Advances in Consumer Research
https://acr-journal.com/
Volume-2 | Issue-6| December 2025
Original Researcher Article

AI in Recruitment: Challenges of Transparency, Accountability and Fairness

Dr. N. Kumaresh [1]*, Dr. R. Mary Angelin [2], Dr. Durgarani M [3], Dr. S. Pramila [4], Dr Komal Nagrani [5], Dr. Gomathi [6]

[1]Associate Professor, Department of MCA, RV Institute of Technology and Management, Bangalore, Karnataka- 560076.
Email ID : kumareshn.rvitm@rvei.edu.in
[2]Assistant Professor, Department of Commerce, BS Abdur Rahman Crescent institute of science and technology, Vandalur, Chennai-600048.
Email ID : angelinmary.r@gmail.com
[3]Associate Professor, School of Business and Management,CHRIST University, Nagasandra, Near Tumkur Road, Bangalore, Karnataka - 560073.
Email ID : durgaranimurali@gmail.com
[4]Associate Professor, School of Commerce, finance and Accountancy, Christ University, Delhi NCR Campus.
Email ID : pramila.s@christuniversity.in
[5]Assistant Professor-III, Department of Business Management, Amity University, Hyderabad.
Email ID : knagrani@hyd.amity.edu
[6]HOD, Department of Business Administration, St. Thomas College of arts and science, koyambedu, Chennai-600107.
Email ID : gomathibaskaran@gmail.com

ABSTRACT

Artificial Intelligence (AI) has rapidly transformed recruitment practices by automating candidate sourcing, screening, assessment, and selection. Organizations widely adopt AI tools to enhance efficiency, minimize hiring time, and improve decision-making. However, these technologies also raise profound concerns regarding transparency, accountability, and fairness. The black-box nature of many AI systems, the risk of bias in algorithmic decision-making, and unclear lines of responsibility create ethical and operational challenges that limit trust and acceptance. This article critically examines the challenges associated with AI-driven recruitment, focusing on transparency, accountability, and fairness. It also explores emerging regulatory responses, organizational strategies, and ethical frameworks that can help ensure a responsible and equitable adoption of AI in hiring. Tools such as fairness metrics, bias audits, and inclusive design practices help maintain impartiality. Ensuring fairness not only protects candidates' rights but also enhances organizational diversity and strengthens employer reputation. Ultimately, fair AI systems support ethical recruitment by promoting equal opportunity for all applicants. The results imply that ethical considerations such as transparency, accountability, and data protection are viewed as universal concerns, cutting across gender lines. This uniformity may also reflect increased access to information, digital exposure, and similar levels of engagement with technology among both genders in the sample. Across all three factors—Transparency, Human Oversight and Privacy & Data Protection—the results consistently show no significant gender-based differences in opinion...

Keywords: *Job Advertising, Resume Screening, Transparency, Accountability and Sourcing.*
.

1. INTRODUCTION:

AI is reshaping the global recruitment landscape. Algorithms now filter resumes, assess personality traits, analyze video interviews, and even predict candidate suitability based on historical hiring patterns. These innovations promise to reduce human workload and increase efficiency. Yet, the introduction of AI into recruitment—a domain traditionally influenced by human intuition, interpersonal understanding, and contextual judgment—raises critical concerns. Proper accountability requires assigning responsibilities across multiple stakeholders, including HR managers, data scientists, policymakers, and technology providers. Organizations must ensure human oversight remains central to all hiring decisions, enabling humans to intervene, correct mistakes, and justify outcomes when necessary. Maintaining audit trails, documenting model behavior, and regularly reviewing system performance are essential components of accountability. When responsibility is clearly outlined, employers can address grievances effectively, meet legal requirements, and uphold ethical hiring standards. Accountability ensures that AI is used as a supportive tool rather than an unquestioned authority. The results imply that ethical considerations such as transparency, accountability, and data protection are viewed as universal **concerns**, cutting across gender lines. This uniformity may also reflect increased access to information, digital exposure, and similar levels of engagement with technology among both genders in the sample.

Three challenges stand out:

1. Transparency: Transparency in AI-based recruitment refers to the clarity and openness with which organizations communicate how artificial intelligence systems operate during the hiring process. A transparent system allows candidates and employers to understand what data is being collected, how it is analyzed, and which algorithmic factors influence decision-making. However, many recruitment tools function as “black boxes,” where the internal logic of the model is hidden due to technical complexity or vendor confidentiality. This lack of visibility creates confusion and mistrust among applicants, who may not know why certain decisions were made. Transparency also empowers organizations to audit the AI system effectively and detect potential biases or errors. When companies disclose AI involvement in screening, provide explanations for decisions, and ensure interpretability of algorithms, they build credibility and demonstrate ethical responsibility. Ultimately, transparency is a foundational requirement for promoting fairness, accountability, and trust in AI-driven hiring.

2. Accountability: Accountability in AI-driven recruitment involves clearly defining who is responsible for decisions made or influenced by artificial intelligence tools. When AI systems autonomously screen resumes, rank candidates, or evaluate video interviews, errors or biases may occur. Without established accountability mechanisms, it becomes difficult to determine whether the employer, the AI vendor, or the system developers should answer for unfair outcomes.

3. Fairness: Fairness in AI recruitment focuses on ensuring that all candidates are evaluated equitably, without discrimination based on gender, race, age, socio-economic background, disability, or other protected characteristics. AI systems may unintentionally perpetuate bias because they learn from historical data that could reflect past discriminatory hiring patterns. For example, if a company’s previous selections favored male candidates, an AI model trained on that data may replicate similar outcomes. Fairness requires the use of diverse, unbiased training datasets, along with techniques to detect and mitigate algorithmic bias. It also involves monitoring model outputs to ensure equitable treatment across demographic groups.

AI in Recruitment: Current Landscape

AI tools in recruitment are used across multiple stages:

Job Advertising and Sourcing: Resume Screening

AI tools are widely used in job advertising and resume screening to enhance efficiency in recruitment. In job advertising and sourcing, algorithms identify suitable candidates by analyzing online profiles, job portals, and social media platforms. They match job descriptions with candidate skills and behaviors, ensuring that job ads reach the most relevant audience. In resume screening, AI filters large volumes of applications by scanning keywords, qualifications, experience, and skill sets. This automated process shortens hiring cycles and reduces manual workloads for HR teams. However, reliance on historical data may introduce bias, making fairness and transparency crucial considerations.

Chatbots: Video-Interview Analytics

AI-powered chatbots improve candidate engagement by answering queries, scheduling interviews, and guiding applicants through the hiring process. They operate 24/7, providing instant support and improving recruitment efficiency. Video-interview analytics go a step further by using AI to evaluate facial expressions, tone, gestures, and speech patterns during video interviews. These tools aim to assess personality traits, communication ability, and job fit. While they promise faster evaluations, they raise ethical issues related to accuracy, cultural differences, and potential bias. Human oversight is essential to ensure fair and valid assessments.

Predictive Analytics

Predictive analytics uses historical and real-time data to forecast candidate suitability, job performance, turnover likelihood, and cultural fit. AI models analyze patterns in previous hiring decisions, employee performance records, and behavioral data to generate predictions about future outcomes. This helps organizations make data-driven decisions and select candidates who are more likely to succeed in specific roles. However, predictive analytics can unintentionally reproduce past biases if the underlying data is skewed. To ensure fairness and reliability, organizations must validate their predictive models, use diverse datasets, and maintain transparency in how predictions are generated.

Transparency Challenges

1. Black-Box Decision-Making

Black-box decision-making refers to the use of AI systems whose internal logic and reasoning processes are not visible or understandable to users. In recruitment, many AI models—especially deep-learning algorithms—analyze data using complex patterns that even developers may not fully interpret. As a result, employers cannot clearly explain why certain candidates were shortlisted, rejected, or ranked. This lack of interpretability creates concerns about hidden biases, fairness, and trust. When decisions lack transparency, candidates may feel unfairly evaluated, and organizations face difficulties defending decisions in legal or ethical reviews.

2. Limited Candidate Awareness

Limited candidate awareness occurs when applicants do not know that AI is being used in the hiring process or how it evaluates their information. Many organizations deploy AI tools for resume scanning, assessments, or video-analysis without adequately informing candidates. As a result, applicants may be unaware of what data is collected, how long it is stored, or what performance criteria are applied. This lack of clarity reduces trust and may create perceptions of unfairness. Transparent communication is essential to help candidates understand their evaluation process and maintain confidence in AI-driven recruitment.

3. Inadequate Information for Employers

Employers often lack detailed insight into how AI recruitment tools function. Many companies purchase off-the-shelf AI systems without knowing what data the model was trained on, which variables influence decisions, or how the algorithm evaluates candidates. This limited understanding makes it difficult for HR teams to interpret outputs, detect bias, or justify decisions when challenged. Without access to internal algorithmic logic or documentation, employers may unintentionally rely on flawed systems. To ensure ethical hiring, organizations must demand greater transparency from vendors, conduct audits, and ensure that HR professionals understand the tool's capabilities and limitations.

4. Proprietary Restrictions

Proprietary restrictions occur when AI vendors refuse to disclose algorithmic details due to intellectual property protections. While companies aim to safeguard their competitive advantage, this secrecy limits transparency and prevents external auditing of the system. As a result, employers cannot fully assess how the AI processes candidate data or identify potential sources of bias. These restrictions create ethical and compliance challenges, especially in high-stakes decisions like hiring. Balancing trade secrets with the need for openness is critical. Regulators increasingly emphasize explainability requirements to ensure that proprietary systems do not undermine fairness and accountability.

Accountability Challenges

1. Data-Driven Errors

Data-driven errors occur when AI recruitment systems produce flawed or biased outcomes because the underlying data used for model training is incomplete, inaccurate, or discriminatory. If historical hiring data reflects gender, racial, educational, or socio-economic biases, the AI will learn and replicate these patterns in its decisions. Errors may also arise from outdated information, inconsistent labeling, or limited sample diversity. As a result, qualified candidates may be unfairly screened out. These errors highlight the importance of high-quality, representative datasets and continuous monitoring to ensure that AI systems deliver reliable and equitable hiring outcomes.

2. Lack of Regulatory Frameworks

The rapid adoption of AI in recruitment has outpaced the development of comprehensive legal and regulatory guidelines. In many countries, there are limited or no specific laws governing how AI systems should be developed, audited, or used in hiring processes. This regulatory gap allows organizations to deploy AI tools without adequate oversight, increasing the risk of bias, discrimination, and privacy violations. The absence of clear standards regarding transparency, explainability, and accountability makes it difficult to evaluate AI practices. Strengthening regulatory frameworks is essential to protect candidates' rights and ensure responsible, ethical use of AI in recruitment.

3. Absence of Human Override Mechanisms

An absence of human override mechanisms occurs when AI recruitment systems make decisions without meaningful human review or intervention. Overreliance on automated short listing, scoring, or ranking may lead to unfair or erroneous outcomes going unnoticed. Without human oversight, candidates who are wrongly filtered out cannot appeal decisions or receive individualized assessments. Human override mechanisms ensure accountability by allowing HR professionals to validate results, correct mistakes, and apply contextual judgment. Maintaining human-in-the-loop processes is critical for fairness, transparency, and ethical hiring, especially when AI systems lack interpretability or consistently produce unclear outcomes.

Fairness Challenges

Bias in Training Data

AI models learn from historical patterns. If a company's past hiring favored male candidates or graduates from specific universities, the AI may embed these preferences. Biases may stem from:

Gendered job descriptions

Skewed performance evaluations

Ethnically imbalanced datasets

Socio-economic factors encoded in resumes

Thus, AI does not simply automate decision-making; it can automate discrimination unless carefully monitored.

Algorithmic Bias and Discrimination

Even when training data appears neutral, algorithms may inadvertently discriminate. For example, AI might use proxies such as postal codes, hobbies, or institutions that correlate with demographic variables, leading to unfair exclusion.

Bias in Video and Facial Recognition Tools

AI video-interview assessments have shown higher error rates when analyzing candidates with darker skin tones, non-Western accents, or disabilities. Emotional recognition algorithms often fail to accurately interpret non-Western facial expressions or cultural norms.

Fairness vs. Efficiency Trade-Off

AI systems optimized for predictive accuracy may not align with fairness objectives. Achieving fairness may require balancing statistical parity, equal opportunity, and predictive validity—goals that do not always coexist harmoniously.

Ethical and Regulatory Responses

Organizational Strategies to Enhance Transparency, Accountability, and Fairness

Transparency: Transparency in AI recruitment involves openly communicating how algorithms function, what data they analyze, and how decisions are generated. It requires providing candidates and employers with clear explanations about the role of AI in screening, scoring, and short listing applicants. Transparent systems help build trust and reduce uncertainty by revealing evaluation criteria and decision-making logic. When organizations disclose AI usage and offer accessible explanations for outcomes, it becomes easier to identify biases, conduct audits, and ensure ethical behavior. Transparency is crucial for maintaining fairness, supporting accountability, and enabling meaningful candidate consent.

Accountability: Accountability ensures that clear responsibility is assigned for every stage of AI-driven recruitment, including data collection, model development, decision-making, and error handling. When AI tools produce biased, inaccurate, or unethical results, organizations must be able to determine who is answerable—the employer, the vendor, the developer, or the HR team. Effective accountability requires documented processes, audit trails, human oversight, and compliance with legal standards. It also involves mechanisms for candidates to challenge or appeal decisions. Establishing strong accountability frameworks prevents blame shifting and ensures that AI supports, rather than replaces, responsible human decision-making.

Fairness and Non-Discrimination: Fairness and non-discrimination aim to ensure that AI recruitment systems evaluate all candidates equitably, regardless of gender, race, age, disability, or socio-economic background. Bias can arise from historical data, algorithmic design, or unintended statistical correlations. Therefore, ensuring fairness requires diverse training datasets, continuous bias testing, and transparent evaluation criteria. Organizations must implement fairness metrics to examine disparities in outcomes and take corrective action when bias is detected. Fair AI systems help promote equal opportunity, improve diversity and inclusion, and strengthen organizational reputation. Upholding fairness is central to ethical recruitment and legal compliance.

Privacy and Data Protection: Privacy and data protection focus on safeguarding candidates' personal information throughout the recruitment process. AI systems collect and process large volumes of data, including resumes, online behavior, and sometimes video or biometric information. Without strong protections, this data may be misused, leaked, or processed without consent. Organizations must comply with data protection laws, minimize unnecessary data collection, ensure secure storage, and maintain transparency about data usage. Candidates should have the right to know what data is being collected and request its deletion. Strong privacy practices help maintain trust, prevent misuse, and support ethical AI deployment.

Human Oversight: Despite their efficiency, AI models can make errors, misinterpret data, or exhibit bias. Human involvement allows HR professionals to review AI outputs, verify accuracy, and apply contextual judgment that algorithms cannot replicate. Oversight mechanisms include human-in-the-loop processes, manual review of flagged cases, and override rights for HR teams. Such involvement ensures fairness, accountability, and transparency, preventing blind reliance on algorithms. Human oversight maintains ethical control and ensures that final hiring decisions align with organizational values and legal standards.

Research Gap

Although artificial intelligence has increasingly been integrated into recruitment processes worldwide, there remains a significant research gap concerning the deeper implications of transparency, accountability, and fairness within AI-driven hiring systems. Existing literature primarily focuses on the technical efficiency, speed, and cost-effectiveness of AI recruitment tools. However, relatively few empirical studies examine how these systems may inadvertently reproduce biases or create opaque decision-making processes that limit candidates' understanding of how they are evaluated. Moreover, while several studies acknowledge ethical concerns, there is limited research that explores the intersectional impact of AI on diverse demographic groups, including gender, age, socio-economic background, and minority communities. Another gap lies in understanding organizational readiness and HR professionals' awareness regarding the ethical deployment of AI tools. There is also insufficient evidence from developing countries such as India, where the adoption of AI in recruitment is rapidly increasing but regulatory and ethical guidelines remain underdeveloped. These gaps highlight the urgent need for comprehensive research that evaluates not only technological functionality but also fairness, auditability, and transparency to ensure responsible and inclusive AI implementation in recruitment.

Importance of the Study

This study is highly significant because the adoption of AI in recruitment is accelerating across industries, yet its ethical, procedural, and fairness-related consequences remain insufficiently understood. This research contributes to a holistic understanding of how AI systems influence candidate experiences, organizational credibility, and compliance with ethical standards. It also highlights the importance of designing recruitment tools that uphold equal opportunity legislation and promote diversity and inclusion. Furthermore, the study is particularly relevant for emerging economies, where AI adoption is growing but regulatory frameworks remain fragmented. Understanding the challenges of AI-driven recruitment can help organizations establish best practices, improve workforce diversity, enhance organizational performance, and build more reliable digital hiring ecosystems. Ultimately, the study offers a foundation for ensuring that AI in recruitment is not only efficient but also ethical, transparent, and equitable.

Statement of the Problem

Despite the rising adoption of AI-based tools in recruitment, organizations continue to face significant challenges related to transparency, accountability, and fairness in the hiring process. Many AI recruitment systems operate as "black boxes," providing limited insight into how decisions such as short listing or rejection are made. This lack of transparency raises concerns among candidates and HR professionals, who cannot fully understand or verify the criteria influencing hiring outcomes. Additionally, the absence of clear accountability mechanisms makes it difficult to identify responsibility when AI-generated decisions result in unfair or discriminatory practices. Biases embedded in training data or algorithms can reinforce existing inequalities, negatively affecting certain demographic groups and undermining organizational diversity goals. Furthermore, the rapid implementation of AI tools, often without adequate regulatory oversight or ethical frameworks, increases the risk of inconsistent, biased, or opaque recruitment practices. This problem is especially serious in fast-growing job markets where organizations depend heavily on digital hiring systems. Therefore, the core problem this study addresses is the urgent need to evaluate and understand the ethical shortcomings and operational challenges of AI-driven recruitment to ensure that hiring processes remain transparent, accountable, and fair for all applicants.

Objectives

To identify the challenges among respondents of different age groups.

To examine the challenges among respondents of different gender.

To outline organizational Strategies to Enhance Transparency, Accountability, and Fairness

Methodology

This study adopted a **descriptive and analytical research design** to examine whether there is a significant difference in respondents' opinions on challenges based on **age groups and gender**. A total of **150 participants** were selected using a **stratified random sampling technique** to ensure adequate representation of different age categories and gender groups. The population consisted of employees from various sectors who have experience with the challenges being investigated, particularly in the context of AI-enabled processes.

A **structured questionnaire** was used as the primary data collection tool. The instrument consisted of close-ended statements measured on a **five-point Likert scale** ranging from "Strongly Agree" to "Strongly Disagree." Descriptive statistics such as mean, standard deviation, and rank scores were also computed. The methodology ensures a systematic approach to determining whether differences in opinion across demographic groups are statistically significant.

Findings and Results

The hypothesis is tested with the help of non-parametric tests Mann-Whitney U test and the Kruskal Wallis test. The results are discussed as below.

AGE

H_0 (Null Hypothesis): There is **no significant difference** in the opinion on Challenges Among respondents of different age groups.

TABLE: 1

DIFFERENCE IN THE OPINION BASED ON THE AGE GROUP

Factors	Age	N	Mean Rank	Test	Result
Transparency	Less than 30	65	71.30	Chi-Square	6.107
	30 to 50	57	79.03	df	2
	More than 50	28	78.07	Asymp. Sig.	.000
	Total	150			
Human Oversight	Less than 30	65	69.39	Chi-Square	4.226
	30 to 50	57	79.94	df	2
	More than 50	28	80.64	Asymp. Sig.	.086
	Total	150			
Privacy and Data Protection	Less than 30	65	73.02	Chi-Square	4.182
	30 to 50	57	75.44	df	2
	More than 50	28	81.39	Asymp. Sig.	.066
	Total	150			

Transparency – Significant Difference (p = 0.000)

The Chi-square test for transparency shows a **significant difference** across age groups (Asymp. Sig. = 0.000 < 0.05). This means the null hypothesis is rejected. The mean ranks show that respondents **30–50 years (79.03)** and **above 50 years (78.07)** scored higher than the younger group **below 30 years (71.30)**. This indicates that **older respondents show stronger expectations or awareness regarding transparency** in AI-based recruitment. Younger respondents may be more accustomed to technology and therefore less concerned or more accepting of automated processes.

Human Oversight – No Significant Difference (p = 0.086)

The significance value (0.086 > 0.05) shows that the **difference among age groups is not statistically significant**. Although the mean ranks increase with age (younger = 69.39, middle = 79.94, older = 80.64), the variation is **not strong enough** to confirm a meaningful difference.
This indicates that **all age groups generally agree on the importance of human oversight**, showing consistent belief that humans should supervise or validate AI decisions.

Privacy and Data Protection – No Significant Difference (p = 0.066)

The significance value (0.066 > 0.05) indicates **no significant difference** in opinions across age groups. The mean ranks show a slight upward trend with age (73.02 → 75.44 → 81.39), but the differences are **not statistically meaningful**.
This suggests that **privacy and data protection concerns are shared across all age groups**, possibly reflecting universal awareness of data risks in digital systems.

Overall Summary

Only Transparency shows a significant age-based difference, with older respondents showing stronger concerns.

Human Oversight and Privacy/Data Protection do not differ significantly across age groups, indicating shared views among all age categories.

H_0 (Null Hypothesis): There is **no significant difference** in the opinion on Challenges Among respondents of different gender.

TABLE 2

DIFFERENCE IN THE OPINION BASED ON THE GENDER

Factors	Gender	N	Mean Rank	Test	Result
Transparency	Male	60	75.92	Mann-Whitney U	20114.000
	Female	90	75.22	Z	.729
	Total	150		Sig.	.206
Human Oversight	Male	60	70.82	Mann-Whitney U	22571.000
	Female	90	78.62	Z	.587
	Total	150		Sig.	.436
Privacy and Data Protection	Male	60	72.43	Mann-Whitney U	24210.000
	Female	90	77.55	Z	-209
	Total	150		Sig.	.878

The results of the Mann–Whitney U tests indicate that there are no significant differences in opinions between male and female respondents across all three factors examined—Transparency, Human Oversight, and Privacy and Data Protection. For Transparency, the mean ranks of males (75.92) and females (75.22) were nearly identical, and the p-value (0.206) confirmed that this difference is not statistically significant. Similarly, in the case of Human Oversight, although females (78.62) had a slightly higher mean rank than males (70.82), the p-value of 0.436 showed that this difference is not meaningful. Finally, for Privacy and Data Protection, the mean ranks of males (72.43) and females (77.55) again reflected only minimal variation, and the very high p-value (0.878) demonstrated that there is no significant gender-based difference in perceptions. Overall, the findings clearly suggest that both male and female respondents hold similar viewpoints regarding all three ethical dimensions assessed.

2. CONCLUSION

The study concluded that **gender does not significantly influence respondents' opinions on key ethical challenges** such as Transparency, Human Oversight, and Privacy & Data Protection. Both male and female participants perceive these issues in a largely consistent manner, highlighting a shared understanding of the importance of ethical safeguards in technological and organizational contexts.. Although the mean ranks show minor variations— males scored slightly higher on Transparency while females scored slightly higher on Human Oversight and Privacy—the differences are not large enough to be considered meaningful. This suggests that both male and female respondents demonstrate **similar levels of awareness, concern, and understanding** about the ethical issues related to technological or organizational challenges. The results imply that ethical considerations such as transparency, accountability, and data protection are viewed as **universal concerns**, cutting across gender lines. This uniformity may also reflect increased access to information, digital exposure, and similar levels of engagement with technology among both genders in the sample.

REFERENCES

1. Asan, O.; Bayrak, A.E.; Choudhury, A. Artificial intelligence and human trust in healthcare: Focus on clinicians. J. Med. Internet Res. 2020, 22, e15154
2. Berk, R.; Heidari, H.; Jabbari, S.; Kearns, M.; Roth, A. Fairness in Criminal Justice Risk Assessments: The State of the Art. Sociol. Methods Res. 2018, 47, 175–210.
3. Buolamwini, J.; Gebru, T. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability and Transparency, New York, NY, USA, 23–24 February 2018; pp. 77–91.
4. Caliskan, A.; Bryson, J.J.; Narayanan, A. Semantics derived automatically from language corpora contain human-like biases. Science 2017, 356, 183–186.
5. Crawford, K.; Paglen, T. Excavating AI: The politics of images in machine learning training sets. AI Soc. 2021, 36, 1105–1116
6. Dastin, J. Amazon scraps secret AI recruiting tool that showed bias against women. In Ethics of Data and Analytics; Auerbach Publications: Boca Raton, FL, USA, 2018; pp. 296–299. [Google Scholar]
7. Eubanks, V. Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor; St. Martin's Press: New York, NY, USA, 2018.
8. Ferrara, E. Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. First Monday 2023, 28.
9. Friedler, S.A.; Scheidegger, C.; Venkatasubramanian, S.; Choudhary, S.; Hamilton, E.P.; Roth, D. A comparative study of fairness-enhancing interventions in machine learning. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; pp. 329–338.
10. Kleinberg, J.; Lakkaraju, H.; Leskovec, J.; Ludwig, J.;

Mullainathan, S. Human decisions and machine predictions. Q. J. Econ. 2018, 133, 237–293.
11. Kleinberg, J.; Ludwig, J.; Mullainathan, S.; Rambachan, A. Algorithmic fairness. AEA Pap. Proc. 2018, 108, 22–27.
12. Kleinberg, J.; Ludwig, J.; Mullainathan, S.; Sunstein, C.R. Discrimination in the Age of Algorithms. J. Leg. Anal. 2018, 10, 113–174.
13. Kleinberg, J.; Mullainathan, S.; Raghavan, M. Inherent trade-offs in the fair determination of risk scores. In Proceedings of the Innovations in Theoretical Computer Science (ITCS), Berkeley, CA, USA, 9–11 January 2017.
14. O'Neil, C. Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy; Broadway Books: New York, NY, USA, 2016.
15. Yan, S.; Kao, H.T.; Ferrara, E. Fair class balancing: Enhancing model fairness without observing sensitive attributes. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online, 26 June–31 July 2020; pp. 1715–1724