Journal of Digital Security and Forensics
January-June 2026 3(1)
www.digitalsecurityforensics.org

Original Article

A CRITICAL REVIEW OF AVAILABLE INFRASTRUCTURE, POLICY FRAMEWORKS, AND ORGANIZATIONAL CULTURE IN THE IMPLEMENTATION OF FAIRNESS-ENHANCING ARTIFICIAL INTELLIGENCE

Gyani Ray [1*], **N Molla** [2]
[1] PhD Scholar, Department of Computer Science and Engineering, Sikkim Professional University, India
[2] Associate Professor, Department of IT, Sikkim Professional University, Sikkim, India

ABSTRACT

The paper is a critical analysis on why infrastructure, policy based, and organizational culture is important in the successful implementation of fairness-improving artificial intelligence (AI). Although the metrics of algorithmic fairness and bias mitigation have been heavily developed, their application in an institutional context is poorly coordinated. The major void in knowledge is how sociotechnical conditions facilitate or limit the process of long-term adoption of fairness beyond the optimization of the technical side. A systematic literature review based on PRISMA helped to identify 150 records, screen them, and then reduce them to 20 studies that fulfilled the inclusion criteria, which were all related to implementation contexts. These results indicate that strong data governance and lifecycle monitoring and auditing systems are the basis of operational fairness, and that enforceable policy mechanisms and culture rooted in leadership play a significant role in the result. The paper concludes that AI-based fairness encompasses a sociotechnical ecosystem and not a set of technical responses.

Keywords: Fairness in AI, Algorithmic Bias, AI Governance, Infrastructure Readiness, Organizational Culture

INTRODUCTION

Artificial intelligence (AI) is rapidly growing and is applied with high-impact in such fields as the healthcare sector, employment background checks, credit rating, and decision-making in the area of public policy. Nevertheless, it has been demonstrated that these systems create systematic effects of bias and unequal results that may reproduce past injustice and contradict moral values of equality and justice Álvarez et al. (2024). Researchers understand that the problem of bias in AI is not solely an issue of technical nature, and stems out of the data, algorithms and sociotechnical, such as policies and organizational practices, where the systems are implemented Papagiannidis (2025). This creates an immediate demand of equity-enhancing AI systems that are able to mitigate against discriminatory result.

Although many algorithmic fairness metrics and bias control methods are designed and existing experiments are conducted in a limited technical context, the majority of the literature does not consider the problem of embedding fairness into the real world Murire (2024). Responsible AI governance emerges in a growing body of literature that suggests that the formal principles (e.g. fairness, accountability, transparency) are broadly embraced but ill-specified in the real world, which creates discrepancies between aspirational principles and real-world performance Pagano et al. (2022). That gap is significant as is presented in the literature: on

*Corresponding Author:
Email address: Gyani Ray (gyaniray@gmail.com)
Received: 09 January 2026; **Accepted:** 12 March 2026; **Published** 02 April 2026
DOI: 10.29121/DigiSecForensics.v3.i1.2026.92
Page Number: 1-6
Journal Title: Journal of Digital Security and Forensics
Journal Abbreviation: J. Digi. Sec. Forensics
Online ISSN: 3048-894X
Publisher: Granthaalayah Publications and Printers, India
Conflict of Interests: The authors declare that they have no competing interests.
Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
Authors' Contributions: Each author made an equal contribution to the conception and design of the study. All authors have reviewed and approved the final version of the manuscript for publication.
Transparency: The authors affirm that this manuscript presents an honest, accurate, and transparent account of the study. All essential aspects have been included, and any deviations from the original study plan have been clearly explained. The writing process strictly adhered to established ethical standards.

the one hand, there are algorithmic methods to implement and identify unfairness, but, on the other hand, there is little knowledge on how organizational and policy structures facilitate or limit their regular use across settings Agarwal et al. (2022).

One of the harsh impediments is the infrastructure preparedness. The good data governance of the AI systems that are to support equality should have an evaluations and surveillance systems that can be extended. However, the nature of such organizations is that they have data silos, superficial tooling and intermittent evaluation pipelines, undermined by the ability to continually assess and adjust the models to equitable levels beyond growth periods Framework Convention on Artificial Intelligence. (2026). This organizational deficiency is very little talked in the technical fairness studies which rather look at ideal infrastructure but fail to give attention to the resource constraint of an organization or the challenge of the integration process.

Artificial intelligence fairness policy frameworks are currently being developed but are largely fragmented and have a loose association with the implementation practices. Although the ethical guidelines, standards, and regulatory plans that promote equity are becoming increasingly popular, a majority of them lack structures of enforceable compliance or realignment with an organization process. This difference in the policies and practice of AI limits the capacity of AI ethics to be established as calculable equities ACM Conference on Fairness, Accountability, and Transparency. (2026). Moreover, the organizational culture is a conclusive aspect in deciding how the element of fairness will use its resources and how the latter is controlled. The extent to which fairness is perceived as a strategic purpose or an administrative liability is formed by the leadership commitment, institutional norms and rewards schemes Toronto Declaration. (2026). However, the field of research of fairness that remains understudied is the cultural life. In general, these results indicate that successful fairness-promoting AI needs a cohesive sociotechnical ecosystem that includes infrastructure, enforceable policy, and enabling organizational culture, instead of using technical solutions.

Figure 1

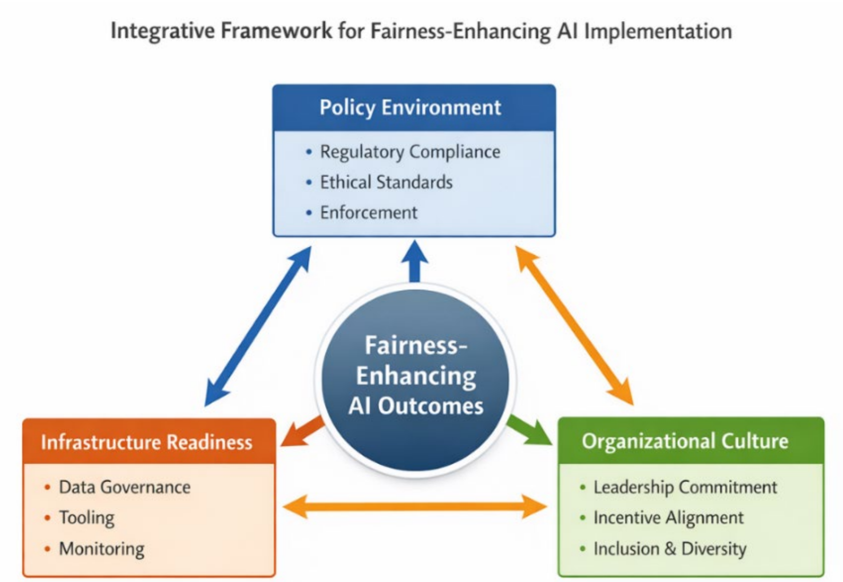

Figure 1 Integrative Sociotechnical Framework for Fairness-Enhancing AI Implementation

METHODOLOGY

This research used systematic literature review (SLR) to bring together empirical and conceptual research on fairness-enhancing AI, and this involved infrastructure, governance frameworks, and organizational culture. The PRISMA guidelines were used to conduct the review to provide clear selection, screening and reporting procedures. The latest articles that were published until January 2026 were found in the largest academic databases, such as Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, Wiley, MDPI, arXiv, and Google Scholar. Pre-established keywords were employed with Boolean operators, which included fairness in AI, algorithmic bias mitigation, AI governance, and organizational culture. Following the elimination of duplicates, studies were filtered using the title, abstract, and full text, according to the inclusion criteria that focused on the implementation contexts. Peer-reviewed publications that were in English were only included. The extraction and analysis of the data were done using narrative synthesis and grouping the results in categories of infrastructure readiness, policy alignment, and organizational culture to find the key enablers, barriers, and gaps in research.

INCLUSION AND EXCLUSION CRITERIA

Articles were screened according to their consideration of at least one of the two main themes of the present review (i) the technical and infrastructural underpinnings of implementing fairness-enhancing AI technologies, including data governance systems, fairness measures and metrics, bias mitigation algorithms, life cycle monitoring tools, and AI auditing systems, and (ii) the governance and organizational aspects that affect the way fairness is implemented, including AI policy models, regulatory frameworks, compliance systems, institutional accountability models, leadership commitment, and organizational culture processes. Both empirical (quantitative, qualitative or mixed-methods) and conceptual or methodological research were included in terms of providing the opportunity to conduct the synthesis of the theoretical advances and the practices of the real-world implementation comprehensively.

Included in the studies were those lacking transparency of methods, not reviewed by peers (with the exception of influential technical reports by reputable institutions), and those dealing only with abstract notions of fairness without mention of implementation settings. English-language publications were taken into consideration only to make sure that there are similarities in conceptual interpretation and the rigor of analysis.

SELECTION PROCESS

To enhance transparency and reproducibility, a structured screening procedure that relied on PRISMA (Preferred Reporting Items to Systematic Reviews and Meta-Analyses) framework was used. First, 150 records have been found in the identified academic databases, such as Scopus, Web of Science, IEEE Xplore, ScienceDirect, SpringerLink, Wiley Online Library, MDPI, arXiv, and Google Scholar. Upon eliminating the redundant records, 115 records were left to be screened in terms of titles and abstracts. At this step, 55 articles were filtered out due to either ignoring fairness-enhancing AI implementation or being irrelevant to the topic of infrastructure, governance, or organizational culture.

Among them, 40 were excluded because they lacked a sufficient methodological description, had a too technical scope that was not connected with sociotechnical and were too limited in their ability to contribute to the implementation aspect of fairness. Finally, it was possible to include 20 studies in the final synthesis. The analyses of these studies were systematically done and thematically classified based on the infrastructure preparedness, policy-regulatory fit, and organizational culture and change management. The conclusion review allowed discovering implementation patterns, barriers, enabling mechanisms, and unresolved gaps in research on the implementation of fairness-promoting AI technologies in a variety of institutional settings.

Figure 2

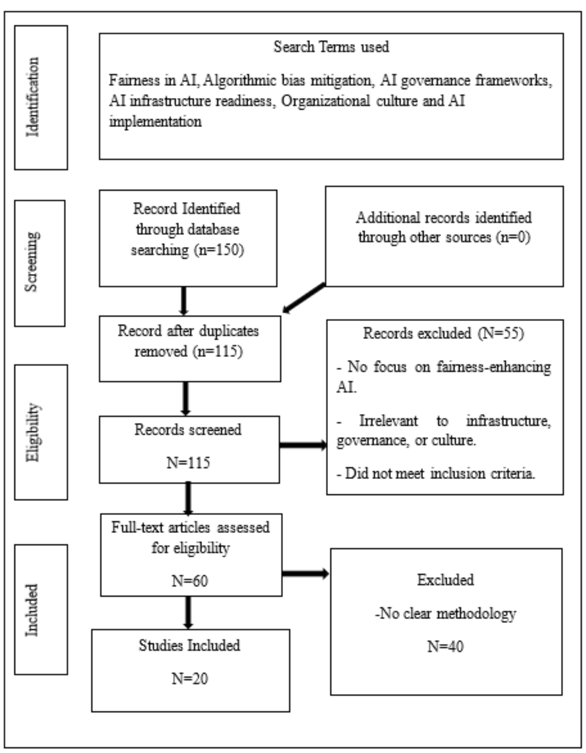

Figure 2 PRISMA Flow Diagram of the Study

RESULTS

The systematic review has been able to identify 20 eligible studies that overall evidence the idea that the introduction of fairness-enhancing AI technologies is being influenced by three closely related themes, which are infrastructure preparedness, policy and regulatory alignment, and organizational culture and governance dynamics. These themes also represent the sociotechnical character of the fairness implementation discovered during the course of the review.

INFRASTRUCTURE READINESS AND TECHNICAL CAPACITY

The initial theme that prevails relates to how infrastructure forms the basis of operationalization of fairness. It has been explicitly pointed out in the analyzed articles that risk metrics and bias reduction methodologies require well-developed systems of data governance, lifecycle management systems, documentation rules, and audit systems Pagano et al. (2022). It is within organizations whose data flow is not organized, reporting is inconsistent, and where they lack constant scrutiny mechanisms that do not abide with fairness interventions in long-term of development Agarwal et al. (2022), Framework Convention on Artificial Intelligence. (2026). Based on several studies, equity testing is normally conducted in the form of an experimental one-off technical test rather than a built-in process of monitoring Mitchell et al. (2019). This means that the results suggest that the concept of infrastructure is the technical infrastructure that assists in converting the abstract principles of fairness into practical forms of inquiry that can be measured and audited

POLICY AND REGULATORY ALIGNMENT

The second theme correlates with the regulatory assistance system and the system of governance. Although ethical codes and AIs principles of non-discrimination, transparency, and accountability tend to be accepted Mittelstadt et al. (2016), Floridi and Cowls (2019), the review indicates areas of inconsistencies in their implementation. Several policy regimes are not binding as far as adhering and are not fully congruent with internal organizational processes Pagano et al. (2022). This kind of policy-practice disparity negatively affects accountability as it does unfair in the institutions Kroll et al. (2017). Based on the studies, AI technologies which encourage fairness have more likelihood of being systematically introduced through the assistance of transparent standards of rule, enforceable reporting frameworks, and institutional oversight systems Murire (2024).

ORGANIZATIONAL CULTURE AND INSTITUTIONAL PRACTICES

The culture of the organization is the third theme discovered to show the role of the culture of the organization in the mediation of the infrastructure and policy. To make a step towards the leadership commitment, the interdisciplinary work, the consciousness of the ethicality and the regular incentives are a must in making fairness central or marginal Murire (2024). The performance-driven organizational cultures that are mainly concerned with efficiency can consider fairness as a peripheral consideration except when made strategic Rakova et al. (2021). Common barriers leading to cultural resistance, poor clarity of ownership of fairness responsibilities, and insufficient expertise within an organization are also identified in the review Madaio et al. (2020). The cultural dimensions are not studied well as compared to the technical discussions giving a clear gap in the study of internal institutional dynamics Selbst et al. (2019).

Altogether, the findings show that AI technologies aimed at enhancing fairness can be the most effective when the infrastructural capacity, regulatory correspondence, and a supportive organizational culture are held together in an integrated system of the sociotechnical ecosystem.

DISCUSSION

This review findings provide an affirmation that fairness-promoting AI technologies cannot be successfully introduced with the help of technical solutions only. Instead, they flourish on a balanced integration of infrastructure, policy framework and culture of the organization. The relative lack of qualified studies will also reveal a research gap; when the measures of algorithmic fairness are widely studied, less literature has tried to show how these tools are being operationalized and situate it within the institution and operationality context Mittelstadt et al. (2016). This fact also aligns with more advanced arguments against AI ethics literature stating that the concept of fairness is often limited to mathematical terms rather than applied and practiced Floridi and Cowls (2019). The needed sociotechnical perspective incites an essential perspective therefore that the fairness outcomes are not only specified by the algorithms, but by the organizational mechanisms, to which they are substantially subjected Barocas et al. (2019).

The dominance of the research done by use of technology would suggest that the theory and the practice of fairness have never been at par. Mitigation plans of the bias and the fairness are increasingly becoming sophisticated, however, only when they are backed by viable data governance, lifecycle monitoring and standardized documentation systems, they can be efficient Mitchell et al. (2019). Devoid of such systems, checks on fairness will result in isolated reviews contrary to continuous accountability systems. The last studies on AI auditing mention that internal audit procedures, open reporting systems, and surveillance are critical to the operation of the goal of fairness Raji et al. (2020). These findings help to substantiate the importance of infrastructure as the layer within which the fairness enhancing technologies can be implemented and put into work on a sustainable basis.

Another factor that is determining but unequal is the policy structures. Despite the fact that international principles and domestic AI rely on the idea of non-discrimination and transparency, their application to enforceable principles remains unbalanced [15]. Regulatory uncertainty applies in organizations that lower the drive to add fairness and rather concentrate on passive compliance, just the minimum before the law Selbst et al. (2019). To make the AI ethics a compelling practice consideration, as governance scholars argue, the institutional enforcement mechanisms should be installed, and accountability channels well defined, to implement ethics in AI practice Kroll et al. (2017).

Organizational culture has been identified to mediate the policy intent and technical implementation. There is high influence of leadership commitment, interdisciplinary team-work and incentive alignment in the prioritizing of fairness or second-fiddle [18]. Performance-oriented cultures that highly respect efficient cultures may perceive fairness as an appendage with the exception of being tactically guaranteed Madaio et al. (2020), Selbst et al. (2019). Overall, the discussion indicates the notion that the use of AI technologies that increase fairness can be productive in case they are supported by a rational sociotechnical ecosystem comprising the ability of the infrastructure, enforceable rules, and moral organizational activities.

CONCLUSION

The paper had performed systematic literature review on the influence of the infrastructure, policy framework, and organizational culture on the successful uptake of AI technologies that enhance fairness using a systematic PRISMA-based literature review. They indicate that the problem of artificial intelligence fairness is not a computational but rather a sociotechnical one. Despite the high level of technical features regarding the mitigation of bias and metrics of fairness development, it should be mentioned that they may be considered only practically effective in case of the strong data governance systems and the lifecycle monitoring systems and auditing systems. It is also indicated in the review that the policy frameworks even after becoming increasingly articulated at the international and national levels still aim to lack clear operationalization; therefore, there is a lack of connection between the ethical values and the realization of them. It was also established that the organizational culture is a critical mediating variable that helps determine the internalization of priorities of fairness, resources mobilization and sustenance in institutions. These things as the commitment of the leaders, the cross-disciplinary teamwork as well as constant incentive systems serve a vital role in making fairness a routine or a token age-old appearance. Overall, the synthesis validates the fact that AI technologies that increase fairness can be effective in the case when the infrastructure capacity, compatibility of the regulations, and ethical principles of the organization collaborate with each other. The emphasis on the limited integrative literature suggests that the future of AI is a gap in research which ought to enable technical innovation to accommodate institutional governance and cultural transformation in a way that AI can be executed in a responsible and sustainable way.

ACKNOWLEDGMENTS

None.

REFERENCES

ACM Conference on Fairness, Accountability, and Transparency. (2026). In Wikipedia.

Agarwal, A., Agarwal, H., and Agarwal, N. (2022). Fairness Score and Process Standardization: Framework for Fairness Certification in Artificial Intelligence systems (arXiv:2201.06952). arXiv.

Barocas, S., Hardt, M., and Narayanan, A. (2019). Fairness and Machine Learning: Limitations and opportunities. https://fairmlbook.org

Floridi, L., and Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review, 1(1).

Framework Convention on Artificial Intelligence. (2026). In Wikipedia. https://en.wikipedia.org/wiki/Framework_Convention_on_Artificial_Intelligence

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., and Yu, H. (2017). Accountable algorithms. University of Pennsylvania Law Review, 165(3), 633–705. https://doi.org/10.2139/ssrn.2765268

Madaio, S., Stark, L., Wortman Vaughan, J., and Wallach, H. (2020). Co-Designing Checklists to Understand Organizational Challenges and Opportunities Around Fairness in AI. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT) (343–353). https://doi.org/10.1145/3351095.3372878

Mitchell, M., et al. (2019). Model Cards for Model Reporting. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT\) (220–229). https://doi.org/10.1145/3287560.3287596

Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., and Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. Big Data & Society, 3(2). https://doi.org/10.1177/2053951716679679

Murire, O. T. (2024). Artificial Intelligence and its Role in Shaping Organizational Work Practices and Culture. Business, 14(12).

OECD. (2019). OECD Principles on Artificial Intelligence.

Pagano, T. P., et al. (2022). Bias and Unfairness in Machine Learning Models: A Systematic Literature Review (arXiv:2202.08176). arXiv.

Papagiannidis, E. (2025). Responsible Artificial Intelligence Governance: A Review. Technological Forecasting and Social Change.

Raji, I. D., et al. (2020). Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAccT) (33–44). https://doi.org/10.1145/3351095.3372873

Rakova, B., Yang, J., Cramer, H., and Chowdhury, R. (2021). Where Responsible AI Meets Reality: Practitioner Perspectives on Enablers for Shifting Organizational Practices. In Proceedings of the ACM CHI Conference on Human Factors in Computing Systems (1–23). https://doi.org/10.1145/3411764.3445518

Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., and Vertesi, J. (2019). Fairness and Abstraction in Sociotechnical Systems. In Proceedings of the ACM Conference on Fairness, Accountability, and Transparency (FAT\) (59–68). https://doi.org/10.1145/3287560.3287598

Toronto Declaration. (2026). In Wikipedia. https://en.wikipedia.org/wiki/Toronto_Declaration

Álvarez, J. M., et al. (2024). Policy Advice and Best Practices on Bias and Fairness in AI. Ethics and Information Technology, 26, Article 31. https://doi.org/10.1007/s10676-024-09746-w