World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(RESEARCH ARTICLE)

Check for updates

Ethical challenges and bias mitigation in Artificial Intelligence systems

Harsh Verma *

Palo Alto Networks, Artificial Intelligence, United States.

World Journal of Advanced Research and Reviews, 2025, 28(03), 2364-2373

Publication history: Received on 12 October 2025; revised on 23 December 2025; accepted on 28 December 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.28.3.3904

Abstract

Artificial Intelligence (AI) has rapidly become central to decision-making in healthcare, finance, governance, and education. Yet this progress has introduced serious ethical challenges, particularly bias, fairness, transparency, and accountability. Recent empirical studies confirm that algorithmic bias remains a persistent issue. Facial recognition systems, for instance, continue to misclassify women and people of color at rates up to 30% higher than white men (Buolamwini & Gebru, 2018; Raji et al., 2024). In healthcare, diagnostic AI models trained on skewed datasets underperform for minority populations, raising concerns about equitable access to care (Chen et al., 2023).

This paper provides a critical evaluation of these ethical issues and explores mitigation measures through a systematic literature review of peer-reviewed articles, conference papers, and policy reports. Findings indicate that AI bias is primarily influenced by disparities in training data, algorithm design, and embedded social inequalities. These prejudices often lead to discriminatory outcomes that reinforce existing inequities. Other significant ethical concerns include the lack of transparency, breaches of privacy, and unclear accountability due to the "black box" nature of many AI systems.

Recent developments highlight promising directions, such as explainable AI, fairness-conscious algorithms, and regulatory frameworks like the EU AI Act (European Commission, 2024) and the NIST AI Risk Management Framework (NIST, 2023). While these initiatives represent meaningful progress, gaps remain in standardizing fairness measures and ensuring global governance. The paper concludes that future research should prioritize interdisciplinary collaboration, robust regulatory frameworks, and continuous monitoring to promote the ethical use of AI.

Keywords*:* Artificial Intelligence; Ethics; Bias Mitigation; Fairness; Explainability; Responsible AI

1. Introduction

Over the past decade, Artificial Intelligence (AI) has shifted from experimental research to a core driver of the digital ecosystem. Advances in computing power, data availability, and algorithmic design have enabled machine learning, deep learning, and generative models to automate complex tasks, analyze vast datasets, and deliver highly accurate predictions. Today, AI underpins critical sectors such as healthcare, finance, education, governance, and transportation, offering efficiency gains, cost reductions, and improved service delivery.

Yet these benefits are shadowed by pressing ethical concerns. Algorithmic bias remains a central challenge, with systems often replicating historical inequalities embedded in training data. Facial recognition technologies, for example, misclassify women and people of color at disproportionately high rates (Buolamwini & Gebru, 2018; Raji et al., 2024), while healthcare models underperform for minority populations (Chen et al., 2023). Beyond bias, the opacity of "black

* Corresponding author: Harsh Verma

box" models undermines transparency and accountability, while generative AI raises risks of misinformation, privacy violations, and content manipulation.

As AI systems grow more autonomous, their social impact intensifies, entrenching divides and amplifying ethical dilemmas. This has fueled calls for fairness-conscious design, explainable AI, and stronger regulatory oversight. Addressing these challenges requires not only technical innovation but also interdisciplinary collaboration and governance frameworks that ensure AI aligns with human values of equity, transparency, and accountability.

1.1. Problem Statement

Despite remarkable technological advances, AI systems often lack the ethical safeguards needed to be truly trustworthy. Bias remains the most pressing issue, as models trained on historical data frequently replicate and even amplify existing inequalities. This has led to discriminatory outcomes in critical areas such as employment, credit scoring, and law enforcement. Amazon's recruitment AI, for instance, penalized résumés with female indicators (Dastin, 2018), while predictive policing tools have disproportionately targeted minority neighborhoods (Lum & Isaac, 2016).

These cases show that AI, though designed to be data-driven and objective, is deeply dependent on the quality and balance of its inputs. Without clear standards for identifying and preventing bias, corporations often prioritize efficiency over fairness, and complex "black box" models obscure accountability. As a result, biased AI systems can undermine equity, erode trust, and have far-reaching consequences for people's rights and opportunities.

1.2. Research Objectives

In response to these challenges, this study seeks to provide a comprehensive examination of ethical issues and bias mitigation strategies in AI systems. The primary objective is to analyze the evolution of ethical concerns in AI and to identify the key factors contributing to bias in modern systems. By reviewing existing literature and empirical studies, the research aims to uncover common patterns and trends that characterize ethical challenges in AI development and deployment.

Another important objective is to evaluate the effectiveness of various bias mitigation techniques that have been proposed and implemented over the years. These include approaches such as data preprocessing, algorithmic adjustments, and post-processing interventions. The study aims to assess how well these methods address bias and whether they can be generalized across different applications and contexts.

1.3. Research Questions

To achieve these objectives, the study is guided by several key research questions. First, it seeks to understand the nature and scope of ethical challenges associated with AI systems, particularly in relation to bias, transparency, and accountability. This involves examining how these challenges have evolved over time and how they manifest in different application domains.

Second, the study investigates the sources of bias in AI systems, with a focus on identifying the underlying factors that contribute to unfair outcomes. This includes analyzing the role of data quality, algorithm design, and human decision-making in shaping AI behavior.

Third, the research evaluates the effectiveness of existing bias mitigation strategies, considering their strengths, limitations, and applicability in real-world scenarios. This involves comparing different approaches and assessing their impact on both fairness and system performance.

2. Literature Review

2.1. Evolution of AI Ethics

We are coming to the point in artificial intelligence (AI) where the ethical debate is moving into the open use of AI technologies. During the initial stages of AI development, the ethical debate was largely theoretical and concerned with the phenomenon of the machines' uprising, job loss, and human oversight of the smart machines. At that time, AI systems were limited in their scope, and the ethical concerns were often overshadowed by the excitement of a new technology and its capabilities.

As AI capabilities grew, especially in terms of machine learning and deep learning techniques, ethical questions were gaining practical relevance. The various risks from the new dependency on data-driven algorithms concern biases, fairness, and transparency. The mid-2010s were a turning point for researchers and practitioners as they started to realise that AI systems are not objective and are as biased as the data they train on. So, we shifted from being concerned about the general ethical questions to more practical issues related to what AI systems will do in the world and for individuals.

AI ethics has gained more traction in recent years, and the focus is on fairness, accountability, transparency, and explainability. The emergence of generative AI systems also plays a role in the ethics conversation because it brings in new problems, such as misinformation, deepfakes, IP, and authenticity. This has also strengthened the demand for robust ethical guidelines for building, using, and managing AI systems. So, AI ethics is an interdisciplinary approach that draws on computer science, law, sociology, and philosophy and seeks to help ensure technological development is in line with human values.

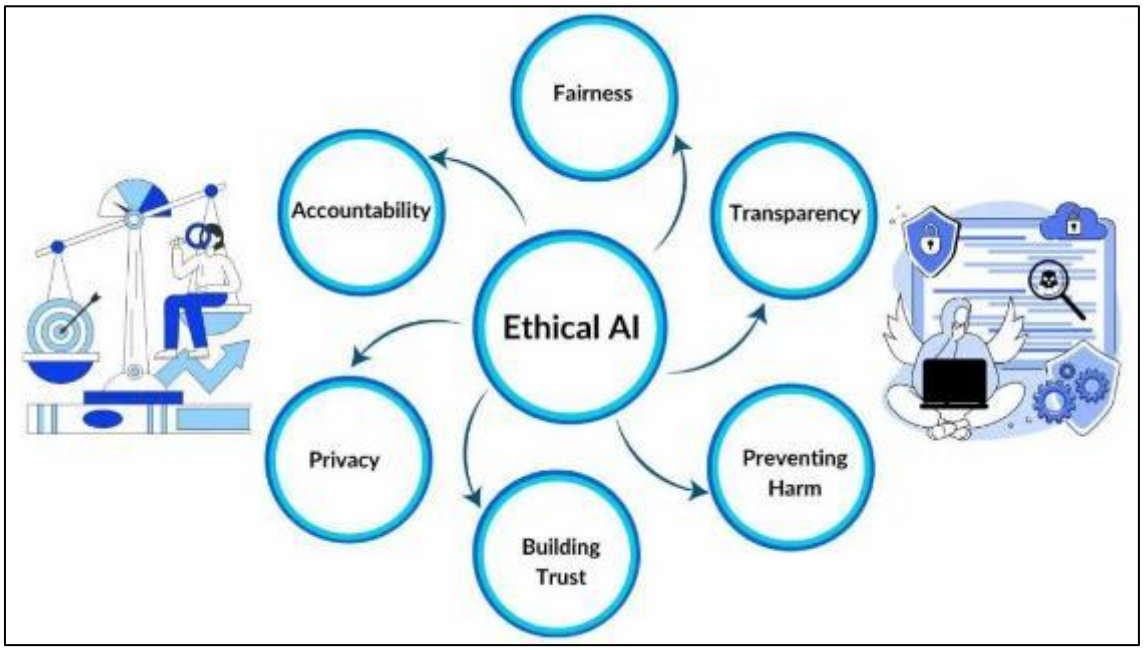

Figure 1 AI Ethics

2.2. Types of Bias in AI

A key concern when working to ensure that artificial intelligence (AI) delivers fair and equitable outcomes is bias. Bias can take many forms, including: data bias, algorithmic bias, and application bias. Data bias is where the training sets used for AI models are not representative, are incomplete, or represent historical discrimination. Given the data-driven nature of machine learning, bias in data can result in biased predictions and decision-making.

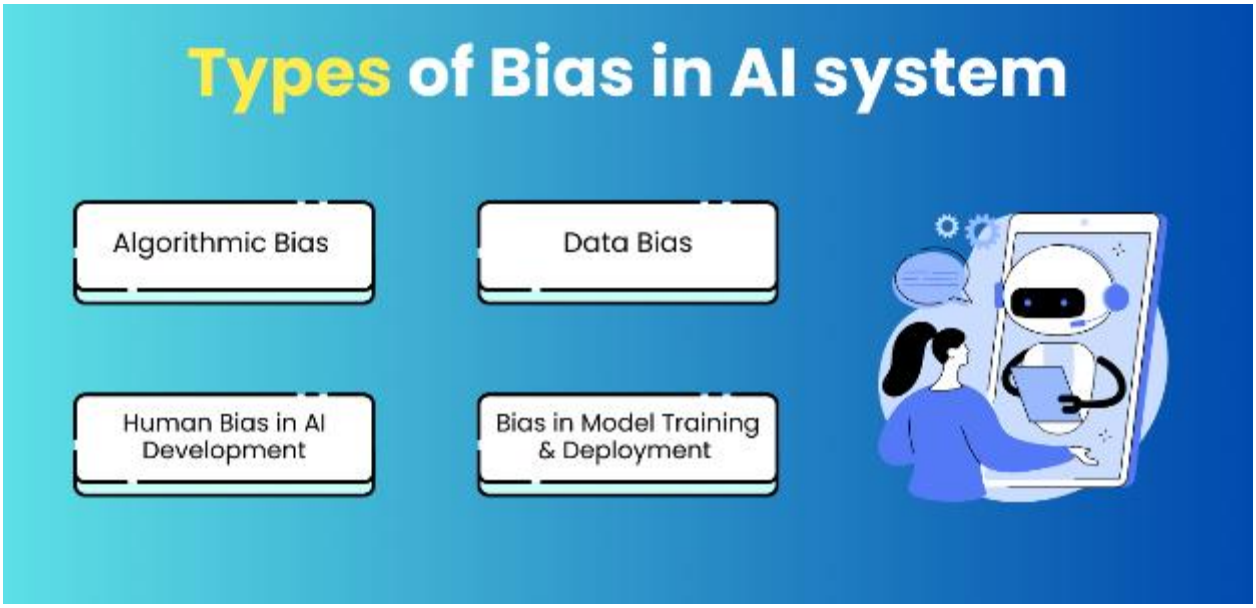

Figure 2 Types Of Bias in AI system

Algorithmic bias occurs when the algorithms themselves are biased. So, while the training data may be "fair", the design of the model, the way the model is trained, and even how we measure the performance of the model can be biased. For instance, models that optimize for overall accuracy may not meet the needs of minorities if they are underrepresented in the training data. This highlights the need for us to consider fairness during model development.

Bias in application occurs when using AI systems. For instance, the wrong use of a model can bias results. For example, an AI system that works well for one group of people may not work well for another. These biases can be intertwined and hard to tease apart. So understanding sources and expressions of bias is essential for reducing bias and increasing fairness.

2.3. Ethical Challenges in AI

AI's rapid adoption has brought ethical challenges that extend beyond technology into society. Privacy is a major concern, amplified by large-scale data breaches such as the Cambridge Analytica scandal, which showed how personal data could be exploited for political manipulation (Isaak & Hanna, 2018). Generative AI adds new risks, enabling deepfakes and misinformation that threaten democratic processes (Floridi, 2023).

Transparency is another critical issue. Many AI systems function as "black boxes," making it difficult to explain or justify decisions in sensitive areas like healthcare, finance, and criminal justice. This lack of clarity undermines trust and accountability. Responsibility is also blurred, as developers, organizations, and users share roles in AI outcomes, complicating regulation and governance.

Finally, biased algorithms can reinforce social inequalities, producing unfair results in hiring, lending, and law enforcement. These examples show that ethical challenges in AI are not abstract they directly affect rights, opportunities, and justice. Addressing them requires proactive measures that combine technical solutions with strong governance and oversight.

2.4. Bias Mitigation Techniques

As awareness of bias and other ethical issues in AI has increased, numerous techniques have emerged to mitigate bias and uphold fairness and accountability in AI systems. These methods can be broadly grouped by the point in the AI development process at which they're employed. Data preprocessing techniques seek to mitigate bias through better representation of training data. These could include data balancing, data anonymisation, and data augmentation to improve diversity and reduce bias.

Modeling methods aim to embed fairness considerations into the model building process. These techniques aim to guide an AI system to make fair decisions by recalibrating the learning objective or incorporating fairness measures into the

training process. By incorporating fairness into the algorithm, these approaches seek to reduce bias while maintaining model accuracy.

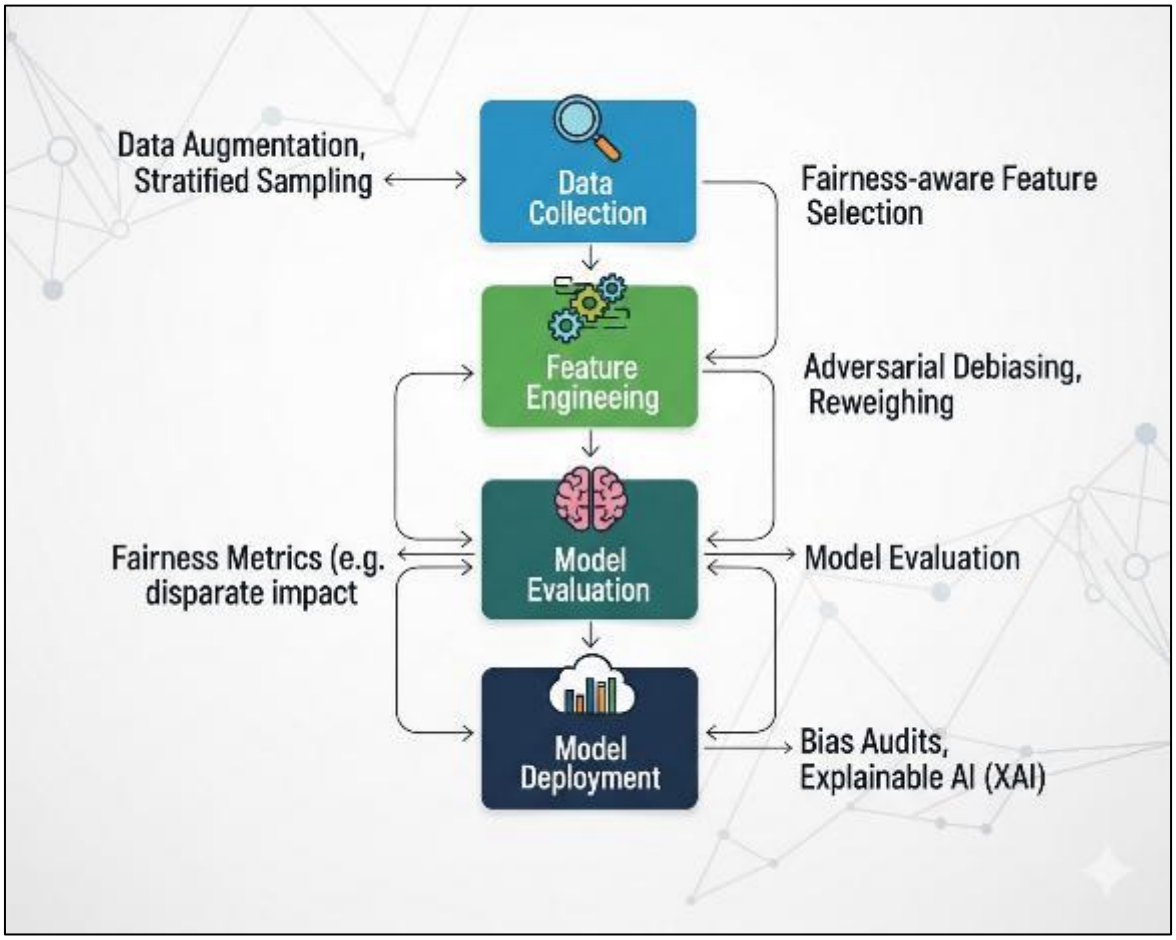

Figure 3 Bias Mitigation Techniques in machine language

Post-processing approaches are used to modify the output of a model to mitigate differences in predictions. These techniques can be used to adjust outcomes to reduce bias, such as by adjusting thresholds or re-weighting predictions. Post-processing methods can be powerful, but are often viewed as a supplement to earlier steps in the process.

Beyond such technical approaches, explainability and transparency are key aspects of bias reduction. Explainable AI methods seek to make the decisions made by models more understandable, allowing users to understand the predictions and recognise potential biases. Through increased transparency, these techniques can help establish trust and accountability.

3. Methodology

3.1. Research Design

The systematic literature review design is used in the study to offer a systematic and comprehensive review of the literature on the topic of ethical issues and solutions to bias in artificial intelligence. The systematic approach is motivated by the need to bring together a rapidly growing body of interdisciplinary research in computer science, ethics, law and the social sciences. Unlike other forms of narrative reviews, the systematic literature review is transparent, replicable and rigorous as a combination of processes are followed when searching, selecting and synthesising research.

The study design followed best practices of systematic reviews by prioritising the specific research questions, systematic search techniques and inclusion and exclusion criteria. The objective of this review was to explore the evolution of ethical considerations in research on AI, and in particular how to handle issues of bias, fairness, accountability and transparency.

There are a number of steps involved in the systematic review process; identification, screening, quality assessment and synthesis. Each of these steps was done in a way that reduces any potential for bias and ensures the studies included are a representative sample of the breadth and quality of the field. This approach can also help to identify research gaps and trends, and give a strong platform for future research on the ethics of designing artificial intelligence.

3.2. Data Sources

We used some of the most respected databases to source data for this review. These include IEEE Xplore, Scopus, Web of Science, and Google Scholar. They have been chosen for their large collection of peer-reviewed academic literature in a very broad field that includes artificial intelligence and ethics.

The strategy used to search for literature was through the use of keywords and operators. These included artificial intelligence, AI ethics, algorithmic bias, machine learning fairness, bias mitigation, responsible AI, and ethical AI systems. These terms were used in conjunction with operators such as AND, OR, and NOT to include and exclude the search results. The search was iterative, and the keywords were refined as the search progressed and topics were identified.

In order to keep the review focused on the most relevant and recent literature, we used filters to limit the search to publications. We also limited the search results to publications in English and included more peer-reviewed journal articles and conference proceedings. We also included in the grey literature reports and policy documents from well-known sources to get a taste of the ethical issues and regulatory trends.

Multidatabase search helped avoid publication bias and gain a variety of perspectives. A multidisciplinary approach allowed us to include technical and ethical considerations in AI research, a complex field.

3.3. Inclusion Criteria

To ensure quality, relevance and consistency of the studies to be incorporated into the review, we applied the following inclusion criteria. The studies had to be peer-reviewed journal articles, conference papers or high-quality reports that explicitly address the topic of ethical concerns to and reducing bias in artificial intelligence systems. The researches chosen had to cover one or more of the following topics: algorithmic bias, fairness, transparency, accountability or ethics of AI systems.

As well as subject matter, the studies had to be published during a particular period. This era was chosen to track the progress of AI ethics since its beginning to date. Studies published before this era or afterward were excluded unless they were deemed essential to settings and foundations of subsequent studies.

We also excluded the studies that were not published in English for the purpose of understanding and interpretation. The studies should also provide sufficient information about the methodology and contribution to research (empirical or theoretical). The studies were excluded if they did not have clear research questions, appropriate research methods or did not contribute to the research question.

The study search was conducted in English by searching titles and abstracts to determine their potential relevance and the full-text was searched for confirmation of the study's inclusion criteria. The two-step approach allowed us to include all relevant high quality studies. Overlapping studies were removed and any disagreement in the study selection process was evaluated for bias.

3.4. Data Analysis

We used a thematic analysis to analyse the selected studies, which is a method of identifying and interpreting patterns, or themes, within data. This is a useful approach for analysing different research and for exploring complex issues such as bias and ethical issues in AI systems.

We analysed the studies by carefully reading each study in order to identify information about the purpose of the study, the methodology used, the findings, and the ethical issues identified. We then grouped the findings into themes according to common themes, such as sources of bias, ethical issues, and how to address them. Coding plans were used to group ideas and themes collectively and find relationships.

The key themes we found include bias in data-driven algorithms, transparency in algorithms, fairness, and regulation. Through the use of thematic analysis, we were also able to identify the trends over the years and find that the AI ethics discussion has shifted from theoretical to practical use and regulation.

Analysis procedures were standardised to improve the validity of the results. This involved repeated reading and interpretation of data and triangulation of results between different studies. Integration of the results was carried out to ensure that each original study is maintained and to provide a consolidated view of the field of study.

These methods used in this paper offer a valid basis for the analysis of ethical concerns and reduction of bias in AI. The systematic literature review and thematic analysis offer a comprehensive and informative summary of research so far, and will guide and direct future development of ethical AI.

4. Results

4.1. Key Ethical Challenges Identified

The analysis of the selected studies shows that the most recurring ethical concerns in artificial intelligence systems are bias, lack of transparency, and weak accountability structures. Bias appears as the dominant issue, often leading to unfair or discriminatory outcomes in decision-making systems. Transparency remains a major challenge due to the "black-box" nature of many AI models, making it difficult to understand how decisions are made. Accountability is also limited, as responsibility for AI-driven decisions is often unclear among developers, organizations, and users.

Table 1 Key Ethical Challenges in AI Systems

Ethical Challenge	Description	Impact on Society
Bias	Systematic unfairness in AI outcomes due to flawed data or design	Discrimination and inequality
Lack of Transparency	Difficulty in understanding how AI systems make decisions	Reduced trust and explainability
Accountability	Unclear responsibility for AI decisions	Legal and ethical uncertainty

4.2. Sources of Bias

The findings indicate that bias in AI systems originates from multiple interconnected sources. Historical data used to train models often reflect existing societal inequalities, which are then replicated by the system. Imbalanced datasets, where certain groups are underrepresented, further amplify biased outcomes. Additionally, human decision-making during model design, such as feature selection and algorithm choice, introduces subjective influences that can unintentionally embed bias into the system.

Table 2 Major Sources of Bias in AI

Source of Bias	Explanation	Example
Historical Data Bias	Training data reflects past societal inequalities	Biased hiring or credit scoring systems
Imbalanced Datasets	Unequal representation of groups in data	Facial recognition errors for minorities
Human/Algorithmic Bias	Decisions made during model design introduce subjectivity	Feature selection favoring certain groups

4.3. Effectiveness of Mitigation Strategies

The review shows that various bias mitigation strategies have been developed, including data preprocessing, fairness-aware algorithms, and post-processing corrections. These methods can reduce bias levels and improve fairness to some extent. However, no single approach completely eliminates bias, as new forms of bias may emerge during deployment

or as data evolves. This highlights the importance of continuous monitoring, evaluation, and updating of AI systems to ensure sustained ethical performance.

Table 3 Bias Mitigation Strategies and Effectiveness

Strategy Type	Description	Effectiveness Level
Pre-processing	Cleaning and balancing datasets before training	Moderate
In-processing	Modifying algorithms to ensure fairness during training	Moderate to High
Post-processing	Adjusting outputs after model prediction	Limited but useful
Continuous Monitoring	Ongoing evaluation and updating of AI systems	Highly Recommended

5. Discussion

This paper has shown that the ethical concerns related to artificial intelligence are not only limited to the technical limitations but are associated with the extensive social system and the mechanism of human decision-making. Historically, the bias in AI systems is not caused by bad algorithms but by the disparities, cultural assumptions, and system imbalances present in the data that trains the systems. By doing so, AI systems are likely to replicate and, indeed, even to increase the already existing inequities, particularly in such a sensitive field as healthcare, work, and criminal justice. This points to the significance of viewing bias as a computational and a socio-technical challenge that ought to be viewed more holistically.

No matter whether a number of bias minimization techniques have been produced including data pre-processing, fairness sensitive algorithms, and post processing adjustments, the fact that biased results still occur indicates that the solutions that are currently in place are not effective enough. These approaches are more likely to address symptoms rather than causes in that they may reduce the observable disparities but leave structural disparities within datasets intact. Furthermore, most mitigation strategies entail trade offs between fairness and model performance, with complicated questions on how fairness ought to be defined and given priority in various situations. In this case, it is worthy to mention that continuous tracking and constant improvement ought to be presented rather than a one-time intervention.

The other huge concern that has been established in the present study is the fact that there are no universally accepted guidelines of fairness and ethical compliance in artificial intelligence systems. Definitions of fairness vary according to various disciplines and regulatory authorities, and, therefore, can lead to inconsistency in their application and assessment. This non-standardization complicates ensuring accountability and transparency as it may be necessary to adjust to the implementation in cross-border situations where the regulatory framework may be extremely different. As the popularity of the AI systems in the whole world is continuously increasing, the need in the coordinated guidelines is getting more and more acute.

The difficulty in solving these issues necessitates the interdisciplinary solution that integrates the knowledge in computer science, ethics, law, and social sciences. The gap between technical innovation and ethical responsibility can be bridged by collaborating, and AI systems can be built and put into practice in a socially-friendly and socially-responsible manner. Lastly, ethical AI will be developed through the ability to incorporate viable technical solutions and comprehensive policy frameworks and deeper understanding of the social contexts in which such technologies are utilized.

6. Conclusion

The results of this paper confirm that ethical dilemmas and biasness in artificial intelligence systems are still burning and ambiguous issues, even with the significant technological progress. Although it can be stated that a considerable amount of effort has been invested into the creation of bias mitigation methods and enhancing the fairness of algorithms, it has not eradicated the threats of discriminatory results, insufficient transparency, and a gap in accountability. Several factors have led to the fact that the creation of AI systems has often surpassed the creation of an elaborate ethical framework, establishing a gap between innovation and responsible use.

Discrimination in AI remains a problem that has various origins, such as lop-sided representations of the historical data, erroneous model construction, and inequalities in the society, which end up being coded in algorithms. Even though the

existing mitigation approaches, like fairness-conscious algorithm and explainable AI models, have led to an increase in the performance and the reliability of systems, they frequently offer a partial solution. This highlights the necessity of a more comprehensive approach that entails a combination of technical, ethical, and institutional approaches.

To have truly ethical AI systems, technical fixes are not enough. It requires a concerted action among researchers, policymakers, industry practitioners, and the society in general. Regulatory control needs to change in line with the advancement of technology in order to make sure that AI systems work with a clear outline of ethical standards. Concurrently, the organizations are supposed to embrace responsible AI practices, which focus on transparency, equity, and responsibility across the system lifecycle.

In the future, it is advisable that research and development be directed towards designing standardized methods of judging fairness and bias and also developing mechanisms of constant monitoring so as to identify and rectify instances of ethical problems in real time. Finally, the development of reliable AI systems will be based on the long-term interdisciplinary cooperation and adherence to the principles of harmonization of technological innovation with human values and societal well-being.

References

[1] Akram, A. (2023). The impact of social media on communication and relationships: A double-edged sword. *Policy Research Journal, 1*(1), 15–24.

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

[3] Allam, Z. (2019). *Cities and the digital revolution: Aligning technology and humanity*. Springer Nature.

[4] Islam, M. S., Verma, H., Khan, L., & Kantarcioglu, M. (2019, December). Secure real-time heterogeneous iot data management system. In *2019 first IEEE international conference on trust, privacy and security in intelligent systems and applications (TPS-ISA)* (pp. 228-235). IEEE.

[5] Ashokan, A., & Haas, C. (2021). Fairness metrics and bias mitigation strategies for rating predictions. *Information Processing & Management, 58*(5), 102646.

[6] Chen, I. Y., Szolovits, P., & Ghassemi, M. (2023). Algorithmic bias in healthcare AI. Nature Medicine, 29(1), 44–50.

[7] Bacchini, F., & Lorusso, L. (2019). Race, again: How face recognition technology reinforces racial discrimination. *Journal of Information, Communication and Ethics in Society, 17*(3), 321–335.

[8] Belkacemi, S. (2022). Artificial intelligence (AI) and its impact on the global economy. *Journal of Financial, Accounting and Management Studies, 9*(2).

[9] Ben-Ishai, S., & Bedford, M. (2021). AI, consumer credit, and discrimination: A comparative look at Canada and the United States. *Corporate and Business Law Journal, 2*, 271.

[10] Boldyreva, E. L., Grishina, N. Y., Duisembina, Y., Boldyreva, E. L., & Grishina, N. Y. (2018). Cambridge Analytica: Ethics and online manipulation with decision-making process. *European Proceedings of Social and Behavioural Sciences, 51*.

[11] Floridi, L. (2023). Ethics of artificial intelligence. Oxford University Press.

[12] Bruckner, M. A. (2018). The promise and perils of algorithmic lenders' use of big data. *Chicago-Kent Law Review, 93*, 3.

[13] Verma, Harsh. (2025). Explainable AI (XAI) for Software Engineering Decision-Making. 10.15680/IJIRCCE.2025.1311002.

[14] Burr, C., & Leslie, D. (2023). Ethical assurance: A practical approach to the responsible design, development, and deployment of data-driven technologies. *AI and Ethics, 3*(1), 73–98.

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

[16] Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56–59.

[17] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), 1–21. https://doi.org/10.1177/2053951716679679

[18] Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

[19] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research, 81, 1–15. https://doi.org/10.48550/arXiv.1807.11485

[20] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of FAT Conference, 149–159. https://doi.org/10.1145/3287560.3287592

[21] Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. Proceedings of FAT Conference, 59–68. https://doi.org/10.1145/3287560.3287598

[22] Shrestha, A. K., Singha, S., Sural, S., Sutton, S., Tahiri, S., Tipper, D., ... & Yu, L. Yu, Xiaoyuan 46 Zhao, Zhilong 236 Zou, Xukai 46.

[23] Raji, I. D., Smart, A., White, R. N., et al. (2020). Closing the AI accountability gap. Proceedings of FAT Conference, 33–44. https://doi.org/10.1145/3351095.3372873

[24] Mitchell, M., Wu, S., Zaldivar, A., et al. (2019). Model cards for model reporting. Proceedings of FAT Conference, 220–229. https://doi.org/10.1145/3287560.3287596

[25] Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An ethical framework for a good AI society. Minds and Machines, 28, 689–707. https://doi.org/10.1007/s11023-018-9482-5

[26] Cath, C. (2018). Governing artificial intelligence: Ethical, legal and technical opportunities. Philosophical Transactions A, 376(2133). https://doi.org/10.1098/rsta.2018.0080

[27] Mehrabi, N., Morstatter, F., Saxena, N., et al. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35. https://doi.org/10.1145/3457607

[28] Caton, S., & Haas, C. (2020). Fairness in machine learning: A survey. arXiv preprint. https://doi.org/10.48550/arXiv.2010.04053

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

[30] Oneto, L., & Chiappa, S. (2020). Fairness in machine learning. arXiv preprint. https://doi.org/10.48550/arXiv.2012.15816

[31] Islam, M. S., Verma, H., Khan, L., & Kantarcioglu, M. (2019, December). Secure real-time heterogeneous iot data management system. In *2019 first IEEE international conference on trust, privacy and security in intelligent systems and applications (TPS-ISA)* (pp. 228-235). IEEE.