Impact Factor (QJIF): 8.47 **E-ISSN: 2583-1615, P-ISSN: 3049-3498**

International Journal of Research in Academic World

Received: 29/November/2025 **IJRAW:** 2026; 5(1):57-60 **Accepted:** 07/January/2026

A Proposed Ethical Framework on AI Mitigation Strategies in Decision Support Systems: Case Study

[*1]Dr. Aparna Vaidyanathan, [2]Dr. Kavita Khobragade and [3]Dr. Deepali Dhainje

[*1, 2, 3]Department of Computer Science, Fergusson College (Autonomous), Pune, Maharashtra, India.

Abstract

Artificial Intelligence (AI) systems have rapidly integrated into critical areas such as healthcare, finance, law enforcement, and education, offering powerful decision making capabilities. As decision making involves factors and emotions, the association rules, references that are applied play an essential role in decision making. Today artificial intelligence raises expectations for faster, more accurate, more rational and fairer decisions with technological advancements. But if these systems behave with their predictions differently than with their parameters. A framework can optimize and enhance the outcome efficiently.

As the primary purpose of this research paper is to examine the bias in the decision-making process of AI systems, the paper focuses on proposing a framework that can optimize the ambiguity that is defined as a systematic error in decision making processes that results in unfair outcomes. In the context of AI, ambiguity can arise from various sources, including data collection, algorithm design, machine learning models. The system can learn and replicate patterns of ambiguity that is present in the data used to train them, resulting in unfair or discriminatory outcomes. It is important to identify and address ambiguity in AI to ensure that these systems are fair and equitable for all users.

This paper proposes a methodology framework, a recommendation for AI systems in medical diagnosis that can assist and behave to the nearest accuracy.

Keywords: Artificial intelligence, ethical decision making, mitigation, framework, medical diagnosis, algorithm, optimization.

Introduction

Artificial Intelligence is increasingly being used to make decisions and predictions affecting most aspects of our lives. Decision support systems (DSS) refer to a broad category of computer systems that assist decision makers to utilize data, models and knowledge to solve semi-structured, ill-structured, or unstructured problems" (Phillips-Wren, 2013; p.5). Traditional approaches to these systems include, for example, rule-based expert systems that simply reflect and communicate the knowledge of experts in a specific subject matter to its users. But the intelligent decision-support systems (IDSS) that utilize AI techniques emerged (Stefan and Carutasu, 2020).

These advances in AI have sparked the power, sophistication and autonomy of these decision-support systems so that they can assist humans in many more areas and (ethical) decision scenarios (Phillips-Wren, 2012).

Pertinent applications span from automated weapons that help soldiers determine whether a certain target shall be hit (Vallor, 2015). Clinical decision-support systems that help healthcare professionals distribute scarce medical resources (Erler and Müller, 2021) to artificial moral advisors that provide concrete moral advice to help users with their personal matters (Savulescu and Maslen, 2015). These developments give rise to the phenomenon of moral enhancement through AI (Lara, 2021), which in the past was mainly addressed in association with biomedical interventions (Savulescu and Maslen, 2015).

Moral enhancement entails interventions that aim to improve an individual's moral capacities, ultimately leading to moral improvement (e.g., better motives, increased understanding of what is right and higher frequency of right actions) (DeGrazia, 2014) and thus, is closely linked to the process and outcome of ethical decision-making. However, that IDSSs purely hold positive implications for individuals' ethical decision-making is viewed critically. Here the proposal of a framework can assist and support a variety of association rule.

Literature Review

Validity refers to the "'appropriateness' of the tools, processes, and data" utilized during research, while reliability involves the replicability of the research process and corresponding results (Leung, 2015; p.325). *Validity measures* consulted in this article include, for example, the reliance on and adoption of established methods for conducting systematic literature reviews such as the ones put forward by Gioia *et al*. (2013) or Webster and Watson (2002). In addition, existing literature reviews such as the ones published by Theurer *et al*. (2018) or Corley and Gioia (2004)

***Corresponding Author:** Dr. Aparna Vaidyanathan

were drawn on as orientation to cross check the legitimacy of utilized tools, data analysis and document writing. Furthermore, the list of keywords that underlie the database search was agreed upon among the co-authors to avoid selection bias, i.e., overlooking terms that are relevant to the topic at hand. Similarly, in line with Leung (2015), triangulation among researchers was conducted by repeatedly consulting generated codes and the derived model among the co-authors as well as with fellow researchers during research colloquia. Furthermore, a preliminary version of this article was presented and discussed with the science community at the international conference "2023 Forum on Philosophy, Engineering & Technology". *Reliability* was warranted by comprehensively documenting the literature search process, which included the disclosure of utilized keywords and databases (see Appendix A – Overview of the literature search process), exclusion criteria as well as the referencing of consulted methodologies. This allows other researchers to replicate or update this study in the future (Brocke *et al*., 2009) so that similar results to the ones sketched in the following sections can be achieved [1].

The related literature and industry press suggest that artificial intelligence (AI)-based decision-making systems may be biased towards gender, which in turn impacts individuals and societies. The information system (IS) field has recognised the rich contribution of AI-based outcomes and their effects; however, there is a lack of IS research on the management of gender bias in AI-based decision-making systems and its adverse effects. Hence, the rising concern about gender bias in AI-based decision-making systems is gaining attention. In particular, there is a need for a better understanding of contributing factors and effective approaches to mitigating gender bias in AI-based decision-making systems. Therefore, this study contributes to the existing literature by conducting a Systematic Literature Review (SLR) of the extant literature and presenting a theoretical framework for the management of gender bias in AI-based decision-making systems [2].

This research paper explores the optimization of decision-making processes through AI-enhanced Bayesian networks. In the context of complex and dynamic environments, traditional decision-making models often fall short due to their inability to adapt and learn from new data. This study proposes a novel framework that combines the adaptive capabilities of reinforcement learning with the probabilistic reasoning and uncertainty management offered by Bayesian networks. By doing so, it aims to create a robust AI support system that can continuously improve decision-making through interaction with its environment. The research methodology involves the development of a hybrid model that utilizes RL algorithms to optimize decision policies and Bayesian networks to update beliefs and handle uncertainty. Experiments conducted in simulated environments demonstrate the system's ability to achieve superior decision quality compared to conventional methods. The proposed system not only adapts to changing conditions but also provides interpretable insights into the decision-making process, enhancing transparency and trustworthiness. This paper contributes to the field by presenting a scalable solution that can be applied across various domains, including healthcare, finance, and autonomous systems, to support human decision-makers in making informed and optimal choices. The findings suggest significant potential for AI-enhanced systems to transform decision-making, enabling more effective and efficient outcomes in real-world applications [3].

This research seeks to achieve the above challenges through a holistic view that will embrace the following three major goals. First, we aim to derive a new generative expert AI system to generate reasonable B2B transaction data. This generative model will generate artificial data so that statistical similarity between the generated data and actual transactions is preserved, and fraud patterns are included deliberately. The process of generating synthetic data will solve the problem of data availability and data privacy and, at the same time, provide the required data diversification for obtaining good-quality models. In this generative AI model, several techniques are employed to optimize both the quality of synthesized data and its applicability. The model must capture the broker structure inherent to the B2B environment and the dynamic pattern of transactions while allowing multiple fraud scenarios to be created, given the emerging threat patterns. This objective contains oversight mechanisms for validating the synthetic data to guarantee its applicability in training fraud detection models. The second business objective is centered on increasing the efficacy of fraud detection models by increasing the efficiency of model building, development, and deployment. We aim to counteract the class imbalance issue while keeping the detector's sensitivity high enough by using the synthetic datasets created by our AI model. The study will analyze diverse model architectures and training methodologies to enhance performance for various businesses. Improving the accuracy of fraud detection also implies the creation of means for minimizing the identified false positives while maintaining the system's detection of fraudulent actions. This balance is important to sustain optimal operations with equally good or better fraud mitigation. The study will analyze the techniques of how the model can be updated and improved regularly with new data and novel fraud scenarios. The third aim focuses on applying and empirically confirming the models derived in practice-oriented business settings [4].

Methodology

Various methods and recommendations for bias mitigate.

The stages where mitigation techniques can be applied include pre-training, training, and post-training.

- Mitigating bias in the pre-training phase is the most effective manner of correcting bias since it transforms the dataset. However, bias may appear after training, hindering developers from dealing with it in the first iteration of the process.
- Training is the most efficient stage for handling bias. These methods are often unsupervised and do not involve adulterating the underlying data set. Not including sensitive features such as gender or race is not enough to mitigate discrimination, considering that other derivative features are introduced. Instead, adding fairness to the objective function is more efficient.
- Post-training is an ideal phase to calculate most of the previously revised metrics. However, mitigating biases in this phase should be the last option.
- Mitigating bias in AI is a complex and multifaceted challenge. However, several approaches have been proposed to address this issue. One common approach is to pre-process the data used to train AI models to ensure that they are representative of the entire population, including historically marginalized groups. This can involve techniques such as oversampling, under-sampling, or synthetic data generation.

- Another approach to mitigate bias in AI is to carefully select the models used to analyze the data. Researchers have proposed using model selection methods that prioritize fairness, such as those based on group fairness or individual fairness. For example, a study by Kamiran and Calders proposed a method for selecting classifiers that achieve demographic parity, ensuring that the positive and negative outcomes are distributed equally across different demographic groups. Another approach is to use model selection techniques that prioritize fairness and mitigate bias. This can be performed through techniques such as regularization, which penalizes models for making discriminatory predictions, or through ensemble methods, which combine multiple models to reduce bias.
- Post-processing decisions are another approach to mitigate bias in AI. This involves adjusting the output of AI models to remove bias and ensure fairness. For example, researchers have proposed post-processing methods that adjust the decisions made by a model to achieve equalized odds, which ensures that false positives and false negatives are equally distributed across different demographic group.

4. Proposed Frame Work

With the increasing use of machine learning models in different areas, it has become important to address the bias problem in these models. This issue can appear in different aspects such as racial, gender or socioeconomic biases leading to unfair outcomes in decision-making processes, for instance, in classification tasks, where models are trained to classify data into different categories. To address this issue, researchers have developed different strategies and techniques to mitigate the bias present in machine learning models. In this article, we explore some of the methods developed to overcome this challenge. The bias problem in classification tasks and the different strategies used for bias mitigation. How these strategies are grouped into categories and a brief introduction of the most representative methods for each one of these categories. How these methods are used to mitigate bias in machine learning models.

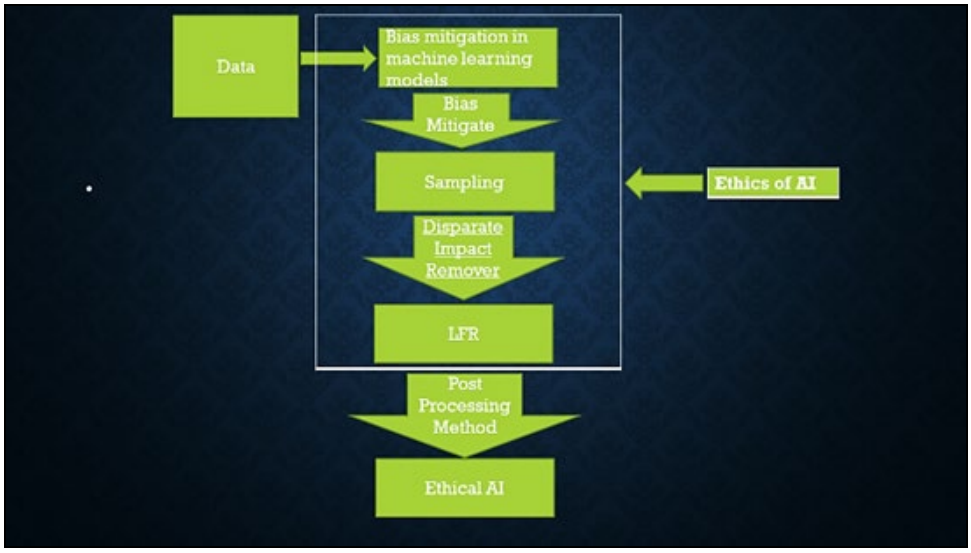

Fig 1:

5. Case Study

Mitigation techniques have a primary goal is to provide a high level of secure, reliable and consistent method for large volumes of data sets with efficiency and speed. Hence there is a need to provide a high level of security and efficiency to this model generated data. The framework thus provides a proposed solution with the use of Machine learning model Bias Mitigation techniques over systematic approach to data to secure the big data generated by predictive models for Medical diagnostic systems. The model develops an overlook of 3 strategies that include pre-processing, in-processing, and post-processing algorithms that can be used in healthcare systems to secure health data and to provide a bias free system. The role of handling mitigation bias in health disease diagnosis will provide free ambiguity decisions using predictive models. Furthermore, the study analyzes the future prospects of applying Bias mitigation strategies using the ML model in healthcare systems with regard to juvenile diabetes, emphasizing its potential to change data security in the healthcare domain.

Securing healthcare data with these algorithms involves using mitigation strategies to enhance clarity, consistency, data integrity, and access control.

6. Conclusion

So the paper suggests that the mitigation technique can provide a secure, reliable and consistent method for evaluating and operating large data to be used in machine learning models. These models can have a separate platform for ethics rules and association to implement a secure and usable model that learns a data set for better solution.

References

1. Poszler F, Lange B. The impact of intelligent decision-support systems on humans' ethical decision-making: A systematic literature review and an integrated framework.
2. Nadeem A, Marjanovic O, Abedin B. Gender bias in AI-based decision-making systems: a systematic literature review.
3. Malikireddy SKR. Generative AI for Fraud Detection in B2B Transactions.
4. Zimek A, Schubert E, Kriegel H. A survey on unsupervised outlier detection in highdimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal*. 2012;5(5):363-387.
5. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD

International Conference on Knowledge Discovery and Data Mining. ACM; 2016. p. 785-794.
6. Kim Y, Park SC, Scornet E. Anomaly detection using ensemble learning and feature extraction from data streams. *Expert Systems with Applications*. 2016;62:300-314.
7. Kalusivalingam AK. Optimizing Decision-Making with AI-Enhanced Support Systems: Leveraging Reinforcement Learning and Bayesian Networks. *International Journal of AI and ML*. 2020;1(2).
8. Kalusivalingam AK, Sharma A, Patel N, Singh V. Leveraging Generative Adversarial Networks and Reinforcement Learning for Business Model Innovation: A Hybrid Approach to AI-Driven Strategic Transformation. *International Journal of AI and ML,*. 2022, 3(9).
9. Ngai EWT, Hu Y, Wong YH, Chen Y, Sun X. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*. 2011;50(3):559-569.
10. Bhuiyan MMR, Rahaman MM, Aziz MM, Islam MR, Das K. Predictive analytics in plant biotechnology: Using data science to drive crop resilience and productivity. *Journal of Environmental and Agricultural Studies*. 2023;4(3):77-83.
11. Rahaman MM, Rani S, Islam MR, Bhuiyan MMR. Machine learning in business analytics: Advancing statistical methods for data-driven innovation. *Journal of Computer Science and Technology Studies*. 2023;5(3):104-111.
12. Fernandes F. Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI: Bias Mitigation Strategies for classification tasks. *Journal of Database Management*. 2020.