American economic review*, *94*(4), 991-1013.

Binns, R. (2018, January). Fairness in machine learning: Lessons from political philosophy. In *Conference on fairness, accountability and transparency* (pp. 149-159). PMLR.

Binns, R. (2024). If the Difference Principle Won't Make a Real Difference in Algorithmic Fairness, What Will? Response to 'Rawlsian Algorithmic Fairness and a Missing Aggregation Property of the Difference Principle'. *Philosophy & Technology*, *37*(4), 119.

Emerson, K. T., Taylor, V. J., Stevens, S. M., Logel, C., & Murphy, M. C. (2024). Stereotype and Social Identity Threat in Context. *Handbook of Prejudice, Stereotyping, and Discrimination*.

Finocchiaro, J., Maio, R., Monachou, F., Patro, G. K., Raghavan, M., Stoica, A. A., & Tsirtsis, S. (2021, March). Bridging machine learning and mechanism design towards algorithmic fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 489-503).

Jibuti, D. (2024). Essays in Experimental Economics: Labor Market Discrimination.

Kroll, J. A. (2015). *Accountable algorithms* (Doctoral dissertation, Princeton University).

Mengel, F. (2024). Experimental Research on Discrimination. *Available at SSRN 4974233*.

Pedulla, D. S., Muñoz, J., Wullert, K. E., & Dias, F. A. (2022). Field experiments and job posting sources: The consequences of job database selection for estimates of racial discrimination. *Sociology of Race and Ethnicity*, *8*(1), 26-42.

Raji, I. D., & Buolamwini, J. (2019, January). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products.

In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 429-435).

Raji, I. D., & Buolamwini, J. (2022). Actionable auditing revisited: Investigating the impact of publicly naming biased performance results of commercial ai products. *Communications of the ACM*, *66*(1), 101-108.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American psychologist*, *52*(6), 613.