Journal of Law, Society and Living Norms
Publisher: CV. Norma Global
E-ISSN: 3123-2566
Volume 1, Number 2, 2025
E-ISSN: 3123-2566

Ethical Principles and Legal Responsibility in Artificial Intelligence: Ensuring Fairness and Accountability in AI Governance

Muhidin[1], Lu Sudirman[2]
[1] Universitas Borobudur
[2] Universits Internasional Batam
Correspondence: muhydyn@email.com[2], dirman_lu@yahoo.com[2]

Article Info
Article history:
Received Nov 19th, 2025
Revised Dec 7th, 2025
Accepted Dec 16th, 2025

Keyword:
Artificial intelligence; Algorithmic bias; AI ethics; Legal responsibility; Fairness by design; AI governance.

ABSTRACT
Artificial intelligence (AI) presents significant ethical and legal challenges, particularly regarding algorithmic bias, opaque decision-making, and the absence of clear liability mechanisms. This study examines how key ethical principles such as transparency, fairness, privacy protection, and accountability can be integrated into a legal responsibility framework to support fair and accountable AI governance. The research uses a normative juridical method with conceptual, statutory, and case approaches to assess the limitations of traditional liability models when applied to autonomous systems. The findings indicate that algorithmic bias can lead to both material and procedural injustice, while gaps in regulation create uncertainty about who should bear responsibility for AI-related harm. This study recommends the application of fairness by design, mandatory model documentation, algorithmic audits, and shared responsibility among developers, operators, and users. These findings highlight the need for adaptive regulation to ensure that the use of AI upholds justice, protects individual rights, and serves the public interest.

INTRODUCTION

The rapid development of artificial intelligence (AI) has reshaped decision-making in areas such as finance, healthcare, public administration, and employment. As AI systems become central to many high-stakes decisions, ethical and legal concerns arise regarding opaque model logic, discrimination embedded in training data, and the difficulty of determining responsibility when automated systems cause harm. Empirical studies show that algorithmic outputs often reproduce or amplify historical inequalities because the underlying data reflects social and institutional biases, which may result in unjust outcomes for individuals or groups (Lendvai & Gosztonyi, 2025). These risks are intensified when models operate as opaque tools that affected parties cannot meaningfully scrutinize or contest.

Scholars have emphasized the need for ethical safeguards that integrate transparency, fairness, privacy protection, and accountability into AI development and deployment. However, legal and technical literature often evolve separately. Xiang (2021) notes that regulatory approaches tend to focus on liability doctrines, while technical research concentrates on detection and mitigation of bias without addressing the question of who should bear responsibility for resulting harm. Policy analyses also highlight that existing frameworks have not fully resolved how responsibility should be allocated when autonomous or semi-autonomous systems malfunction or produce discriminatory decisions (Australian Human Rights Commission, 2020). As a result, governance debates remain fragmented, and there is limited integration between ethical principles and enforceable legal mechanisms.

This study addresses that gap by examining how ethical principles can guide the construction of legal responsibility in the context of AI systems. Three research questions frame the analysis: how ethical principles should shape accountability mechanisms for AI, how algorithmic bias affects material and procedural justice, and what form of liability and governance framework is most appropriate for

autonomous systems. The objective is to develop a coherent analytical foundation that links ethics, bias, and legal responsibility in a way that supports fair and accountable AI governance.

The contribution of this research is twofold. Theoretically, it connects normative ethical requirements with the legal doctrines that regulate responsibility for harm, thereby clarifying why fairness, transparency, and accountability must be embedded in AI regulation. Practically, it offers structured recommendations based on recent academic findings, including the need for model documentation duties, algorithmic audits, and shared responsibility among developers, operators, and users (Munifah et al., 2024). These contributions aim to support the development of adaptive governance that protects individual rights and ensures that AI technologies serve the public interest.

RESEARCH METHODS

This study uses a normative juridical research method because the issues examined relate to ethical principles, liability doctrines, and regulatory needs in governing artificial intelligence. The analysis focuses on how existing legal concepts interact with the technical realities of AI systems and how ethical principles should inform the development of accountability mechanisms.

A conceptual approach is used to examine key ethical foundations such as fairness, transparency, privacy protection, and accountability, along with their relevance to legal responsibility. A statutory approach is applied to review existing legal instruments, including national regulations on data protection and international policy frameworks that address algorithmic governance. A case approach is used to analyse documented instances in which AI systems have generated discriminatory or harmful outcomes. These cases help illustrate the limitations of traditional liability models when applied to autonomous or semi-autonomous systems. A comparative approach is also employed to assess regulatory developments in other jurisdictions, particularly frameworks that introduce mechanisms such as model documentation, algorithmic audits, and risk-based classification of AI systems.

The research relies on primary legal materials such as legislation and official policy documents, secondary sources including academic journal articles and scholarly books, and tertiary sources that provide technical background relevant to algorithmic systems. The analysis is conducted using qualitative, deductive reasoning to develop an integrated understanding of how ethical principles and legal doctrines can be aligned in regulating AI.

These approaches are used because the core issues examined in this study are algorithmic bias, ethical principles, and legal responsibility which requires analysis of both normative concepts and documented cases. The case approach supports the identification of how algorithmic harms manifest in practice, while the comparative approach allows the study to evaluate regulatory developments in other jurisdictions that may guide the refinement of domestic legal frameworks.

RESULTS AND DISCUSSION

Ethical Principles as the Foundation of AI Responsibility

The rapid adoption of artificial intelligence has created new forms of risk that cannot be fully addressed by traditional regulatory concepts. Ethical principles provide the initial foundation for determining how AI systems should operate and how responsibility should be assigned when harm occurs. Transparency is essential because opaque or “black-box” systems limit the ability of individuals to understand or challenge automated decisions that affect their rights. This problem is widely discussed in the literature on algorithmic accountability, which highlights that unclear model logic undermines both fairness and the legitimacy of decision-making processes (Xiang, 2021).

Fairness is another fundamental requirement in AI governance. Research shows that AI systems may reproduce or intensify discrimination if training data reflects historical inequities or institutional bias (Lendvai & Gosztonyi, 2025). This creates a situation in which individuals subjected to automated decisions may face unequal treatment without any meaningful procedural safeguards. Privacy protection also serves as a basic ethical condition because AI systems depend heavily on large datasets, often containing sensitive personal information. Without adequate safeguards, the use of such data risks violating privacy rights and expanding unequal surveillance practices. Taken together, these ethical principles function as the normative basis for determining how legal structures should address harms, risks, and responsibilities in AI systems.

Algorithmic Bias and Its Implications for Justice

Algorithmic bias is one of the most documented sources of injustice in AI systems. Bias can originate from imbalanced training data, flawed labeling processes, or design choices made by developers without awareness of underlying discriminatory patterns. Lendvai and Gosztonyi (2025) show that legal systems struggle to respond when automated systems produce discriminatory outcomes because the harm arises indirectly through algorithmic processes rather than direct human intention.

Real-world cases further illustrate how algorithmic systems can generate inequitable outcomes. One of the most widely cited examples is the COMPAS recidivism prediction tool, which an investigative analysis showed to produce disproportionately higher false-positive rates for Black defendants compared to white defendants, even when controlling for prior criminal history (Angwin et al., 2016). Similarly, Amazon discontinued its automated hiring model after internal evaluation revealed that the system consistently downgraded applications from women because it had been trained on historical hiring data dominated by male candidates (Dastin, 2018). These cases demonstrate that algorithmic systems can amplify historical biases and therefore require legal scrutiny that addresses not only technical flaws but also their social and legal consequences.

The implications for justice are significant. Material injustice occurs when individuals are denied opportunities, misclassified, or subjected to wrongful decisions. Procedural injustice arises when affected parties cannot access the reasoning behind AI-generated outcomes or challenge the decision through effective remedies. Public policy analysis also demonstrates that many existing regulatory frameworks still lack clear mechanisms to prevent or correct such harms (Australian Human Rights Commission, 2020). These findings confirm that algorithmic bias is not only a technical flaw but also a legal problem, and therefore requires a structured approach to accountability.

Limitations of Traditional Liability Models

Traditional liability doctrines were designed for human decision-makers, not autonomous systems. Strict liability, negligence, and product liability all face limitations when applied to AI. Negligence requires proof of a breach of duty, yet harms generated by opaque machine-learning models often lack identifiable human error. Product liability assumes stable and predictable product behavior, which does not align with models that continuously update or adapt.

Comparative legal analysis shows that uncertainty regarding responsibility becomes more pronounced when multiple actors contribute to the design, deployment, and operation of AI. Xiang (2021) notes that technical methods for addressing bias do not inherently resolve questions about who should be held legally accountable. Developers may design the model, but operators choose how it is implemented, and users may apply it in contexts not originally intended. This interdependence makes sole-responsibility frameworks increasingly inadequate.

As a result, current doctrines cannot fully address situations in which injuries arise from algorithmic decision processes, especially when the causal chain is diffuse or partially concealed within automated systems. In the Indonesian context, existing legal instruments also show limitations in addressing harms arising from autonomous decision-making. Liability standards in the Civil Code, which are based on fault and product responsibility, assume a stable relationship between human action and the resulting damage, an assumption that does not fully align with adaptive machine-learning systems. The Personal Data Protection Law provides an important foundation for regulating data processing and privacy, but it does not yet address accountability for discriminatory automated outputs or decisions generated without meaningful human oversight. Policy analysis in Indonesia similarly notes that regulatory gaps persist in determining responsibility when harm emerges indirectly through algorithmic processes rather than intentional human conduct (Hidayati, 2024).

Toward a Fair and Accountable AI Governance Framework

Recent scholarship emphasizes the need for governance mechanisms that integrate ethical principles with enforceable legal duties. Munifah et al. (2024) demonstrate that combining technical controls with legal requirements helps reduce the gap between ethical theory and practical accountability. Three governance strategies emerge from the literature. The first is fairness-by-design, which requires developers to identify and mitigate risks of discrimination during the design phase. This includes validating datasets, documenting assumptions, and implementing mechanisms to monitor disparate impacts. The second is mandatory documentation and algorithmic auditing, which forms the

evidentiary basis for determining responsibility when harm occurs. Audits provide transparency and allow regulators or courts to examine whether an AI system has been operated responsibly. The third is shared liability, a model that distributes responsibility across developers, operators, and users according to their role and level of control in the AI lifecycle.

These mechanisms address both ethical and legal concerns by strengthening transparency, reducing the likelihood of discriminatory outcomes, and providing structured routes for accountability. They also support the development of adaptive regulation, which is necessary for technologies that evolve more quickly than statutory frameworks. Comparative developments in other jurisdictions offer additional insight into how accountability for AI can be strengthened through binding regulatory structures. The European Union's Artificial Intelligence Act adopts a risk-based framework that imposes stringent obligations on high-risk AI systems, including requirements for technical documentation, human oversight, post-market monitoring, and transparency toward affected individuals. These obligations reflect a shift toward proactive governance that embeds ethical safeguards directly into enforceable legal duties. Although Indonesia operates within a different legal and institutional context, the EU model provides a relevant reference point for designing accountability mechanisms that address both technical risks and the legal implications of automated decision-making (European Parliament & Council, 2024).

CONCLUSION

Artificial intelligence introduces ethical and legal challenges that cannot be addressed through traditional responsibility frameworks alone. The findings of this study show that algorithmic bias has the potential to create both material and procedural injustice, particularly when opaque model logic and imbalanced training data affect decisions in high-stakes domains. Ethical principles such as transparency, fairness, privacy protection, and accountability therefore become essential foundations for assessing how AI should be governed.

The analysis also demonstrates that existing liability doctrines encounter significant limitations when applied to autonomous or semi-autonomous systems. Questions of responsibility become fragmented between developers, operators, and users, while the dynamic nature of machine-learning models complicates efforts to attribute fault or identify clear causal chains. These conditions highlight the need for regulatory mechanisms that incorporate ethical considerations into enforceable legal duties.

This study recommends three key governance approaches. The first is fairness by design, which requires early identification and mitigation of risks related to discrimination. The second is mandatory model documentation and algorithmic auditing to support transparency and provide an evidentiary basis for determining responsibility. The third is shared liability, which distributes responsibility according to each actor's role within the AI lifecycle. Collectively, these strategies support the development of adaptive AI regulation that protects individual rights, promotes fair outcomes, and ensures that the benefits of AI are realized without compromising public trust or justice. Embedding these mechanisms within binding legal standards is essential to prevent both material and procedural injustice as AI systems become increasingly integrated into public and private decision-making.

REFERENCES

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: Risk assessments in criminal sentencing. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Australian Human Rights Commission. (2020). Addressing the problem of algorithmic bias. Australian Human Rights Commission. https://humanrights.gov.au/sites/default/files/document/publication/final_version_technical_paper_addressing_the_problem_of_algorithmic_bias.pdf

Bharati, R. (2025). Bias and fairness in AI algorithms: Legal standards and ethical guidelines. SSRN. https://ssrn.com/abstract=5378211

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. Humanities and Social Sciences Communications, 10, 567. https://doi.org/10.1038/s41599-023-02079-x

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK0AH

European Parliament & Council. (2024). Regulation laying down harmonised rules on artificial

intelligence (Artificial Intelligence Act). EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

Hanna, M. G. (2025). Ethical and bias considerations in artificial intelligence and machine learning in the medical domain. Artificial Intelligence Review. https://www.sciencedirect.com/science/article/pii/S0893395224002667

Hidayati, S. (2024). Civil liability for clients who suffer losses in using artificial intelligence services. Jurnal Lex Publica, 7(1), 101–112. https://dinastires.org/JLPH/article/download/894/844

International Journal for Legal Studies. (2025). Legal accountability of algorithmic bias: Examining the role of enforcement and transparency. International Journal for Legal Studies. https://international.appihi.or.id/index.php/IJLS/article/view/521

Lendvai, K., & Gosztonyi, G. (2025). Algorithmic bias as a core legal dilemma in the age of artificial intelligence: Conceptual basis and the current state of regulation. Laws, 14(3), 41. https://doi.org/10.3390/laws14030041

Min, A., et al. (2024). Artificial intelligence and bias: Challenges, issues and regulatory perspectives. International Journal for Scientific Research. https://ijsr.internationaljournallabs.com/index.php/ijsr/article/download/1477/976

Munifah, D., Komariah, E., & Muchlis, A. (2024). Ethical challenges in AI-driven decision-making: Addressing bias and accountability in business applications. Jurnal Manajemen Informatika, 15(1), 45–53. https://jmi.stekom.ac.id/index.php/jmi/article/view/48

Pfeiffer, M. J. (2023). First, do no harm: Algorithms, AI, and digital product liability. arXiv. https://arxiv.org/abs/2311.10861

Wachter, S. (2022). The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law. arXiv. https://arxiv.org/abs/2205.01166

Xiang, A. (2021). Reconciling legal and technical approaches to algorithmic bias. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3650635