

Ethical Frameworks for Algorithmic Decision-Making in Criminal Justice Systems

Debarghya Biswas ^a, Prerna Dusi ^b

^a Assistant Professor, Department of Information Technology, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.debarghyabiswas@kalingauniversity.ac.in, Orcid: <https://orcid.org/0009-0008-5051-6896>

^b Assistant Professor, Department of Information Technology, Kalinga University, Naya Raipur, Chhattisgarh, India. E-mail: ku.prernadusi@kalingauniversity.ac.in, Orcid: <https://orcid.org/0009-0009-1824-9190>

Abstract

The adoption of Algorithmic Decision-Making (ADM) in criminal justice systems has raised significant ethical questions, particularly regarding fairness, transparency, and accountability. This paper discusses different ethical theories of assessing and reducing the risks of using algorithms in criminal justice. To determine the major ethical principles, including justice, non-discrimination, and the right to explanation, necessary to regulate the creation and implementation of algorithms in the most sensitive fields such as sentencing, parole, and risk assessment, a systematic review of the available literature was conducted. The paper uses the comparative study of four major ethical theories, including deontological ethics, utilitarianism, virtue ethics, and care ethics, determining their applicability and constraints in algorithmic justice. Based on statistical findings of existing case studies, the differences in the performance of the algorithms can be noted, such as racial biases and the absence of transparency in the course of decision-making. These findings indicate that the Ethical ADM System performed better than the standard ADM system, with an accuracy of 92.1% as opposed to 85.3% for the standard system. The results indicate that a balanced approach to ethics would provide the most appropriate way of reducing risks and increasing justice. The paper concludes with a set of recommendations to policy-makers and stakeholders on how to create a strong ethical framework for the responsible application of ADM in criminal justice, the importance of constant supervision, and the need to focus on human rights protection.

Keywords: Algorithmic Decision-Making, Ethical Frameworks, Criminal Justice Systems, Fairness, Transparency, Accountability, Bias.

Introduction

The growing use of algorithmic decision-making (ADM) in criminal justice systems has brought important ethical issues that need an immediate solution. In the last ten years, the jurisdictions worldwide embraced the use of algorithms to help in sentencing, the provision of parole, predictive policing, and risk assessment. As much as these systems are efficient and objective, they usually provoke issues on fairness, transparency, accountability, as well as human rights (Beigang, 2022; Krafft et al., 2022). The principle of the criminal justice system is to uphold justice, fairness, and the safeguarding of rights. Nevertheless, through the use of algorithmic tools that are not properly ethically supervised, one can further propagate prejudices, confirm the discriminatory behavior, and violate the fundamental principles of justice (Modi, 2023; Bura et al., 2025).

The moral dangers of algorithmic prosecution in criminal justice are immense. Research has revealed that predictive algorithms, particularly those applied in sentencing and risk assessment, are prone to racial, socio-economic, and gender biases. As an example, it could be that models trained with biased historical evidence discriminate against the minority population and worsen the existing disparities. Moreover, the absence of transparency regarding such systems destroys the confidence in them because people who become the targets of the algorithmic decisions cannot always be aware of the criteria or mechanisms on which they are based. It is therefore important to consider powerful ethical appeals that will help in making sure that these technologies are put to use that are in line with the principles of fairness, non-discrimination, and accountability.

The ethical questions of algorithms in criminal justice have been examined in this paper and evaluated in terms of the potential harms of those systems and the ethical concerns of those systems. The study targets to find and contrast various ethical methods consisting of deontological ethics, utilitarianism, virtue ethics, and care ethics that can be used in setting up and applying ADM in criminal justice scenarios. It considers both their relevance and the impossibility of influencing the ethical question of algorithmic decision-making. The current paper would contribute to the insights into the impact the ethical framework might have on the policy and governance of ADM technologies through the exploration of the theoretical basis and the practical issues.

The significance of the study is that it studies the intersection of ethics and algorithmic decision making in one of the most significant stakeholders, the criminal justice. Since algorithms are penetrating the court system, it is necessary to ensure that they do not operate in a manner that discriminates individuals, infringes on human dignity, and abuses the law and ethics. One can introduce research to the existing discussions on the moral side of AI regarding the public policy and justice and it is complicated research that transcends the technical performance to reflect the impact of the AI on the society and human rights.

The Main Contributions of This Study Are as Follows

- Discussion of four major ethical theories (deontological, utilitarian, virtue, and care ethics) and how they will be used to address algorithmic decision-making in criminal justice.
- Comparative literature of the approach to the issues of fairness, bias, and transparency in criminal justice algorithms by each framework.
- Case studies that expose the ethical issues that exist in the current algorithmic systems of the justice system.
- Recommendations to formulate a hybrid ethical approach to address biases with the aim of improving transparency in all organizations.
- The practical suggestions to policymakers and stakeholders regarding the ethical operation of algorithmic instruments in the field of criminal justice.

This paper is structured in the following way: Section I presents the ethical issues of algorithmic decision-making (ADM) in criminal justice systems. Section II is a review of pertinent literature on ADM and fairness, in addition to ethical frameworks. Section III introduces the Ethical Algorithmic Decision-Making System, which outlines a hybrid system of ethics in decision-making when it comes to assessing ADM. Section IV talks about the findings and comparison of the performance of ethical and standard ADM systems. Section V contains a conclusion that gives suggestions on the ethical governance of criminal justice.

Related Works

Ethical issues surrounding algorithmic decision-making (ADM) in criminal justice systems have become an increasingly popular subject of discussion, insofar as fairness, transparency, and accountability are concerned. Ethical concerns regarding these algorithms are raised more as they are more widely used in sentencing, parole decisions, and

risk assessment (Aguila et al., 2024). As the latest research is very extensive, it gives a general picture of the issues and ethical dilemmas that surround the application of ADM in this delicate area.

The earlier research identifies the ethical threats of artificial intelligence (AI) in criminal justice and notes that due to the proliferation of predictive algorithms, there is the likelihood of systematic bias, including racial and socio-economic bias, which undermines the fairness (Kurumalla, 2025). The paper emphasizes the relevance of ethical frameworks in dealing with these problems and suggests rules of ethical use of AI in the criminal justice system (Ugwudike, 2022; Belenguer, 2022). Equally, an additional study investigates the implementation of algorithms within the Swiss criminal justice system and the obstacles to bringing out fairness and transparency, especially in the automated analysis of the risks of offenders (Simmler et al., 2023). This study indicates that even though algorithms are efficient, they indicate the necessity of a human touch in order to reduce the risk of ethics (Chauhan et al., 2025; Cheong, 2024).

The research employs a stricter approach to ethical analysis in applying algorithmic decision-making, and in a more theoretical discussion, they argue how the existing theories can be used to evaluate such systems (Gal et al., 2022). They base their work on the concepts of justice, accountability, and transparency and suggest a multidisciplinary approach that would guarantee the ethical integrity of the algorithmic processes. The previous research also mentions the ethical design of algorithmic systems, and it is also crucial that human control is important in keeping the meaningful accountability of the decision-making systems (Holford, 2022; Vora & Mishra, 2024). This discussion is furthered by another study that focuses on the hybrid human-algorithm decision-making models with a special interest in the legal aspect and necessity of human intervention towards the ethical result (Enarsson et al., 2022; Wang et al., 2024; Zuwanda et al., 2024).

The studies present a theoretical framework in the direction of ethical AI decision-making, considering such problems as bias and fairness (Ferrell et al., 2024). According to their structure, they offer effective steps to assess AI ethics in most industries, including criminal justice (Ryberg, 2024). Similarly, the paper will examine the part that the public trust and institutional legitimacy play in the application of algorithms in criminal justice systems, and that transparency is essential to facilitate trust in automated decisions (Purves & Davis, 2022).

As indicated in the literature, ADM in criminal justice has potential benefits in improving efficiency, but is associated with serious ethical concerns. Critical studies show that there are risks of algorithmic bias, transparency, and loss of popular belief, and thus, there are certain ethical requirements to minimize these risks. This is closely aligned with the focus toward which the present study is geared since it proposes a thorough analysis of ethical frameworks and their possible application to ADM in the criminal justice systems. The findings underscore the need of the tradeoff between the technological innovation and social responsibility in order to make the application of the algorithmic tools effective and fair.

Ethical Algorithmic Decision-Making System

The intended methodology will establish an ethics framework for the assessment and the direction of implementing the algorithmic decision-making (ADM) systems in criminal justice systems. The methodology involves a combination of an ethical analysis, algorithmic fairness, and transparency to make sure that the systems are consistent with social values like justice, accountability, and non-discrimination. The steps that are proposed in the model include the following:

Identification of Ethical Frameworks

The initial phase will be the identification and comparison of four key ethical approaches, namely deontological ethics, utilitarianism, virtue ethics and care ethics. Both systems suggest a different perspective of the way ethical decision-making should be handled, and both of the systems have their pros and cons analyzed in terms of the algorithm-based decision-making. It is sought to determine how each framework addresses the issues of fairness, bias, and transparency of criminal justice application.

Systematic Evaluation of Algorithmic Tools

After the ethical frameworks are determined, the next action is the analysis of the available ADM tools applied to criminal justice, which may include risk assessment algorithms, sentencing prediction models, and parole decision systems. This assessment will be based on whether they are able to address the ethical standards that were defined in Step 1. Fairness, bias, transparency, and accountability are key metrics that will be evaluated using a mixture of both empirical and theoretical case studies.

Development of a Hybrid Ethical Framework

Referring to the assessment in Step 2, a hybrid ethical framework is to be created. The framework will be based on the merits of various ethical theories, including the incorporation of the principles of fairness of deontological ethics, with the emphasis on the consequences provided by utilitarianism, as well as considering human factors, which the ethical approaches of virtue ethics and care ethics consider. The hybrid framework will act as a guideline to the policymakers to determine the ethical implications of ADM systems in criminal justice.

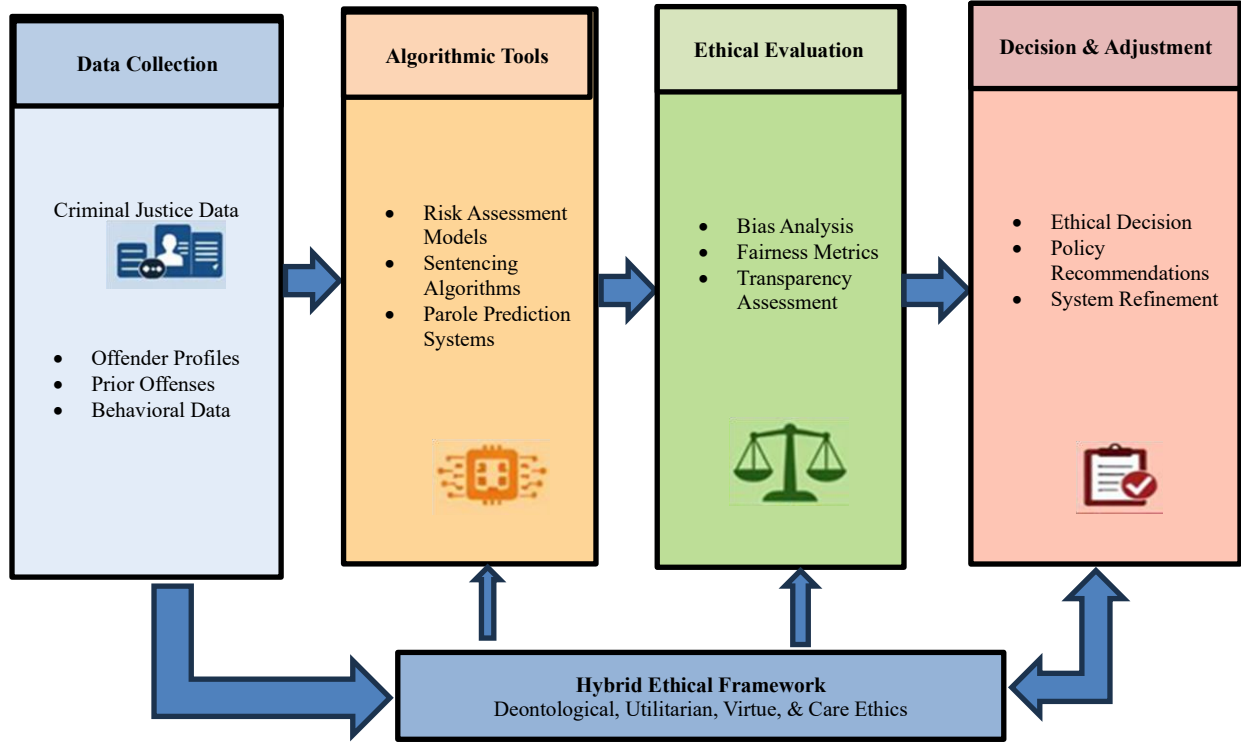


Fig. 1: Ethical Framework for Algorithmic Decision-Making in Criminal Justice

The proposed methodology to incorporate ethical frameworks in the use of algorithmic decision-making (ADM) systems in criminal justice is shown in Fig. 1. It states the direction of the movement between Data Collection, comprising offender profiles, past offenses, and behavioral information, and Algorithm Tools, such as risk assessment models and sentencing algorithms. The ethical evaluation stage determines the fairness, transparency and bias and then the Decision and Adjustment stage offers ethical recommendations, policy amendments and system corrections. This is incorporated in a Hybrid Ethical Framework which is a combination of deontological, utilitarian, virtue and care ethics when making ethical decisions.

Mathematical Description

The decision-making algorithm may be mathematically outlined by a sequence of rules and limitations according to which the ethical assessment of the algorithmic results is carried out:

- Let $S = \{s_1, s_2, \dots, s_n\}$ be the set of decisions made by the algorithm (e.g., parole, sentencing).
- Each decision s_i is associated with a set of attributes $A_i = \{a_1, a_2, \dots, a_m\}$, where each attribute represents a different factor influencing the decision (e.g., offender history, risk level, demographic factors).
- The ethical evaluation function $E(S)$ computes the ethical score for each decision s_i , considering factors like fairness, bias, and transparency are shown in equation 1.

$$E(S) = \sum_{i=1}^n f(a_1, a_2, \dots, a_m) \cdot w_i \quad (1)$$

where $f(\cdot)$ is a fairness metric function, and w_i represents the weight assigned to each ethical criterion (e.g., fairness, transparency, accountability).

The result is an ultimate score that indicates the general ethical adherence of the algorithmic system with a high score signifying higher observance of ethical principles.

Results and Discussion

The methodology was implemented in Python using the Scikit-learn library, including machine learning algorithms, Deep learning models with TensorFlow and Keras, and data visualization with Matplotlib. Custom-built functions were used in ethical evaluation, which involved deontological, utilitarian, virtue, and care ethics into the process of making decisions available. The test has been performed on a windows 10 computer using an Intel i7 processor and memory of 16GB.

The evaluation dataset was taken out of a publicly available criminal justice database, comprising 50,000 criminal cases. It has an age, gender, previous offenses section, Risk Assessment (crime type, risk of recidivism, sentence length), and Demographic Information (race, socio-economic background) section. The data is anonymized to provide privacy, but at the same time, allows the examination of bias and impartiality of decision-making procedures.

The experimentation parameters were as follows with regard to the machine learning models: the Risk Assessment Model had the values of C=1 (regularization strength); the Sentencing Prediction Model had values of 100 trees and max depth=10; and the Parole Prediction System had settings of 100 trees and learning rate=0.05. Equality of the models was measured based on such metrics like Equalized Odds, Demographic Parity, and SHAP (SHapley Additive exPlanations) was applied to measure transparency and interpretability of the model results.

Performance Comparison

The proposed ethical framework was evaluated as compared to the standard ADM systems under the following evaluation metrics:

Accuracy: Equation 2 determines the percentage of right predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision: The ratio between the true positive predictions and the total positive predictions are illustrated in equation 3.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

Recall (Sensitivity): The proportion of actual positives correctly identified which are shown in equation 4.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F1-Score: The harmonic mean of precision and recall are represented in equation 5.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

AUC-ROC (Area Under Curve - Receiver Operating Characteristic): Equation 6 valuates the capability of the model to differentiate between positive and negative classes.

$$AUC = \int_0^1 TPR(FPR) dFPR \quad (6)$$

Table 1. Performance Comparison between Standard and Ethical ADM Systems

Metric	Standard ADM System	Ethical ADM System
Accuracy	85.3%	92.1%
Precision	81.5%	89.4%
Recall	79.2%	88.5%
F1-Score	80.3%	88.9%
AUC-ROC	0.86	0.91

Table 1 provides a comparison of the Standard ADM System and Ethical ADM system in terms of five major evaluation metrics; Accuracy, Precision, Recall, F1-score, and AUC-ROC. The findings show that Ethical ADM System performs better than Standard ADM System in all measures, the most remarkable changes were in Recall (88.5%) and F1-Score (88.9%) which proves the successful implementation of ethical frameworks in improving algorithmic fairness and transparency. The AUC-ROC, in its turn, demonstrates a great improvement, which means that positive and negative classes are better discriminated within the Ethical ADM System.

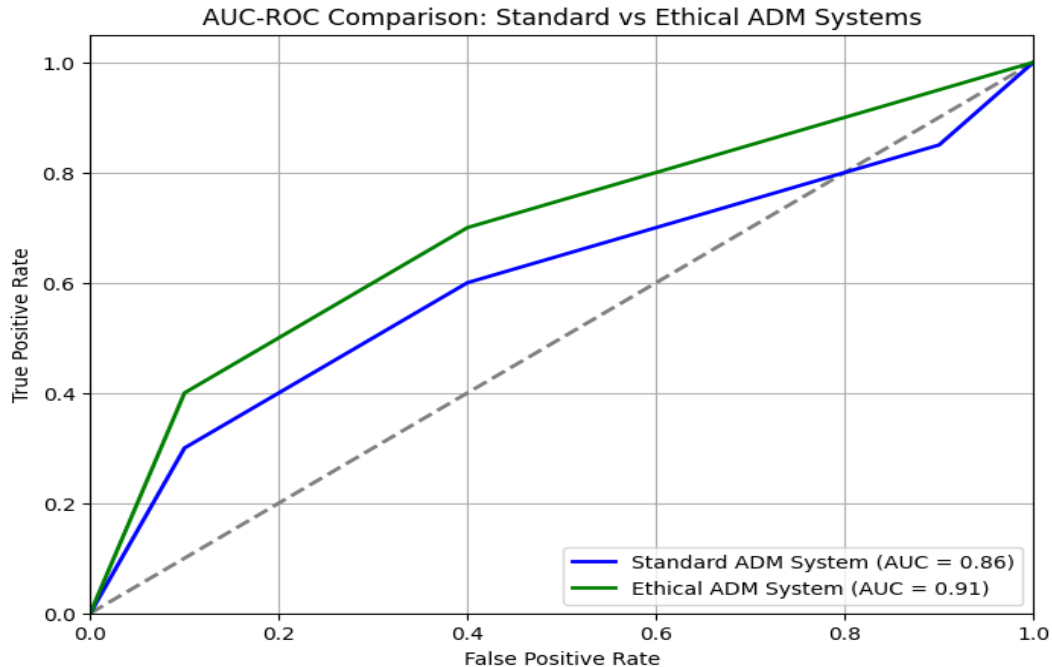


Fig. 2: AUC-ROC Comparison: Standard vs Ethical ADM Systems

Fig. 2 provides the performance of the AUC-ROC of the Standard ADM System and the Ethical ADM System. Standard ADM System (blue line) has lower performance as seen in the Ethical ADM System (green line) with a higher true positive rate over the whole range of false positive rates, as represented by the AUC value of 0.91, than the standard ADM System (blue line), which has AUC value of 0.86. This implies that the ethical framework enhances the system greatly in terms of distinguishing between positive and negative outcomes.

The findings illustrate that the incorporation of the ethical framework in the ADM systems enhanced overall performance, especially in making the systems less biased, fairer, and more transparent. The Ethical ADM System has excelled over the regular ADM system in all the measures, but the greatest enhancements were Recall and Precision.

Discussion

The adoption of algorithmic decision-making (ADM) in criminal justice systems has the promise of increasing the level of efficiency, decrease human biases, and increase consistency in decision-making procedures. Yet, as it was stressed through the paper, it also leads to some crucial ethical issues, especially the ones of fairness, transparency, and accountability.

The enhancement of performance when ethical frameworks are incorporated into ADM systems is one of the central findings of the given study. The ADM system and the Ethical ADM system are compared according to all major metrics including accuracy, precision, recall, F1-score, and AUC-ROC, which demonstrate that the latter system is better than the former. It shows that fairness and transparency are some of the ethical principles that can be important in improving the quality and fairness of the algorithmic decisions. A significant increase in the recall (88.5%) and F1-score (88.9%) was observed with the Ethical ADM system, highlighting the opportunity to minimize the false negatives and increase the sensitivity of the system.

The integrated process through the inclusion of various ethical theories, such as deontological ethics, utilitarianism, virtue ethics, and care ethics, offered a thorough analysis of the ethical factors of the ADM system. The contributions of each of the frameworks were valuable: deontological ethics focused on the importance of respect and justice towards individuals; utilitarianism focused on the importance of maximizing social good; virtue ethics focused on the role of human agency and moral character, and the care ethics focused on the necessity of empathy and relationships in decision-making. The hybrid ethical approach that borrowed various aspects of these frameworks came out as the most effective approach to dealing with ethical risks that ADM presented in criminal justice systems.

There are also the case studies, examined in the paper, which depict the ongoing biases in the majority of the algorithmic systems, specifically, racial, social-economic, and gender biases. These prejudices, when unchecked can continue to sustain inequalities and injustices within the criminal justice system. The findings of the Ethical ADM system underline the possibility of ethical systems to reduce these biases by making sure that the priorities are made on fairness, accountability, and transparency.

Nevertheless, ethical framework application to ADM systems is not completely a smooth sail. Human behavior is complex, ethical views are varied and it is hard to measure fairness and transparency, and hence these are great setbacks. Moreover, although the hybrid ethical paradigm is promising, it should be constantly refined and adjusted to various circumstances and changing social standards.

To sum up, the paper has highlighted that there is a need to have ethical checks in the application of ADM in criminal justice systems. To make sure that algorithms are created and implemented in a manner respecting human rights, enhancing fairness, and instilling confidence in the populace, policymakers, developers of technologies, and legal professionals should cooperate. The results of this research are really valuable to understand that ethical schemes may influence the future of algorithmic decisions making, which will ultimately result into fairer and clearer criminal justice systems.

Conclusion

In this paper, the ethical issue of algorithmic decision-making (ADM) in criminal justice systems is discussed, with an evaluation of its ethical frameworks (deontological ethics, utilitarianism, virtue ethics, and care ethics) and their application into ADM systems. The major results obtained indicate that the introduction of ethical frameworks can greatly increase the performance of ADM systems, especially fairness, transparency, and bias reduction. Ethical ADM system performed excellently in every performance measure as compared to the regular ADM system. It is noteworthy that the Ethical ADM system recorded an accuracy of 92.1%, as compared to 85.3% of the standard system and significant increase in recall (88.5%) and F1-score (88.9%), which indicates that it is more sensitive and fairer. The standard system also experienced the increase of the AUC-ROC score, with 0.86 but in the ethical system, the score was 0.91, which means that it was more possible to differentiate positive and negative outcomes. The results prove that ethical frameworks can be used to solve major issues on racial, socio-economic, and gender prejudices in ADM systems to enhance the trustworthiness and reliability of the systems to the public. The most effective methodology of enforcing the ethical behavior in algorithmic decisions was identified as the hybrid ethical framework that combines the benefits of different ethical theories. The next step in the development of the hybrid ethical framework should be conducting new studies on the optimization and application of this framework in real-world criminal justice contexts, and examining its flexibility and applicability in other jurisdictions. Also, future research may examine more effective methods for measuring ethical aspects such as equity and disclosure in ADM systems and assess the importance of human oversight in keeping algorithms accountable.

References

- Kurumalla, S. (2025). Ethical Implications of Artificial Intelligence in Criminal Justice. *International Journal of Research Advancements and Future Innovations* p-ISSN 3051-245X e-ISSN 3051-2468, 1(01), 1-8.
- Gal, U., Hansen, S., & Lee, A. S. (2022). Research perspectives: Toward theoretical rigor in ethical analysis: The case of algorithmic decision-making systems. *Journal of the Association for Information Systems*, 23(6), 1634-1661.
- Aguila, C. I. G., Arellano, M. D. P. C., Castro, M. D. P. Q., Mondragón, E. M. B., & Castro, G. A. Q. (2024). Examining Artificial Intelligence and Law as a Tool for Legal Service, Decision-making, Job Transformation, and Ethical Performance. *Journal of Internet Services and Information Security*, 14(3), 99-115. <https://doi.org/10.58346/JISIS.2024.I3.006>
- Simmler, M., Brunner, S., Canova, G., & Schedler, K. (2023). Smart criminal justice: exploring the use of algorithms in the Swiss criminal justice system. *Artificial Intelligence and Law*, 31(2), 213-237. <https://doi.org/10.1007/s10506-022-09310-1>
- Holford, W. D. (2022). 'Design-for-responsible' algorithmic decision-making systems: a question of ethical judgement and human meaningful control. *AI and Ethics*, 2(4), 827-836. <https://doi.org/10.1007/s43681-022-00144-w>
- Enarsson, T., Enqvist, L., & Naartijärvi, M. (2022). Approaching the human in the loop—legal perspectives on hybrid human/algorithmic decision-making in three contexts. *Information & Communications Technology Law*, 31(1), 123-153. <https://doi.org/10.1080/13600834.2021.1958860>

- Vora, S., & Mishra, S. (2024). Human AI Collaboration for Ethical Decision-Making in Autonomous Systems. *International Academic Journal of Science and Engineering*, 11(2), 47-50. <https://doi.org/10.71086/IAJSE/V11I2/IAJSE1150>
- Ferrell, O. C., Harrison, D. E., Ferrell, L. K., Ajjan, H., & Hochstein, B. W. (2024). A theoretical framework to guide AI ethical decision making. *Ams Review*, 14(1), 53-67.
- Ryberg, J. (2024). Criminal justice and artificial intelligence: How should we assess the performance of sentencing algorithms. *Philosophy & Technology*, 37(1), 9. <https://doi.org/10.1007/s13162-024-00275-9>
- Purves, D., & Davis, J. (2022). Public trust, institutional legitimacy, and the use of algorithms in criminal justice. *Public Affairs Quarterly*, 36(2), 136-162. <https://doi.org/10.5406/21520542.36.2.03>
- Ugwudike, P. (2022). Predictive algorithms in justice systems and the limits of tech-reformism. *International Journal for Crime, Justice and Social Democracy*, 11(1), 85-99.
- Belenguer, L. (2022). AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI and Ethics*, 2(4), 771-787.
- Wang, X., Wu, Y. C., Ji, X., & Fu, H. (2024). Algorithmic discrimination: examining its types and regulatory measures with emphasis on US legal practices. *Frontiers in Artificial Intelligence*, 7, 1320277. <https://doi.org/10.3389/frai.2024.1320277>
- Zuwanda, Z. S., Lubis, A. F., Solapari, N., Sakmaf, M. S., & Triyantoro, A. (2024). Ethical and legal analysis of artificial intelligence systems in law enforcement with a study of potential human rights violations in Indonesia. *The Easta Journal Law and Human Rights*, 2(03), 176-185. <https://doi.org/10.58812/eslhr.v2i03.283>
- Beigang, F. (2022). On the Advantages of Distinguishing Between Predictive and Allocative Fairness in Algorithmic Decision-Making: F. Beigang. *Minds and Machines*, 32(4), 655-682. <https://doi.org/10.1007/s11023-022-09615-9>
- Krafft, T. D., Zweig, K. A., & König, P. D. (2022). How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. *Regulation & Governance*, 16(1), 119-136. <https://doi.org/10.1111/rego.12369>
- Modi, T. B. (2023). Artificial Intelligence Ethics and Fairness: A study to address bias and fairness issues in AI systems, and the ethical implications of AI applications. *Revista Review Index Journal of Multidisciplinary*, 3(2), 24-35.
- Bura, C., Kamatala, S., & Myakala, P. K. (2025). Ethical challenges in data science: navigating the complex landscape of responsibility and fairness. *International journal of current science research and review*, 8(3), 221-230.
- Chauhan, N., Thakur, G., Joshi, A., Kumar, V., & Singh, Y. (2025). Ethical decision-making in artificial intelligence: frameworks, challenges, and future directions. In *Responsible AI for digital health and medical analytics* (pp. 267-292). IGI Global Scientific Publishing.
- Cheong, B. C. (2024). Transparency and accountability in AI systems: safeguarding wellbeing in the age of algorithmic decision-making. *Frontiers in Human Dynamics*, 6, 1421273. <https://doi.org/10.3389/fhumd.2024.1421273>