# Integrating Human Judgment to Address Algorithmic Bias in HR Practices in the Indonesian Context

Emeralda Ayu Kusuma
*Sekolah Tinggi Ilmu Ekonomi Indonesia Surabaya; Surabaya, Indonesia*
E-Mail: emeraldaayukusuma@stiesia.ac.id



***ABSTRACT***

*Algorithmic decision-making is increasingly adopted in variance accounted for to streamline processes like recruitment, performance evaluations, and promotions, but it raises ethical concerns about bias, fairness, and transparency. This study aims to examine the risks, fairness challenges, and the role of human judgment in these systems. Using a qualitative literature review, the research analyzes existing studies to explore the benefits and ethical implications of algorithms in human resources. The findings reveal that algorithms enhance efficiency and objectivity but can perpetuate biases from historical data, potentially leading to unfair outcomes, particularly for marginalized groups. Human judgment is critical to ensure ethical decisions, addressing nuances like cultural fit that algorithms may overlook. In contexts like Indonesia, where cultural values influence workplace dynamics, tailored approaches are essential. The study concludes that organizations should implement regular audits, transparency protocols, and training for professionals to oversee algorithms effectively. Future research should develop fairness-focused algorithms and hybrid models integrating human oversight, especially in non-Western settings, to promote inclusive practices. This review contributes to balancing technological innovation with ethical considerations, fostering equitable human resources practices globally.*

**Submitted:**
**SEPTEMBER 2025**

**Accepted:**
**DECEMBER 2025**

***Keywords**: Algorithmic Decision-Making, Bias, Fairness, HR, Human Judgment, Transparency.*

***ABSTRAK***

*Pengambilan keputusan algoritmik semakin banyak diadopsi dalam sumber daya manusia untuk menyederhanakan proses seperti rekrutmen, evaluasi kinerja, dan promosi, tetapi hal ini menimbulkan kekhawatiran etis tentang bias, keadilan, dan transparansi. Studi ini bertujuan untuk mengkaji risiko, tantangan keadilan, dan peran penilaian manusia dalam sistem ini. Dengan menggunakan tinjauan pustaka kualitatif, penelitian ini menganalisis studi yang ada untuk mengeksplorasi manfaat dan implikasi etis algoritma dalam sumber daya manusia. Temuan ini mengungkapkan bahwa algoritma meningkatkan efisiensi dan objektivitas tetapi dapat melanggengkan bias dari data historis, yang berpotensi menyebabkan hasil yang tidak adil, terutama bagi kelompok terpinggirkan. Penilaian manusia sangat penting untuk memastikan keputusan etis, mengatasi nuansa seperti kesesuaian budaya yang mungkin diabaikan oleh algoritma. Dalam konteks seperti Indonesia, di mana nilai-nilai budaya memengaruhi dinamika tempat kerja, pendekatan yang disesuaikan sangat penting. Studi ini menyimpulkan bahwa organisasi harus menerapkan audit rutin, protokol transparansi, dan pelatihan bagi para profesional untuk mengawasi algoritma secara efektif. Penelitian di masa mendatang harus mengembangkan algoritma yang berfokus pada keadilan dan model hibrida yang mengintegrasikan pengawasan manusia, terutama di lingkungan non-Barat, untuk mendorong praktik inklusif. Tinjauan ini berkontribusi pada keseimbangan inovasi teknologi dengan pertimbangan etika, serta mendorong praktik sumber daya manusia yang adil secara global.*

***Kata kunci**: Pengambilan Keputusan Algoritmik, Bias, Keadilan, SDM, Penilaian Manusia, Transparansi.*

**JIMKES**

Jurnal Ilmiah Manajemen Kesatuan
Vol. 13 No. 6, 2025
pp. 5169-5178
IBI Kesatuan
ISSN 2337 – 7860
E-ISSN 2721 – 169X
DOI: 10.37641/jimkes.v13i6.4147

## INTRODUCTION

The integration of algorithmic decision-making in Human Resources (HR) has emerged as a transformative force in organizations worldwide. As Artificial Intelligence (AI) and Machine Learning (ML) technologies become more sophisticated, HR professionals are increasingly relying on algorithms to streamline various functions such as recruitment, employee evaluation, and performance management (Vallor, 2016; Binns, 2018). While algorithmic systems promise greater efficiency, consistency, and objectivity, concerns about their potential risks, particularly regarding fairness, transparency, and bias, have sparked significant debate in both academic and professional circles (Burrell, 2016; O'neil, 2017; Eubanks, 2018). Therefore, it is crucial to explore the intersection of these technologies with human judgment to understand the broader implications of their application in HR (Binns, 2018; Noble, 2018).

The first question that arises in the context of algorithmic decision-making in HR is why this issue matters. The use of algorithms can lead to more streamlined processes, but it also raises important ethical and legal concerns, especially related to discrimination and bias in hiring and promotion decisions (Angwin et al., 2016). Inaccuracies in the data used to train these algorithms, or the way these systems are designed, can perpetuate existing inequalities in the workforce, particularly with regard to gender, race, and age (Binns, 2018; Noble, 2018). Given the growing reliance on these systems, understanding their risks is critical for ensuring that HR departments do not inadvertently contribute to discriminatory practices (O'neil, 2017; Dastin, 2022).

In response to these concerns, scholars have examined various facets of algorithmic decision-making, with a particular focus on mitigating bias and ensuring fairness in AI systems (Rahwan et al., 2019; Chouldechova & Roth, 2020). Many argue that while algorithms can significantly enhance HR processes, their application must be carefully managed to prevent reinforcing stereotypes or perpetuating existing inequalities (Eubanks, 2018; Holstein et al., 2019). Some researchers have proposed strategies such as auditing algorithms for bias and increasing transparency in how decisions are made (Crawford, 2016; Binns, 2018). However, other scholars emphasize that algorithms should not replace human judgment entirely, but instead complement it, particularly when it comes to interpreting the nuances of complex HR decisions (Crawford, 2016; Hilton et al., 2019).

The second key issue to consider is how these algorithms interact with human judgment and the potential consequences of relying solely on automated systems for HR decisions. While algorithms can assist in analyzing large volumes of data efficiently, they often lack the contextual understanding that human evaluators bring to decision-making. Binns (2018) and Dastin (2022) highlight that human judgment plays an essential role in interpreting subtle factors that are difficult for algorithms to capture, such as organizational culture, emotional intelligence, and ethical considerations. According to Tufekci (2015) and Crawford (2016), the combination of both algorithmic insight and human oversight appears to be the most viable approach for ensuring that HR decisions are both effective and fair.

The objective of this article is to critically examine the risks, fairness, and ethical considerations associated with algorithmic decision-making in HR, with a specific focus on the role of human judgment in mitigating these challenges. This study aims to address the gap in the literature regarding how organizations can balance the use of algorithms with the need for human oversight, ensuring fairness and equity in HR decisions. Through a comprehensive literature review and analysis of recent case studies, this article will provide practical recommendations for HR professionals seeking to integrate AI technologies while maintaining fairness and transparency in their practices.

This research is based on a qualitative approach, analyzing both theoretical frameworks and empirical findings to assess the current state of algorithmic decision-making in HR. The primary contribution of this article lies in its exploration of the synergy between algorithmic efficiency and human oversight, providing a roadmap for organizations looking to navigate the complexities of AI implementation in HR. By

addressing the risks and challenges posed by algorithmic decision-making, the article contributes to the growing body of literature on AI ethics, fairness, and human-centered design.

## LITERATURE REVIEW

### Risks of Algorithmic Decision-Making

The primary risk associated with algorithmic decision-making in HR is the potential for algorithmic bias. As noted by Angwin et al. (2016), algorithms often replicate and reinforce biases that exist in the data they are trained on. For instance, an algorithm designed to predict employee performance might favor individuals with certain demographic characteristics if the training data predominantly consists of successful employees who belong to a particular gender, race, or age group. This form of bias, known as algorithmic bias, can exacerbate existing inequalities in the workplace and lead to discriminatory outcomes (O'neil, 2017; Raji & Buolamwini, 2019; Madanchian et al., 2023). Several scholars have highlighted the importance of auditing algorithms for bias to ensure that HR decisions do not disproportionately disadvantage underrepresented groups (Kroll, 2015; De-Arteaga et al., 2019; Saeidnia et al., 2024). However, as pointed out by Mittelstadt et al. (2016), addressing algorithmic bias requires not only technical solutions but also a broader understanding of the social and ethical implications of using these systems in HR.

Another significant risk is the potential for over-reliance on automated systems, which may lead to the dehumanization of the HR process. While algorithms are adept at processing data, they often fail to capture the complexities of human behavior, including emotional intelligence, organizational fit, and interpersonal dynamics (Pc & Varughese, 2024). Over-relying on algorithms without human oversight can result in HR decisions that lack nuance and fail to account for the broader context of an individual's career trajectory (Shukla et al., 2023). This has prompted calls for a balanced approach where human judgment plays a central role in the decision-making process, especially when the stakes are high, such as in hiring or promotion decisions (Tufekci, 2015).

### Fairness in Algorithmic Decision-Making

Fairness in algorithmic decision-making is a critical concern for researchers and practitioners alike. The concept of fairness encompasses several dimensions, including procedural fairness (ensuring that the algorithm's decision-making process is transparent and accountable) and distributive fairness (ensuring that the outcomes of algorithmic decisions are equitable for all groups) (Friedman & Nissenbaum, 1996; Chouldechova, 2017; Barocas et al., 2023). One of the primary challenges is ensuring that algorithms do not reproduce or amplify societal biases. As highlighted by O'Neil (2017), algorithms can unintentionally favor certain demographic groups, particularly if the data they are trained on reflects historical inequalities. To ensure fairness, scholars recommend adopting more inclusive data practices and conducting regular audits of algorithms to detect and mitigate any biases that may emerge (Kroll, 2015; Noble, 2018). Furthermore, researchers have suggested that fairness can be promoted by increasing transparency in the algorithmic process and making it easier for users to understand how decisions are being made (Shin et al., 2024).

Fairness also requires that algorithms be designed with diversity and inclusion in mind. For example, algorithms can be programmed to actively promote diversity by prioritizing candidates from underrepresented groups or ensuring that performance evaluation criteria are free from bias (Wachter et al., 2017; Lipton, 2018). However, as noted by Rahwan et al. (2019), achieving fairness in algorithmic decision-making is a complex and ongoing process that requires continuous monitoring, adjustment, and human oversight. The development of fairness-enhancing algorithms is an area of active research, with scholars proposing various techniques, such as fairness constraints and fairness-aware machine learning models, to address these challenges (Kroll, 2015; Mittelstadt et al., 2016).

**Human Judgment in Algorithmic Decision-Making**

Despite the growing reliance on algorithms, human judgment remains a crucial element in the decision-making process. While algorithms excel at analyzing large datasets and identifying patterns, they are often limited by their inability to understand the broader context in which decisions are made. Human judgment plays an important role in interpreting algorithmic outputs, incorporating contextual factors, and ensuring that decisions align with organizational values and ethical considerations (Crawford & Paglen, 2021). In particular, HR professionals are needed to provide the ethical oversight required to ensure that algorithmic decisions do not lead to harmful or unintended consequences.

Several scholars argue that algorithms should complement, rather than replace, human decision-making in HR (Rahwan et al., 2019). For example, while an algorithm can help identify candidates with the right qualifications, human judgment is essential in evaluating how well a candidate fits with the organizational culture or how their unique skills align with the company's long-term goals (Grgic-Hlaca et al., 2018). Furthermore, human judgment is needed to assess the ethical implications of algorithmic decisions and ensure that these decisions are made with fairness and transparency in mind (Tufekci, 2015; Selbst et al., 2019; Mehrabi et al., 2021).

Algorithmic decision-making has the potential to improve HR practices by increasing efficiency and reducing bias; it also presents significant challenges related to fairness, transparency, and human oversight. The literature suggests that a balanced approach, where algorithms and human judgment work together, is essential for ensuring that HR decisions are both effective and equitable. The following sections will explore these issues in greater detail and propose practical recommendations for HR professionals seeking to integrate algorithmic systems while maintaining fairness and accountability.

## RESEARCH METHODS

This study employs a qualitative literature review approach to examine algorithmic decision-making in Human Resources (HR), focusing on its risks, fairness, and the role of human judgment. This method was chosen because it enables a deep exploration of complex issues, such as the ethical implications of AI in HR, which are challenging to capture through quantitative methods. A literature review facilitates the synthesis of existing research, identifying trends, and uncovering gaps to guide future studies (Fink, 2019). By analyzing relevant scholarly articles, industry reports, and case studies from disciplines including HR, AI ethics, and technology management, this research provides a comprehensive understanding of how algorithmic systems are transforming HR practices and the challenges they present.

Data collection involved a systematic search of academic databases such as Google Scholar, JSTOR, Scopus, and PubMed, using keywords like "algorithmic decision-making in HR," "AI and bias in recruitment," "algorithmic fairness," "human judgment in AI," and "ethics of HR algorithms." The search was restricted to English-language articles published between 2013 and 2024 to ensure relevance to recent technological advancements. Over 100 articles were initially identified, with 40 selected based on strict inclusion criteria, prioritizing peer-reviewed articles from reputable journals, as well as relevant industry reports addressing fairness, bias, and transparency in HR algorithms. Articles focusing solely on technical aspects without ethical implications were excluded to maintain alignment with the research objectives.

The data analysis utilized a thematic synthesis approach, ideal for qualitative literature reviews, to identify and organize key themes such as the application of algorithms in HR, risks of bias, fairness challenges, and the role of human judgment (Flick, 2022). This process involved coding concepts like "bias," "fairness," and "transparency," which were grouped into broader themes to provide an integrated perspective. The synthesis compared scholars' viewpoints, identified areas of consensus, and highlighted gaps, such as the limited research on hybrid models in HR.

To ensure reliability and validity, the data collection and analysis processes were conducted systematically and transparently, with clear inclusion criteria and reliance on credible sources. However, the literature review approach has limitations, including potential selection bias in article choice and the absence of empirical data for generalizing findings. Despite these constraints, this review establishes a robust foundation for understanding algorithmic decision-making in HR and offers directions for future empirical research.

## RESULTS

### Risks and Fairness of Algorithmic Decision-Making in HR

Algorithmic decision-making systems in HR frequently inherit and amplify biases embedded in historical data. When training datasets reflect past discriminatory hiring or promotion practices, algorithms systematically disadvantage certain demographic groups, including women, ethnic minorities, and older workers. A prominent example is Amazon's scrapped recruiting tool (2014–2018), which downgraded resumes containing the word "women's" because it was trained predominantly on male-dominated resumes from the previous decade (Dastin, 2022). This demonstrates that algorithms do not eliminate human bias but can institutionalise it at scale (Angwin et al., 2016; O'Neil, 2017).

Beyond gender and racial bias, algorithmic systems can produce intersectional discrimination that is difficult to detect without deliberate auditing. De-Arteaga et al. (2019) found that seemingly neutral biographical signals in online professional profiles led to significant bias against women in occupations historically dominated by men. Similarly, Buolamwini and Gebru (2018) showed that commercial facial analysis systems exhibit higher error rates for darker-skinned and female faces, which can indirectly affect HR tools that incorporate image or video analysis. These compounded errors disproportionately harm already marginalised applicants (Raji & Buolamwini, 2019; Mehrabi et al., 2021).

Over-reliance on algorithmic outputs without human oversight further escalates risk by reducing accountability. Mittelstadt et al. (2016) argue that opaque "black-box" models make it nearly impossible to trace why certain candidates are rejected, creating due process violations. Selbst et al. (2019) add that abstraction mismatches between technical systems and social contexts cause algorithms to misinterpret legally protected characteristics as legitimate proxies. Consequently, organisations using unexamined algorithmic tools may face legal liability and reputational damage while perpetuating systemic inequality (Eubanks, 2018; Saeidnia et al., 2024).

Defining and operationalising fairness remains one of the most contentious issues in algorithmic HR systems. Different mathematical definitions of fairness are often mutually incompatible, forcing developers to choose which groups to prioritise (Barocas et al., 2023). This impossibility theorem means that satisfying one fairness criterion can violate another, creating persistent trade-offs that cannot be fully resolved technically (Sweeney, 2013). Procedural fairness is equally difficult to achieve because most high-performing models in recruitment and performance evaluation are inherently opaque. Commercial HR platforms rarely disclose their training data or full feature weights, making external auditing challenging (Kroll, 2015; Burrell, 2016). Even when explanations are provided, they are often post-hoc and oversimplified, failing to reveal the true decision logic (Lipton, 2018). Lack of transparency undermines candidates' ability to contest adverse decisions and erodes trust in the hiring process (Shin et al., 2024).

Contextual and cultural differences further complicate global fairness efforts. Most fairness research and mitigation techniques have been developed and tested in Western organisational settings, potentially ignoring collectivist values, relational hiring practices, or local equality norms prevalent in countries such as Indonesia (Xin et al., 2018). Warna et al. (2024) note that accountability mechanisms effective in individualistic cultures may fail in high-context societies where decisions are influenced by harmony and seniority

rather than purely meritocratic criteria. The scarcity of non-Western case studies therefore represents a critical gap in achieving universally applicable fairness standards.

**The Role of Human Judgment in Algorithmic Decision-Making**

The role of human judgment emerges as a vital component in mitigating the risks associated with algorithmic decision-making. Crawford and Paglen (2021) emphasize that while algorithms excel at processing large datasets and identifying patterns, they often fail to capture contextual nuances, such as organizational culture, emotional intelligence, or ethical considerations. Human oversight is essential for interpreting algorithmic outputs and ensuring that decisions align with organizational values. For instance, in recruitment, HR professionals can assess a candidate's cultural fit or potential for growth, factors that algorithms may overlook (Ulrich & Dulebohn, 2015). The literature underscores that a hybrid approach, combining algorithmic efficiency with human judgment, is optimal for achieving fair and effective HR outcomes. This synergy allows organizations to leverage data-driven insights while addressing subjective elements that require human expertise. However, the review reveals a gap in exploring how human judgment is implemented across diverse cultural contexts, particularly in regions like Southeast Asia, where HR practices may be influenced by unique social dynamics (Xin et al., 2018).

To provide a structured overview of these findings, Table 1 summarizes the key themes identified in the literature, highlighting their implications and key references. The table organizes the results into four core areas: efficiency in HR processes, risks of bias and discrimination, fairness challenges, and the role of human judgment. It serves as a concise reference for understanding the consensus and gaps in the current research, reinforcing the need for ongoing monitoring and human oversight in algorithmic HR systems.

**Table 1.** Key Findings on Algorithmic Decision-Making in HR

| Theme | Findings | Key References |
|---|---|---|
| Efficiency in HR processes | Algorithms improve speed and data-driven decision-making, especially in recruitment and performance evaluation. | Angwin et al., 2016; Binns, 2018 |
| Bias and Discrimination | Algorithms can perpetuate or amplify biases based on historical data. | Angwin et al., 2016; O'Neil, 2017 |
| Fairness in Algorithms | Fairness issues arise when algorithms unintentionally discriminate against marginalized groups. Need for fairness-aware models. | Kroll, 2015; Green & Viljoen, 2020 |
| Role of Human Judgment | Human judgment is essential for interpreting algorithmic outputs and ensuring fairness, particularly for subjective elements such as organizational fit. | Lepri & Cialdini, 2020; Crawford & Paglen, 2021 |

Although this study is qualitative and does not involve statistical hypothesis testing, the synthesis of findings enables inferential insights into patterns across industries and contexts. For instance, the consistent emphasis on algorithmic bias across studies, as noted by Angwin et al. (2016), suggests a broad consensus on the need for structured auditing mechanisms to ensure fairness. Similarly, the recurring theme of human judgment, as highlighted by Sargeant (2023), indicates that algorithms should support, rather than replace, human decision-making in HR. These patterns underscore the importance of integrating ethical considerations into algorithmic design. However, the limited representation of non-Western perspectives in the literature suggests a need for future research to explore how cultural and regional factors influence the application and impact of algorithms in HR, particularly in diverse settings like Indonesia.

The findings reveal that algorithmic decision-making in HR offers significant potential for improving efficiency and objectivity but is fraught with challenges related to bias, fairness, and the need for human oversight. Algorithms can streamline processes but risk perpetuating historical inequalities if not carefully managed. Fairness remains a complex issue, requiring intentional design and continuous monitoring, while human judgment is indispensable for ensuring ethical and context-sensitive decisions. These insights highlight

the importance of a balanced approach that leverages algorithmic strengths while addressing their limitations through human expertise and robust governance frameworks.

## DISCUSSION

The findings of this literature review underscore the transformative potential of algorithmic decision-making in Human Resources (HR), while highlighting its inherent risks and ethical challenges. Algorithms offer significant advantages in enhancing efficiency and objectivity in processes like recruitment and performance evaluations, as they can process vast datasets quickly and reduce human error. However, as scholars like O'Neil (2017) argue, these systems can perpetuate biases embedded in historical data, such as gender or racial disparities, if not carefully managed. For instance, Amazon's abandoned AI hiring tool, which favored male candidates due to biased training data, illustrates how algorithms can reinforce systemic inequities. This aligns with Distributive Justice Theory, which emphasizes that fair systems must promote equity and avoid harm (Rawls, 1971). The reliance on historical data underscores a critical paradox: while algorithms are often perceived as impartial, they can inadvertently amplify existing inequalities, particularly in hiring and promotion decisions, necessitating a deeper examination of their design and application.

The issue of fairness remains a central concern in the adoption of algorithmic systems in HR. Kroll (2015) emphasizes that achieving fairness requires both procedural transparency and distributive equity, ensuring that algorithms do not disproportionately disadvantage marginalized groups. For example, algorithms used in recruitment may favor candidates from dominant demographic groups if trained on biased datasets, as seen in cases where systems penalized resumes with terms associated with underrepresented groups. To address this, fairness-aware algorithms and regular audits are essential to promote inclusivity. However, the literature reveals a gap in exploring fairness in non-Western contexts, such as Indonesia, where cultural factors like collectivism or regional diversity may influence HR practices. Incorporating these perspectives could enhance the global applicability of fairness frameworks, ensuring that algorithms account for diverse workforce dynamics and local ethical norms.

Human judgment plays an indispensable role in mitigating these risks and ensuring ethical HR practices. Crawford and Paglen (2021) highlight that algorithms lack the contextual understanding needed to assess factors like organizational culture or emotional intelligence, which are critical in HR decisions. For instance, while an algorithm can identify qualified candidates based on data, HR professionals are better equipped to evaluate cultural fit or long-term potential. This synergy between algorithmic efficiency and human oversight is crucial for balancing objectivity with ethical considerations. In Indonesia, where interpersonal relationships often influence workplace dynamics, human judgment can ensure that algorithmic decisions align with local values, such as communal harmony. This human-in-the-loop approach not only mitigates bias but also fosters trust among employees, as decisions reflect both data-driven insights and contextual sensitivity.

These findings are significant for HR practitioners and organizations adopting algorithmic systems. First, organizations must implement robust governance frameworks, including regular audits and transparency protocols, to ensure fairness and accountability. Training HR professionals to understand and oversee algorithmic systems is critical, particularly in contexts like Indonesia, where cultural nuances may require tailored approaches. Second, the development of fairness-aware algorithms should prioritize diversity and inclusion, actively countering biases in historical data. Finally, future research should explore the long-term impact of algorithms on workforce diversity and organizational culture, especially in non-Western settings, to address gaps in the current literature. By integrating algorithmic efficiency with human oversight and ethical governance, organizations can foster inclusive HR practices that balance technological innovation with fairness and equity.

## CONCLUSION

This literature review highlights the complex role of algorithmic decision-making in Human Resources (HR), revealing its potential to enhance efficiency and objectivity in processes like recruitment and performance evaluations, while also posing significant ethical challenges. Algorithms streamline HR tasks by processing large datasets, but they risk perpetuating biases embedded in historical data, leading to unfair outcomes in hiring and employee evaluations, particularly for marginalized groups. Human judgment remains essential to ensure ethical and context-sensitive decisions, addressing nuances like organizational culture and emotional intelligence that algorithms often overlook. By balancing algorithmic efficiency with human oversight, organizations can leverage technology to improve HR practices while maintaining fairness and inclusivity.

The findings have important implications for HR practitioners, particularly in designing governance frameworks that prioritize fairness through regular audits and transparency. In contexts like Indonesia, where cultural values such as collectivism shape workplace dynamics, tailoring algorithms to local norms is crucial for equitable outcomes. However, this study's reliance on a literature review limits its ability to provide empirical evidence, and the predominance of Western perspectives may overlook unique regional challenges. Future research should focus on empirical studies to assess the long-term impact of algorithms on workforce diversity, particularly in non-Western settings. Developing fairness-aware algorithms and hybrid models that integrate human oversight, along with training programs for HR professionals to manage AI ethically, will be critical to fostering inclusive and equitable HR practices.

## REFERENCES

[1] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. ProPublica. Retrieved on March 6, 2025, from https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

[2] Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. Cambridge: MIT Press.

[3] Binns, R. (2018). Fairness in machine learning: A survey. *ACM Computing Surveys (CSUR)*, *51*(3), 1–35.

[4] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Conference on Fairness, Accountability and Transparency*, *1*(1), 77–91.

[5] Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12.

[6] Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *Big Data*, *5*(2), 153–163.

[7] Chouldechova, A., & Roth, A. (2020). A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, *63*(5), 82–89.

[8] Crawford, K. (2016). Artificial intelligence's white guy problem. *The New York Times*. Retrieved on March 6, 2025, from https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.

[9] Crawford, K., & Paglen, T. (2021). Excavating AI: The politics of images in machine learning training sets. *AI & Society*, *36*(4), 1105–1116.

[10] Dastin, J. (2022). Amazon scraps secret AI recruiting tool that showed bias against women. In *Ethics of data and analytics* (pp. 296–299). Auerbach Publications.

[11] De-Arteaga, M., Romanov, A., Wallach, H., Chayes, J., Borgs, C., Chouldechova, A., Geyik, S., Kenthapadi, K., & Kalai, A. T. (2019). Bias in bios: A case study of semantic representation bias in a high-stakes setting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, *(1)*, 120–128.

[12] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. New York: St. Martin's Press.

[13] Fink, A. (2019). *Conducting research literature reviews: From the internet to paper*. London: Sage Publications.

[14] Flick, U. (2022). *An introduction to qualitative research*. London: Sage Publications.

[15] Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems (TOIS)*, *14*(3), 330–347.

[16] Grgic-Hlaca, N., Redmiles, E. M., Gummadi, K. P., & Weller, A. (2018). Human perceptions of fairness in algorithmic decision making: A case study of criminal risk prediction. In *Proceedings of the 2018 World Wide Web Conference*, 903–912.

[17] Hilton, N. Z., Ham, E., & Green, M. M. (2019). Adverse childhood experiences and criminal propensity among intimate partner violence offenders. *Journal of Interpersonal Violence*, *34*(19), 4137–4161.

[18] Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., & Wallach, H. (2019). Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16.

[19] Kroll, J. A. (2015). *Accountable algorithms*. Princeton: Princeton University Press.

[20] Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, *16*(3), 31–57.

[21] Madanchian, M., Taherdoost, H., & Mohamed, N. (2023). AI-based human resource management tools and techniques: A systematic literature review. *Procedia Computer Science*, *229*(1), 367–377.

[22] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1–35.

[23] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, *3*(2), 1–21.

[24] Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York: New York University Press.

[25] O'Neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown.

[26] Pc, N., & Varughese, A. (2024). Emotional intelligence and interpersonal dynamics in the workplace: Importance of psychological contract. *Организационная Психология*, *14*(2), 44–57.

[27] Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., & Jackson, M. O. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

[28] Raji, I. D., & Buolamwini, J. (2019). 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*, *1*, 429–435.

[29] Rawls, J. (1971). *A theory of justice*. Cambridge: Harvard University Press.

[30] Saeidnia, H. R., Hashemi Fotami, S. G., Lund, B., & Ghiasi, N. (2024). Ethical considerations in artificial intelligence interventions for mental health and well-being: Ensuring responsible implementation and impact. *Social Sciences*, *13*(7), 381–395.

[31] Sargeant, H. (2023). Algorithmic decision-making in financial services: Economic and normative outcomes in consumer credit. *AI and Ethics*, *3*(4), 1295–1311.

[32] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, *1*, 59–68.

[33] Shin, D., Lim, J. S., Ahmad, N., & Ibahrine, M. (2024). Understanding user sensemaking in fairness and transparency in algorithms: Algorithmic sensemaking in over-the-top platform. *AI & Society*, *39*(2), 477–490.

[34] Shukla, R. P., Mandhanya, Y., Mishra, S., Jahagirdar, R., Dari, S. S., & Vij, R. (2023). Cognizant prognostication: An in-depth comparative study of machine learning models for predictive employee turnover analysis in the realm of human resources analytics. In *International Conference on Intelligent Systems Design and Applications*, *1*, 196–204.

[35] Sweeney, L. (2013). Discrimination in online ad delivery. *Communications of the ACM*, *56*(5), 44–54.

[36] Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. *Colorado Technology Law Journal*, *13*(1), 203–218.

[37] Ulrich, D., & Dulebohn, J. H. (2015). Are we there yet? What's next for HR? *Human Resource Management Review*, *25*(2), 188–204.

[38] Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford: Oxford University Press.

[39] Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, *7*(2), 76–99.

[40] Warna, W., Hamzani, U., & Rusmita, S. (2024). Accountability and transparency of village fund budget management. *Jurnal Ilmiah Manajemen Kesatuan*, *12*(5), 1821–1830.

[41] Xin, D., Ma, L., Liu, J., Macke, S., Song, S., & Parameswaran, A. (2018). Accelerating human-in-the-loop machine learning: Challenges and opportunities. *Proceedings of the Second Workshop on Data Management for End-to-End Machine Learning*, *1*, 1–4.

*Risks, Fairness, and the Role of Human Judgment*

# 5178