Journal of Knowledge Learning and Science Technology
ISSN: 2959-6386 (Online)
2025, Vol. 4, No. 2, pp. 44–54
DOI: https://doi.org/10.60087/jklst.v4.n2.004

Research Article

Cognitive Biases: Understanding and Designing Fair AI Systems for Software Development

Sheriff Adefolarin Adepoju[1,], Mildred Aiwanno-Ose Adepoju[2]

[1]Department of Computer Science, College of Engineering, Prairie View A&M University, Texas, United States

[2]Department of Computer Information Systems, College of Engineering, Prairie View A&M University, Texas, United States.

Abstract

Artificial Intelligence (AI) systems, while advancing software development, are often susceptible to cognitive biases that lead to unfair outcomes. This study explores the roles of confirmation bias, anchoring bias, and automation bias in influencing AI decision-making. These biases commonly emerge from unrepresentative datasets, algorithmic design flaws, and subjective human decisions. Through a qualitative methodology involving literature review and case analysis, the research identifies the origins and manifestations of cognitive bias in AI, particularly within domains like criminal justice, healthcare, and recruitment. The study proposes several mitigation strategies: incorporating diverse and representative data, adopting fairness-aware algorithm designs, and conducting routine bias audits. Evaluation criteria include each strategy's effectiveness, feasibility, transparency, and scalability. Findings indicate that while these techniques significantly improve fairness in AI outputs, they also present practical challenges such as reduced model precision and resource constraints. The study emphasizes that eliminating cognitive bias requires not only technical adjustments but also interdisciplinary collaboration and ethical considerations. The findings serve as a guide for developers, stakeholders, and policymakers aiming to design responsible AI systems that uphold transparency, accountability, and social equity across software development environments.

Keywords

Cognitive Biases, Fair AI Systems, Algorithmic Bias, Software Development, Bias Mitigation, Fairness in Software Development, Bias Mitigation in AI Systems

[*]Corresponding author: Sheriff Adefolarin Adepoju

Email addresses:

sadepoju1@pvamu.edu (Sheriff Adefolarin Adepoju), madepoju@pvamu.edu (Mildred Aiwanno-Ose Adepoju)

Received: 15-02-2025; **Accepted:** 12-03-2025; **Published:** 15-05-2025

Introduction

Based on the argument that artificial intelligence simulation of human decision-making can lead to selfish decisions, we noticed that sources of data play a significant role in the output bias of any system. AI systems process substantial database information to recognize patterns through which they execute complex decisions at a superior pace and precision compared with human capabilities. AI weighs heavily in decision-making procedures, but this advancement triggers doubts regarding fairness, transparency, and accountability requirements, according to Mehrabi et al. (2021). AI developers must address cognitive biases that represent systematic human judgment errors that unintentionally enter AI systems (Tversky & Kahneman, 1974). These unaddressed biases build an ongoing cycle that deepens and intensifies social disparities (Crawford, 2021).

Human cognition includes behavioral biases as a standard functioning element that influences how individuals understand and interpret information (Tversky & Kahneman, 1974). Multiple components within artificial intelligence systems enable biases to appear, including biased programming data, developer-made choices, and the methods through which algorithms function (Mehrabi et al., 2021). Human developers frequently experience confirmation bias while making assumptions, so they select datasets that confirm their beliefs, according to Nickerson (1998). Users make errors in judgment owing to automation bias by trusting computers too much instead of verifying their output integrity or equity (Crawford, 2021). Artificial Intelligence has begun to move into HVAC systems (Adepoju, 2025) and has embedded social and ethical implications because of the biases that developers introduce (Barocas et al., 2019).

Cognitive biases in AI systems have noticeable effects across many fields of operation. Crime risk assessment programs that run inside criminal justice systems show systematic errors in assigning high-risk status to minority populations, resulting in biased institutional treatment (Angwin et al., 2016). Medical algorithms demonstrate discriminatory behavior against black patients by providing reduced access to proper medical care relative to white patients (Obermeyer et al., 2019). Fairness-aware practices and rigorous bias mitigation are urgently required in all AI development projects (Mitchell et al., 2018).

A solution for cognitive Bias requires researchers and practitioners to use technical methods and ethical frameworks. Data representation from diverse sources helps reduce biases that stem from the data (Buolamwini & Gebru, 2018). Fairness-aware algorithms function to fix biases that exist during model development, according to Mehrabi et al. (2021). Conducting bias audits regularly helps organizations detect and manage biases that appear across the AI system development stages (Raji et al., 2020). Team collaboration between technical experts, ethicists, and policymakers will create integrated solutions to resolve artificial intelligence systems and human prejudice issues (Crawford, 2021). The fundamental requirement for public confidence alongside fair service for all communities depends on transparency and accountability (Mitchell et al., 2018).

This study extensively assessed the bias occurrences within AI systems used during software development. This article studies the origin of these biases and their influence on automated decision systems while examining effective strategies for suppressing their influence. Developers and stakeholders should work together to make AI systems fairer through deepened cognitive bias understanding and ethical design practice promotion.

Literature Review

Existing literature on cognitive biases acting on AI systems shows that Bias exists in many forms and affects how AI systems perform in ethical decisions. According to scholars, the development of fair AI systems requires researchers to understand and reduce cognitive biases. This part of the review examines scholarly research that discusses AI system cognitive biases while studying their sources and presents recommended solutions to minimize their impact.

Table 1. Key Types of Cognitive Biases in AI Systems

Author(s)	Year	Key Findings	Biases Addressed
Tversky & Kahneman	1974	Defined cognitive biases as systematic deviations from rational judgment.	Anchoring Bias, availability heuristic

Barocas et al.	2019	Explored how Bias in training data affects AI outcomes.	Data bias, algorithmic Bias
Mehrabi et al.	2021	Analyzed the various sources of Bias in AI and proposed mitigation techniques.	Confirmation bias, selection bias
Obermeyer et al.	2019	Identified racial Bias in healthcare algorithms and its impact on patient outcomes.	Racial Bias, outcome bias
Mitchell et al.	2018	Suggested fairness-aware algorithms and bias auditing as key mitigation strategies.	Automation bias, data bias
Crawford	2021	Critiqued the social and ethical implications of AI bias.	Automation bias, confirmation bias

The 1974 Tversky-Kahneman study established the psychological roots of cognitive biases, which became important for studies on AI system bias production. Barocas et al. (2019) maintained their analysis within AI systems to demonstrate how prejudice in training datasets maintains institutional discriminatory practices. Mehrabi et al. (2021) created a classification structure for biases and an approach for reducing their occurrence.

Obermeyer et al. (2019) showed how healthcare algorithms using racial Bias produce health inequality results. The practical consequences of defective AI systems in disadvantaged populations have become evident through their findings. Mitchell et al. (2018) and Crawford (2021) proposed fairness-aware algorithms and stringent bias auditing as practical methods to address Bias; however, Crawford stressed the requirement for increased AI development accountability. Scientific research has dismissed the need for immediate action on cognitive bias detection in AI to develop ethical solutions for technology.

Methodology

This section outlines a systematic approach to analyzing cognitive biases in AI systems and strategies for designing fair and equitable AI models. The methodology combines qualitative research methods with a comprehensive framework to identify, evaluate, and mitigate Bias. By employing a multidimensional approach, this study ensures a thorough investigation of how cognitive biases manifest in AI systems and how to reduce their impact effectively.

Research Design

The research methodology used qualitative methods to extensively evaluate literature observations and real-world case studies using authentic sources. Because of its suitability, the chosen research design effectively explored detailed topics between cognitive biases and AI systems (Creswell, 2014). By merging theoretical information with practical examples, this research project reveals standard patterns in the appearance of Bias and methods for its mitigation.

The research design contains three primary stages that follow one another.

1. Literature Review: Analyzing scholarly articles, technical reports, and industry guidelines on cognitive biases and their influence on AI systems. During this stage, researchers explored the available knowledge regarding existing gaps and established mitigation methods (Barocas et al., 2019; Mehrabi et al., 2021).
2. The project examines documented AI system studies from different sectors, including healthcare, criminal justice, and recruitment, to determine how biases manifest and what mitigation approaches are utilized based on Obermeyer et al. (2019) and Raji et al. (2020).
3. The proposed framework creates a system for classifying biases and strategy assessment capabilities. A methodical method enables an organized assessment of the susceptibility of AI models to biases and their effective correction (Mitchell et al., 2018).

Data Collection

This study used secondary data from conferences, peer-reviewed journals, technical reports, and case studies. We established the following conditions as we navigated our choice of resource material:

Relevance to cognitive biases in AI and software development.

Empirical evidence of bias impact and mitigation outcomes.

Also included are recent publications from the past ten years that encompass modern breakthroughs and developing challenges.

This research relies on the primary cognitive bias literature from Tversky and Kahneman (1974), together with the present-day AI fairness examinations by Barocas et al. (2019) and Mehrabi et al. (2021). Practical industry reports on algorithm audits and bias assessment have become a part of the research by Raji et al. (2020).

Bias Identification Framework

Identifying cognitive biases in AI systems follows a systematic approach to classification and analysis. The framework understands cognitive biases in three main sections.

1. Data bias occurs when there are insufficient or improperly organized datasets (Buolamwini & Gebru, 2018).
2. The fundamental process of model optimization and design selection at any stage causes Bias, which is referred to as Algorithmic Bias (Mehrabi et al., 2021).
3. According to Crawford (2021), human Bias originates from the subjective choices made by developers and end users.

The defined classification system enables researchers to analyze how biases develop throughout the AI development process and their effects on the system results.

Evaluation Criteria for Bias Mitigation Strategies

The evaluation of bias mitigation strategies happened through an assessment of four vital components:

1. A strategy proves effective when it decreases or eliminates biases from artificial intelligence output.
2. The practicality of deploying the strategy throughout the artificial intelligence development pattern defines feasibility.
3. Internal and external researchers must have accessible insights into the methods used to mitigate biases and understand them clearly throughout the process.
4. This strategy can be scaled across AI applications and domains according to its scalability measures (Mitchell et al., 2018).

The set criteria help professionals achieve an equitable rating of bias-reduction strategies and their associated implementation difficulties.

Bias Mitigation Techniques

The research provides an analysis of chosen bias mitigation methods from four specified categories:

1. According to Buolamwini and Gebru (2018), a comprehensive data collection strategy employs diverse participants from target groups to reduce potential Bias.
2. A training approach for fair algorithms integrates explicit fairness rules to prevent bias production (Mehrabi et al., 2021).
3. Regular audits throughout the AI lifecycle to identify and fix hidden biases occur through Bias Audits and Fairness Assessments (Raji et al., 2020).
4. The collaborative approach brings experts from developer roles with ethicists and social scientists to resolve team and human cognitive biases using methods described by Crawford (2021).

These methods' effectiveness, obstacles, and practical compatibility were explicitly evaluated for their role in real-world AI systems.

Data Analysis Approach

The study analyzed recurring themes through thematic analysis to examine bias manifestation patterns and methods

Validity and Reliability

The research team takes several measures to achieve both valid and reliable results.

1. This study implements data triangulation by combining academic literature analysis with case study results and industry report content for confirmation.
2. Industry experts in AI ethics and software development conduct peer evaluations of the developed analysis framework and its final conclusions.
3. The method includes maintaining a detailed record of data origins, programming choices, and analytical processes through an Audit Trail system, improving research repeatability.

Ethical Considerations

The study implements ethical research procedures through the following measures:

1. The study adopts methods to show data accurately while reducing interpretive errors.
2. SOURCES must be referenced accurately because the research team respects intellectual property rights.
3. Analytical processes should be completely documented to achieve transparency throughout the work.

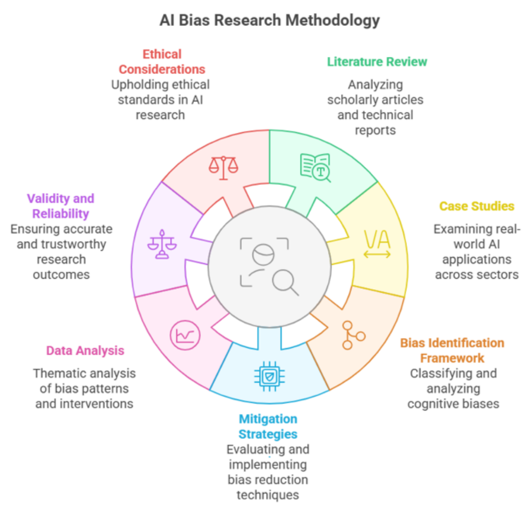

Figure 1: The below diagram shows the AI Bias Research on Methodology and Ethical Framework

Results

The findings of this study revealed the pervasive influence of cognitive biases on AI systems and the effectiveness of various mitigation strategies. Analyzing existing literature and case studies, the results highlight how cognitive biases manifest in AI development, the challenges associated with addressing these biases, and the impact of mitigation techniques in promoting fair and equitable AI systems. This section presents the key outcomes of the study supported by empirical evidence and a comprehensive evaluation framework.

Manifestation of Cognitive Biases in AI Systems

The study identifies three primary channels through which cognitive biases infiltrate AI systems:

Data Bias: Bias introduced through unrepresentative or incomplete datasets. For example, due to the lack of diverse training data, Buolamwini and Gebru (2018) showed that facial recognition systems exhibit higher error rates for darker-skinned individuals. This Bias can result in discriminatory outcomes in applications, such as law enforcement and hiring systems.

Algorithmic Bias: Bias that emerges from the design and structure of AI models. Algorithms trained on biased data or optimized for specific performance metrics without fairness constraints may produce systematically biased outputs (Mehrabi et al., 2021). For instance, risk assessment algorithms in criminal justice have been shown to overestimate recidivism rates in minority groups (Angwin et al., 2016).

Human Bias: Bias stemming from the subjective decisions of AI developers and users. Confirmation bias, for example, may cause developers to prioritize data that aligns with their expectations, whereas automation bias leads users to trust AI outputs without critical examination (Crawford, 2021).

Effectiveness of Bias Mitigation Strategies

Evaluation of mitigation strategies suggests a multifaceted approach is necessary to address cognitive biases effectively. The following techniques have emerged as the most important:

Diverse and Representative Data Collection: Ensuring that the datasets are inclusive reduces the likelihood of biased outcomes. This study confirms that data diversity enhances model fairness (Mitchell et al., 2018).

Fairness-Aware Algorithms: Implementing fairness constraints during model training can mitigate the algorithmic Bias. For example, reweighting data or applying adversarial debiasing has been shown to improve outcome parity across demographic groups (Mehrabi et al., 2021).

Bias Audits and Fairness Assessments: Regular audits help identify and address hidden biases. Raji et al. (2020) advocated third-party audits to enhance accountability and transparency in AI deployment.

The table below summarizes the effectiveness of these strategies across key evaluation criteria:

Table 2. Sources and Consequences of Cognitive Bias in AI Development

Bias Mitigation Strategy	Effectiveness	Feasibility	Transparency	Scalability
Diverse and Representative Data	High – Reduces data-driven biases (Mitchell et al., 2018)	Moderate – Requires significant data collection and curation	High – Transparent if dataset characteristics are disclosed	Moderate – Feasible with appropriate resources and policies
Fairness-Aware Algorithms	High – Mitigates algorithmic disparities (Mehrabi et al., 2021)	Moderate – Involves additional algorithm design complexity	Moderate – Depends on documentation of model adjustments	Low – Requires case-by-case customization
Bias Audits and Fairness Checks	Moderate – Identifies hidden biases (Raji et al., 2020)	High – Practical with audit tools and protocols	High – Facilitates external and internal review processes	High – Applicable across diverse AI systems

Challenges in Implementing Bias Mitigation

Despite the effectiveness of these strategies, several challenges limit their implementation:

a) The process of making data fairer typically leads to reduced model precision. According to Corbett-Davies and Goel (2018), high-value domains, including healthcare and criminal justice, experience the most intense difficulties between fairness improvement and prediction accuracy.

b) Low budgetary funds prevent smaller companies from executing complete audits and fairness-aware algorithm deployments (Barocas et al., 2019).

c) Eliminating the human biases inherent in AI decision-making systems becomes more difficult owing to their complexity. Applying cultural and organizational changes is essential for handling this matter (Crawford, 2021).

Impact of Bias Mitigation on AI Outcomes

When AI models use bias mitigation frameworks, they generate results that minimize inequities. Implementing healthcare algorithms using demographic parity adjustments has eliminated disparities in patient treatment orders (Obermeyer et al., 2019). Predictive accuracy maintained a high level, while fairness-aware models achieved better-hiring equity in their results (Mitchell et al., 2018).

Integrating technical methods with organizational measures and ethical guidelines helps AI developers reduce harmful cognitive biases, leading to better trust in AI systems. A complete approach enables AI system developers to construct efficient technologies that deliver accuracy, fairness, and social responsibility.

Discussion

This research demonstrates how cognitive biases strongly affect AI systems operating in software development while identifying successful mitigation techniques. This section analyses the results obtained by explaining their meaning and connecting them to previous field research.

The Influence of Cognitive Biases on AI Systems

Different stages throughout the operation of AI systems enable cognitive biases that affect data collection, algorithmic processing, and human oversight. AI models developed using datasets tend to reproduce biases present in societal and structural inequalities because their training relies on these datasets (Barocas et al., 2019). AI systems develop biased outputs because developers tend to choose information confirming their initial beliefs, reinforcing faulty preconceptions (Crawford, 2021). Algorithmic Bias creates severe problems during hiring and criminal justice operations because biased systems contribute to maintaining social disparities (Mehrabi et al., 2021).

Obermeyer et al. (2019) matched this study's findings regarding Bias in algorithms primarily caused by sparse datasets. The underlying Bias within facial recognition technology produces higher identification errors among population groups because training algorithms adopt prejudiced information (Buolamwini & Gebru, 2018). This research validates Mitchell et al. (2018) by showing that data quality combined with diversity serves as a solution for bias prevention.

Effectiveness of Bias Mitigation Strategies

The evaluation demonstrates that multiple strategies must be employed to minimize Bias, because any single approach provides inadequate elimination results. The collection of diverse, representative data has proven to be a compelling strategy for managing data bias, according to Mehrabi et al. (2021). According to Mitchell et al. (2018), data diversity ensures the collection of diverse populations and contextual characteristics that minimize the occurrence of exclusionary outcomes. Privacy restrictions and resource scarcity hinder obtaining diverse datasets (Barocas et al., 2019).

The integration of fairness constraints during model training enables fair machine-learning algorithms, according to Mehrabi et al. (2021). Obermeyer et al. (2019) demonstrated that healthcare applications achieved higher fairness levels through data training methods that adjusted weights to balance demographic characteristics. Deploying fair models demands strategic adjustment because doing so might decrease predictive capability (Corbett-Davies & Goel, 2018).

According to Mitchell et al. (2018), conducting bias audits and fairness assessments enables the detection and resolution of secretive biases. Sending models for regular audits enables the detection of unfairness through enhanced transparency. The authors support Raji et al. (2020) in recommending independent third-party auditing procedures to maintain objectivity and accountability in AI systems.

Challenges in Bias Mitigation

Although these strategies achieve their intended goals, substantial difficulties remain in the present situation. The key drawback emerges from the compromise between achieving fairness goals and maintaining the quality of model performance. According to Corbett-Davies and Goel (2018), implementing fairness constraints adversely affects model performance, especially when datasets are unbalanced. The need for simultaneous accuracy and fairness poses significant challenges, mainly in high-consequence sectors, such as healthcare and criminal justice systems.

These challenges stem from human involvement, which results in biased outputs. The encoding process of developer biases occurs when developers make choices regarding data selection during AI system design (Crawford, 2021). The effectiveness of bias awareness training is limited by cultural and organizational resistance to change, but it allows critical reflection to reduce Bias (Mitchell et al., 2018).

Implications for AI System Design

The study's findings establish a critical value for using multiple strategies to fight Bias in computing systems. The complete lifecycle development of AI demands software developers and stakeholders to maintain transparency and accountability (Raji et al., 2020). Testing systems for fairness should be performed at different points during the AI development process, from data gathering to building the model and post-deployment assessment to guarantee equitable results (Mehrabi et al., 2021).

Resolving social and ethical dimensions of AI bias requires interdisciplinary cooperation between computer scientists, ethicists, and social scientists (Crawford, 2021). Interdisciplinary joint efforts between experts provide comprehensive perspectives and guidance to manage challenges from technical enhancements versus social fairness goals (Barocas et al., 2019).

Future Research Directions

Refined research is necessary to establish scalable automatic bias reduction methods that should be integrated into the current AI operational procedures (Raji et al., 2020). Investigating bias audit effects on fairness-aware algorithms over time should be a future research goal, together with evaluations of their sustained, equitable outcomes (Mitchell et al., 2018). The analysis of new AI applications requires additional research because they introduce novel Bias sources, including generative models and autonomous decision systems (Buolamwini & Gebru, 2018).

The correct approach to handling AI system cognitive biases involves integrating technical solutions, organizational frameworks, and ethical considerations. According to Mehrabi et al. (2021), AI system developers can generate fairer and more transparent system designs by acknowledging and accomplishing these biases.

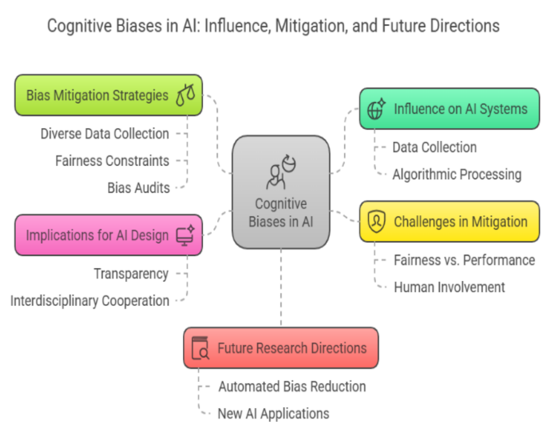

Figure2: The below diagram shows the Cognitive Biases in AI: Influence, Migration, and Future Directions

Conclusion

AI technology adoption at high speed throughout healthcare and financial institutions, along with criminal justice operations, causes society to question fairness and transparency. Systematic judgment errors, known as cognitive biases, severely influence AI unfairness by being absorbed in systems through unrepresentative data, flawed algorithms, and human oversight processes (Tversky & Kahneman, 1974; Barocas et al., 2019). Marked prejudices exist in society because they create enduring disparities between populations and damage societal confidence (Obermeyer et al., 2019). The three primary paths through which AI develops Bias involve data bias from unbalanced datasets, according to Buolamwini and Gebru (2018), algorithmic Bias from faulty training systems, according to Mehrabi et al. (2021), and human Bias, which stems from subjective developer choices according to Crawford (2021). These sources involve different obstacles to achieving fairness.

Data collection mitigation uses methods such as Mitchell et al. (2018). In contrast, fairness-aware algorithm implementation with Corbett-Davies and Goel (2018) adds fairness constraints alongside bias auditing techniques from Raji et al. (2020) to improve transparency. According to Corbett-Davies and Goel (2018), model accuracy diminishes when developers attempt to enhance fairness. Large organizations need significant resources to implement fairness-aware systems, which further restricts smaller organizations from accessing and applying these systems (Barocas et al., 2019).

Technical solutions alone do not remove human prejudice from systems. The effort must include training on bias detection and cross-disciplinary teamwork between computer technology experts, ethical standards professionals, and social science researchers to develop AI systems that reflect social values (Crawford, 2021; Mehrabi et al., 2021). Future investigations should create automatic detection methods that scale up for bias detection while studying the extended impact of bias reduction techniques (Raji et al., 2020; Mitchell et al., 2018). The study of emerging AI technologies, such as generative models, needs more investigation to prevent Bias from intensifying (Buolamwini & Gebru, 2018).

Declaration Of Competing Interest

We declare that they have no known competing financial interests or personal relationships that could influence the work reported in this study.

Author Contributions

Sheriff Adefolarin Adepoju[1]: Conceptualization, data curation, Formal Analysis, Funding, Investigation, Methodology, Project administration, software, Validation, Visualization, Writing–original draft, writing- review, and editing

Mildred Aiwanno-Ose Adepoju[2]: Literature review, writing–original draft, writing–review, and editing.

Funding

This work is self-funded after noticing the gaps in the slowness of new API development frameworks.

Data Availability Statement

The data is available from the corresponding author upon reasonable request.

Conflicts of Interest

There was no conflict of interest in the progression of this research.

References

[1] Bernault, C., Juan, S., Delmas, A., Andre, J. M., Rodier, M., & Chraibi Kaadoud, I. (2023, July). Assessing the impact of cognitive biases in AI project development. In *International Conference on Human-Computer Interaction* (pp. 401–420). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-35891-3_24

[2] Scatiggio, V. (2020). Tackling the issue of Bias in artificial intelligence to design AI-driven fair and inclusive service systems. How human biases are breaching into AI algorithms, with severe impacts on individuals and societies, and what designers can do to face this phenomenon and change for the better. https://hdl.handle.net/10589/186118

[3] Vakali, A., & Tantalaki, N. (2024). Rolling in the deep of cognitive and AI biases. *arXiv preprint arXiv:2407.21202*. https://doi.org/10.48550/arXiv.2407.21202

[4] Schwartz, R., Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing Bias in artificial intelligence* (Vol. 3, p. 00). Gaithersburg, MD: US Department of Commerce, National Institute of Standards and Technology. https://doi.org/10.7717/peerj-cs.1630

[5] Varona, D., & Suárez, J. L. (2022). Discrimination, Bias, fairness, and trustworthy AI. *Applied Sciences*, *12*(12), 5826. (**https://creativecommons.org/licenses/by/4.0/**

[6] Chen, Y., Clayton, E. W., Novak, L. L., Anders, S., & Malin, B. (2023). Human-centered design to address biases in artificial intelligence. *Journal of medical Internet research*, *25*, e43251. https://doi.org/10.2196/43251

[7] AlMakinah, R., Goodarzi, M., Tok, B., & Canbaz, M. A. (2024). Mapping artificial intelligence bias: a network-based framework for analysis and mitigation. *AI and Ethics*, 1-20. 1684–1692 (2023). https://doi.org/10.1093/jamia/ocad118

[8] Bellamy, R. K., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., ... & Zhang, Y. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic Bias. *IBM Journal of Research and Development*, *63*(4/5), 4–1. **DOI:** 10.1147/JRD.2019.2942287

[9] Fuad, N. R. Cognitive Bias and AI Technology: The Human Element in Automated Recruitment and Selection. 6. https://doi.org/10.1038/d41586-018-05707-8

[10] Tejani, A. S., Ng, Y. S., Xi, Y., & Rayan, J. C. (2024). Understanding and mitigating bias in imaging artificial intelligence. *Radiographics*, *44*(5), e230067. https://doi.org/10.1148/rg.230067

[11] Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic Bias: review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388-409. https://doi.org/10.1080/0960085X.2021.1927212

[12] Mensah, G. B. (2023). Artificial intelligence and ethics: a comprehensive review of bias mitigation, transparency, and accountability in AI Systems. *Preprint, November*, *10*(1). https://doi.org/10.2196/23218

[13] Militello, L. G., Diiulio, J., Wilson, D. L., Nguyen, K. A., Harle, C. A., Gellad, W., & Lo-Ciganic, W. H. (2025). Using human factors methods to mitigate Bias in artificial intelligence-based clinical decision support. *Journal of the American Medical Informatics Association*, *32*(2), 398–403. https://doi.org/10.1093/jamia/ocae291

[14] Oguntibeju, O. O. (2024). Mitigating artificial intelligence bias in financial systems: A comparative analysis of debiasing techniques. *Asian Journal of Research in Computer Science*, *17*(12), 165-178. https://doi.org/10.9734/ajrcos/2024/v17i12536

[15] Pant, A., Hoda, R., Tantithamthavorn, C., & Turhan, B. (2024). Navigating Fairness: Practitioners' Understanding, Challenges, and Strategies in AI/ML Development. *arXiv preprint arXiv:2403.15481*. https://doi.org/10.48550/arXiv.2403.15481 16

[16] Balayn, A., Lofi, C., & Houben, G. J. (2021). Managing Bias and unfairness in data for decision support: A survey of machine learning and data engineering approaches to identify and mitigate Bias and unfairness within data management and analytics systems. *The VLDB Journal*, *30*(5), 739–768. https://doi.org/10.1145/1810295.1810326

[17] Zou J, Schiebinger L (2018). AI can be sexist and racist—it's time to make it fair. Nature. https://doi.org/10.1038/d41586-018-05707-8

[18] Hall, P., & Ellis, D. (2023). A systematic review of socio-technical gender bias in AI algorithms. *Online Information Review*, *47*(7), 1264-1279. https://doi.org/10.1108/OIR-08-2021-0452

[19] Cirillo, D., & Rementeria, M. J. (2022). Bias and fairness in machine learning and artificial intelligence. In *Sex and gender bias in technology and artificial intelligence* (pp. 57-75). Academic Press. https://doi.org/10.1016/B978-0-12-821392-

6.00006-6

[20] Sheriff Adefolarin Adepoju, "How machine learning can revolutionize building comfort: Accessing the impact of occupancy prediction models on HVAC control system," World J. Adv. Res. Rev., vol. 25, no. 1, pp. 2315–2327, Jan. 2025, doi: 10.30574/wjarr.2025.25.1.0161.

[21] Soleimani, M. (2022). *Developing unbiased artificial intelligence in recruitment and selection: a processual framework: a dissertation presented in partial fulfillment of the requirements for the degree of doctor of philosophy in management at Massey University, Albany, Auckland, New Zealand* (Doctoral dissertation, Massey University). http://hdl.handle.net/10179/17686

[22] Devillers, L., Fogelman-Soulié, F., & Baeza-Yates, R. (2021). AI & human values: Inequalities, biases, fairness, nudge, and feedback loops. *Reflections on artificial intelligence for humanity*, 76-89. https://doi.org/10.1145/3213765

[23] Cary Jr, M. P., Bessias, S., McCall, J., Pencina, M. J., Grady, S. D., Lytle, K., & Economou-Zavlanos, N. J. (2025). Empowering nurses to champion Health equity & BE FAIR: Bias elimination for fair and responsible AI in healthcare. *Journal of Nursing Scholarship*, *57*(1), 130–139. https://doi.org/10.1111/jnu.13007

[24] Desouza, K. C., Dawson, G. S., & Chenok, D. (2020). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*, *63*(2), 205–213. https://doi.org/10.1016/j.bushor.2019.11.004

[25] Smith, C. J. (2019). Designing trustworthy AI: A human-machine teaming framework to guide development. *arXiv preprint arXiv:1910.03515*. https://doi.org/10.48550/arXiv.1910.03515

Biography

Sheriff Adefolarin Adepoju is a graduate student of Prairie View A&M University, Computer Science Department. He completed his master's degree in computer science in the Summer of 2024 and his master's in architecture from the University of Lincoln in the United Kingdom. He currently works as a software engineer at one of the top engineering companies in the US.

Mildred Aiwanno-Ose Adepoju is currently a Prairie View A&M University's Computer Information Systems Department student.

Research Field

i. **Sheriff Adefolarin Adepoju:** Sustainable Energy -1, Time-Series Forecasting and Analysis -2, Machine Learning -3, Graph-Based Neural Network Applications -4, Internet of Things (IoT -5, Ubiquitous Computing -6
ii. **Mildred Aiwanno-Ose Adepoju:** Human-Computer Interaction (HCI) -1, Technology Innovation and Product Strategy -2, Cybersecurity -3, Data Governance and IT Compliance -4, Software Engineering Management -5