EDUSCHOOL JOURNAL OF SOCIOLOGY AND SOCIAL WORK (EJSSW)

(Open Access, Double-Blind Peer Reviewed Journal)

ISSN Online: ISSN Print

Algorithmic Bias and Social Stratification

Elizabeth

Assistant Professor and Head, PG Department of Social Work, KE College, Mannanam, India.

Article information

Received: 5th November 2025
Received in revised form: 6th December 2025
Accepted: 10th January 2026
Available online: 27th February 2026

Volume:1
Issue: 1
DOI: https://doi.org/10.5281/zenodo.18783908

Abstract

This paper examines how machine learning algorithms reproduce and amplify existing social inequalities across three critical domains: housing, employment, and criminal justice systems. Drawing on critical algorithm studies and digital sociology theories, this research demonstrates that algorithmic systems, far from being neutral technical tools, encode and perpetuate historical patterns of discrimination based on race, class, and gender. Through systematic analysis of algorithmic decision-making processes, this paper reveals how biased training data, opaque algorithmic architectures, and feedback loops create self-reinforcing cycles of disadvantage. The findings indicate that algorithmic systems deployed in these domains disproportionately harm marginalized communities by limiting access to housing, restricting employment opportunities, and intensifying surveillance and punishment. This research contributes to understanding the sociotechnical mechanisms through which algorithms become instruments of social stratification, and calls for increased algorithmic accountability, transparency, and justice-oriented design practices.

Keywords:- Algorithmic Bias, Social Stratification, Machine Learning, Discrimination, Housing Algorithms, Employment Algorithms, Criminal Justice Algorithms, Digital Inequality

Introduction

The rapid proliferation of algorithmic decision-making systems across social institutions represents one of the most consequential technological transformations of the 21st century. From determining who receives housing loans to selecting job candidates and assessing criminal recidivism risk, algorithms now mediate access to fundamental life opportunities. These systems are frequently promoted as objective, efficient, and fair alternatives to human decision-making, promising to eliminate the biases and inconsistencies that plague human judgment (O'Neil, 2016). However, a growing body of evidence suggests that algorithmic systems, far from transcending human prejudice, systematically reproduce and amplify existing patterns of social inequality along lines of race, class, gender, and other axes of marginalization (Benjamin, 2019; Noble, 2018).

This paper investigates the mechanisms through which machine learning algorithms function as instruments of social stratification, focusing specifically on three domains where algorithmic systems have become deeply embedded: housing, employment, and criminal justice. These domains are particularly significant because they fundamentally shape life chances and social mobility. Access to stable housing, meaningful employment, and freedom from carceral control are not merely individual concerns but structural determinants of social position that have historically been distributed unequally along racial and class lines (Massey & Denton,

1993; Western, 2006; Wilson, 1996). The introduction of algorithmic systems into these already-stratified domains raises critical questions about whether technology exacerbates or mitigates existing inequalities.

The central argument advanced in this paper is that algorithmic bias is not merely a technical problem of flawed code or incomplete data, but a sociotechnical phenomenon embedded within broader systems of power and inequality. Algorithms encode the priorities, assumptions, and biases of their creators and the societies that produce them (Eubanks, 2018). When trained on historical data that reflects past discrimination, algorithms learn to replicate discriminatory patterns. Moreover, the opacity of many algorithmic systems what Pasquale (2015) terms the "black box society" makes it difficult to identify, challenge, or remedy algorithmic harms. This combination of embedded bias, self-reinforcing feedback loops, and limited accountability creates conditions under which algorithms not only reproduce but actively amplify social stratification.

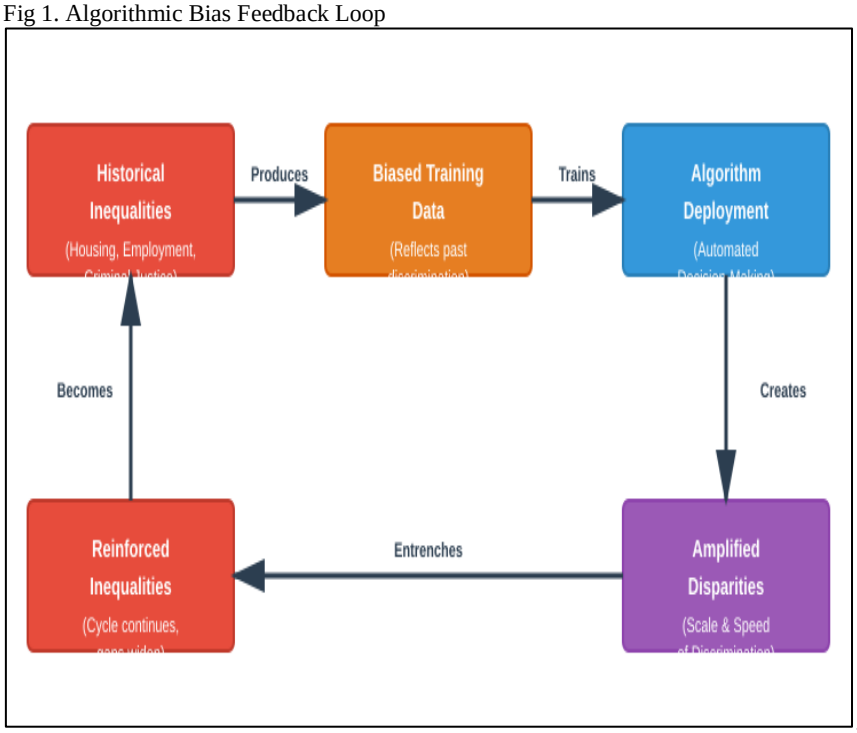
Theoretical Framework

This analysis draws on several interconnected theoretical traditions. First, critical algorithm studies provides a framework for understanding algorithms not as neutral technical artifacts but as socially constructed systems that reflect and reinforce existing power relations (Gillespie, 2014; Kitchin, 2017). This perspective emphasizes that algorithms are embedded in social contexts and shaped by human decisions about what to measure, how to measure it, and what outcomes to optimize. Second, theories of institutional discrimination and structural racism illuminate how apparently neutral rules and procedures can systematically disadvantage certain groups (Bonilla-Silva, 2006; Feagin, 2006). These theories help explain how algorithms can produce discriminatory outcomes even in the absence of explicit discriminatory intent.

Third, the concept of technological redlining (Benjamin, 2019) extends historical analyses of spatial segregation and resource allocation to the digital domain, showing how algorithmic systems create new forms of exclusion that map onto existing patterns of racial and economic inequality. Finally, theories of social reproduction (Bourdieu & Passeron, 1977) and cumulative disadvantage (DiPrete & Eirich, 2006) help explain how algorithmic systems contribute to the intergenerational transmission of inequality by limiting opportunities for marginalized groups while preserving advantages for dominant groups.

The feedback loop model (Figure 1) conceptualizes algorithmic bias as a cyclical process in which historical inequalities produce biased training data, which trains discriminatory algorithms, which generate decisions that amplify disparities, which in turn become the new historical baseline for future algorithmic systems. This self-reinforcing cycle is particularly pernicious because it operates at scale and speed, rapidly entrenching patterns that might have taken decades to develop through human decision-making alone. Understanding this cyclical process is essential for developing interventions that can disrupt rather than reinforce patterns of algorithmic discrimination.

Fig 1. Algorithmic Bias Feedback Loop



Note: Each cycle amplifies existing inequalities through automated systems

Figure 1. Algorithmic Bias Feedback Loop: The cyclical process through which algorithms reproduce and amplify social inequalities.

Algorithmic Bias in Housing Systems

Housing algorithms operate across multiple stages of the residential process, from property valuation and mortgage lending to tenant screening and eviction prediction. These systems have profound implications for residential segregation and wealth accumulation, particularly given the centrality of homeownership to wealth building in the United States (Shapiro, 2017). Algorithmic mortgage lending systems, while ostensibly designed to expand access to credit, have been shown to charge higher interest rates to minority borrowers even when controlling for creditworthiness (Bartlett et al., 2022). This digital redlining mirrors historical practices of racial exclusion while operating through seemingly objective mathematical models.

Tenant screening algorithms exemplify how algorithmic systems can exclude marginalized populations from housing opportunities. These systems aggregate data from eviction records, credit reports, and criminal background checks to generate risk scores that landlords use to evaluate prospective tenants (Christin, 2017). However, this approach systematically disadvantages populations who face higher rates of eviction and criminalization due to structural inequalities. Low-income tenants, particularly Black and Latinx renters, are more likely to have negative rental histories not because they are inherently unreliable but because they face housing precarity resulting from employment instability, inadequate social safety nets, and discriminatory practices (Desmond, 2016). Algorithmic tenant screening codifies these structural disadvantages as individual risk factors, creating barriers to housing stability that perpetuate cycles of poverty and displacement.

Property valuation algorithms similarly encode racial biases. Automated valuation models used in real estate and lending decisions have been found to systematically undervalue properties in predominantly Black and Latinx neighborhoods, even when controlling for property characteristics (Howell & Korver-Glenn, 2018). This algorithmic devaluation has material consequences for wealth accumulation, as homeowners in these neighborhoods face lower returns on their property investments and reduced ability to leverage home equity. The result is a technological reproduction of the appraisal gap that has historically depressed property values in minority communities, demonstrating how algorithms can serve as mechanisms for maintaining racialized wealth hierarchies.

Algorithmic Bias in Employment Systems

Employment algorithms have proliferated across the hiring process, from resume screening and candidate ranking to personality assessment and performance prediction. These systems are marketed as tools for increasing efficiency and reducing human bias in hiring decisions (Raghavan et al., 2020). However, research reveals that employment algorithms often reproduce and amplify discriminatory hiring patterns. Resume screening algorithms have been shown to systematically disadvantage candidates from underrepresented backgrounds, penalizing applicants who attended historically Black colleges and universities or who have gaps in employment history associated with caregiving responsibilities that disproportionately fall to women (Ajunwa, 2020).

The problem is compounded by the widespread use of training data derived from past hiring decisions. When algorithms are trained to identify candidates who resemble historically successful employees, they inevitably learn to prefer candidates who match the demographic profile of existing workers. In industries and organizations where leadership is predominantly white and male, this approach produces algorithms that systematically favor white male candidates regardless of actual qualifications (Dastin, 2018). The algorithmic reproduction of workplace homogeneity not only limits opportunities for women and minorities but also perpetuates organizational cultures that fail to benefit from diverse perspectives and experiences.

Personality assessment algorithms introduce additional concerns about discriminatory screening. These systems claim to identify candidates with desirable personality traits and cultural fit, but research suggests they may encode class-based and culturally-specific norms about appropriate behavior and communication styles (Rivera, 2015). Algorithms that prioritize traits like assertiveness or individual achievement may disadvantage candidates from cultural backgrounds that emphasize collective values or indirect communication. Moreover, the use of facial recognition and voice analysis in video interview screening raises concerns about racial and gender bias, as these technologies have been shown to perform less accurately for women and people of color (Buolamwini & Gebru, 2018). The result is a multi-layered system of algorithmic exclusion that operates throughout the employment lifecycle.

Algorithmic Bias in Criminal Justice Systems

The deployment of risk assessment algorithms in criminal justice systems represents perhaps the most consequential application of predictive analytics to social stratification. These algorithms are used at multiple

decision points in the criminal justice process, including pretrial detention, sentencing, and parole decisions (Angwin et al., 2016). Advocates argue that algorithmic risk assessment can reduce incarceration rates and eliminate human bias in judicial decision-making. However, empirical analysis reveals that these systems often exacerbate rather than mitigate racial disparities in criminal justice outcomes.

The COMPAS recidivism risk assessment algorithm, which has been adopted by courts across the United States, exemplifies the problem of algorithmic discrimination in criminal justice. ProPublica's investigation of COMPAS revealed that the algorithm produces significantly different error rates for Black and white defendants: Black defendants who did not reoffend were nearly twice as likely as white defendants to be incorrectly classified as high risk, while white defendants who did reoffend were more likely to be incorrectly classified as low risk (Angwin et al., 2016). These disparate error rates have profound implications for liberty and life chances, as risk scores influence decisions about bail, sentencing length, and parole eligibility.

The mechanisms producing these biased outcomes are multiple and interconnected. First, risk assessment algorithms are typically trained on historical criminal justice data that reflects decades of racially discriminatory policing, prosecution, and sentencing practices (Alexander, 2010). Patterns of over-policing in communities of color mean that Black individuals are more likely to have criminal records and encounters with law enforcement that serve as inputs to risk algorithms, independent of actual criminal behavior. Second, many risk assessment instruments include socioeconomic factors such as employment history, neighborhood characteristics, and social networks as predictive variables (Harcourt, 2007). While these factors may correlate with recidivism, they also reflect structural inequalities rather than individual dangerousness. The result is algorithms that effectively penalize individuals for experiencing poverty and racial segregation.

Predictive policing algorithms extend these concerns to the front end of the criminal justice system. These systems analyze historical crime data to identify areas and individuals at high risk for criminal activity, directing police resources accordingly (Ferguson, 2017). However, because historical crime data reflects patterns of police deployment rather than actual crime rates, predictive policing algorithms tend to direct officers to the same neighborhoods that have been historically over-policed, creating a feedback loop of surveillance and enforcement in minority communities. This algorithmic reinforcement of discriminatory policing patterns not only increases the likelihood of arrest and incarceration for residents of these communities but also generates new data that further entrenches the perception of these areas as high-crime, regardless of actual criminal behavior.

Table 1. Examples of Algorithmic Bias Across Domains

Domain	Algorithm Type	Bias Mechanism	Impact
Housing	Mortgage lending algorithms, tenant screening systems	Training data reflects historical redlining; proxy variables encode race/class	Higher interest rates for minorities, housing instability, wealth inequality
Employment	Resume screening, personality assessments, video interview analysis	Learns from homogeneous past hires; penalizes non-traditional backgrounds	Reduced opportunities for women and minorities, workplace homogeneity
Criminal Justice	Recidivism risk assessment, predictive policing	Trained on discriminatory policing data; socioeconomic factors as proxies	Higher false positive rates for Black defendants, concentrated surveillance

Critical Evaluation

The evidence presented demonstrates that algorithmic systems in housing, employment, and criminal justice systematically disadvantage marginalized populations. However, several important limitations and counterarguments merit consideration. First, proponents of algorithmic decision-making argue that these systems, despite their flaws, may still produce more equitable outcomes than human decision-makers who operate with conscious or unconscious biases (Kleinberg et al., 2018). This argument suggests that the appropriate comparison is not between algorithmic systems and a hypothetical perfectly fair system, but between algorithms and the actually existing human decision-making processes they replace. While this perspective has merit, it risks naturalizing algorithmic bias by treating human bias as inevitable rather than addressing the social conditions that produce both human and algorithmic discrimination.

Second, technical researchers have proposed various methods for detecting and mitigating algorithmic bias, including fairness constraints in algorithm design, adversarial debiasing techniques, and post-hoc auditing procedures (Barocas & Selbst, 2016). These technical interventions can reduce some forms of bias and represent

important advances in algorithmic accountability. However, technical fixes alone are insufficient to address algorithmic discrimination because they do not challenge the underlying social inequalities that produce biased training data or the power relations that shape which algorithmic systems are developed and deployed. Moreover, different definitions of fairness can be mathematically incompatible, meaning that optimizing for one fairness criterion may necessarily involve trade-offs with others (Chouldechova, 2017).

Third, the analysis presented here focuses primarily on discrimination based on race, class, and gender, but algorithmic systems can reproduce bias along multiple axes of inequality simultaneously. Intersectional approaches reveal that individuals who occupy multiple marginalized identities face compounded forms of algorithmic discrimination that cannot be understood by analyzing single categories in isolation (Crenshaw, 1991). Additionally, this paper has emphasized bias in algorithmic outputs, but equally important questions concern the distributional effects of algorithmic automation: who benefits from efficiency gains, whose labor is displaced, and who bears the costs of algorithmic errors. A complete analysis of algorithmic inequality must attend to these questions of political economy alongside concerns about discriminatory decision-making.

Implications for Policy and Practice

The findings presented in this paper carry significant implications for policy, organizational practice, and future research. First, there is an urgent need for regulatory frameworks that mandate transparency and accountability in algorithmic systems that affect access to fundamental opportunities. This includes requirements for algorithmic impact assessments prior to deployment, ongoing auditing for discriminatory outcomes, and mechanisms for meaningful appeal and redress when individuals are harmed by algorithmic decisions (Kaminski & Malgieri, 2020). Such frameworks must be accompanied by enforcement mechanisms with sufficient power to impose meaningful consequences for algorithmic discrimination.

Second, organizations deploying algorithmic systems must move beyond narrow technical definitions of fairness to embrace justice-oriented design practices that center the needs and experiences of marginalized communities (Costanza-Chock, 2020). This includes participatory design processes that involve affected communities in decisions about whether and how to deploy algorithmic systems, as well as ongoing monitoring for disparate impacts. Organizations should also consider whether algorithmic automation is appropriate for high-stakes decisions that fundamentally affect human dignity and life chances, or whether some domains require human judgment with appropriate safeguards against bias.

Third, addressing algorithmic bias requires confronting the underlying social inequalities that algorithms encode. Technical interventions to reduce algorithmic discrimination will have limited effect if they do not address the structural conditions that produce disparities in employment, housing, and criminal justice involvement. This means that efforts to create fairer algorithms must be accompanied by broader efforts to address economic inequality, residential segregation, and discriminatory policing. Ultimately, truly equitable algorithmic systems may require transformation of the social institutions in which they are embedded rather than mere technical optimization of existing systems.

Conclusion

This paper has examined how machine learning algorithms reproduce and amplify social inequalities in housing, employment, and criminal justice systems. The analysis reveals that algorithmic bias is not a mere technical glitch but a systematic pattern embedded in the design, training, and deployment of these systems. Algorithms trained on historical data inevitably learn historical patterns of discrimination, while opacity and lack of accountability make it difficult to identify and remedy algorithmic harms. The feedback loop model demonstrates how these systems create self-reinforcing cycles of disadvantage that rapidly entrench existing inequalities.

The proliferation of algorithmic decision-making in domains that fundamentally shape life chances represents a critical juncture for social stratification research and social justice advocacy. These systems have the potential to either exacerbate existing inequalities or, if properly designed and regulated, to mitigate some forms of discrimination. However, realizing the latter possibility requires moving beyond technocratic fantasies of algorithmic objectivity to confront the social and political choices embedded in algorithmic systems. It requires transparency about how these systems work, accountability for their impacts, and participation by affected communities in decisions about their deployment.

Future research should continue to investigate the mechanisms through which algorithms produce and amplify inequality across different domains and populations. Longitudinal studies examining the cumulative effects of algorithmic discrimination over individual life courses and across generations would illuminate how these systems shape long-term trajectories of stratification. Additionally, comparative research examining algorithmic governance in different national and institutional contexts could identify conditions under which

algorithmic systems can be deployed more equitably. Ultimately, ensuring that algorithmic systems serve rather than undermine social justice requires sustained critical attention to the relationship between technology and inequality, and a commitment to designing technological futures that prioritize equity and human flourishing.

References

Ajunwa, Ifeoma. 2020. "The Paradox of Automation as Anti-Bias Intervention." Cardozo Law Review 41: 1671–1742.

Alexander, Michelle. 2010. The New Jim Crow: Mass Incarceration in the Age of Colorblindness. New York: The New Press.

Angwin, Julia, Jeff Larson, Surya Mattu, and Lauren Kirchner. 2016. "Machine Bias: There's Software Used across the Country to Predict Future Criminals. And It's Biased against Blacks." ProPublica, May 23, 2016.

Barocas, Solon, and Andrew D. Selbst. 2016. "Big Data's Disparate Impact." California Law Review 104: 671–732.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. 2022. "Consumer-Lending Discrimination in the FinTech Era." Journal of Financial Economics 143 (1): 30–56. https://doi.org/10.1016/j.jfineco.2021.05.047

Benjamin, R. 2019. Race after Technology: Abolitionist Tools for the New Jim Code. Polity Press.

Bonilla-Silva, E. 2006. Racism without Racists: Color-Blind Racism and the Persistence of Racial Inequality in the United States. Rowman & Littlefield.

Bourdieu, P., and J. C. Passeron. 1977. Reproduction in Education, Society and Culture. Sage.

Buolamwini, J., and T. Gebru. 2018. "Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification." Proceedings of Machine Learning Research 81: 1–15.

Chouldechova, A. 2017. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments." Big Data 5 (2): 153–163.

Christin, A. 2017. "Algorithms in Practice: Comparing Web Journalism and Criminal Justice." Big Data & Society 4 (2): 1–14.

Costanza-Chock, S. 2020. Design Justice: Community-Led Practices to Build the Worlds We Need. MIT Press.

Crenshaw, K. 1991. "Mapping the Margins: Intersectionality, Identity Politics, and Violence against Women of Color." Stanford Law Review 43 (6): 1241–1299.

Dastin, J. 2018. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women." Reuters, October 10, 2018.

Desmond, M. 2016. Evicted: Poverty and Profit in the American City. Crown Publishers.

DiPrete, T. A., and G. M. Eirich. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." Annual Review of Sociology 32: 271–297.

Eubanks, V. 2018. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.

Feagin, J. R. 2006. Systemic Racism: A Theory of Oppression. Routledge.

Ferguson, A. G. 2017. The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement. NYU Press.

Gillespie, T. 2014. "The Relevance of Algorithms." In Media Technologies: Essays on Communication, Materiality, and Society, edited by T. Gillespie, P. J. Boczkowski, and K. A. Foot, 167–194. MIT Press.

Harcourt, B. E. 2007. Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age. University of Chicago Press.

Howell, J., and E. Korver-Glenn. 2018. "Neighborhoods, Race, and the Twenty-First-Century Housing Appraisal Industry." Sociology of Race and Ethnicity 4 (4): 473–490.

Kaminski, M. E., and G. Malgieri. 2020. "Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations." International Data Privacy Law 11 (2): 125–144.

Kitchin, R. 2017. "Thinking Critically about and Researching Algorithms." Information, Communication & Society 20 (1): 14–29.

Kleinberg, J., J. Ludwig, S. Mullainathan, and C. R. Sunstein. 2018. "Discrimination in the Age of Algorithms." Journal of Legal Analysis 10: 113–174.

Massey, D. S., and N. A. Denton. 1993. American Apartheid: Segregation and the Making of the Underclass. Harvard University Press.

Noble, S. U. 2018. Algorithms of Oppression: How Search Engines Reinforce Racism. NYU Press.

O'Neil, C. 2016. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishers.

Pasquale, F. 2015. The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.

Raghavan, M., S. Barocas, J. Kleinberg, and K. Levy. 2020. "Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency: 469–481.

Rivera, L. A. 2015. Pedigree: How Elite Students Get Elite Jobs. Princeton University Press.

Shapiro, T. M. 2017. Toxic Inequality: How America's Wealth Gap Destroys Mobility, Deepens the Racial Divide, and Threatens Our Future. Basic Books.

Western, B. 2006. Punishment and Inequality in America. Russell Sage Foundation.

Wilson, W. J. 1996. When Work Disappears: The World of the New Urban Poor. Knopf.