Int J Ethics Soc 2026; 7(4): 1-11

10.22034/ijethics.7.4.424

INTERNATIONAL JOURNAL OF Ethics & Society

Review Article

Artificial Intelligence and Ethical Decision-Making: An Analysis of the Social Implications of Algorithms in Justice and Discrimination

Maya Bialik[1*], Charles Fadel[1], Wayne Holmes[2]
1. Department of Education, Center for Curriculum Redesign, Boston University, USA.
2. UCL Institute of Education, University College London, London, UK.

Corresponding Author: Maya Bialik, Department of Education, Center for Curriculum Redesign, Boston University, USA. E-mail: mbialik@bu.edu

Received 16 Sep 2025 **Accepted** 25 Oct 2025 **Online Published** 04 Jan 2026

Abstract

Introduction: The increasing proliferation of artificial intelligence systems and algorithmic decision-making in domains such as criminal justice, employment, resource allocation, and public policymaking has brought about wide-ranging ethical and social implications. Although these systems are often deployed with the aim of enhancing efficiency and neutrality, scientific research shows that algorithms can reproduce or intensify structural discrimination and historical inequalities. This review article adopts an interdisciplinary approach to survey the scholarly literature on the ethics of algorithmic decision-making and to analyze the role of algorithms in justice, discrimination, transparency, responsibility, and accountability. The article focuses on explicating ethical theoretical frameworks, examining social and political consequences, and identifying gaps and contradictions in prior research.
Material and Methods: It is a narrative review article that focuses on the existing scholarly literature to analyze the ethical dimensions of algorithmic ethical decision-making. Previous studies were analyzed and then the subject was explained.
Conclusion: The findings of this review indicate that addressing the ethical challenges of artificial intelligence requires moving beyond purely technical solutions and paying serious attention to social, institutional, and political contexts. Lack of transparency, diffusion of responsibility, and weak accountability mechanisms can lead to the legitimization of unjust algorithmic decisions. The article concludes that the development and deployment of fair artificial intelligence necessitate ethics-centered design, stakeholder participation, effective institutional oversight, and interdisciplinary, context-sensitive research, so that this technology can serve to reduce discrimination and promote social justice.
Keywords: Artificial Intelligence, Ethical Decision-Making, Justice, Discrimination

How to Cite: Bialik M, Fadel C, Holmes W. Artificial Intelligence and Ethical Decision-Making: An Analysis of the Social Implications of Algorithms in Justice and Discrimination, Int J Ethics Soc. 2026;7(4): 1-11. doi: 10.22034/ijethics.7.4.424

INTRODUCTION

With the rapid expansion of artificial intelligence technologies, algorithms are increasingly employed in social, economic, and legal decision-making processes. Today, these systems play a significant role in areas such as human resource recruitment, the criminal justice system, credit and loan allocation, medical diagnosis, and social surveillance. Although artificial intelligence has been introduced with the promise of increasing efficiency, accuracy, and neutrality, recent research demonstrates that algorithms are not neutral; rather, they can reproduce or even exacerbate existing social inequalities and forms of discrimination [1].

Copyright © 2026 Bialik M. et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License(http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Algorithmic decision-making is directly intertwined with fundamental ethical concepts such as justice, fairness, accountability, and transparency. The data used to train artificial intelligence systems often reflect historical unequal structures and social biases; as a result, algorithms can reproduce these biases in a systematic and large-scale manner [2]. From this perspective, artificial intelligence should be understood not merely as a technical tool, but as a socio-technical phenomenon with profound ethical and political consequences.

Ethical challenges related to artificial intelligence have attracted considerable attention in the philosophy of technology and interdisciplinary studies. Researchers emphasize that the lack of transparency in algorithmic decision logic (the "black box" problem), the difficulty of assigning moral and legal responsibility, and the disproportionate impact of these technologies on marginalized groups constitute some of the most pressing contemporary concerns in this field [3, 4]. Particularly in criminal justice systems and public policymaking, the use of algorithms can have serious consequences for individuals' fundamental rights.

Accordingly, the present article focuses on the existing scholarly literature to analyze the ethical dimensions of algorithmic decision-making and to examine the social implications of artificial intelligence in reproducing or mitigating justice and discrimination. The aim of this article is to identify theoretical frameworks, empirical findings, and key challenges in the field of AI ethics, in order to provide a more comprehensive understanding of the role of algorithms in shaping power structures and inequalities in contemporary societies.

MATERIAL AND METHODS

It is a narrative review article that focuses on the existing scholarly literature to analyze the ethical dimensions of algorithmic ethical decision-making. Previous studies were analyzed and then the subject was explained.

DISCUSSION

Ethical Analysis of Algorithmic Decision-Making Based on Theoretical Frameworks

The expansion of algorithmic decision-making in socially sensitive domains such as criminal justice, employment, financial resource allocation, and public services has raised fundamental questions about the ethical legitimacy of these systems. Unlike human decision-making, which is formed within social, cultural, and normative contexts, algorithmic decision-making is based on statistical models and machine learning techniques that often lack contextual understanding of human values. Therefore, ethical analysis of this form of decision-making requires drawing on normative ethical and applied ethics frameworks in order to systematically examine its consequences, limitations, and value conflicts [4].

- Utilitarianism and Algorithmic Optimization: One of the most common implicit frameworks in the design of artificial intelligence systems is utilitarianism. In this approach, an ethical action is one that produces the greatest benefit for the greatest number of people. Algorithms are often optimized based on criteria such as accuracy, efficiency, and error reduction, which align with utilitarian logic [3]. However, critics argue that an exclusive focus on aggregate outcomes can lead to the neglect of the rights of minorities and vulnerable groups. For example, a crime risk prediction algorithm may exhibit high overall accuracy, yet in practice result in systematic discrimination against specific racial groups. From a utilitarian perspective, such discrimination might be considered justifiable, whereas from an ethical standpoint it constitutes a violation of justice and equality [1].

- Deontology and Fundamental Individual Rights: The deontological framework, rooted in Kantian philosophy, emphasizes adherence to moral principles and duties independent of outcomes. From this perspective, ethical decision-making must always respect human dignity, autonomy, and individuals' fundamental rights. In the context of algorithmic decision-making, this framework raises essential questions: Do individuals have the right to know the logic behind decisions made about them? Is the use of algorithms without individuals' consent or awareness ethically acceptable? [5] One of the main challenges within this framework is the "black box" problem. Complex deep learning algorithms often lack transparency and limit explainability of decisions. This conflicts with the deontological principle of respect for individual autonomy, as people cannot meaningfully challenge decisions that affect their lives [4].
- Virtue Ethics and the Role of Human Actors: Virtue ethics focuses not on rules or consequences, but on the character and virtues of moral agents. In the context of artificial intelligence, this approach shifts attention from the algorithm itself to the designers, developers, and organizations that deploy it. The central question is which moral virtues-such as responsibility, fairness, and prudence-should be institutionalized in the design and use of algorithms [6]. This framework highlights that algorithms are not inherently ethical or unethical; rather, they reflect human values and decisions. Consequently, focusing solely on technical fixes without considering organizational culture and economic incentives will not be sufficient to resolve ethical issues.
- Algorithmic Justice and Contemporary Theories of Justice: In recent years, the concept of "algorithmic justice" has emerged as an interdisciplinary framework inspired by theories of social justice, particularly the ideas of John Rawls. This approach emphasizes that algorithmic decision-making should be designed in ways that reduce inequalities or at least do not exacerbate them [2]. Studies show that many technical definitions of fairness-such as statistical parity or error balance-cannot by themselves guarantee social justice, because they overlook historical and structural contexts of discrimination. Therefore, ethical analysis of algorithmic decision-making requires integrating technical criteria with broader social and ethical considerations [7].
- Limitations of Classical Theoretical Frameworks: Although utilitarianism, deontology, and virtue ethics provide important conceptual tools for ethical analysis, none of them alone can fully address the complexities of algorithmic decision-making. Algorithmic decisions are often collective, impersonal, and large-scale-characteristics that classical ethical theories have paid relatively little attention to. For this reason, many researchers emphasize the need to develop hybrid and applied ethical frameworks that can simultaneously address consequences, individual rights, and institutional virtues [3].

Reproduction or Intensification of Structural Discrimination by Algorithms

The widespread deployment of algorithms and artificial intelligence–based systems in social decision-making has reinforced the perception that technology can serve as an objective, neutral, and error-free substitute for human judgment. However, a growing body of scientific evidence indicates that algorithms are not only non-neutral but, in many cases, contribute to the reproduction or even intensification of existing structural discriminations within societies. These forms of discrimination are rooted in historical,

social, economic, and cultural inequalities and are transmitted into algorithmic systems through data, computational models, and institutional contexts [1].

Structural discrimination refers to persistent and systematic patterns of inequality that are embedded in institutions, laws, and social practices. Algorithms are typically trained on historical data-data that are themselves products of these unequal structures. Consequently, when algorithms identify and generalize existing statistical patterns in the data, they effectively carry past discrimination into the future. This phenomenon, commonly referred to as "data-driven bias," constitutes one of the most significant mechanisms through which algorithmic discrimination is reproduced [7].

Research across various domains has demonstrated that algorithms can generate tangible discriminatory outcomes. For instance, in the criminal justice system, risk assessment algorithms used to predict recidivism often assign higher risk scores to individuals from specific racial groups, even when their actual behavior does not differ meaningfully from that of other groups. This not only leads to unjust judicial decisions but also reinforces existing social stereotypes [8]. Similarly, in the employment domain, résumé-screening algorithms may unconsciously exclude women or ethnic minorities because they rely on past patterns of success that have often been male-centered or elitist.

Another factor that amplifies structural discrimination through algorithms is the scalability and automation of decision-making. Unlike human decisions, which are limited and case-specific, algorithmic decisions can be applied broadly and continuously. This characteristic means that even minor biases in design or data can produce wide-ranging and systematic social consequences. In other words, algorithms do not merely reproduce discrimination; they stabilize and institutionalize it [4].

From a critical perspective, some scholars argue that a narrow focus on technical fixes-such as optimizing statistical fairness metrics-is insufficient to address structural discrimination. Such approaches often neglect social contexts, power relations, and the historical dimensions of inequality. Noble [2] demonstrates how search engines, through racist and sexist representations, contribute to the reproduction of dominant ideologies-a phenomenon rooted in the economic and political structures of the technology industry rather than merely in technical errors.

Moreover, algorithms can conceal and legitimize discrimination by creating an "illusion of objectivity." Decisions made by automated systems are often presented as scientific and impartial outcomes, while in reality they are underpinned by a set of human choices, normative assumptions, and institutional interests. This dynamic reduces opportunities for contestation and accountability and shifts structural discrimination from a visible social level to a hidden technical one [9].

Overall, the scholarly literature shows that algorithms do not operate in a vacuum; rather, they are developed and deployed within unequal social structures. Without critical attention to these contexts, the use of artificial intelligence can reinforce existing inequalities and generate new forms of discrimination. Addressing algorithmic discrimination therefore requires an approach that goes beyond technical adjustments-one that places the analysis of power structures, the participation of marginalized groups, and ethics-centered regulation at the forefront.

The Issue of Transparency, Responsibility, and Accountability in Algorithmic Decision-Making

The expanding use of algorithmic decision-making systems in socially sensitive domains has raised fundamental questions concerning transparency, responsibility, and accountability. In many cases, algorithms are employed in processes with direct and profound consequences for individuals' rights, opportunities, and social standing, including criminal justice, public service allocation, credit assessment, and employment. Yet the internal logic of these systems is often not fully understandable to users, regulatory bodies, or even their developers-a situation commonly described as the "black box problem" [9].

Transparency, as a core principle of AI ethics, refers to the ability to understand, explain, and scrutinize the logic underlying algorithmic decisions. The absence of transparency not only undermines public trust in these technologies but also facilitates the reproduction of discrimination and inequality, as biased decisions can be hidden behind technical complexity. Studies show that many machine learning algorithms, particularly deep learning models, are designed in ways that severely limit interpretability, thereby making ethical and legal evaluation of their outcomes difficult [10].

The lack of transparency is directly linked to the issue of responsibility. In human decision-making, it is usually possible to identify the responsible agent and hold them accountable in cases of error or discrimination. In algorithmic decision-making, however, a chain of actors-including algorithm designers, data providers, deploying organizations, and policymaking institutions-is involved. This diffusion of responsibility poses a serious challenge to determining who should be held accountable for unjust or discriminatory outcomes [4]. As a result, a form of "responsibility gap" emerges, which can allow social harms to persist without effective avenues for redress.

Accountability, as a complement to transparency and responsibility, requires that individuals and institutions affected by algorithmic decisions be able to challenge those decisions, demand explanations, and, when necessary, seek correction or compensation. However, many automated systems are implemented in ways that deprive users of meaningful opportunities for contestation. Researchers [5] show that even within advanced legal frameworks-such as the European Union's General Data Protection Regulation-the practical realization of a "right to explanation" faces significant conceptual and technical limitations.

From a social justice perspective, the absence of transparency and accountability has uneven consequences for marginalized groups. These groups are often more frequently subjected to automated decisions while simultaneously having less access to legal, technical, and institutional resources for contestation and redress. Consequently, algorithms can entrench existing inequalities in power and access, further deepening the gap between decision-makers and those subject to decisions [2].

In response to these challenges, various ethical and policy frameworks have emphasized the necessity of developing responsible artificial intelligence. These frameworks identify transparency, responsibility, and accountability as key principles that must be embedded throughout the entire lifecycle of an algorithm-from design and training to deployment and evaluation [3]. Nonetheless, critics caution that an exclusive focus on technical solutions such as "explainable AI," without reforming institutional and economic structures, cannot by itself guarantee justice and accountability.

In sum, the scholarly literature indicates that issues of transparency, responsibility, and accountability in algorithmic decision-making are not merely technical challenges, but are deeply rooted in power relations, organizational

interests, and social structures. Effectively addressing these issues requires an interdisciplinary approach that integrates ethics, law, social sciences, and engineering, and that critically evaluates algorithms as social actors with serious ethical consequences rather than relying unquestioningly on their presumed objectivity.

Research Gaps and Contradictions in the Literature on Artificial Intelligence and Ethical Decision-Making

Despite the significant growth of research in the field of AI ethics and algorithmic decision-making, the existing literature continues to face substantial conceptual, methodological, and empirical gaps that hinder the development of a comprehensive and coherent understanding of the social implications of these technologies. One of the most important research gaps is the lack of alignment between technical studies and social-critical research. While computer science research primarily focuses on developing quantitative metrics to measure fairness, reduce bias, or increase algorithmic accuracy, studies in the social sciences and applied ethics emphasize historical contexts, power relations, and the structural consequences of algorithmic decision-making. These two research streams often proceed in parallel without effective dialogue, resulting in fragmented and sometimes inconsistent findings [4].

Another major gap concerns the limited empirical scope of existing studies. A substantial portion of the literature focuses on specific case studies or simulated data and pays less attention to examining the long-term and real-world impacts of algorithm use across diverse social contexts. This issue is particularly significant for marginalized groups, as the discriminatory effects of algorithms often emerge gradually and cumulatively and are not easily observable in short-term studies [2]. Consequently, there remains a lack of clear understanding of the structural and intergenerational effects of algorithmic decision-making.

A notable conceptual contradiction is also evident in the definition and interpretation of "algorithmic justice." In technical literature, justice is often defined as a set of statistical criteria-such as error rate parity or predictive balance-whereas in philosophical and social literature, justice is understood as a normative and context-dependent concept tied to histories of inequality, power distribution, and individuals' fundamental rights. This divergence in the understanding of justice has led to situations in which solutions proposed in some technical studies are socially insufficient or even problematic [1].

From a geographical and cultural perspective, there are also significant gaps in the literature. Most empirical and policy-oriented studies on AI ethics have been conducted in developed countries, particularly the United States and Europe. This geographical concentration has resulted in limited attention to the cultural, legal, and economic contexts of other regions of the world-especially developing countries. While algorithms are deployed globally, their consequences are highly dependent on social and institutional contexts, and generalizing existing findings without regard to these differences can be misleading [3].

Another contradiction evident in the literature is the tension between technological optimism and critical approaches. Some studies present artificial intelligence as a potential tool for reducing human bias and increasing fairness, whereas critical research demonstrates that the same technologies can reinforce and legitimize existing forms of discrimination. These two narratives are often presented without sufficient efforts at integration or dialogue, and relatively few studies have examined the conditions under

which artificial intelligence can genuinely contribute to the promotion of social justice [9].

Finally, another significant gap relates to issues of governance and accountability. Although numerous ethical principles for artificial intelligence have been proposed, there is limited empirical evidence regarding the practical effectiveness of these frameworks in reducing discrimination and enhancing accountability. Many of these principles remain at an abstract level and have not been translated into concrete policies and operational practices. This gap between theory and practice constitutes one of the central challenges in the realization of ethics-centered artificial intelligence [5].

Overall, the existing literature indicates that research on artificial intelligence and ethical decision-making, despite considerable advances, still requires more interdisciplinary approaches, deeper empirical studies, and greater attention to diverse social and cultural contexts. Addressing these gaps and contradictions is not only essential for the theoretical development of the field, but also a fundamental condition for the responsible design and deployment of algorithms in the pursuit of justice and the reduction of discrimination.

Social and Political Implications and Ethics-Centered Design of Artificial Intelligence

The growing expansion of algorithmic decision-making systems has produced not only technical and economic consequences, but also profound social and political effects that are directly connected to concepts such as justice, power, democracy, and social equality. Algorithms now function as mechanisms for distributing opportunities, resources, and risks, thereby playing an active role in shaping social structures and in reproducing or transforming inequalities. Research shows that algorithmic decision-making can blur traditional boundaries between the technical and political spheres, turning technology into an influential actor in social life [11].

From a social perspective, one of the most significant implications of artificial intelligence is the transformation of patterns of access to opportunities and services. When algorithms are used in employment, education, credit evaluation, or the provision of public services, their decisions can meaningfully affect individuals' life trajectories. In the absence of ethical and social oversight, these systems can entrench existing structural discriminations and even legitimize them, as algorithmic decisions are often perceived as "scientific" and "neutral." This situation can lead to diminished public trust, a sense of injustice, and the weakening of social cohesion [2].

At the political level, algorithmic decision-making has substantial implications for governance and democracy. The widespread use of algorithms in public policymaking, social surveillance, and population management has given rise to new forms of power that often operate outside traditional mechanisms of democratic accountability. Researcher [9] argues that the "black box society" emerges from the concentration of power in the hands of institutions that design and control algorithms, while citizens lack effective means of oversight. This concentration of power can undermine transparency, reduce public participation, and constrain opportunities for protest and civic resistance.

Moreover, algorithmic decision-making can blur the boundaries of political responsibility. When public policies and decisions are delegated to automated systems, the accountability of elected officials and governmental institutions is weakened, leading to a form of "responsibility shifting" to technology. This dynamic not only conflicts with principles of democratic governance, but can also have particularly negative consequences for marginalized groups,

who are more frequently subjected to automated decisions than others [4].

In response to these social and political implications, the concept of "ethics-centered design" has been proposed as a preventive and normative approach to the development of artificial intelligence. Ethics-centered design is grounded in the principle that human values-such as justice, transparency, respect for human dignity, and accountability-should be incorporated from the earliest stages of algorithm design, rather than being added retroactively after social harms have occurred [3]. By emphasizing the participation of diverse stakeholders, including affected groups, this approach seeks to reduce the gap between technical development and social consequences.

However, critical literature warns that if ethics-centered design remains limited to voluntary guidelines or abstract frameworks, it may become a symbolic tool that serves more to legitimize technology than to bring about real change in power relations. From this perspective, ethics-centered design must be accompanied by binding legal mechanisms, institutional oversight, and public accountability in order to effectively constrain the negative social and political impacts of artificial intelligence [12].

In sum, the social and political implications of algorithmic decision-making demonstrate that artificial intelligence is not merely a neutral technology or an efficient tool, but an integral part of contemporary power and governance structures. Ethics-centered design, when implemented in an interdisciplinary and context-sensitive manner, can provide a pathway for aligning technological development with the values of social justice and democracy. Achieving this goal requires that AI ethics be treated not as a peripheral addition, but as a foundational element in the policymaking, design, and deployment of algorithms.

Future Research Directions in Artificial Intelligence and Ethical Decision-Making

Despite the significant expansion of the scholarly literature on AI ethics, existing analyses indicate that research in this field is still in a transitional phase and requires new, deeper, and more interdisciplinary orientations. One of the most important future research directions is moving beyond an exclusive focus on the technical aspects of algorithms toward context-sensitive social analyses. Many current studies treat algorithmic bias and injustice as problems that can be resolved through data correction or model optimization, whereas critical research has shown that these phenomena are rooted in broader social, economic, and political structures. Accordingly, future research should systematically examine the links between algorithms and structural inequalities and analyze artificial intelligence as part of systems of power [2, 11].

Another key direction involves the development of long-term empirical research on the real-world consequences of algorithmic decision-making. A significant portion of the existing literature is based on limited case studies or theoretical analyses, and there is relatively little data on the cumulative and long-term effects of algorithms on the lives of individuals and social groups. Future studies can draw on qualitative methods, ethnographic approaches, and longitudinal research to demonstrate how algorithmic decisions, over time, contribute to the reproduction or transformation of patterns of justice and discrimination [4].

From a conceptual perspective, one of the priorities of future research is rethinking the definition of "algorithmic justice." Dominant technical definitions of justice, which focus on statistical metrics, are often unable to capture the ethical and social complexities of justice. Future research should integrate theories of social justice-particularly critical and post-structuralist

approaches-with technical studies in order to provide more comprehensive and context-sensitive frameworks for evaluating algorithmic decision-making [1].

Another key orientation is a stronger focus on the participation of stakeholders and marginalized groups in both research processes and algorithm design. To date, much of the research has been conducted from the perspectives of developers, policymakers, or dominant institutions, while the voices of those most affected by algorithmic decisions have been less frequently heard. Future research, by adopting participatory and justice-oriented approaches, can help redefine the role of users and citizens in the governance of artificial intelligence [12].

From a policy and governance perspective, the future of research in AI ethics requires critical evaluation of the effectiveness of existing ethical frameworks and principles. Although numerous documents emphasize transparency, responsibility, and accountability, there is limited evidence regarding the extent to which these principles are actually implemented in real-world systems. Future research should address questions such as under what conditions ethical frameworks can lead to institutional change and reductions in discrimination, and which factors hinder their effective implementation [3].

Finally, one of the most important future research directions is the strengthening of interdisciplinary approaches. The complexity of ethical issues related to algorithmic decision-making is such that they cannot be adequately analyzed from the perspective of a single discipline. Integrating insights from computer science, moral philosophy, sociology, law, and political studies can lead to the production of more comprehensive knowledge that is both theoretically rich and practically applicable. Such an approach enables a shift from abstract ethics to applied, context-sensitive ethics and facilitates the responsible development of artificial intelligence [9].

Overall, future research directions in artificial intelligence and ethical decision-making should focus on deepening social analysis, expanding empirical methods, enhancing stakeholder participation, and strengthening the effective linkage between theory and practice. Only through such an approach can it be ensured that technological advances serve social justice rather than exacerbate existing forms of discrimination.

CONCLUSION

The systematic review of the scholarly literature demonstrates that algorithmic decision-making and the widespread applications of artificial intelligence in social, legal, and political domains constitute not merely a technical or managerial challenge, but a deeply ethical and social issue intertwined with fundamental concepts such as justice, discrimination, power, and accountability. Contrary to technologically optimistic narratives that portray algorithms as neutral, objective, and efficient tools, research evidence indicates that these systems often reflect and reinforce existing unequal social structures and can lead to the reproduction of historical and institutional forms of discrimination.

The ethical analysis of algorithmic decision-making based on various theoretical frameworks shows that none of the classical ethical theories-ranging from utilitarianism to deontology and virtue ethics-can, on its own, fully address the complexities of this phenomenon. Utilitarianism, with its focus on outcome optimization, risks neglecting the rights of minorities; deontology faces challenges related to algorithmic transparency and explainability; and virtue ethics, while emphasizing the role of human actors, is insufficient without institutional reform. Consequently, the literature emphasizes the necessity of hybrid and interdisciplinary approaches that can simultaneously address

outcomes, individual rights, and social contexts.

One of the key findings of this article is that algorithmic discrimination cannot be reduced solely to data errors or technical flaws. Algorithms are designed and deployed within contexts shaped by power relations, economic interests, and structural inequalities, and these contexts play a decisive role in shaping their social consequences. While technical solutions-such as statistical fairness metrics or explainable AI-are necessary, they are not sufficient to guarantee justice and fairness without attention to social and political contexts.

The issue of transparency, responsibility, and accountability, as a central axis of AI ethics, has been examined in this article as a multilayered challenge. The lack of transparency in algorithmic decision-making logic, combined with the diffusion of responsibility among designers, organizations, and policymakers, has led to the emergence of a form of accountability gap that can conceal and entrench discriminatory outcomes. This situation is particularly dangerous for marginalized groups, whose access to mechanisms of contestation and legal redress is more limited.

Furthermore, the examination of gaps and contradictions in the literature reveals that future research must move beyond the dichotomy between technological optimism and radical critique toward empirical, context-sensitive, and participatory analyses. The absence of long-term studies, limited geographical focus, and the gap between abstract ethical principles and their practical implementation are among the challenges that researchers and policymakers must address.

Ultimately, this article argues that realizing fair and ethics-centered artificial intelligence requires a shift in perspective from “algorithms as tools” to “algorithms as social actors.” Ethics-centered design can contribute to reducing discrimination and promoting social justice only when it is accompanied by stakeholder participation, effective institutional oversight, and democratic accountability. This approach requires that ethics be embedded not as an afterthought, but as a foundational element throughout the entire lifecycle of artificial intelligence-from design to deployment and evaluation.

Overall, the findings of this article indicate that the future of artificial intelligence depends not only on technical advances, but also on the ethical, political, and social choices made by societies. Only through a critical, interdisciplinary, and justice-oriented approach can artificial intelligence be ensured to become a tool for reducing inequalities rather than a force that deepens and legitimizes existing structural discrimination.

ETHICAL CONSIDERATIONS

Ethical issues (such as plagiarism, conscious satisfaction, misleading, making and or forging data, publishing or sending to two places, redundancy and etc.) have been fully considered by the writers.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interests.

FUNDING DECLARATION

This research did not receive any grant from funding agencies in the public, commercial, or non-profit sectors.

REFERENCES

1. Barocas S, Selbst A D. Big data’s disparate impact. California Law Review, 2016; 104(3): 671–732. https://doi.org/10.15779/Z38BG31
2. Noble S U. Algorithms of oppression: How search engines reinforce racism. New York University Press. 2018.
3. Floridi L, Cowls J, Beltrametti M, Chatila R, Chazerand P, Dignum V, Luetge C, Madelin R, Pagallo U, Rossi F, Schafer B, Valcke P, Vayena E. AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 2018; *28*(4): 689–707. https://doi.org/10.1007/s11023-018-9482-5
4. Mittelstadt B D, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data &

Society, 2016; 3(2). https://doi.org/10.1177/2053951716679679
5. Wachter S, Mittelstadt B, Floridi L. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. International Data Privacy Law, 2017; 7(2): 76–99. https://doi.org/10.1093/idpl/ipx005
6. Coeckelbergh M. AI ethics. MIT Press. 2020.
7. Friedman B, Nissenbaum H. Bias in computer systems. ACM Transactions on Information Systems, 1996; 14(3): 330–347. https://doi.org/10.1145/230538.230561
8. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks. ProPublica. 2016.
9. Pasquale F. The black box society: The secret algorithms that control money and information. Harvard University Press. 2015. https://www.jstor.org/stable/j.ctt13x0hch
10. Burrell J. How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society, 2016; 3(1): 1–12. https://doi.org/10.1177/2053951715622512
11. Beer D. The social power of algorithms. Information, Communication & Society, 2016; 20(1): 1–13. https://doi.org/10.1080/1369118X.2016.1216147
12. Whittlestone J, Nyrup R, Alexandrova A, Dihal K, Cave S. Ethical and societal implications of algorithms, data, and artificial intelligence: A roadmap for research. London: Nuffield Foundation. 2019.