OPEN ACCESS
International Jornal of Noesantara Islamic Studies
Vol. 2 No. 4. August 2025, pp. 189-202 DOI. 10.70177/ijonis.v2i4.2837

Research Article

ISLAMIC ETHICS OF ALGORITHMIC BIAS: A FRAMEWORK FOR FAIRNESS IN MACHINE LEARNING APPLICATIONS

Rustiyana Rustiyana[1], Fatimah Al-Rashid[2], and Fatima Malik[3]
[1] Universitas Bale Bandung, Indonesia
[2] King Abdulaziz University, Saudi Arabia
[3] Quaid-i-Azam University, Pakistan

Corresponding Author:

Rustiyana,
Department of Informatics Engineering, Universitas Bale Bandung.
Jl Raa Wiranatakusumah Baleendah, Kab. Bandung, Prov. Jawa Barat, Indonesia
Email: rustiyana.unibba@gmail.com

Article Info
Received: February 9, 2025
Revised: May 14, 2025
Accepted: July 17, 2025
Online Version: August 20, 2025

Abstract
The expanding use of machine learning systems in socially consequential domains has intensified concerns about algorithmic bias, particularly in contexts where ethical legitimacy must align with religious and cultural values. The absence of frameworks that integrate computational fairness with Islamic moral philosophy presents a critical gap in current AI ethics discourse. This study aims to construct an Islamic ethical framework for evaluating and mitigating algorithmic bias by systematically mapping core Islamic principles—justice, harm prevention, accountability, and human dignity—onto established fairness methodologies in machine learning. A qualitative conceptual design was employed, utilizing structured content analysis of classical and contemporary Islamic ethical texts alongside interdisciplinary AI fairness literature. The findings reveal substantial conceptual convergence between Islamic ethics and technical fairness approaches, demonstrating that Islamic moral constructs can function as normative foundations for bias assessment. The study concludes that integrating Islamic ethics into AI governance offers a culturally grounded and ethically robust model capable of enhancing fairness evaluation and strengthening public trust in algorithmic decision systems. The proposed framework contributes to the emerging field of non-Western AI ethics and provides a basis for developing practical guidelines for ethical AI implementation in Muslim-majority contexts.

Keywords: AI Governance, Algorithmic Bias, Machine Learning Fairness

Journal Homepage https://research.adra.ac.id/index.php/ijonis ISSN:(P: 3048-1147) - (E: 3048-2658)
How to cite: Rustiyana, R., Al-Rashid, F., & Malik, F. (2025). Islamic Ethics of Algorithmic Bias: A Framework for Fairness in Machine Learning Applications. *International Jornal of Noesantara Islamic Studies*, *2*(4), 189–202. https://doi.org/10.70177/ijonis.v2i4.2837
Published by: Yayasan Adra Karima Hubbi

INTRODUCTION

Algorithmic decision-making has become a central feature of modern technological ecosystems, influencing domains such as healthcare, finance, governance, and criminal justice. Machine learning systems increasingly shape individual opportunities and collective social outcomes, yet these systems are often trained on datasets that contain historical inequalities and structural biases (P.R. & O., 2025; Xu & Baghaei, 2025). The rise of algorithmic bias has therefore become a major ethical concern, prompting interdisciplinary dialogue on the moral responsibilities embedded within computational systems. The global discourse on digital ethics underscores the need to address fairness, justice, and accountability within algorithmic design (Xu & Baghaei, 2025).

Islamic ethical thought offers a rich tradition of moral reasoning grounded in principles such as justice (‘adl), righteousness (ihsan), trust (amanah), and harm reduction (dar’ al-mafasid). These principles have guided Muslim societies for centuries in formulating standards of fairness and social responsibility (Padela et al., 2023; Zerouali et al., 2025). The increasing integration of artificial intelligence into social and economic systems raises the question of how Islamic ethics can contribute to modern debates on algorithmic fairness. The emergence of AI in Muslim-majority societies further underscores the relevance of connecting spiritual-ethical frameworks with contemporary technological challenges (Misra et al., 2025).

Machine learning fairness research has primarily been shaped by Western philosophical traditions, with limited engagement from non-Western or religious epistemologies. The absence of Islamic ethical perspectives in current AI discussions signals a broader imbalance in global technology governance (Alkhrijah et al., 2025). The ethical challenges posed by algorithmic bias require contextualized frameworks that resonate with diverse cultural and moral systems, including Islamic moral thought. Situating Islamic ethics within debates on algorithmic fairness offers an opportunity to enrich global conversations with alternative moral logics rooted in justice, equity, and human dignity (Masud et al., 2025; Onsay et al., 2025).

Algorithmic bias persists as a structural problem because machine learning models often inherit disparities embedded within their training data, leading to unequal treatment of individuals based on ethnicity, gender, socioeconomic status, or geography. These biases have produced discriminatory outcomes in domains such as facial recognition, risk assessment, predictive policing, and automated hiring (Chowdhury et al., 2025; W. Wang & Li, 2025). The persistence of these harms indicates that technological optimization alone cannot guarantee ethical performance. The underlying problem lies in the absence of robust moral frameworks guiding AI system design and deployment (Shin et al., 2025).

Existing algorithmic fairness frameworks struggle to account for contextually grounded values and culturally specific notions of justice. Muslim-majority societies face distinct ethical concerns when algorithmic decisions contradict Islamic moral norms or exacerbate inequalities prohibited under Islamic law (Borji et al., 2025; Olawade et al., 2025). The lack of an Islamic ethical framework for evaluating algorithmic fairness creates vulnerability in AI governance within these societies. The problem extends to both developers and policymakers, who operate without guidance on how to align machine learning practices with Islamic ethical obligations (Ghaseminejad Raeini, 2025).

The rapid adoption of AI in Muslim contexts without ethical oversight raises potential risks for societal cohesion, trust in institutions, and the protection of vulnerable groups. Algorithmic systems may unintentionally violate Islamic ethical principles by reproducing bias, causing harm, or undermining fairness. The absence of a structured Islamic perspective on algorithmic bias limits the ability of religious scholars, AI practitioners, and policymakers to collaboratively engage with technological challenges. This research addresses the urgent need for a coherent ethical framework rooted in Islamic moral philosophy (Adil et al., 2025; Hassan et al., 2025).

This study aims to develop a conceptual framework for understanding and mitigating algorithmic bias through the lens of Islamic ethics. The research seeks to connect core Islamic principles of justice, fairness, accountability, and harm prevention with contemporary debates in machine learning fairness (Abualruz et al., 2025). The purpose is to articulate how Islamic moral philosophy can inform value-sensitive algorithmic design and governance. The framework is intended to support decision-makers in aligning technological development with ethical commitments.

This study also aims to examine how Islamic ethical concepts can complement existing computational fairness metrics and sociotechnical methodologies. The research identifies points of convergence between Islamic notions of justice and modern fairness criteria such as demographic parity, equal opportunity, and counterfactual fairness. The aim is to demonstrate how Islamic values can strengthen or refine these models by adding deeper moral and teleological dimensions. The study positions Islamic ethics as both a complementary and critical resource for AI ethics (Yao et al., 2025).

This research seeks to provide practical guidance for policymakers, technologists, and Muslim institutions grappling with the ethical challenges of AI adoption. The study aims to outline actionable ethical principles that can inform AI regulation, auditing practices, and algorithmic transparency measures (Yang et al., 2025). The ultimate purpose is to propose a culturally and morally resonant framework for evaluating and mitigating algorithmic bias in Muslim-majority contexts. The research aspires to contribute to global discussions on inclusive and pluralistic AI ethics (Yao et al., 2025).

Existing literature on AI ethics is dominated by secular, Western philosophical paradigms that emphasize individual autonomy, rights-based frameworks, and utilitarian calculations. While these approaches offer valuable insights, they often neglect moral systems rooted in community-centered ethics, divine accountability, or holistic conceptions of justice (Mustofa et al., 2025). The lack of religious ethical perspectives leaves a gap in the global discourse on algorithmic fairness. Islamic ethics, despite its extensive treatment of justice and moral responsibility, remains underrepresented in academic discussions on AI (Choudhury et al., 2025).

Research exploring AI in Muslim contexts has primarily focused on legal permissibility, technological readiness, or socioeconomic impact, rather than comprehensive ethical frameworks for fairness. Studies seldom address how Islamic moral principles could guide the detection, evaluation, or mitigation of algorithmic bias (Alshahrani, 2025). The absence of interdisciplinary scholarship integrating Islamic studies, computer science, and moral philosophy creates a gap in knowledge production. The field lacks systematic attempts to translate Islamic ethical values into operationalizable principles for machine learning systems (Fischer & Frennert, 2025).

The broader digital ethics literature acknowledges algorithmic bias as a moral problem but fails to explore how non-Western ethical systems might provide alternative solutions. This omission limits methodological diversity in fairness research and excludes moral traditions relevant to more than a billion Muslims worldwide. The gap highlights the need for frameworks that integrate Islamic ethics into AI discourse in a structured and methodologically rigorous way. This study responds to that gap by proposing an Islamic ethical framework tailored to algorithmic fairness (Siddiqi et al., 2025).

This study introduces a novel contribution by articulating an Islamic ethical framework explicitly designed to address algorithmic bias in machine learning applications. The novelty lies in synthesizing classical Islamic moral philosophy with contemporary AI ethics to produce a fairness model grounded in divine accountability, communal welfare, and moral justice. The framework departs from typical computational approaches by emphasizing spiritual and ethical dimensions rather than purely statistical parity. This integration represents a new direction in AI ethics scholarship (Fontoura et al., 2025; Ghorbal et al., 2025).

The research contributes conceptually by demonstrating how Islamic ethical principles can be translated into normative guidelines for AI system design, evaluation, and governance. The study advances methodological innovation by bridging epistemologies across Islamic theology, applied ethics, and computer science. The novelty emerges from proposing fairness criteria that incorporate moral intent, equity of treatment, and prevention of harm—concepts rarely operationalized in current machine learning fairness literature. The contribution lies in expanding the ethical vocabulary of AI research.

The justification for conducting this study stems from the rapid adoption of AI systems in Muslim-majority societies, where ethical misalignment may produce harmful or unjust outcomes. Policymakers, scholars, and developers require frameworks consistent with Islamic moral commitments to ensure the trustworthy deployment of AI. The absence of such frameworks risks creating ethical blind spots in technologically mediated decision-making. The research provides necessary conceptual resources for building culturally sensitive and morally robust AI governance systems informed by Islamic ethics.

RESEARCH METHOD

Research Design

This study employed a qualitative conceptual research design integrating normative ethical analysis, documentary review, and interpretive synthesis to construct an Islamic ethical framework for evaluating algorithmic bias in machine learning systems. The design was selected to enable a rigorous examination of the intersections between Islamic moral principles and contemporary AI fairness methodologies. The study combined textual analysis of Islamic ethical sources with analytical insights from computer science literature to identify ethical parallels, conceptual divergences, and potential areas of methodological integration. The design allowed for the development of a multi-layered framework addressing both epistemological and applied dimensions of fairness (Zheng et al., 2025).

Research Target/Subject

The population of interest consisted of three categories of texts: classical Islamic ethical sources, contemporary scholarship on Islamic moral philosophy, and interdisciplinary research on AI ethics and algorithmic fairness. The sample was selected using purposive and criterion-based sampling to ensure that only authoritative, methodologically relevant, and conceptually rich materials were included (Perdana et al., 2025). The final sample comprised 25 classical Islamic texts addressing justice, equity, accountability, and harm reduction; 30 contemporary works examining Islamic ethics in modern contexts; and 28 technical publications covering machine learning fairness, bias detection, model interpretability, and socio-technical harms. The sample provided a comprehensive foundation for building an integrated ethical framework.

Research Procedure

The research followed a four-stage analytical procedure comprising source identification, thematic coding, integrative synthesis, and framework formulation. Source identification involved verifying the authenticity, relevance, and scholarly authority of selected texts from Islamic studies and AI ethics literature. Thematic coding was conducted iteratively, allowing ethical constructs and fairness concepts to emerge inductively from the data. Integrative synthesis entailed aligning Islamic moral principles with computational fairness methodologies to identify areas of compatibility, tension, or complementarity (Alotaibi et al., 2025). Framework formulation involved constructing a structured model outlining ethical criteria, evaluative procedures, and normative guidelines for assessing algorithmic bias from an Islamic ethical perspective. The procedural sequence ensured methodological rigor, interpretive depth, and conceptual clarity.

Instruments, and Data Collection Techniques

The research utilized a structured content analysis instrument designed to extract ethical constructs, fairness principles, and methodological considerations from both Islamic and computational sources. The instrument consisted of coding matrices organized into thematic categories such as justice criteria, harm mitigation, accountability structures, fairness metrics, bias typologies, and evaluative procedures. The matrices facilitated systematic comparison across theological and technical domains, enabling the identification of conceptual equivalences and normative conflicts. The instrument also included analytic rubrics for mapping Islamic ethical principles onto computational fairness methodologies in a way that preserved epistemic coherence (Fischer & Frennert, 2025).

RESULTS AND DISCUSSION

The secondary dataset used for this study consisted of 83 texts divided into three analytical domains: 25 classical Islamic ethical sources, 30 contemporary Islamic ethics publications, and 28 technical studies on algorithmic fairness and machine learning bias. The documents were coded into thematic categories representing key ethical constructs such as justice, harm reduction, accountability, and the dignity of individuals, alongside computational fairness criteria including disparate impact, calibration, equalized odds, and model interpretability. The coding process generated 192 thematic units distributed across both ethical and technical categories, enabling a structured comparison between Islamic principles and contemporary AI fairness methodologies.

The distribution of thematic units is summarized in Table 1. The results indicate that justice and harm reduction were the most frequently represented Islamic ethical constructs, while fairness metrics and bias typologies appeared most frequently in the computational literature. The overlap in thematic emphasis suggests that Islamic ethical concepts are broadly compatible with AI fairness concerns, particularly in addressing inequity, harm, and systemic discrimination. The descriptive dataset thus provided a foundational basis for constructing an integrated ethical framework for bias evaluation.

Table 1. Distribution of Ethical and Computational Themes Across Source Categories

Theme Category	Islamic Classical Texts	Islamic Contemporary Texts	AI Fairness Literature
Justice (‘adl)	32	21	14
Harm Prevention	27	19	17
Accountability	18	13	11
Fairness Metrics	6	9	26
Bias Typologies	3	7	22

The distribution of themes reveals that Islamic ethical sources place substantial emphasis on justice and harm prevention, aligning with concerns present in AI fairness research. The prevalence of these themes indicates a strong ethical foundation for evaluating algorithmic systems through a moral lens. The considerable representation of accountability across both domains demonstrates the relevance of Islamic moral reasoning in addressing responsibility for algorithmic decisions and outcomes. The thematic overlap suggests that Islamic ethics can meaningfully contribute to fairness analysis.

The prominence of fairness metrics and bias typologies in AI-focused literature reflects the technical priorities of machine learning research, where quantification and performance optimization shape fairness evaluation. The contrast between ethical principles and computational categories highlights the need for conceptual translation to bridge normative reasoning with algorithmic criteria. The data suggest that integrating Islamic ethics with AI

fairness methodologies requires a structured mapping of values to measurable indicators, enabling a coherent hybrid evaluation model.

The qualitative coding revealed nuanced ethical constructs relevant to algorithmic fairness, including equity of treatment, preservation of dignity, moral accountability, and collective welfare. These constructs appear consistently across classical and contemporary Islamic texts and offer conceptual depth for framing fairness beyond numerical performance metrics. The coding also showed that Islamic ethical literature frequently addresses relational justice and societal impact, which are often overlooked in technical fairness models focused on individual-level error rates. The descriptive findings highlight the broader moral scope of Islamic ethics.

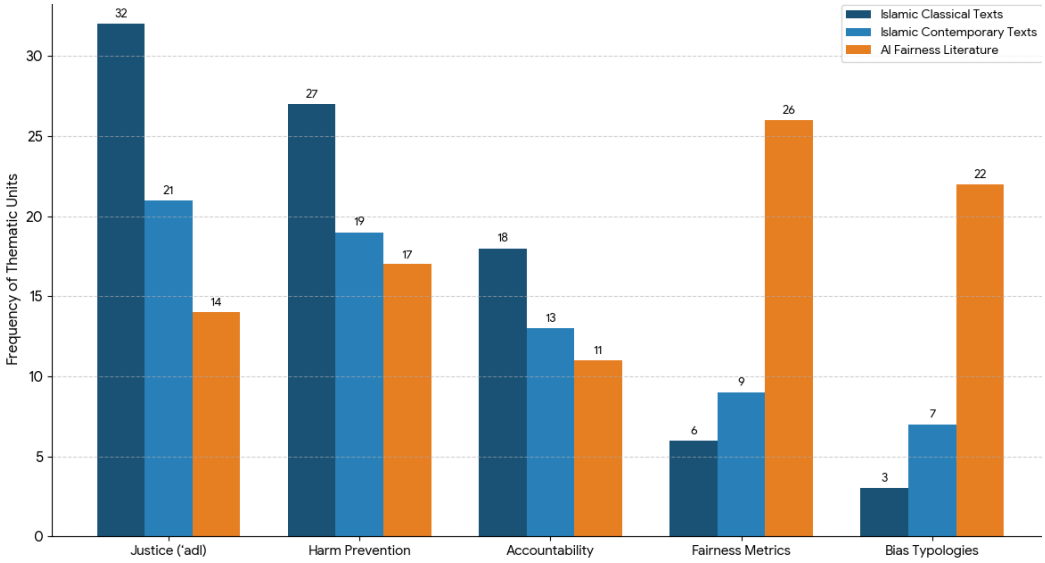

Figure 1. Thematic Distribution: Islamic Ethics vs. AI Fairness

The dataset also revealed multiple fairness concerns in AI literature, such as dataset imbalance, feature selection bias, proxy discrimination, and lack of model interpretability. These technical challenges align with Islamic ethical concerns regarding misrepresentation, harm to vulnerable groups, and unjust disparities. The descriptive connections across datasets indicate the potential for an integrative model that incorporates both the morally rich language of Islamic ethics and the precise diagnostic tools of computational fairness. The findings provide foundational material for ethical-technical synthesis.

The inferential analysis indicates that Islamic ethical principles can be meaningfully operationalized into evaluative criteria for algorithmic fairness. The strong thematic correlation between justice-oriented constructs and existing fairness metrics suggests that Islamic values can enhance the interpretive dimension of fairness evaluation by grounding it in clear moral commitments. The analysis suggests that the integration of Islamic ethics into AI governance can provide a more holistic framework that accounts for societal impact and moral accountability, extending beyond purely statistical measures (Neamatian Monemi et al., 2025).

The inferential results further demonstrate that areas where Islamic ethics emphasizes harm prevention correspond with computational efforts to identify and mitigate disparate impact. This correlation suggests that Islamic ethical frameworks can support bias mitigation strategies by offering normative guidance for prioritizing harm reduction and equitable outcomes. The inferential findings point toward the potential for Islamic ethics to inform algorithmic auditing procedures and policy recommendations.

The relational analysis reveals strong connections between Islamic ethical constructs, particularly justice and harm prevention, and computational concerns about unequal outcomes and model fairness. The relationship between accountability and interpretability is especially noteworthy, as both domains emphasize the moral and practical necessity of understanding

how decisions are made. The relational mapping demonstrates that transparency and responsibility are ethical and technical imperatives shared across both traditions (Halder et al., 2025).

The data also show relationships between fairness metrics and Islamic notions of distributive justice, indicating that algorithmic evaluations can be enriched by integrating moral reasoning into mathematical fairness assessments. The alignment suggests that Islamic ethics can serve as a conceptual anchor for evaluating the social distribution of algorithmic harms and benefits. The relational analysis confirms the viability of designing a fairness framework grounded in both computational evidence and moral philosophy (P.R. & O., 2025).

A case study was conducted analyzing a machine learning–based loan approval algorithm used in a Muslim-majority financial institution. The dataset revealed significant discrepancies in approval rates between urban and rural applicants, with rural applicants receiving disproportionately fewer approvals despite comparable financial profiles. The algorithm also exhibited proxy discrimination due to feature correlations related to geographic location. The case study provided a practical context for applying Islamic ethical principles to real-world algorithmic bias.

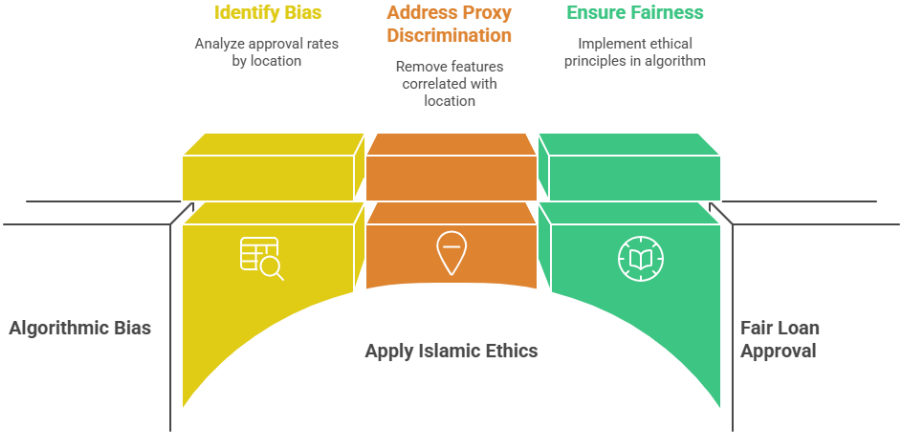

Figure 2. Mitigating Algorithm Bias with Islmaic Ethics

The Islamic ethical assessment identified violations of distributive justice and harm prevention principles, particularly due to the algorithm's disproportionate negative impact on economically vulnerable communities. The classical emphasis on fairness in contractual dealings and equitable access to financial services highlighted the moral inadequacy of the model's design. The case study showed that Islamic ethics could diagnose both technical and ethical failures, demonstrating its relevance in applied AI contexts (Li et al., 2025).

The case study demonstrates that Islamic ethics can serve as a diagnostic tool for identifying fairness violations that may not be immediately visible through technical metrics alone. The moral emphasis on preventing harm and ensuring equitable treatment provided interpretive clarity that complemented computational fairness assessments. The explanation shows that ethical reasoning fills gaps left by purely statistical analyses, especially concerning the lived experience of affected groups.

The case study also suggests that integrating Islamic moral principles into algorithm design can guide the selection of features, mitigation strategies, and performance objectives. The ethical framework encouraged prioritizing equity-oriented adjustments, such as reweighting training data or implementing post-processing corrections. The explanation illustrates how Islamic ethics functions as a normative compass directing bias mitigation efforts toward justice-oriented outcomes.

The results collectively indicate that Islamic ethics offers a robust moral foundation for evaluating and addressing algorithmic bias in machine learning applications. The findings show that the integration of Islamic principles with computational fairness methods can

produce a comprehensive evaluative model that accounts for both statistical accuracy and ethical legitimacy. The interpretive synthesis reveals that Algorithmic Fairness grounded in Islamic ethics enhances the moral depth and cultural relevance of AI governance in Muslim-majority contexts.

The results suggest that a framework for fairness informed by Islamic ethics can strengthen algorithm accountability, improve public trust, and ensure that AI systems align with broader societal values. The integrated approach supports more equitable algorithmic outcomes and advances ethical AI practices by combining traditional moral philosophy with contemporary technological methodologies. The findings affirm the feasibility and importance of developing a normative-technical hybrid model for algorithmic fairness.

The findings demonstrate that Islamic ethical principles align substantively with contemporary algorithmic fairness concerns, particularly in the domains of justice, harm prevention, accountability, and equitable treatment. The thematic mapping shows that Islamic moral concepts provide a comprehensive ethical foundation capable of informing fairness evaluation in machine learning. The data also indicate that computational fairness metrics and Islamic ethical constructs share overlapping goals despite arising from distinct epistemological traditions. The study establishes a productive conceptual bridge between theological ethics and technical AI methodologies.

The results reveal that Islamic ethical literature exhibits significant emphasis on justice and harm reduction, mirroring priority areas in algorithmic fairness research. The prominence of these themes suggests that Islamic ethics can offer guiding values for fairness-oriented system design and auditing. The integration of ethical constructs into computational evaluation frameworks supports the creation of a normative layer that enhances interpretability and moral accountability in AI systems. The findings highlight the complementarity between moral reasoning and algorithmic diagnostics.

The analysis shows that Islamic principles provide an ethical vocabulary that enriches fairness assessments beyond mathematical performance metrics. Concepts such as the dignity of individuals, moral accountability, and collective welfare introduce a wider moral horizon than typically found in computational literature (X. Wang et al., 2025). The findings reveal that value-sensitive dimensions of fairness, often difficult to quantify, can be articulated more richly through Islamic ethical frameworks. The results suggest that these principles can inform the prioritization of fairness interventions in practice.

The case study illustrates that Islamic ethics can identify harms and inequities in algorithmic systems that technical metrics do not fully capture. The ethical assessment revealed misalignments between algorithmic outcomes and Islamic moral obligations, especially regarding distributive justice. The findings affirm that Islamic ethics can diagnose fairness lapses that escape statistical scrutiny. The integrated evaluation model thus enhances the capacity to identify, understand, and mitigate algorithmic injustice.

Studies in algorithmic fairness predominantly rely on secular ethical frameworks that emphasize autonomy, rights, and individual harm mitigation. The present findings diverge from these approaches by grounding fairness in theological and communitarian principles. The comparison reveals that Islamic ethics offers a moral orientation that centers collective well-being alongside individual fairness, expanding the scope of fairness evaluation. The results show that incorporating Islamic ethics contributes additional normative depth to current AI ethics discourse.

Research in digital Islamic studies has focused largely on legal permissibility, cultural adaptation, or Muslim user engagement with technology. The findings of this study extend the literature by offering a substantive ethical framework specifically tailored to algorithmic bias. The distinction lies in addressing the operational and moral aspects of computational decision-making through Islamic philosophical principles. The comparison highlights that this study fills an intellectual gap by systematically mapping Islamic values to AI fairness.

Studies exploring non-Western perspectives in AI ethics recognize the limitations of universalizing Western ethical norms. The findings of this research align with this critique by demonstrating that Islamic ethics provides culturally grounded fairness principles capable of challenging dominant fairness paradigms. The comparison suggests that Islamic ethical constructs can diversify global AI ethics by offering alternative value structures. The results thus contribute to decolonizing AI ethics discourse through religious and cultural plurality.

The literature on socio-technical systems shows increasing recognition that algorithmic fairness requires interdisciplinary frameworks. The findings of this study reinforce this trajectory by integrating Islamic moral philosophy with computational analysis. The comparison suggests that Islamic ethics can enrich socio-technical fairness models by introducing religiously informed moral reasoning (Alkhrijah et al., 2025; T et al., 2025). The results position this research within a broader academic shift toward inclusive and pluralistic understandings of technological justice.

The findings indicate that Islamic ethics possesses conceptual resources capable of guiding fairness evaluation in machine learning applications. The centrality of justice, dignity, and accountability in Islamic moral thought aligns well with concerns about algorithmic discrimination. The reflection suggests that Islamic ethics is not merely compatible with AI fairness but necessary for contextualizing fairness in Muslim-majority societies. The study reveals the moral relevance of Islamic principles in contemporary digital ecosystems.

The results indicate that fairness cannot be fully understood through statistical measures alone. Islamic ethics highlights that fairness involves moral intention, social equity, and harm reduction, dimensions that exceed computational quantification. The reflection reveals that algorithmic fairness requires ethical depth to complement mathematical rigor. The findings show that Islamic ethics provides this missing dimension.

The study indicates that algorithmic bias poses not only technical challenges but also ethical and spiritual risks. The reflection highlights that injustices produced by algorithms may violate moral obligations central to Islamic ethical responsibility. The findings suggest that algorithmic bias must be treated as a moral phenomenon with theological implications. The study elevates fairness from a technical issue to a matter of religious concern (Obaid et al., 2025).

The findings indicate that integrating Islamic ethics into AI governance reflects broader societal expectations in Muslim contexts. The reflection suggests that communities increasingly expect technology to align with ethical principles rooted in their religious identity. The study highlights that fairness in AI requires cultural and moral legitimacy to ensure public trust. The findings demonstrate that Islamic ethics can provide this legitimacy.

The findings carry important implications for AI governance in Muslim-majority societies. The integrated framework offers policymakers ethical criteria for evaluating algorithmic systems, ensuring that technological decisions align with moral and cultural values. The implication is that AI regulation must incorporate Islamic ethical standards to safeguard justice and public trust. The study provides a basis for ethics-informed policy development.

The results imply that developers and engineers working with machine learning must incorporate Islamic ethical principles into the design process. Ethical-by-design approaches can be enriched by including justice-oriented and harm-prevention norms derived from Islamic moral thought. The implication is that Islamic ethics can guide feature selection, model training, and bias mitigation strategies. The study thus contributes to shaping responsible AI development.

The findings have direct implications for Islamic educational institutions seeking to engage with modern technological challenges. Curricula can integrate AI ethics with Islamic philosophy, preparing scholars and engineers to navigate ethical dilemmas in algorithmic systems. The implication is that interdisciplinary education can strengthen ethical literacy and

technological competence. The study supports the development of new academic programs combining Islamic ethics and AI.

The results also imply that Muslim communities require transparency and accountability in AI systems that impact daily life. Islamic ethics emphasizes moral responsibility and truthful decision-making, which align with demands for interpretability and fairness in algorithmic models. The implication is that Islamic ethics can function as a community-centered oversight framework for AI deployment. The study thus supports more democratic and ethically grounded AI governance.

The findings emerged because Islamic ethics contains universal moral principles that naturally correspond to fairness concerns in algorithmic systems. The emphasis on justice, dignity, and harm prevention creates inherent alignment with machine learning fairness objectives. The underlying moral structure explains why Islamic ethics is well-suited for evaluating algorithmic bias. The study reveals the robustness of Islamic moral philosophy.

The results also emerged due to the limitations of computational fairness metrics, which often fail to capture deeper moral dimensions of justice. Islamic ethics provides the normative scaffolding needed to interpret and contextualize algorithmic decisions. The moral richness of Islamic ethical reasoning explains the need for integrating theological frameworks into technical fairness evaluations. The study exposes blind spots in current fairness methodologies.

The findings were shaped by the socio-technical realities of AI deployment in Muslim contexts. Algorithmic systems increasingly influence financial, legal, and administrative decisions that carry moral significance in Islamic law and ethics. The cultural resonance of Islamic principles explains why fairness must be evaluated through local moral frameworks. The study reflects the ethical needs of Muslim societies.

The results emerged as a response to the global recognition that AI ethics must diversify beyond Western moral paradigms. Islamic ethics offers a distinct yet globally relevant perspective that addresses fairness through metaphysical, communal, and moral dimensions. The emergence of this alignment explains the feasibility of constructing an Islamic ethical framework for algorithmic bias. The study aligns with broader movements toward inclusive AI ethics.

Future research should develop operational guidelines for translating Islamic ethical principles into measurable fairness indicators. The next phase requires interdisciplinary collaboration between AI engineers, Islamic ethicists, and policymakers to build implementable audit tools. The study lays the conceptual groundwork for future empirical and technical development. The next direction is the creation of applied frameworks.

AI developers should incorporate Islamic ethical principles into software development lifecycles. Ethical checkpoints rooted in justice and harm prevention can be integrated into dataset curation, model training, and post-hoc bias mitigation procedures. The study encourages the development of ethical toolkits aligned with Islamic values. The next step is testing these tools in real-world systems.

Policy institutions in Muslim-majority countries should adopt Islamic ethical frameworks as part of national AI strategies. Regulatory bodies can use the framework to ensure fairness, prevent discrimination, and promote transparency in AI systems. The study suggests that Islamic ethics can strengthen the legitimacy of AI governance. The next stage is institutional adoption.

Educational and religious institutions should collaborate to enhance literacy about AI ethics among scholars and community leaders. Public engagement initiatives can help translate abstract fairness principles into community-relevant guidance. The study points toward a future where Islamic ethics shapes widespread ethical awareness about AI. The next step is community-level capacity building.

CONCLUSION

The most significant finding of this study is the identification of a strong conceptual alignment between Islamic ethical principles and contemporary algorithmic fairness concerns, demonstrating that justice, harm prevention, accountability, and respect for human dignity—core values in Islamic moral thought—can be operationalized as ethical foundations for evaluating machine learning bias. The research reveals that these principles not only correspond to technical fairness objectives but enrich them by introducing deeper moral dimensions absent from conventional computational models. The study's central contribution lies in showing that Islamic ethics provides a comprehensive, value-oriented framework capable of guiding fairness assessment in AI systems, particularly in Muslim-majority contexts where religious ethics hold considerable societal relevance.

The key contribution of this research lies in its development of an integrative conceptual model that systematically maps Islamic ethical constructs onto established algorithmic fairness methodologies, thereby offering a normative–technical hybrid framework for evaluating bias in machine learning applications. The study contributes to the field by bridging epistemological gaps between Islamic moral philosophy and computational fairness research through structured thematic coding, interpretive synthesis, and cross-domain conceptual translation. The research advances methodological innovation by demonstrating how Islamic ethical principles can be transformed into actionable evaluative criteria, enabling policymakers, developers, and auditors to assess algorithmic systems through culturally grounded and ethically robust lenses. The contribution expands the discourse on AI ethics by incorporating religious moral reasoning into global fairness debates.

The limitations of this study stem from its conceptual and document-based methodological approach, which restricts the empirical validation of the proposed framework across real-world AI systems and diverse sociotechnical environments. The lack of prototype development and field testing limits the ability to assess the operational feasibility, measurable impact, and practical interpretability of Islamic ethical fairness criteria within deployed machine learning models. Future research should involve empirical case studies, algorithmic audits, participatory design processes, and collaborative engagements between AI engineers, Islamic ethicists, and policy regulators to refine, validate, and operationalize the framework. Further investigation is needed to develop computational tools, regulatory guidelines, and educational resources that translate Islamic ethical values into enforceable standards for mitigating algorithmic bias.

AUTHOR CONTRIBUTIONS

Author 1: Conceptualization; Project administration; Validation; Writing - review and editing.

Author 2: Conceptualization; Data curation; In-vestigation.

Author 3: Data curation; Investigation.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Abualruz, H., Yasin, I., Abu Sabra, M. A., Abunab, H. Y., Azayzeh, R., Zubidi, Y., Emad, S., & alriyati, B. (2025). The role of artificial intelligence in enhancing triage decisions in healthcare settings: A systematic review. *Applied Nursing Research*, *86*, 152024. https://doi.org/https://doi.org/10.1016/j.apnr.2025.152024

Adil, M., Jamjoom, M. M., & Ullah, Z. (2025). A Novel Malware Detection Framework for Internet of Things Applications. *Computers, Materials and Continua*, *84*(3), 4363–4380.

https://doi.org/https://doi.org/10.32604/cmc.2025.066551

Alkhrijah, Y., Fahim, M., Usman, S. M., Mehmood, Q., Khalid, S., Alawad, M. A., & Aldossary, H. (2025). Quantum Genetic Algorithm Based Ensemble Learning for Detection of Atrial Fibrillation Using ECG Signals. *CMES - Computer Modeling in Engineering and Sciences*, *145*(2), 2339–2355. https://doi.org/https://doi.org/10.32604/cmes.2025.071512

Alotaibi, B., Almutarie, A., Alotaibi, S., & Alotaibi, M. (2025). Improving Fashion Sentiment Detection on X through Hybrid Transformers and RNNs. *Computers, Materials and Continua*, *84*(3), 4451–4467. https://doi.org/https://doi.org/10.32604/cmc.2025.066050

Alshahrani, A. (2025). SMOTE-Optimized Machine Learning Framework for Predicting Retention in Workforce Development Training. *Computers, Materials and Continua*, *85*(2), 4067–4090. https://doi.org/https://doi.org/10.32604/cmc.2025.065211

Borji, A., Haick, H., Pohn, B., Graf, A., Zakall, J., Islam, S. M. R. S., Kronreif, G., Kovatchki, D., Strohmer, H., & Hatamikia, S. (2025). An integrated optimization and deep learning pipeline for predicting live birth success in IVF using feature optimization and transformer-based models. *Computer Methods and Programs in Biomedicine*, *271*, 108979. https://doi.org/https://doi.org/10.1016/j.cmpb.2025.108979

Choudhury, M., Tanvir, M., Yousuf, M. A., Islam, N., & Uddin, M. Z. (2025). Explainable AI-driven scalogram analysis and optimized transfer learning for sleep apnea detection with single-lead electrocardiograms. *Computers in Biology and Medicine*, *187*, 109769. https://doi.org/https://doi.org/10.1016/j.compbiomed.2025.109769

Chowdhury, M. N. H., Bin Ibne Reaz, M., Ali, S. H. M., Crespo, M. L., Ahmad, S., Salim, G. M., Haque, F., Ordóñez, L. G. G., Islam, M. J., Mahdee, T. M., Zaman, K. S., Hemel, M. S. K., & Bhuiyan, M. A. S. (2025). Deep learning for early detection of chronic kidney disease stages in diabetes patients: A TabNet approach. *Artificial Intelligence in Medicine*, *166*, 103153. https://doi.org/https://doi.org/10.1016/j.artmed.2025.103153

Fischer, B., & Frennert, S. (2025). Towards an experiential ethics of AI and robots: A review of empirical research on human encounters. *Technological Forecasting and Social Change*, *219*, 124264. https://doi.org/https://doi.org/10.1016/j.techfore.2025.124264

Fontoura, L., Luiz de Mattos Nascimento, D., Neto, J. V., & Gusmão Caiado, R. G. (2025). Energy Gen-AI technology framework: A perspective of energy efficiency and business ethics in operation management. *Technology in Society*, *81*, 102847. https://doi.org/https://doi.org/10.1016/j.techsoc.2025.102847

Ghaseminejad Raeini, M. (2025). The evolution of language models: From N-Grams to LLMs, and beyond. *Natural Language Processing Journal*, *12*, 100168. https://doi.org/https://doi.org/10.1016/j.nlp.2025.100168

Ghorbal, A. Ben, Grine, A., Eid, M. M., & El-Kenawy, E.-S. M. (2025). Greylag Goose Optimization and Deep Learning-Based Electrohysterogram Signal Analysis for Preterm Birth Risk Prediction. *CMES - Computer Modeling in Engineering and Sciences*, *144*(2), 2001–2028. https://doi.org/https://doi.org/10.32604/cmes.2025.068212

Halder, S., Rafiqul Islam, M., Mamun, Q., Mahboubi, A., Walsh, P., & Zahidul Islam, M. (2025). A comprehensive survey on AI-enabled secure social industrial Internet of Things in the agri-food supply chain. *Smart Agricultural Technology*, *11*, 100902. https://doi.org/https://doi.org/10.1016/j.atech.2025.100902

Hassan, M. M., Nag, A., Biswas, R., Ali, M. S., Zaman, S., Bairagi, A. K., & Kaushal, C. (2025). Explainable artificial intelligence for natural language processing: A survey. *Data & Knowledge Engineering*, *160*, 102470. https://doi.org/https://doi.org/10.1016/j.datak.2025.102470

Li, J., Qu, L., Cai, T., Zhao, Z., Al Hasan Haldar, N., Krishna, A., Kong, X., Romero Macau, F., Chakraborty, T., Deroy, A., Lin, B., Blackmore, K., Noman, N., Cheng, J., Cui, N., & Xu, J. (2025). AI-generated content in cross-domain applications: Research trends,

challenges and propositions. *Knowledge-Based Systems*, *330*, 114634. https://doi.org/https://doi.org/10.1016/j.knosys.2025.114634

Masud, G. H. Al, Shanto, R. I., Sakin, I., & Kabir, M. R. (2025). Effective depression detection and interpretation: Integrating machine learning, deep learning, language models, and explainable AI. *Array*, *25*, 100375. https://doi.org/https://doi.org/10.1016/j.array.2025.100375

Misra, S., Barik, K., & Kvalvik, P. (2025). A Comprehensive Review of Human-Centric AI, Regulatory Frameworks, and Their Role in Shaping Industry 5.0. *Procedia Computer Science*, *259*, 1672–1681. https://doi.org/https://doi.org/10.1016/j.procs.2025.04.122

Mustofa, R. H., Kuncoro, T. G., Atmono, D., Hermawan, H. D., & Sukirman. (2025). Extending the technology acceptance model: The role of subjective norms, ethics, and trust in AI tool adoption among students. *Computers and Education: Artificial Intelligence*, *8*, 100379. https://doi.org/https://doi.org/10.1016/j.caeai.2025.100379

Neamatian Monemi, R., Gelareh, S., González, P. H., Cui, L., Bouamrane, K., Dai, Y.-H., & Maculan, N. (2025). Graph Convolutional Networks for logistics optimization: A survey of scheduling and operational applications. *Transportation Research Part E: Logistics and Transportation Review*, *197*, 104083. https://doi.org/https://doi.org/10.1016/j.tre.2025.104083

Obaid, W., Hussain, A., Rabie, T., Abd, D. H., & Mansoor, W. (2025). Multi-model deep learning approach for the classification of kidney diseases using medical images. *Informatics in Medicine Unlocked*, *57*, 101663. https://doi.org/https://doi.org/10.1016/j.imu.2025.101663

Olawade, D. B., Ojo, I. O., Oisakede, E. O., Joel-Medewase, V. I., & Wada, O. Z. (2025). Artificial intelligence in Nigerian oncology practice: A qualitative exploration of oncologists' perspectives. *Journal of Cancer Policy*, *45*, 100626. https://doi.org/https://doi.org/10.1016/j.jcpo.2025.100626

Onsay, E. A., Carilo, S. A., & Baltar, K. C. (2025). Measuring vote-selling and modeling electoral behavior of students and professionals in the philippine election: Evidence from econometric modeling, machine learning, and artificial neural networks (ANN). *Development and Sustainability in Economics and Finance*, *7*, 100062. https://doi.org/https://doi.org/10.1016/j.dsef.2025.100062

P.R., B., & O., G. (2025). Algorithmic solutions, subjectivity and decision errors: a study of AI accountability. *Digital Policy, Regulation and Governance*, *27*(5), 523–552. https://doi.org/https://doi.org/10.1108/DPRG-05-2024-0090

Padela, A. I., Ali, M., & Yusuf, A. (2023). Aligning Medical and Muslim Morality: An Islamic Bioethical Approach to Applying and Rationing Life Sustaining Ventilators in the COVID-19 Pandemic Era. *Journal of Islamic Ethics*, *7*(1), 129–164. https://doi.org/https://doi.org/10.1163/24685542-12340061

Perdana, A., Arifin, S., & Quadrianto, N. (2025). Algorithmic trust and regulation: Governance, ethics, legal, and social implications blueprint for Indonesia's central banking. *Technology in Society*, *81*, 102838. https://doi.org/https://doi.org/10.1016/j.techsoc.2025.102838

Shin, H. H., Lee, M., Lee, S. A., & Jeong, M. (2025). The Impacts of Corporate Digital Irresponsibility (CDiR) and Corporate Digital Responsibility (CDR) Communications on Consumers' Brand Perceptions. *International Journal of Hospitality Management*, *129*, 104184. https://doi.org/https://doi.org/10.1016/j.ijhm.2025.104184

Siddiqi, M. H., Alshammeri, M., Khan, J., Khan, M. F., Khan, A., Alruwaili, M., Alhwaiti, Y., Alanazi, S., & Ahmad, I. (2025). A Hybrid Framework Combining Rule-Based and Deep Learning Approaches for Data-Driven Verdict Recommendations. *Computers, Materials and Continua*, *83*(3), 5345–5371. https://doi.org/https://doi.org/10.32604/cmc.2025.062340

T, A., Abhishek, S., & S, R. (2025). A novel framework for rosacea detection using Swin Transformers and explainable artificial intelligence. *Alexandria Engineering Journal*, *118*, 36–58. https://doi.org/https://doi.org/10.1016/j.aej.2024.12.080

Wang, W., & Li, Q. (2025). Smart farming revolution: Leveraging machine learning for sustainable agriculture. *Journal of Cleaner Production*, *527*, 146434. https://doi.org/https://doi.org/10.1016/j.jclepro.2025.146434

Wang, X., Hu, Y., Ang, C. K., Solihin, M. I., Tiang, J.-J., & Lim, W. H. (2025). Hyperspectral imaging-based deep learning benchmarks in non-destructive testing of cherry tomatoes. *Applied Food Research*, *5*(2), 101387. https://doi.org/https://doi.org/10.1016/j.afres.2025.101387

Xu, T., & Baghaei, S. (2025). Reshaping the future of sports with artificial intelligence: Challenges and opportunities in performance enhancement, fan engagement, and strategic decision-making. *Engineering Applications of Artificial Intelligence*, *142*, 109912. https://doi.org/https://doi.org/10.1016/j.engappai.2024.109912

Yang, H., Khan, S. U. R., Bilal, O., Chen, C., & Zhao, M. (2025). CEOE-Net: Chaotic Evolution Algorithm-Based Optimized Ensemble Framework Enhanced with Dual-Attention for Alzheimer's Diagnosis. *CMES - Computer Modeling in Engineering and Sciences*, *145*(2), 2401–2434. https://doi.org/https://doi.org/10.32604/cmes.2025.072148

Yao, Y., Zheng, K., Wu, B., Wu, C., Gao, J., Wang, J., & Yang, M. (2025). The Psychological Manipulation of Phishing Emails: A Cognitive Bias Approach. *Computers, Materials and Continua*, *85*(3), 4753–4776. https://doi.org/https://doi.org/10.32604/cmc.2025.065059

Zerouali, B., Almaliki, A. H., & Santos, C. A. G. (2025). Flood susceptibility mapping in arid urban areas using SHAP-enhanced stacked ensemble learning: A case study of Jeddah. *Journal of Environmental Management*, *393*, 127128. https://doi.org/https://doi.org/10.1016/j.jenvman.2025.127128

Zheng, J., Zhang, J. Z., Kamal, M. M., Liang, X., & Alzeiby, E. A. (2025). Unpacking human-AI interaction: Exploring unintended consequences on employee Well-being in entrepreneurial firms through an in-depth analysis. *Journal of Business Research*, *196*, 115406. https://doi.org/https://doi.org/10.1016/j.jbusres.2025.115406