Artificial Intelligence, Gender Bias, and Human Rights: Challenges for Modern Justice Systems

Elizabeta Imeraj[1]

Flutur Shabani[2]

Marco Cavina[3]

[1]Assoc. Prof. Dr., "Barleti" University, Tirana, Albania
[2]PhD (C.), "Barleti" University, Tirana, Albania
[3]Prof. Dr., University of Bologna, Bologna, Italy

Received: 17 March 2026 | ***Accepted: 23 October 2026*** | ***Published: 20 November 2026***

© 2026 Imeraj et al.

DOI: 10.56345/ijrdv13n319

Abstract

Artificial intelligence (AI) has become increasingly integrated into modern justice systems, influencing decision-making in criminal justice, hiring, surveillance, healthcare, and public administration. Although AI promises efficiency and consistency, a growing body of research documents risks of algorithmic bias and unequal performance across demographic groups. This paper examines the relationship between artificial intelligence, gender bias, and human rights within contemporary justice systems through a qualitative, interdisciplinary analysis combining a narrative review of scholarly literature with doctrinal analysis of relevant international and European legal and ethical frameworks. It considers how biased datasets, discriminatory or uneven algorithmic outcomes, and unequal technological representation may affect fairness, equality, privacy, due process, and access to justice. Particular attention is given to documented evidence on facial-analysis systems and automated recruitment, while broader claims concerning institutional legitimacy and human-centered justice are treated as normative implications rather than empirical findings. The paper also evaluates regulatory responses, including Regulation (EU) 2024/1689 (the EU Artificial Intelligence Act) and UNESCO's Recommendation on the Ethics of Artificial Intelligence. It concludes by identifying transparency, accountability, representative data, independent evaluation, and meaningful human oversight as central safeguards for rights-respecting AI governance.

Keywords: ***artificial intelligence, gender bias, human rights, justice systems, algorithmic discrimination, AI ethics***

1. Introduction

Artificial intelligence (AI) has transformed many sectors of society, including healthcare, finance, education, employment, and criminal justice. Governments and private institutions increasingly rely on AI-driven systems to automate decisions, analyze data, and improve efficiency. In justice systems, AI technologies are used for predictive policing, facial recognition, surveillance, sentencing recommendations, and risk assessment tools designed to support judicial decision-making. These systems are often presented as objective and efficient alternatives to human judgment because they can process large amounts of information quickly and identify patterns that may otherwise remain unnoticed. However, the

growing integration of AI into public institutions has also raised serious concerns regarding discrimination, accountability, fairness, and the protection of fundamental human rights. One of the most significant concerns surrounding AI involves gender bias embedded within algorithmic systems. Although AI is frequently described as neutral technology, algorithms are shaped by the data and assumptions used in their development. Since machine learning systems rely heavily on historical datasets, they often reproduce the inequalities already present in society. As a result, discrimination can become automated and normalized through technological processes that appear scientifically objective. This creates a particularly dangerous form of bias because discriminatory outcomes may be hidden behind the perception of technological neutrality. Documented examples illustrate how these risks can arise in practice. Amazon discontinued an experimental recruitment system after internal testing showed that a model trained on historical resumes learned patterns that disadvantaged applications containing indicators associated with women (Dastin, 2018). In facial analysis, Buolamwini and Gebru (2018) found substantial intersectional performance disparities across three commercial gender-classification systems: error rates for darker-skinned women reached as high as 34.7%, compared with a maximum of 0.8% for lighter-skinned men. These findings concern specific systems and tasks and should not be generalized to all AI applications; nevertheless, they demonstrate how unrepresentative data and model design can produce materially unequal error rates. These inaccuracies are especially concerning in law enforcement and surveillance contexts, where algorithmic errors may affect individual freedoms, reputations, and access to justice. The impact of gender bias extends beyond isolated technical errors and influences broader social and institutional structures. In criminal justice systems, predictive policing technologies may disproportionately target communities already affected by inequality because they rely on historical policing data shaped by discriminatory practices. Risk assessment systems used in sentencing or bail decisions may also reinforce existing inequalities by incorporating social factors linked to race, gender, or economic status. These mechanisms create a plausible risk that AI may reproduce or intensify pre-existing inequalities when historical patterns are encoded into data or proxy variables. The extent of that effect, however, depends on the particular system, dataset, institutional setting, and safeguards in place. Another important issue concerns the lack of diversity within the technology industry itself. AI systems are designed by human developers whose assumptions, experiences, and priorities influence technological outcomes. When women and marginalized groups remain underrepresented in AI research and development, there is a greater risk that discriminatory patterns will go unnoticed during system design and testing. This demonstrates that algorithmic bias is not only a technical problem but also a structural and institutional one connected to unequal representation and power within the technology sector. The relationship between AI and human rights has therefore become a major issue in contemporary legal and ethical debates. AI systems can threaten rights related to equality, privacy, dignity, and non-discrimination, particularly when used in surveillance and automated governance. AI-powered monitoring technologies enable unprecedented levels of data collection and social control, raising concerns about excessive surveillance and the erosion of personal autonomy. At the same time, opaque algorithmic systems undermine procedural fairness because individuals often cannot understand, challenge, or question decisions that directly affect their lives. Transparency and accountability remain central challenges in AI governance. Many algorithmic systems function as “black boxes,” meaning that their internal reasoning processes are difficult to interpret even by experts. This lack of explainability weakens accountability because institutions may rely on automated systems without fully understanding how decisions are produced. As a result, victims of algorithmic discrimination may struggle to seek legal remedies or prove that unfair treatment has occurred. The growing dependence on complex AI systems therefore raises broader concerns regarding democratic oversight and public trust in institutions. Recognizing these risks, governments and international organizations have begun developing legal and ethical frameworks aimed at regulating AI technologies. However, effective governance requires more than regulation alone. Addressing gender bias and protecting human rights in AI systems demands interdisciplinary cooperation between legal experts, policymakers, technologists, ethicists, and civil society organizations. Transparent algorithms, inclusive datasets, independent oversight, and meaningful human control are essential to ensuring that AI contributes to justice and equality rather than reinforcing existing forms of discrimination. The following analysis therefore distinguishes between documented empirical findings about particular AI systems and broader normative arguments about how justice institutions ought to govern AI. The latter are evaluated through human-rights and legal principles rather than presented as empirically established consequences.

2. Methodology

This study adopts a qualitative, interdisciplinary methodology combining a narrative literature review with doctrinal legal analysis. The literature-based component synthesizes peer-reviewed scholarship and selected institutional reports on gender bias, algorithmic discrimination, facial analysis, automated recruitment, risk assessment, transparency, and

human oversight. The legal component examines relevant international and European standards, including the Universal Declaration of Human Rights, the Convention on the Elimination of All Forms of Discrimination against Women, UNESCO's Recommendation on the Ethics of Artificial Intelligence, and Regulation (EU) 2024/1689 (the Artificial Intelligence Act). The paper is therefore conceptual and legal in orientation rather than an original empirical study: it does not generate a new dataset, test an algorithm, or estimate causal effects. Documented empirical findings are attributed to the studies from which they derive and are described within their specific scope. By contrast, propositions concerning dignity, legitimacy, democratic accountability, and the desirable role of human judgment are treated as normative legal and ethical arguments. This distinction is maintained throughout the analysis to avoid presenting broader policy implications as if they were direct empirical findings.

3. Artificial Intelligence and Gender Bias

Gender bias in artificial intelligence refers to systematic discrimination that disadvantages individuals based on gender identity or expression through automated decision-making systems. Such bias may emerge during several stages of AI development, including data collection, dataset labeling, algorithmic design, model training, testing, and deployment (Ferrara, 2024). Although AI systems are frequently presented as neutral and objective technologies, they are fundamentally shaped by the human choices, social structures, and historical data upon which they are built. Consequently, artificial intelligence can unintentionally reproduce and reinforce existing inequalities present within society. One of the primary sources of gender bias in AI originates from biased datasets. Machine learning algorithms rely heavily on historical information to identify patterns and make predictions. However, historical data often reflects longstanding social, economic, and institutional inequalities related to gender discrimination. When such biased information is integrated into AI systems without proper safeguards, discriminatory outcomes become normalized and automated (Lima et al., 2023). In this context, algorithms may perpetuate stereotypes while presenting decisions as technologically objective and scientifically accurate. A concrete example is Amazon's experimental recruitment model. Reporting based on interviews with people familiar with the project indicated that the system, trained on resumes submitted over a ten-year period, learned to penalize some indicators associated with women, including the term "women's," before the project was abandoned (Dastin, 2018). The case does not establish that all automated hiring systems discriminate against women; rather, it illustrates how historical training data can encode past workforce imbalances into model outputs. Facial recognition technologies provide another important example of gender bias in AI systems. Buolamwini and Gebru (2018) evaluated three commercial gender-classification systems and found that darker-skinned women were the most frequently misclassified subgroup, with error rates reaching 34.7%, whereas the maximum error rate for lighter-skinned men was 0.8%. Their study concerned commercial gender classification rather than every form of facial recognition. The relevance to justice and law-enforcement settings is therefore risk-based: where facial-analysis technologies are used in consequential contexts, unequal error rates can increase the possibility of misidentification and unequal treatment. The implications of gender bias become even more concerning when AI technologies are integrated into criminal justice systems. Governments increasingly use predictive policing algorithms, risk assessment tools, and automated surveillance systems to improve efficiency and reduce crime rates. Scholarly and policy literature has raised concerns that systems trained on historically generated policing or judicial data may reproduce patterns contained in those data (Rafanelli, 2022). This is best understood as a context-dependent risk rather than a universal outcome of predictive systems. Risk assessment algorithms used in sentencing and bail decisions also raise significant ethical and legal concerns. Nevertheless, scholars have demonstrated that algorithmic systems may indirectly incorporate gender, racial, and socioeconomic biases through proxy variables such as employment status, education, neighborhood, or family background (Chouldechova, 2017). Because these factors are often connected to broader social inequalities, algorithmic predictions may disadvantage already vulnerable populations. This creates tension between technological efficiency and the principle of equality before the law. Another critical factor contributing to gender bias in AI involves the underrepresentation of women and marginalized groups within technology industries and AI research communities. According to West et al. (2019), the lack of diversity among developers, engineers, and decision-makers influences how algorithms are designed, tested, and evaluated. Homogeneous development teams may fail to identify discriminatory outcomes affecting groups whose experiences differ from their own. Consequently, biases embedded within AI systems may remain unnoticed until harmful consequences emerge in real-world applications. The absence of gender diversity within technological development also affects broader institutional priorities. AI systems are often optimized according to efficiency and profitability objectives rather than fairness, inclusivity, or human rights protection. Feminist scholars argue that technological systems frequently reflect patriarchal power structures that prioritize dominant

social perspectives while marginalizing women and minority voices (Criado-Perez, 2019). Therefore, addressing gender bias in AI requires not only technical improvements but also structural changes within the institutions responsible for technological innovation. Transparency and accountability remain central challenges in addressing algorithmic discrimination.

4. Human Rights Implications

Gender bias in artificial intelligence directly affects several internationally recognized human rights, including the rights to equality, non-discrimination, privacy, dignity, autonomy, and access to justice. As AI technologies increasingly shape decisions within public institutions and private sectors, concerns have emerged regarding their compatibility with international human rights standards. Although artificial intelligence is frequently promoted as a neutral and objective tool capable of improving efficiency and consistency, growing evidence demonstrates that algorithmic systems may perpetuate structural inequalities and reinforce discriminatory practices against women and marginalized communities (Ferrara, 2024). From a legal perspective, automated decisions that produce unjustified differential treatment may engage equality and non-discrimination guarantees, depending on the applicable legal framework, the protected ground, and the circumstances of the decision. AI-driven systems used in employment, education, healthcare, banking, and criminal justice can reinforce historical patterns of exclusion and inequality. Hiring algorithms, for example, may prioritize male candidates if trained on historical employment data reflecting male-dominated workplaces (West et al., 2019). Similarly, healthcare algorithms trained using incomplete or biased datasets may fail to account for gender-specific medical needs, leading to unequal treatment and reduced quality of care for women (Criado-Perez, 2019). Such outcomes demonstrate that algorithmic systems can institutionalize discrimination under the appearance of technological objectivity. The human rights implications become particularly serious when AI technologies are integrated into justice systems and law enforcement practices. Predictive policing tools, facial recognition technologies, and automated surveillance systems are increasingly used to identify potential threats, monitor populations, and support criminal investigations. However, these technologies often disproportionately target marginalized communities because they rely on historical data shaped by discriminatory policing practices (Rafanelli, 2022). Women, ethnic minorities, migrants, and economically disadvantaged populations may therefore experience heightened surveillance and unequal treatment. Facial recognition technologies present one of the clearest examples of how AI can threaten both equality and privacy rights. Buolamwini and Gebru (2018) documented marked accuracy disparities in commercial gender-classification systems, with the highest error rates affecting darker-skinned women. The study itself did not measure downstream legal harms such as detention or denial of services. Those consequences should therefore be understood as foreseeable risks when inaccurate biometric systems are deployed in high-stakes contexts, rather than as outcomes established by that particular study. This distinction is important when connecting technical performance evidence to human-rights analysis.

Privacy rights are also increasingly threatened by AI-powered surveillance systems. The widespread collection and processing of biometric data, personal information, online activity, and behavioral patterns have raised concerns regarding mass surveillance and excessive state control. AI technologies can monitor individuals continuously through facial recognition cameras, social media analysis, predictive analytics, and geolocation tracking systems. While governments often justify these practices in the name of public safety and national security, critics argue that excessive surveillance undermines the right to privacy and creates chilling effects on freedom of expression and association (UNESCO, 2021). Women and marginalized communities may be disproportionately affected by AI surveillance technologies because they are often subject to greater social monitoring and institutional control. For example, predictive policing systems trained on biased arrest records may repeatedly target specific neighborhoods or demographic groups, thereby reinforcing cycles of surveillance and criminalization. This creates what scholars describe as “feedback loops,” in which discriminatory practices generate biased data that further justifies discriminatory policing strategies (Chouldechova, 2017). Consequently, AI systems may perpetuate systemic inequality while appearing scientifically neutral and objective. Another major human rights concern involves transparency and accountability in algorithmic decision-making. This lack of explainability presents serious challenges for justice systems because procedural fairness requires that individuals understand how decisions affecting their rights and freedoms are made. In democratic legal systems, transparency is essential for ensuring accountability, due process, and access to justice. Individuals subjected to automated decisions must have the ability to challenge unfair outcomes and seek effective legal remedies. However, when AI systems are opaque or highly complex, affected individuals may struggle to identify whether discrimination has occurred or how decisions were reached. This undermines procedural justice because accountability mechanisms become limited or ineffective.

The issue of accountability is particularly problematic when governments and private corporations rely on proprietary AI technologies developed by external companies. In many cases, algorithms used in criminal justice, hiring, or financial services are protected as trade secrets, preventing public scrutiny and independent evaluation. Critics argue that this creates a dangerous imbalance between technological power and democratic oversight because institutions may deploy systems that significantly affect human rights without sufficient transparency or public accountability (West et al., 2019). Furthermore, the use of AI in judicial decision-making raises important philosophical and ethical concerns regarding the role of human judgment in justice systems. Algorithms can process data and identify patterns, but they lack moral reasoning, empathy, and contextual understanding. Justice requires more than efficiency or statistical prediction; it also involves ethical considerations related to fairness, proportionality, dignity, and social responsibility (Rafanelli, 2022). Excessive reliance on automated systems may therefore undermine the human-centered foundations of democratic legal institutions. Ultimately, the human rights implications of gender bias in AI demonstrate that technological innovation cannot be separated from ethical responsibility and democratic values. Artificial intelligence has the potential to improve public services and increase efficiency, but without appropriate safeguards, it may also reinforce structural discrimination, undermine procedural fairness, and weaken public trust in legal institutions. Ensuring that AI systems remain consistent with human rights principles is therefore essential for building just, inclusive, and accountable societies in the digital era.

5. Challenges for Modern Justice Systems

Modern justice systems face growing difficulties in regulating artificial intelligence while preserving fairness, accountability, transparency, and fundamental human rights. Although AI technologies promise greater efficiency and consistency in legal decision-making, their rapid integration into courts, law enforcement, and public administration has introduced complex ethical and legal concerns. Predictive policing systems, facial recognition technologies, sentencing algorithms, surveillance tools, and automated risk assessments are increasingly used to support judicial and administrative decisions. However, the speed of technological development has often exceeded the capacity of legal systems to establish effective safeguards and oversight mechanisms (Ferrara, 2024). As a result, justice institutions are confronted with the challenge of balancing innovation with democratic values and the protection of individual rights. One of the most significant concerns surrounding AI governance is the lack of transparency in algorithmic decision-making. Many machine learning systems operate as “black boxes,” meaning that their internal reasoning processes are difficult to interpret even by experts (Rudin, 2019). While these systems can analyze large volumes of data quickly, they frequently fail to explain how specific conclusions or recommendations are reached. In legal contexts, this creates serious problems because procedural fairness requires that decisions affecting individuals’ rights and freedoms be understandable and open to challenge. If defendants, lawyers, or judges cannot fully understand how an algorithm reached a decision, meaningful accountability becomes difficult to achieve. This challenge is intensified by the limited technical expertise within many judicial institutions. Judges, lawyers, and policymakers may not possess the specialized knowledge necessary to evaluate whether algorithms operate fairly or whether datasets contain hidden discriminatory assumptions (Rafanelli, 2022). Consequently, algorithmic decisions may acquire excessive authority simply because they appear technologically sophisticated or statistically objective. In practice, this can create a form of “automation bias,” where human decision-makers place undue trust in AI-generated recommendations even when those systems contain significant errors or biases. Transparency problems are further complicated by the role of private technology companies in developing AI systems. Many algorithms used in criminal justice and public administration are protected as proprietary trade secrets, limiting independent scrutiny and public oversight. Courts and governments may therefore rely on systems whose internal mechanisms remain inaccessible to defendants, researchers, or civil society organizations. Critics argue that such opacity undermines democratic accountability and weakens public trust in justice institutions because citizens cannot effectively challenge decisions made through hidden computational processes (West et al., 2019). In democratic societies, justice must not only be fair but also appear fair, and opaque technologies risk undermining this principle. Another major challenge involves the absence of comprehensive legal frameworks capable of regulating AI accountability effectively. Most anti-discrimination laws and constitutional protections were developed before the emergence of modern machine learning technologies and therefore often assume discrimination results from intentional human actions rather than automated computational systems (UNESCO, 2021). However, algorithmic discrimination frequently occurs indirectly through biased data patterns, proxy variables, or unequal statistical outcomes. Existing legal systems are not always prepared to address these new forms of discrimination, creating uncertainty regarding liability and accountability. Although ethical principles such as fairness, transparency, and human oversight are widely recognized, many countries have not yet translated these standards into binding and enforceable legislation. Regulation (EU) 2024/1689, the

European Union Artificial Intelligence Act, establishes a risk-based regulatory framework. It classifies specified uses in areas including employment, law enforcement, access to essential services, migration and border control, and the administration of justice as high-risk when the conditions in the Regulation are met. In the justice context, the Regulation expressly recognizes risks of bias, error, and opacity and provides that AI may support judicial decision-making but should not replace the human decision-maker. Nevertheless, even advanced regulatory models face implementation difficulties because technological development evolves faster than legislative processes. Governments often struggle to update laws quickly enough to address emerging risks and applications. Global disparities in AI governance create additional complications. Some states prioritize economic competitiveness, national security, or technological innovation over human rights protections, resulting in uneven regulatory standards across jurisdictions. This fragmented landscape allows corporations and governments to exploit regulatory gaps or deploy controversial technologies with limited oversight. As AI systems increasingly operate across borders, the absence of internationally coordinated standards makes effective governance more difficult and weakens protections against discrimination and abuse. Biased training data remains one of the most persistent challenges in AI governance. Machine learning systems depend on large datasets to identify patterns and generate predictions, yet these datasets often reflect historical inequalities connected to gender, race, ethnicity, and socioeconomic status. When biased historical information is incorporated into AI systems, discriminatory outcomes may become automated and normalized (Lima et al., 2023). Rather than eliminating bias, AI may reinforce and legitimize pre-existing inequalities under the appearance of scientific neutrality. The consequences of biased data are especially visible in predictive policing systems. These technologies analyze historical crime and arrest records to predict where crimes may occur or who may be involved in criminal activity. However, because policing data already reflects patterns of over-surveillance in marginalized communities, predictive systems can reproduce patterns present in the records on which they are trained. The broader literature on algorithmic fairness shows that predictive accuracy and group-level fairness criteria can conflict, especially where underlying outcome rates differ (Chouldechova, 2017). In policing contexts, this supports concern about feedback effects, but the magnitude and direction of such effects must be established for the particular system and deployment setting rather than assumed in advance. In this way, AI can unintentionally perpetuate structural inequalities while appearing objective and evidence-based. Bias in AI systems also affects women and minority populations through underrepresentation in training datasets. Researchers have demonstrated that women, racial minorities, migrants, and economically disadvantaged groups are often insufficiently represented in data used to train algorithms. This underrepresentation contributes to higher error rates and discriminatory outcomes in technologies such as facial recognition, healthcare diagnostics, and employment screening systems. These problems reveal that AI systems are not socially neutral technologies but products of broader institutional inequalities embedded within society itself. Another significant challenge concerns the reduction of human judgment within justice systems. Although AI can process information quickly and identify statistical correlations, it cannot fully understand social context, moral responsibility, or human experience. Justice involves more than efficiency and prediction; it requires empathy, ethical reasoning, and individualized consideration of circumstances (Rudin, 2019). Excessive reliance on automated systems risks transforming individuals into data profiles or risk categories rather than recognizing their dignity and unique social realities. This concern is particularly important in criminal justice contexts where AI influences sentencing, bail decisions, parole evaluations, and risk assessments. Algorithms may encourage standardized and data-driven decision-making at the expense of judicial discretion and contextual understanding. Overreliance on statistical predictions may also reinforce deterministic assumptions about human behavior, reducing opportunities for rehabilitation and individualized justice. In this sense, the expansion of AI within justice systems raises broader philosophical questions regarding the appropriate relationship between technology, law, and human judgment. Furthermore, public trust in justice institutions may decline if citizens perceive algorithmic systems as unfair, discriminatory, or insufficiently accountable. Trust is essential for the legitimacy of democratic legal systems, and perceptions of hidden bias or technological injustice can weaken confidence in courts and public institutions. Ensuring accountability and transparency is therefore not only a legal necessity but also a democratic imperative.

6. International and Legal Responses

As artificial intelligence expands across public and private sectors, governments and international organizations have increasingly sought legal and ethical frameworks to regulate AI systems and reduce algorithmic discrimination. Regulation (EU) 2024/1689 (the Artificial Intelligence Act) is particularly relevant to justice systems because it treats certain AI uses in the administration of justice as high-risk and subjects high-risk systems to requirements concerning risk management, data governance, documentation, transparency, human oversight, accuracy, robustness, and

cybersecurity. This legal classification reflects a regulatory judgment about potential impacts on fundamental rights; it should not be read as empirical proof that every judicial AI system causes discrimination. Implementation nevertheless remains challenging. Rapid technological change can outpace legal updates, while enforcement requires significant institutional capacity and technical expertise (Floridi, 2023). There are also concerns that large technology companies may be better positioned than governments to comply with complex regulations, creating imbalances in oversight. Alongside regional laws, international organizations have developed ethical frameworks for AI governance. UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) emphasizes principles such as human dignity, fairness, transparency, inclusivity, accountability, and sustainability. It also calls for human rights impact assessments, protection of vulnerable groups, and greater diversity in AI development teams to reduce systemic bias. The framework further highlights the importance of human oversight, ensuring AI supports rather than replaces human decision-making in sensitive areas such as justice and public services. Similarly, the United Nations, through the OHCHR, has warned that AI may threaten privacy, equality, and freedom of expression if insufficiently regulated, and has called for restrictions on high-risk technologies such as facial recognition in public surveillance (OHCHR, 2021). Effective AI governance requires interdisciplinary cooperation among legal experts, computer scientists, ethicists, policymakers, and civil society (Floridi & Cowls, 2019). While technologists focus on fairness and explainability, legal and ethical experts help translate these principles into enforceable regulations. Human-centered AI governance emphasizes that systems should prioritize dignity, transparency, accountability, and inclusivity over efficiency alone. Transparency is essential to ensure that individuals can understand and challenge algorithmic decisions, while explainability supports fairness and public trust (Rudin, 2019). Inclusivity is also crucial, as underrepresentation of women and marginalized groups in AI development contributes to biased outcomes (West et al., 2019). Increasing diversity in AI teams helps reduce these risks. Despite these efforts, global coordination remains limited. Countries differ in regulatory priorities, and multinational corporations often exert strong influence over AI governance, raising concerns about accountability and democratic control (Zuboff, 2019). These fragmented approaches make it difficult to establish consistent global standards. In this sense, frameworks such as the EU AI Act and UNESCO guidelines represent important developments, although significant challenges remain in implementation, enforcement, and international coordination. Effective AI governance requires continuous human oversight, inclusive development, and strong regulatory systems to ensure that technological innovation aligns with human rights, equality, and democratic values.

7. Conclusion

Artificial intelligence has become an increasingly influential part of modern justice systems and public institutions, shaping decision-making in areas such as criminal justice, employment, healthcare, surveillance, and public administration. While AI technologies offer important benefits, including efficiency, speed, and data-processing capabilities, the literature and legal frameworks reviewed in this study identify significant risks related to gender bias, discrimination, and the protection of human rights. Algorithmic systems are not inherently neutral or objective because they are developed using historical data and social structures that often reflect existing inequalities. As a result, AI can reproduce and even intensify patterns of discrimination against women and marginalized groups. The reviewed literature identifies multiple mechanisms through which gender bias may emerge, including biased datasets, opaque algorithmic systems, unequal technological representation, and insufficient regulatory oversight. Documented studies show unequal outcomes in particular applications, including commercial gender classification and automated recruitment, while scholarship on policing and judicial risk assessment identifies additional context-dependent risks associated with historical data and proxy variables. These outcomes undermine fundamental human rights principles such as equality, dignity, privacy, non-discrimination, and access to justice. Particularly in justice systems, excessive reliance on automated decision-making may weaken procedural fairness and reduce trust in democratic institutions. The study also highlighted that the challenges associated with AI governance cannot be resolved solely through technical solutions. Addressing algorithmic discrimination requires broader structural reforms involving interdisciplinary cooperation among policymakers, legal scholars, technologists, ethicists, and civil society organizations. Improving dataset diversity and representativeness is essential because AI systems learn from historical patterns that may contain social and institutional biases. At the same time, transparency and explainability remain crucial for ensuring accountability. Many AI systems continue to function as "black boxes," making it difficult for individuals to understand or challenge decisions affecting their rights and freedoms. Independent audits, explainability standards, and public oversight mechanisms are therefore necessary to strengthen fairness and democratic accountability. As a normative legal and governance principle, meaningful human oversight is also central to the frameworks examined here. Although AI systems can support decision-making processes, they should not replace

human judgment in areas involving legal rights, liberty, and social welfare. Maintaining meaningful human control over AI-assisted decisions is therefore essential for protecting procedural fairness and human dignity. However, important challenges remain regarding implementation, enforcement, and international coordination. Rapid technological development frequently outpaces legal systems, while multinational technology corporations continue to exert significant influence over AI development and governance. Ultimately, the future impact of artificial intelligence will depend on how societies choose to regulate and govern these technologies. Without strong legal safeguards, inclusive technological development, and continuous human rights evaluation, AI systems risk reinforcing structural inequalities and weakening democratic values. Conversely, responsible governance can ensure that artificial intelligence contributes to more equitable, transparent, and accountable institutions. In this sense, protecting equality, dignity, and fairness must therefore remain central priorities in the development and application of AI technologies within modern justice systems and society as a whole.

8. AI Use Statement

The author(s) declare that artificial intelligence (AI) tools were used during the preparation of this manuscript. Specifically, AI-assisted tools were used for language editing, grammatical correction, and improvements in clarity and readability. The intellectual content, analysis, interpretation of results, and final approval of the manuscript remain the sole responsibility of the author(s).

References

Ayata, Z. (2025). Gender bias in AI systems: A critical analysis of regulatory frameworks and policy responses. Białostockie Studia Prawnicze. https://bsp.uwb.edu.pl/article/view/1734

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15.

Chauhan, P., & Kaur, G. (2022). Gender bias and artificial intelligence: A challenge within the periphery of human rights. Hasanuddin Law Review, 8(1), 46–59. https://doi.org/10.20956/halrev.v8i1.3569

Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big Data, 5(2), 153–163. https://doi.org/10.1089/big.2016.0047

Criado-Perez, C. (2019). Invisible women: Data bias in a world designed for men. Abrams Press.

Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters.

European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

European Parliament. (2024). Artificial Intelligence Act: Parliament adopts landmark law. European Parliament. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

Ferrara, E. (2024). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. Sci, 6(1), 3. https://doi.org/10.3390/sci6010003

Floridi, L. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities. Oxford University Press.

Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

Ho, J. Q. H., Hartanto, A., Koh, A., & Majeed, N. M. (2025). Gender biases within artificial intelligence and ChatGPT: Evidence, sources of biases and solutions. Computers in Human Behavior: Artificial Humans, 4, 100145.

Lima, R. M. de, Pisker, B., & Corrêa, V. S. (2023). Gender bias in artificial intelligence: A systematic review of the literature. Journal of Telecommunications and the Digital Economy. https://jtde.telsoc.org/index.php/jtde/article/view/690

Lütz, F. (2022). Gender equality and artificial intelligence in Europe: Addressing direct and indirect impacts of algorithms on gender-based discrimination. ERA Forum, 23, 33–52. https://doi.org/10.1007/s12027-022-00709-6

Office of the United Nations High Commissioner for Human Rights (OHCHR). (2021). The right to privacy in the digital age. United Nations. https://www.ohchr.org/

Rafanelli, L. M. (2022). Justice, injustice, and artificial intelligence: Lessons from political theory and philosophy. Big Data & Society. https://doi.org/10.1177/20539517221080676

Temperman, J., & Quintavalla, A. (Eds.). (2023). Artificial intelligence and human rights. Oxford University Press.

The use of artificial intelligence in the administration of justice: Suggested framework of ethical principles and reasoning of judges in the use of intelligent systems. (2026). Laws, 15(2), 20. https://doi.org/10.3390/laws15020020

United Nations. (1948). Universal Declaration of Human Rights. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations. (1979). Convention on the Elimination of All Forms of Discrimination Against Women. https://www.un.org/womenwatch/daw/cedaw/

Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

West, S. M., Whittaker, M., & Crawford, K. (2019). Discriminating systems: Gender, race, and power in AI. AI Now Institute. https://ainowinstitute.org/reports.html

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. Public Affairs.