Published by OTS Canadian Journal
Available at: *https://journal.canadian-ots.ca*

ISSN: 2817-1756
Vol-4, Issue-7; July, 2025

Research Article
Open Access

Bias in AI Decision-Making: Ethical Implications for Hiring and Healthcare Algorithms

Rafael Gutierrez

School of Business, Australia

Article History
Received: June, 05,2025
Revised: June, 22,2025
Accepted: July 08, 2025

Doi: https://doi.org/10.58840/ke81n154

Abstract:

Artificial Intelligence (AI) has revolutionized decision-making in critical sectors such as hiring and healthcare. While AI promises enhanced efficiency, scalability, and data-driven objectivity, emerging evidence shows that these systems can inherit and even amplify societal biases, leading to unfair outcomes. This article explores the ethical implications of AI-driven bias in hiring and healthcare algorithms. It examines the sources of bias, the consequences of algorithmic discrimination, and the regulatory and ethical frameworks needed to ensure transparency, accountability, and equity. Case studies illustrate real-world impacts, and policy recommendations are provided for developers, organizations, and regulators seeking to create responsible AI systems.

Keywords: *Algorithmic Bias***,** *Ethical AI***,** *Fairness In Machine Learning***,** *AI In Hiring Practices***,** *Healthcare Algorithms***,** *Discrimination In Automated Decision-Making***.**

1. Introduction

Artificial intelligence (AI) algorithms are increasingly deployed to support high-stakes decision-making processes that significantly impact individuals' lives. From screening job applicants to diagnosing patients, these systems promise enhanced efficiency, scalability, and data-driven precision. However, as AI becomes more deeply integrated into organizational and institutional workflows, concerns are intensifying about its capacity to replicate—and even amplify—existing human biases (Obermeyer, Powers, Vogeli, & Mullainathan, 2019).

Algorithmic bias typically stems from historical inequalities embedded in training data, flawed development methodologies, or unexamined institutional practices. When such biases are encoded into AI systems, they can lead to unfair or discriminatory outcomes, particularly in critical areas such as hiring and healthcare. In recruitment, biased AI tools may favor certain demographics over others, undermining diversity and perpetuating exclusion. In healthcare, algorithmic disparities can result in unequal diagnosis or treatment recommendations, adversely affecting patient safety and trust (Raji & Buolamwini, 2019).

The urgency of addressing algorithmic bias is underscored by high-profile failures and growing societal reliance on automation. For instance, an AI resume screening tool developed by a major tech company was found to systematically downgrade applications from women due to biased historical data reflecting male-dominated hiring trends. Similarly, a medical algorithm used in U.S. hospitals was discovered to underestimate the health needs of Black patients compared to white patients with similar clinical profiles, raising serious ethical and equity concerns (Binns, 2018). These examples illustrate that algorithmic bias is not merely a technical flaw but a social and institutional problem requiring interdisciplinary solutions. Ethical AI design demands the collaboration of computer scientists, ethicists, legal scholars, policymakers, and impacted communities. Fairness must be embedded not only in data selection and model training but also in how algorithms are implemented, evaluated, and governed (Ferrara, 2024).

Furthermore, regulatory landscapes are evolving to keep pace with these challenges. The European Union's AI Act and various national frameworks are beginning to set legal boundaries around high-risk AI applications, with a strong emphasis on human oversight, transparency, and accountability. Meanwhile, organizations are adopting responsible AI principles to preempt reputational, legal, and operational risks (Shan, 2025).

This article examines the underlying causes and real-world consequences of algorithmic bias, with a focus on employment and healthcare as case studies. It proposes an ethical framework for mitigating bias, built on the pillars of fairness, accountability, transparency, and inclusivity. Ultimately, the goal is to ensure that AI not only augments decision-making but also upholds the values of justice and human dignity in the digital age.

2. Sources of Bias in AI Systems

Bias in AI systems can be introduced at multiple levels, often reinforcing and compounding one another. Understanding these layers is essential to mitigating discriminatory outcomes in sensitive sectors like hiring and healthcare (Nouri et al., 2025).

2.1 Data Bias

AI systems are inherently reliant on the quality and representativeness of their training data. When data reflect historical inequalities, systemic discrimination is likely to be reproduced in algorithmic predictions. For instance, if a recruitment dataset primarily includes successful male candidates, the model may learn to associate masculinity with employability. This issue was exemplified in Amazon's now-scrapped AI recruiting tool, which downgraded résumés containing the word "women’s" (e.g., “women’s chess club captain”), because past data reflected male-dominated hiring (Macru, 2025).

In healthcare, data bias can have even graver consequences. Many machine learning models are trained on datasets that underrepresent racial and ethnic minorities, leading to reduced diagnostic accuracy for these populations. A well-known example is a widely-used algorithm in U.S. hospitals that systematically underestimated the medical needs of Black patients, because it used historical healthcare expenditure as a proxy for illness severity—a metric influenced by unequal access to care (Rajkomar et al., 2024).

2.2 Algorithmic Bias

Bias can also be introduced through the algorithm's internal logic. Design decisions such as feature weighting, model architecture, and training objectives can reflect implicit assumptions or reinforce societal biases (Chen et al., 2021). For example, using zip codes or educational background as features in a predictive hiring model can serve as proxies for race or socioeconomic status, unintentionally disadvantaging certain groups. Algorithms that prioritize accuracy without fairness constraints may also amplify majority class dominance, marginalizing minority groups whose patterns deviate from the norm (Naveuler, 2025). Moreover, technical oversights—like insufficient fairness auditing, lack of adversarial testing, or uncalibrated thresholds—can entrench discriminatory outcomes even in otherwise well-designed models.

2.3 Human and Institutional Bias

Human bias is deeply embedded in institutional practices, and it inevitably influences AI development and deployment. This includes the assumptions of developers, who may unconsciously embed their own worldview into the design. For instance, a lack of diversity in development teams can lead to blind spots, where the needs and contexts of marginalized communities are overlooked (Mehrabi et al., 2019).

Organizational priorities also shape algorithmic outcomes. When companies prioritize efficiency, profit, or speed over fairness and inclusivity, algorithmic systems are likely to reflect those skewed incentives. Furthermore, ethical oversight is often deprioritized in fast-paced tech environments, contributing to a culture where biased algorithms can go unchecked until harm is done (Sadiq, Kanabi, Tahir, & Nader, 2025).

3. Ethical Implications in Hiring

The growing reliance on AI in hiring processes promises to increase efficiency and reduce human biases. Tools now scan résumés, evaluate speech and facial expressions in video interviews, and even predict cultural fit. However, these technologies often fail to deliver fair outcomes and instead introduce new forms of discrimination (Raghavan, Barocas, Kleinberg, & Levy, 2019). A key case that exemplifies these issues is Amazon's AI recruiting tool, which was scrapped after it was revealed to penalize applications that included words like "women's chess club" or came from all-women colleges. The model, trained on résumés submitted over a decade—mostly from male candidates—learned to downgrade terms statistically associated with female applicants. This case highlighted a fundamental problem: machine learning models often learn from biased data and reinforce those biases rather than correct them (Bayz, 2024).

Another troubling development is the use of AI-powered video analysis tools, such as those developed by companies like HireVue and Pymetrics. These tools assess candidates based on micro-expressions, eye contact, vocal pitch, and word choice—metrics that are not only scientifically questionable but also culturally and neurologically biased (Scassa, 2024). Neurodivergent individuals, such as those with autism spectrum disorder, may struggle to perform according to the expected norms, and non-native speakers may be penalized for their accents or speech patterns. Critics have argued that these systems favor conformity over diversity, undermining efforts toward inclusive hiring (Chen, Johansson, & Sontag, 2020).

Ethical Concerns

- **Lack of Transparency**: Many hiring algorithms function as "black boxes." Candidates receive no explanation for rejections, and employers may not fully understand how decisions are made. This opacity hinders accountability and raises questions about procedural fairness.
- **Disparate Impact**: Even if a system is not explicitly programmed to discriminate, it may still **disproportionately affect members of protected groups**. For example, predictive models that associate leadership traits with assertive communication styles may penalize individuals from cultures where modesty or indirectness is valued.
- **Loss of Human Dignity and Autonomy**: Automating hiring decisions can strip the process of human judgment, reducing individuals to data points. This mechanization not only undermines the candidate's dignity but also makes it difficult to address context-specific nuances such as career gaps due to caregiving, migration, or disability.

To address these concerns, scholars and ethicists emphasize the importance of algorithmic accountability, human oversight, and inclusive design in hiring AI. Employers must ensure that these tools are audited for bias, that candidates are given avenues for recourse, and that AI complements rather than replaces human judgment in recruitment.

4. Ethical Implications in Healthcare

Artificial Intelligence (AI) is rapidly transforming healthcare by enhancing diagnostic accuracy, streamlining administrative processes, and enabling predictive analytics. From triaging patients in emergency rooms to predicting disease outbreaks, AI-driven tools are becoming integral to decision-making. However, alongside these benefits lie critical ethical challenges—especially when algorithms inherit or exacerbate pre-existing inequalities in healthcare systems (Shukur, 2023).

A well-known 2019 study published in Science revealed that a healthcare algorithm widely used in the United States systematically underestimated the health needs of Black patients. The algorithm used healthcare expenditure as a proxy for health needs, assuming that higher spending correlated with greater need. However, due to structural racism and unequal access, Black patients often spend less—not because they are healthier, but because they face barriers to care. Consequently, the algorithm incorrectly prioritized White patients over Black patients for advanced care programs, despite comparable or greater need in the latter group (Kleinberg, Ludwig, Mullainathan, & Sunstein, 2018).

Similarly, AI-based dermatology applications have been criticized for poor diagnostic accuracy on darker skin tones. A 2021 review published in The Lancet Digital Health found that most dermatological datasets were composed predominantly of light-skinned individuals, leading to algorithmic blind spots in detecting conditions like melanoma or eczema in people of color. This highlights a widespread issue in AI development: training data that fails to represent population diversity (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016).

Ethical Concerns

- **Equity of Care**: When algorithms are biased, they can produce skewed treatment recommendations, misdiagnose patients, or fail to identify high-risk individuals—particularly in marginalized or underserved communities. These disparities not only harm individuals but also entrench systemic health inequalities.
- **Accountability and Liability**: AI complicates traditional frameworks of medical accountability. When a misdiagnosis or inappropriate treatment results from an algorithm's recommendation, determining who is responsible—the healthcare provider, software developer, or institution—becomes difficult. This ambiguity poses legal and ethical challenges, particularly in malpractice and liability cases.
- **Informed Consent and Transparency**: Patients are often unaware of the extent to which AI tools inform their care. This raises concerns about autonomy, trust, and consent. Ethical healthcare demands that patients be fully informed not only of treatment options but also the role that automated systems play in their diagnosis or prognosis.

To mitigate these risks, algorithmic audits, bias testing, and regulatory oversight are essential. Additionally, AI models must be trained on diverse and representative datasets, and healthcare professionals should receive training on AI ethics to ensure equitable and responsible use.

5. Regulatory and Ethical Frameworks

As the adoption of AI systems accelerates, regulatory bodies and institutions worldwide are grappling with the challenge of governing algorithmic decision-making—particularly in high-stakes sectors like employment and healthcare, where the cost of error or bias can be profound. A patchwork of emerging laws, guidelines, and ethical frameworks seeks to ensure that AI technologies are deployed in a manner that is equitable, transparent, and accountable (Abdalla, Younis, & Azeez, 2023).

5.1 Regulatory Developments

The European Union's Artificial Intelligence Act, first proposed in 2021 and nearing implementation, is one of the most comprehensive regulatory efforts to date. It classifies AI systems by levels of risk—from minimal to unacceptable—and imposes rigorous requirements on high-risk applications such as biometric identification, recruitment algorithms, and clinical decision support tools (Ahmad & Balisany, 2023). Key mandates include:

- **Data quality and documentation standards**
- **Human oversight requirements**
- **Algorithmic transparency and explainability**
- **Post-market monitoring obligations**

Organizations that fail to comply may face fines of up to €30 million or 6% of global turnover, underscoring the EU's strong commitment to ethical AI. In the United States, the Algorithmic Accountability Act—reintroduced in 2022—seeks to mandate impact assessments for AI systems used in housing, employment, healthcare, and credit. These assessments would examine bias risks, data integrity, and discriminatory impacts, and require organizations to explain how they mitigate potential harms. While the Act has not yet passed, it reflects growing bipartisan interest in tech regulation as a civil rights issue (Buolamwini & Gebru, 2018).

5.2 Ethical Principles and Institutional Guidance

Several non-binding frameworks from academic, industry, and multilateral organizations also help guide AI ethics. These include:

- The **OECD AI Principles**, adopted by over 40 countries, which promote inclusive growth, human-centered values, and robustness**.**
- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, which advocates for transparency, algorithmic fairness, and user agency.

Core ethical principles shared across these frameworks include:

- **Fairness**: Algorithms should not perpetuate or amplify discrimination. This involves using representative training data, de-biasing techniques, and regular audits.
- **Transparency**: AI decision-making processes must be interpretable. Stakeholders should understand how inputs lead to outcomes, particularly when decisions affect rights or access to services.

- **Accountability**: Clear lines of responsibility should be established, ensuring that developers, deployers, and regulators can be held accountable for outcomes.
- **Human Oversight**: Even in highly automated systems, critical decisions should include meaningful human review, especially where fundamental rights are at stake.

By embedding these principles into both technical design and organizational policy, stakeholders can foster responsible AI that aligns with democratic values and public trust.

6. Recommendations for Mitigating AI Bias

Addressing bias in AI systems—particularly in sensitive domains such as hiring and healthcare—requires proactive, multidisciplinary approaches. While technology itself is neutral, the contexts in which it is developed and deployed are not. Mitigating bias involves not only refining technical systems but also reshaping organizational practices and cultural norms (Ormzyar, 2023).

6.1 Inclusive and Representative Data

Bias often originates in training data that is incomplete, imbalanced, or historically prejudiced. To counter this, organizations must conduct comprehensive audits of datasets before model training. This includes identifying underrepresented groups, assessing demographic distributions, and applying techniques such as oversampling, synthetic data generation, or reweighting to achieve more equitable representation. For example, healthcare models should include data from diverse racial, gender, and socioeconomic groups to ensure accurate diagnostics for all populations (Faeq, 2025). Additionally, ongoing data updates are critical. AI systems must evolve with society, and outdated or static datasets risk entrenching past injustices.

6.2 Bias Audits and Impact Assessments

Independent bias audits and impact assessments are essential for identifying and correcting discriminatory outcomes. These assessments should be transparent, standardized, and repeatable, allowing organizations to benchmark progress over time. Open-source toolkits like IBM's AI Fairness 360, Microsoft's Fairlearn, and Google's What-If Tool provide technical means to detect disparities in outcomes across demographic groups (Chouldechova & Roth, 2020). Such audits should be required not just before deployment, but throughout the lifecycle of the AI system. Regulatory bodies could mandate third-party certification, similar to financial or environmental audits, to ensure credibility (Weller, 2019).

6.3 Human-Centered Design

AI should serve as a complement to human expertise, not a replacement. Human-centered design emphasizes user experience, contextual understanding, and ethical foresight. In hiring, this means giving recruiters final decision-making authority while using AI to surface candidates fairly. In healthcare, AI should inform clinicians—not dictate treatments—preserving professional autonomy and patient rights (Kakai, 2023). Embedding ethical reflexivity into product development cycles, including ethical checklists and red-teaming exercises, can help mitigate unintended consequences.

6.4 Transparency and Communication

Trust in AI hinges on clarity and explainability. Users and affected individuals should be able to understand how decisions are made and on what basis. This includes explaining model logic, input features, and risk factors in non-technical language. Tools like LIME (Local Interpretable Model-Agnostic Explanations) or SHAP (SHapley Additive exPlanations) can provide insight into model decisions (Ali, 2024). Transparency also includes disclosure policies, where organizations inform users when AI is being used, especially in high-impact areas like employment or diagnosis.

6.5 Interdisciplinary Collaboration

Ethical AI development requires input beyond computer science. Collaboration with ethicists, sociologists, legal scholars, clinicians, HR professionals, and members of affected communities ensures that multiple perspectives are incorporated. This promotes not only technical robustness, but also social legitimacy (Raji et al., 2020). Establishing ethics review boards, AI governance committees, and community advisory groups can institutionalize this collaboration and make ethical oversight a standard practice.

7. Conclusion

AI holds transformative potential in reshaping hiring and healthcare practices, offering opportunities for efficiency, personalization, and accessibility. However, this promise is tempered by the risk of embedding and amplifying societal biases when these technologies are developed or deployed without sufficient safeguards. As demonstrated, biases in data, algorithm design, and institutional priorities can lead to unjust outcomes, particularly for already marginalized groups.
Ethical AI design requires more than algorithmic fairness—it demands a holistic commitment to human values, transparency, and accountability. This includes rigorous oversight, inclusive data practices, interdisciplinary collaboration, and ongoing evaluation. Regulatory frameworks such as the EU's AI Act and the U.S. Algorithmic Accountability Act mark important steps, but industry stakeholders must also take responsibility. Ultimately, ensuring fairness in AI is not only a technical challenge but a moral imperative. As algorithmic systems become more embedded in high-stakes decision-making, the future of equitable governance will depend on our collective ability to prioritize justice, protect human dignity, and foster trust in digital innovation.

References

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453.

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. AAAI/ACM Conference on AI Ethics and Society.

Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability and Transparency.

U.S. Congress. (2022). Algorithmic Accountability Act of 2022.

European Commission. (2021). Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence.

Ferrara, E. (2024). *Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies*. *Sci*, 6(1), 3.

Shan, D. K. (2025). Legal Foundations in Business: Key Factors Influencing Modern Practices. OTS Canadian Journal, 4(1), 8-16.

Nouri, S., et al. (2025). *Bias in medical AI: Implications for clinical decision-making*. *Journal of Medical Ethics*, 51(2), 147–158.

Macru, K. (2025). Customer Preferences and Advertising Trends: A Symbiotic Connection. OTS Canadian Journal, 4(1), 35-41.

Rajkomar, A., et al. (2024). *Bias recognition and mitigation strategies in healthcare AI*. *NPJ Digital Medicine*, 8, 69.

Chen, R. J., et al. (2021). *Algorithm Fairness in Medicine and Healthcare*. *arXiv:2110.00603*.

Naveuler, K. (2025). What Drives Purchases? A Closer Look at Shopping Determinants. OTS Canadian Journal, 4(1), 42-46.

Mehrabi, N., et al. (2019). *A Survey on Bias and Fairness in Machine Learning*. *arXiv:1908.09635*.

Sadiq, G. J., Kanabi, I. S., Tahir, S. R., & Nader, A. S. (2025). The Impact of Marketing Crises on Marketing Recovery. OTS Canadian Journal, 4(1), 47-65.

Raghavan, M., Barocas, S., Kleinberg, J., & Levy, K. (2019). *Mitigating Bias in Algorithmic Hiring: Evaluating Claims and Practices*. *arXiv:1906.09208*.

Bayz, H. Q. (2024). The Role and Impact of the International Criminal Court in Global Justice. OTS Canadian Journal, 3(5).

Scassa, T. (2024). *Ethical and bias considerations in AI-ML within pathology and medical domains*. *Clinical Informatics Review*, 12(4), 209–223.

Chen, I. Y., et al. (2020). *Why is my classifier discriminatory? NeurIPS 33*, 19589–19602.

Shukur, I. (2023). Invisalign vs Braces: Which is Right for You? A Guide from an Orthodontist: Case of Blanca Dental Care. OTS Canadian Journal, 2(4).

Kleinberg, J., Ludwig, J., Mullainathan, S., & Sunstein, C. R. (2018). *Algorithms as discrimination detectors*. *Journal of Legal Analytics, 10*, 113–174.

Mittelstadt, B. D., et al. (2016). *The ethics of algorithms: Mapping the debate*. *Big Data & Society, 3*(2), 1–21.

Abdalla, K. R., Younis, B. J., & Azeez, R. J. (2023). Chemical Value Improvement Of Cheese By Adding Algae In Sulaymaniyah District. OTS Canadian Journal, 2(4).

Obermeyer, Z., et al. (2019). *Dissecting racial bias in a healthcare algorithm*. *Science, 366*(6464), 447–453.

Ahmad, A. F., & khalid Balisany, W. M. (2023). Sustainable Tourism Management and Ecotourism as a tool to evaluate tourism's contribution to the Sustainable Development Goals and local community. OTS Canadian Journal, 2(4).

Buolamwini, J., & Gebru, T. (2018). *Gender Shades: Intersectional disparities in commercial gender classification*. *PMLR, 81*, 1–15.

Ormzyar, N. I. M. (2023). The Mediation Role of Student Engagement Between the Influence of English Language Anxiety and Academic Achievement in Higher Education. OTS Canadian Journal, 2(2).

Faeq, D.K. Narcissistic leadership, workplace bullying, turnover intention, and creative performance: a study of nurses. BMC Nurs 24, 898 (2025). https://doi.org/10.1186/s12912-025-03479-x

Chouldechova, A., & Roth, A. (2020). *A survey of fair machine learning*. *Annual Review of Statistics and Its Application, 6*, 143–163.

Weller, A. (2019). *Transparency: Motivations and challenges*. In *Explainable AI* (pp. 23–40).

Kakai, L. R. (2023). NATO; From Regional Military Force to International "Peace keeping". OTS Canadian Journal, 2(6).

Ali, A. O. (2024). Unveiling Violence and Masculinity through a Psychoanalytic Study of Sam Shepard's Curse of the Starving Class (1977), Sarah Kane'Blasted (1995) and The Body of a Woman as a Battlefield in Bosnian War (1996) by Matei Visniec. OTS Canadian Journal, 3(9), 1-42.

Raji, I. D., et al. (2020). *Closing the AI accountability gap: Framework for algorithmic auditing*. *FAT* '20, 33–44.