Mitigating Artificial Intelligence Bias in Financial Systems: A Review of Debiasing Techniques

ABSTRACT

The use of Artificial Intelligence (AI) in industries such as Banking, Financial Services, and Insurance (BFSI) raises concerns over bias and fairness in AI-driven decision-making systems. Despite the promise of AI to optimize business processes and enhance customer satisfaction, inherent biases embedded in AI systems can lead to unintended discrimination, particularly in sensitive areas such as loan approvals, recruitment, and fraud detection. This study investigates the origins of AI bias, how it happens in business processes, and the challenges it poses to ethical and transparent decision-making. Upon reviewing existing literature, the research explores the various types of biases—including cognitive, algorithmic, and representation biases—and their impact on AI systems in the BFSI sector. Furthermore, the study evaluates current debiasing techniques, such as pre-processing, fairness-aware models, and post-processing, highlighting their limitations in balancing fairness with predictive accuracy. The findings underscore the need for a more comprehensive approach to mitigating bias and promoting fairness, suggesting that a comprehensive framework for addressing these challenges is essential for the development of ethical AI systems. Ultimately, this research aims to contribute to the creation of more equitable AI systems by offering insights into best practices for data governance, debiasing strategies, and regulatory oversight, with a focus on minimizing bias and fostering trust in AI-driven decision-making.

Keywords: Bias in Financial Services, AI Bias, Algorithmic Fairness, Debiasing Techniques, Ethical AI, AI Transparency

1. INTRODUCTION

Artificial Intelligence (AI), in the 21st century, has become more popular and a common theme in every sphere of our lives [1]. This belief is further reinforced with the wide acceptance of Generative Pre-trained Transformer tools such as ChatGPT, Gemini and Copilot [2] and massive development of diverse AI models trained on specific datasets [3]. The application of AI in Banking, Financial Services, and Insurance (BFSI) is a common feature and has been applied to predict better customer choices, customize, and deliver seamless solutions to customers and help businesses achieve competitive advantage [4]. AI and Technology adoption in firms is expected to enhance business operations, growth and make better decisions.

AI-driven decision-making has shown a tendency to produce uneven results, yet research on this topic remains limited. These biases, which frequently manifest in areas like racial profiling,

credit assessments, and facial recognition, pose significant hurdles for ensuring fairness in the way businesses use AI. Despite its benefits, it is prone to biases and errors, particularly in the BFSI sectors. For example, biased loan decisions have been observed even without explicit discriminatory programming [5]. Similarly, gender bias has surfaced in career-related ads especially in STEM fields that raise concerns about AI-driven decision-making processes.

Biased AI projections can negatively impact consumers, leading to dissatisfaction, reduced customer loyalty, and lower profitability for firms [6]. Biases can exist within the algorithms' code, even when they fail to make decisions. When data scientists overlook the societal context of AI applications, bias is further introduced into business processes [7]. Automated choices in AI systems have also led to discriminatory outcomes and undesirable ads [8], which reflect technological flaws rather than human error.

To tackle these ongoing problems, it's crucial for businesses and researchers to take a deeper, more thoughtful approach when evaluating AI systems.

1.2 IMPORTANCE OF THE PROBLEM

Recent studies have highlighted how pervasive bias is within AI algorithms, and how this can worsen societal disparities [1]. The vast influence of AI is visible in everything from businesses to public institutions. For instance, Amazon's recruitment tool, which exhibited significant gender bias, had to be shut down in 2015 [9]. The tool was trained on a dataset that overrepresented men, inadvertently causing it to favor male candidates. This serves as a clear example of how the quality of data fed into AI systems can introduce unintended biases.

As AI learns from the data it's given, poor-quality or biased data can lead to unintended outcomes. To combat this, researchers have proposed various techniques to minimize bias and promote fairness during the AI development process. Each technique comes with its own set of advantages and drawbacks, making it essential to take a closer look to truly grasp the effectiveness of each technique.

In this study, the origins of AI bias and the impact it has, examining not only biases in the data and algorithms but also those introduced by users. Ultimately, the goal is to contribute to the development of more responsible and ethical AI systems by addressing the root causes and solutions for bias and fairness.

1.3. RESEARCH OBJECTIVE

The persistent issue of AI failing to produce unbiased outcomes highlights a problem that still requires significant attention. This research aims to address these challenges by attempting to answer the following key questions:

- What are AI biases, and how can fairness be ensured when integrating AI into business processes?
- How can biases and potential vulnerabilities in AI systems, particularly within the BFSI sector, be effectively addressed?
- How can the current debiasing techniques and approaches address the problem of bias and fairness in AI decision making?

1.4. RESEARCH QUESTIONS

This study will explore the challenges posed by AI bias and vulnerabilities, particularly within the sensitive context of Banking, financial services, and Insurance sectors (BFSI). To achieve these objective, the following primary and secondary research questions will be addressed:

- What are the types of AI biases and fairness observed in business processes?
- What are the ethical and regulatory requirements for ensuring fairness in AI systems within BFSI sectors?
- How can BFSI organizations identify and address vulnerabilities in their AI systems to prevent bias?

Secondary Questions:

- What are the best practices for data governance and quality assurance in the BFSI sector to mitigate bias and discrimination?
- How effective are current debiasing techniques in addressing bias and promoting fairness in Algorithmic decision making?
- What are the limitations and challenges associated with implementing debiasing techniques, particularly in BFSI sectors?
- How can debiasing techniques be combined with other approaches to create fair AI systems?

2. LITERATURE REVIEW

2.1. ORIGIN OF AI SYSTEMS & BIAS

The origins of Artificial Intelligence (AI) can be traced back to 1950, when British mathematician Alan Turing developed a test to determine whether a machine could replicate human cognitive abilities to recognize patterns [10]. AI gained further attention in 1956 when John McCarthy, a computer scientist, brought together academics and industry experts from around the world to discuss the potential of machines that could process data and imitate human behavior [11]. The ability to share and process data on a global scale became a reality with advancements in computing power which transformed businesses and reshaped marketplaces.

Algorithmic bias can be traced to 1976, when Joseph Weizenbaum posited that bias could arise from the instructions issued to the computer or from the data used to train the system. Earliest computers were designed to think and mimic human reasoning and make deductions to reflect human thinking. By following rules based on the assumptions, thought process of humans on how problems should be solved, bias could be introduced unintentionally or by design choices [1].

As AI systems become more intricate, analyzing algorithmic bias has become increasingly challenging. Decisions made by individual designers, engineers or teams can become buried within layers of code, and over time, the influence of these choices on the program's behavior

might be overlooked. These biases could, in turn, create new patterns as technology interacts with society. Additionally, biases can shape how society adapts to the data algorithms rely on. For instance, if an algorithm detects a higher rate of arrests in a certain area, it may increase police presence, potentially leading to even more arrests [6].
Concerns about algorithmic impact have led companies like Google and Microsoft to form groups addressing fairness and transparency. Google's initiatives include community oversight of algorithm outcomes. The study of algorithmic fairness has grown into a dedicated research field with its own conference called Fairness, Accountability, and Transparency (FAccT).

2.2. DEFINITIONS

2.2.1 PROTECTED ATTRIBUTES AND PRIVILEGE GROUPS

Protected attributes are qualities like race, gender, age, and religion that are legally safeguarded against discrimination. These characteristics must be handled with care in decision-making to prevent bias, ensuring fairness in processes like hiring, where choices should be based on merit rather than these sensitive attributes.

Privilege groups are people who, due to their social standing or protected attributes, have access to opportunities and resources not available to others [12]. For example, white men or individuals from affluent backgrounds often benefit from systemic advantages in areas like education and employment, a result of historical inequalities like racism and sexism. On the other hand, non-privileged groups face barriers to opportunities, discrimination, and disadvantages due to their protected attributes [13]. Due to factors such as race, color, disability and demographics location, such people may face bias and discrimination.

2.2.2. DISPARATE IMPACT
When AI decisions negatively affect a specific group, even if there was no intent to discriminate, disparate impact is said to occur. Possible causes include when AI models are trained on biased data or when algorithms unintentionally reinforce societal biases. Facial recognition systems, in the past, have been less accurate in identifying individuals with darker skin tones, leading to biased outcomes in law enforcement [1]. Loan and mortgage approval algorithms might unfairly deny loans to marginalized groups, such as those with lower incomes or people of color.
In the study by [14], beyond biased data, a lack of diversity in the team involved in the development of AI systems could introduce disparate impact because they may be less likely to spot and address potential biases.

2.2.3. GRANDFATHERED DATA
As important as historical data is when making decisions, they may raise ethical concerns. In the past, long before regulations on privacy and ethical practices were set, data collected before the introduction of new privacy laws and ethical standards may not meet requirements of current regulations. Grandfathered data could pose potential problems if they risk violation of individual rights. [5] advised on the importance of taking extra steps when dealing with historical data. Data anonymization, pseudonymization should be basic steps taken to ensure that historical data meets today's regulatory requirements when used. As data gets trained, new data might have been trained on biased and non-compliant historical data, then bias becomes propagated into new datasets.

2.2.4. INDIVIDUAL AND GROUP FAIRNESS
Group fairness in AI systems encompasses the objective of treating different groups equally or proportionally. Individual fairness, in contrast, pertains to the assurance that comparable

individuals receive similar treatment from AI systems, irrespective of their group affiliations. This objective can be accomplished through approaches such as similarity-based or distance-based measures, which aim to guarantee that individuals with similar characteristics or attributes are treated comparably by the AI system [15].

2.3. BIAS AND FAIRNESS IN AI SYSTEMS

Fairness and bias are closely related concepts, yet they have distinct meanings. Despite these differences, the two concepts are closely intertwined. Addressing bias is a fundamental step toward achieving fairness in Artificial Intelligence**.** Bias refers to consistent errors in an algorithm's outputs where results diverge from what is accurate [12]. In contrast, fairness in AI aims to eliminate any form of discrimination based on factors like race, gender, age, or religion. One important difference between these concepts is that bias can occur without intent, while fairness is an outcome that requires deliberate effort. Bias may arise from various issues, such as inaccurate data, interest of AI designers or poorly designed algorithms. On the other hand, achieving fairness demands proactive measures to ensure that no individual or group faces unjust treatment.

Moreover, bias can be categorized as either positive or negative. Positive bias occurs when an algorithm tends to favor a certain group, whereas negative bias results in discrimination against a group. Fairness, however, is primarily concerned with addressing and eliminating negative bias, striving to ensure that all individuals receive equitable treatment.

When biased data is fed into a system, the output is likely to mirror that bias [16]. In the insurance sector, biases have also emerged, such as when premium calculations were influenced by religious affiliations rather than gender [17]. Furthermore, bias can manifest in dynamic pricing models and targeted promotions, where certain groups might be unfairly favored by the algorithms [18].This indicates that bias can deeply embed itself within algorithms, particularly when they are trained on skewed datasets.

Type of Bias	Description	Examples
Cognitive Bias	Bias that stems from human decision-making that affects the development of AI systems leading to unintentional discrimination during programming and coding.	This arises from assumptions during algorithm development or biased training data, leading to discriminatory outcomes in marketing or consumer predictions.
Algorithmic Bias	Bias that stems from design and implementation of AI systems which end up unintentionally favoring certain outcomes. This results in unfair results especially due to biased data.	Biases in AI can manifest in consumer choice prediction and can be categorized into observable (e.g., purchasing patterns) and unobservable (e.g., hidden pricing) biases.
Representation Bias	This bias happens when training data doesn't reflect the real-world diversity. This results in unfair outcomes for certain groups because the AI model does not account for the full diversity.	In banking, a credit scoring model trained mostly on higher-income data may discriminate against lower-income applicants, reinforcing financial inequalities.

Confirmation Bias	Bias that stems from AI systems favoring data that aligns with existing assumptions, further reinforcing pre-existing patterns and excluding diverse perspectives.	Loan approval algorithms that favor applicants from wealthier areas may perpetuate disparities by consistently denying loans to lower-income neighborhoods.
Sampling Bias	Bias that occurs when the sample used to train an AI model or conduct research is not randomly selected and therefore is not representative of the broader population	A study on health outcomes that only surveys people in urban areas, ignoring rural populations, leading to results that don't reflect the experiences of rural residents.

Table 1: Different types of AI biases.

Fairness in AI is a complicated and multi-dimensional issue that has sparked extensive discussions in both academic and industry circles. Fairness, at its core, means that AI systems operate without bias or discrimination [14]. However, reaching this fairness is no easy task; it involves a thorough examination of the various types of biases that can emerge and strategies for addressing them.

Type of Fairness	Description	Examples
Group Fairness	Ensures that all demographic groups (e.g., race, gender, age) are treated equally, preventing algorithms from amplifying historical inequalities.	In banking, a loan approval system should provide equal opportunities to all applicants regardless of their racial or ethnic background to prevent worsening financial gaps in marginalized communities.
Individual Fairness	Focuses on treating individuals fairly based on their personal characteristics, rather than group identity.	In credit scoring, individuals should be evaluated based on their financial behaviors, not external factors like race or socioeconomic status, to avoid unfair penalties, such as lower credit limits.
Counterfactual Fairness	Ensures AI decisions remain consistent across hypothetical scenarios, even when sensitive attributes (e.g., race, gender) are changed.	In financial services, it would test whether a loan approval decision would be the same if factors like race or gender were different, helping to detect and eliminate bias.
Demographic Parity	Aims for equal distribution of positive outcomes (e.g., job offers, loan approvals) among different demographic groups, regardlessof their	In hiring, if 30% of applicants are women, then 30% of job offers should ideally go to women. Tech companies like Google and Facebook aim to implement demographic parity in their

	representation in the population.	hiring practices to ensure diversity in job offers.
Procedural Fairness	Focuses on ensuring fairness in the decision-making process itself, emphasizing transparency and accountability in AI systems.	In banking, procedural fairness ensures loan applicants are informed about how decisions are made and can challenge unfavorable outcomes.
Causal Fairness	Ensures that AI decisions are based on legitimate causal factors, rather than irrelevant correlations, promoting fairness in decision-making.	In hiring, Unilever's algorithm prioritizes candidates' qualifications over factors like address or socioeconomic status, mitigating biases and promoting a fairer selection process.

Table 2: Different types of AI Fairness

3. METHODOLOGY

A systematic review is a widely utilized methodology across multidisciplinary fields. In recent years, its application has extended into business, management, and accounting, where it is used to analyze the large volume of data dispersed across the internet. This approach offers a structured, reproducible, and quantifiable framework for synthesizing and providing a comprehensive understanding of specific domains [19]. In conducting the literature review for this study, we adhered to established guidelines from seminal review articles [20],which informed our process for identifying sources of AI bias within BFSI sectors.

To ensure a comprehensive and academically rigorous review, we used the Scopus database to source relevant publications, as it offers extensive access to scholarly resources that facilitate a deeper understanding of the topic at hand. Research papers indexed in Scopus were selected based on stringent criteria to ensure academic reliability and validity [21]. A combination of strategic keywords and database searches was employed to filter relevant literature. Specifically, the following Boolean search strategy was used: ALL ((“AI Bias*” OR “artificial intelligence bias” OR “algorithm bias*”) AND (“Bias*” OR “Risk*”)), resulting in the identification of 884 documents. Figure 1 illustrates the inclusion and exclusion criteria used for selecting relevant papers in this systematic review.

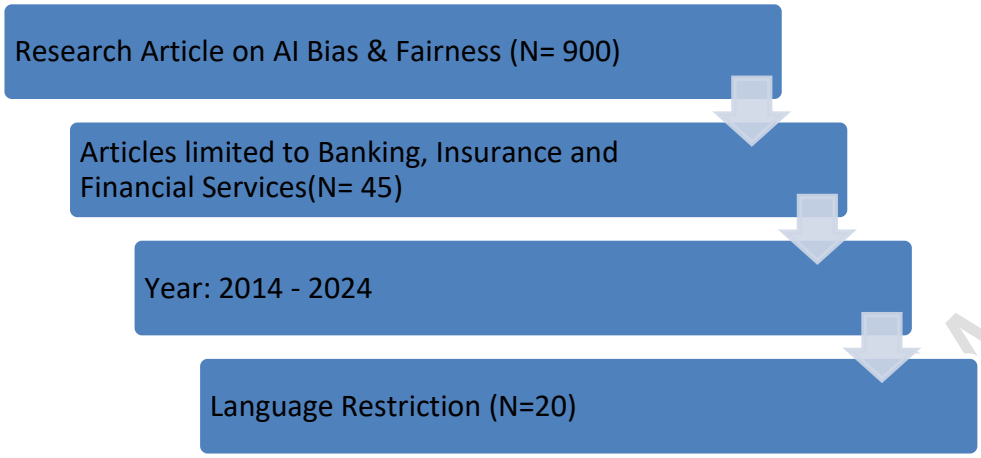

Fig 1-

4. FINDINGS FROM STUDY

4.1. BIAS IN CREDIT SCORING

Recent literature on bias in credit scoring systems highlights the tension between the necessity of traditional models and the growing concern over their potential to perpetuate systemic inequalities. Traditional credit scoring models, which rely on historical financial data, are critical for assessing credit risk, but critics argue that they may disproportionately disadvantage marginalized groups. Historical data often reflects societal biases—such as racial or socioeconomic inequalities—that can lead to unfair lending practices [19]. Consequently, there is increasing scrutiny over the fairness of these models, particularly regarding their impact on underrepresented populations [20]. To address these concerns, some researchers and companies have proposed the use of non-traditional data, such as social media activity and online behavior, to create more inclusive credit scoring systems. Proponents argue that such data provides a more holistic view of an individual's creditworthiness, especially for those without traditional credit histories. However, critics caution that these data sources may introduce new biases, as they can reflect societal prejudices and lead to discriminatory practices. Additionally, the complex and opaque nature of many machine learning algorithms used in credit scoring, often referred to as the "black box" problem, further complicates efforts to ensure transparency and accountability in these systems. As a result, there are calls for greater regulation to address algorithmic bias, though some industry stakeholders warn that overly stringent regulations may hinder innovation.

4.2. BIAS IN STOCK MARKET TRADING

Bias in stock market trading, both human and algorithmic, has garnered significant attention in academic and financial research due to its implications for decision-making, market efficiency, and fairness. Despite the long-standing assumption of rationality and efficiency in financial markets, a growing body of literature highlights the pervasive influence of cognitive and behavioral biases, such as overconfidence and herd behavior, which lead traders to make irrational decisions [4]. Studies have demonstrated that biases like overconfidence often result

in excessive trading, while herd behavior can amplify market trends, leading to price bubbles and crashes [21]. This evidence suggests that financial markets are not always efficient, as cognitive distortions can disrupt rational expectations.

In addition to human biases, the rise of algorithmic trading systems introduces a new dimension of concern. Algorithmic trading, which leverages historical data to optimize trading strategies, has the potential to both mitigate and exacerbate biases in market behavior. Research has shown that algorithms trained on biased or incomplete datasets can unintentionally replicate and amplify existing inequalities, influencing trading outcomes in ways that disadvantage certain market participants [22]. Proponents of algorithmic trading argue that these systems improve market liquidity and reduce human error, offering faster and more accurate decision-making processes. However, critics contend that algorithms can perpetuate structural inequalities by reinforcing biases embedded in the training data, leading to disparities in access to trading opportunities and market outcomes. Thus, the challenge is to develop algorithms that not only improve market efficiency but also address biases that may undermine fairness and equity.

4.3. BIAS IN FRAUD DETECTION

The growing concern over bias in AI-driven fraud detection systems has prompted significant research into various strategies for mitigating these biases while maintaining system efficacy. Despite advancements in debiasing techniques, the literature indicates that current methods still have notable limitations in terms of their effectiveness and adaptability across different machine learning (ML) models. A key challenge lies in achieving a balance between fairness, transparency, and predictive accuracy, as biases in AI systems can lead to discriminatory outcomes that disproportionately affect certain demographic groups.

Pre-processing methods, which adjust the input data to remove biased elements before training, are among the most widely adopted techniques. These methods aim to reduce historical biases, such as those related to race or gender, in the data used to train fraud detection models. While pre-processing can reduce biased outcomes, it is critiqued for oversimplifying complex relationships within the data, potentially leading to a loss of predictive power [23]. By altering data to remove biases, valuable fraud-related information may be inadvertently discarded, resulting in increased false positives or false negatives.

Another prominent strategy is the development of fairness-aware ML models, such as adversarial debiasing and fairness-constrained optimization, which introduce fairness constraints during the learning phase. These models aim to ensure that outcomes are equitable across demographic groups, but they often face trade-offs between fairness and accuracy. For example, adversarial debiasing can reduce fraud detection accuracy, particularly for certain demographic groups, as fairness constraints limit the model's ability to detect fraud in real-time [24]. Furthermore, fairness-aware models struggle with generalization across different types of fraud, leading to inconsistent performance across datasets [25]. Post-processing techniques, which adjust model outputs to achieve fairness, represent another solution but are often criticized for addressing bias too late in the process, after the model has already generated potentially skewed predictions. These methods, while improving fairness, can undermine transparency and accountability, eroding trust in AI systems used in high-stakes applications like fraud detection.

5. DISCUSSION

5.1. IMPLICATIONS FOR LITERATURE

This study contributes to the growing body of literature on AI biases within organizations, with a specific focus on the banking, financial services, and insurance (BFSI) sectors. It outlines various AI biases that can arise in these industries and offers a novel framework to better understand and mitigate these biases.

AI's ability to process large amounts of data and make decisions akin to human cognitive functions presents both opportunities and challenges. Therefore, AI systems in BFSI sectors must be critically examined for potential biases that may affect financial inclusion, risk assessment, and customer interactions [26].

5.2. IMPLICATIONS FOR MANAGERIAL AND BUSINESS PRACTICE

The findings of this study offer practical guidance for managers and policymakers in the banking, financial services, and insurance sectors seeking to address AI-related biases. The analysis suggests two primary approaches to mitigating biases. First, AI systems can be employed to identify and correct human biases in decision-making, such as in credit scoring or risk assessment models. Second, AI models themselves must be carefully constructed and monitored to ensure that they do not perpetuate societal biases or generate new forms of prejudice [27]. In the BFSI industries, this involves addressing biases in lending, underwriting, fraud detection, and customer service, where algorithmic errors may disproportionately affect certain demographic groups.

To effectively mitigate AI biases, organizations in the BFSI sectors should consider the following strategies: (1) identify contexts where AI can help eliminate bias, as well as situations where AI may unintentionally amplify bias, (2) establish robust policies and techniques for detecting and addressing biases in AI systems, and (3) promote fact-based, transparent discussions around human biases in decision-making processes. Furthermore, operational procedures should be designed to improve the quality of data used in AI models, including more diverse data sampling, frequent audits of algorithms and models (internally or through third-party audits), and greater engagement with marginalized communities to ensure inclusivity. Clear guidelines on fairness metrics in decision-making algorithms will also be crucial for maintaining consumer trust in financial products and services.

Moreover, BFSI organizations must recognize the importance of a multidisciplinary approach to addressing AI biases. By prioritizing ethical AI practices and customer well-being, BFSI organizations can build trust and strengthen their relationships with consumers.

5.3. FUTURE RESEARCH DIRECTIONS

While this study contributes to the understanding of AI bias in the BFSI sectors, it also highlights several avenues for future research. One key limitation is the focus of the literature review on business, management, and accounting contexts. Future studies could expand the scope to explore AI bias in other sectors, where similar issues related to algorithmic fairness are prevalent. In addition, more in-depth research on AI bias in the areas of consumer behavior and pricing bias can shed light on how financial institutions may inadvertently introduce discrimination in digital lending or online insurance models.

Another promising area for future research is AI bias in fraud detection systems, which could disproportionately target certain demographic groups, and remains an area requiring further investigation. Additionally, examining the ethical challenges associated with AI bias in customer service automation and claims processing could provide valuable insights into the role of fairness in consumer relations.

Future research could also focus on AI bias in data security, exploring how biases in datasets used for fraud detection or credit scoring may introduce security vulnerabilities, with consequences for both institutions and consumers.

Ultimately, addressing AI bias in the BFSI sectors requires a multidisciplinary approach, combining expertise from computer science, economics, ethics, law, and organizational behavior. Future research should continue to focus on developing solutions that reduce bias in AI systems and promote fairness, ensuring that AI technologies are used to improve financial services in ways that benefit all stakeholders.

6. CONCLUSION

The banking, financial services, and insurance (BFSI) sectors are undergoing significant digital transformation, with AI technologies being increasingly deployed to enhance decision-making processes and drive organizational success. However, as with other industries, several challenges and flaws have emerged in the application of AI within this context. This paper has highlighted the need for responsible AI practices that prioritize fairness, transparency, and accountability for firms, customers, and stakeholders. Although AI has the potential to revolutionize decision-making in BFSI organizations, it remains in an early stage, with substantial vulnerabilities that need to be addressed to ensure ethical and unbiased outcomes.

As AI continues to evolve in the BFSI space, it is evident that while AI systems can process vast amounts of data and improve operational efficiency, they cannot replicate the nuanced, emotional intelligence of human decision-making, particularly in areas such as customer relations, claims processing, and financial advising. Furthermore, the study of AI bias within the BFSI sector is still in its infancy, offering vast opportunities for innovation, research, and policy development. As financial institutions continue to implement AI, there is a critical need for ongoing research to identify, understand, and mitigate biases. Developing robust frameworks, policies, and tools to ensure fairness will be key to fostering trust, minimizing risks, and enhancing the overall customer experience in the digital economy. Future advancements in AI should aim to balance technological efficiency with ethical responsibility, ensuring that AI solutions benefit all stakeholders in the BFSI ecosystem.

REFERENCES

[1] J. Buolamwini, Unmasking AI: My Mission to Protect What Is Human in a World of Machines, Random House, 2023, p. 336.

[2] O. Oguntibeju, M. Adonis and J. Alade, "Systematic Review of Real-Time Analytics and Artificial Intelligence Frameworks for Financial Fraud Detection," International Journal of Advanced Research in Computer and Communication Engineering, vol. 13, no. 9, 2024.

[3] R. Kitchin and G. McArdle, "What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets," Big Data & Society, 2016.

[4] R. M. Gonzales and C. A. Hargrea, "How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system," International Journal of Information Management Data Insights, vol. 2, no. 2, 2022.

[5] K. Ukanwa and R. Rust, "Algorithmic Bias in Service," USC Marshall School of Business Research Paper, p. 69, 30 11 2021.

[6] F. Teleaba, S. Popescu, M. Olaru and D. Pitic, "RISKS OF OBSERVABLE AND UNOBSERVABLE BIASES IN ARTIFICIAL INTELLIGENCE USED FOR PREDICTING CONSUMER CHOICE," Amfiteatru Economic, vol. 23, no. 56, pp. 102-119, 2021.

[7] L. Yarger, F. C. Payton and B. Neupane, "Algorithmic equity in the hiring of underrepresented IT job candidates," Emerald Insight, vol. 44, no. 2, pp. 383-395, 2020.
[8] A. Datta, M. C. Tschantz and A. Datta, "Automated Experiments on Ad Privacy Settings," 2015.
[9] J. Dastin, "Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women," in Ethics of Data and Analytics, Auerbach Publications, 2022.
[10] G. Batra, A. Queirolo and N. Santhanam, "Artificial intelligence: The time to act is now," 208. [Online]. Available: https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/Artificial%20intelligence%20The%20time%20to%20act%20is%20now/Artificial-intelligence-The-time-to-act-is-now.pdf. [Accessed 01 10 2024].
[11] H. Herath and M. Mittal, "Adoption of artificial intelligence in smart cities: A comprehensive review," International Journal of Information Management Data Insights, vol. 2, no. 1, 2022.
[12] D. Pessach and E. Shmueli, "Improving fairness of artificial intelligence algorithms in Privileged-Group Selection Bias data settings," Expert Systems with Applications, vol. 185, no. 21, 2021.
[13] A. Booth, A. Sutton, M. Clowes and M. M.-S. James, Systematic Approaches to a Successful Literature Review, SAGE Publications Ltd, 2022, p. 424.
[14] E. Ferrara, "Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies," vol. 6, no. 1, 2024.
[15] A. G. Fergusson, "Predictive policing and reasonable suspicion," Emory Law Journal, p. 259, 2012.
[16] M.-H. Huang and R. T. Rust, "A strategic framework for artificial intelligence in marketing," Journal of the Academy of Marketing Science, vol. 49, no. 1, 2021.
[17] S. Akter, Y. K. Dwivedi, S. Sajib, K. Biswas, R. J. Bandara and K. Michael, "Algorithmic bias in machine learning-based marketing models," Journal of Business Research, vol. 144, pp. 201-216, 2022.
[18] A. Israeli and E. Ascarza, "Algorithmic Bias in Marketing," 2022.
[19] D. Gough, S. Oliver and J. Thomas, An Introduction to Systematic Reviews, SAGE Publications Ltd;, 2017.
[20] C. Durach, J. Kembro and A. Wieland, "A new paradigm for systematic literaturereviews in supply chain management," Journal of Supply Chain Management, vol. 53, no. 4, pp. 67-85, 2017.
[21] P. Kumar, L. Hollebeek, A. Kar and J. Kukk, "Charting the intellectual structure of customer experience research," Marketing Intelligence & Planning, 2022.
[22] J. Banasik, J. Crook and L. Thomas, "Sample Selection Bias in Credit Scoring Models," The Journal of the Operational Research Society, vol. 54, no. 8, pp. 822-832, 2004.
[23] A. C. B. Garcia, M. G. P. Garcia and R. Rigobon, "Algorithmic discrimination in the credit domain: what do we know about it?," AI & SOCIETY, vol. 39, p. 2059–2098, 2023.
[24] A. Bouteska, M. Harasheh and M. Z. Abedin, "Revisiting overconfidence in investment decision-making: Further evidence from the U.S. market," Research in International Business and Finance, vol. 66, 2023.
[25] M. A. Wibowo, N. K. Indrawati and S. Aisjah, "The impact of overconfidence and herding bias on stock investment decisions mediated by risk perception," International Journal of Research in Business and Social Science, vol. 12, no. 5, pp. 174-184.

[26] B. O. Adelakun, B. O. Antwi, D. T. Fatogun and O. P. Olaiya, "Enhancing audit accuracy: The role of AI in detecting financial anomalies and fraud," Finance & Accounting Research Journal, vol. 6, no. 6, 2024.
[27] J. Pombal, A. F. Cruz, J. Bravo, P. Saleiro, M. A. Figueiredo and P. Bizarro, "Understanding Unfairness in Fraud Detection through Model and Data Bias Interactions," KDD'22 Workshop on Machine Learning in Finance, 2022.
[28] A. Chouldechova and A. Roth, "A snapshot of the frontiers of fairness in machine learning," Communications of the ACM, vol. 63, no. 5, pp. 82-89, 2020.
[29] T. Panch, H. Mattie and R. Atun, "Artificial intelligence and algorithmic bias: Implications for health systems," Journal of Global Health, vol. 9, no. 2, 2019.
[30] J. Silberg and J. Manyika, "Notes from the AI frontier: Tackling bias in AI (and in humans)," 06 2019. [Online]. Available: https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/tackling%20bias%20in%20artificial%20intelligence%20and%20in%20humans/mgi-tackling-bias-in-ai-june-2019.ashx. [Accessed 11 2024].
[31] D. Gough, S. Oliver and J. Thomas, An Introduction to Systematic Reviews, SAGE Publications Ltd, 2017.
[32] M. Petticrew and H. Roberts, Systematic Reviews in the Social Sciences: A Practical Guide, 1st Edition ed., Wiley-Blackwell, 2006.
[33] S. Ness, T. R. Xuan and O. O. Oguntibeju, "Influence of AI: Robotics in Healthcare," Asian Journal of Research in Computer Science, vol. 17, no. 5, pp. 222-237, 2024.
[34] "Algorithmic Bias? An Empirical Study into Apparent Gender-Based Discrimination in the Display of STEM Career Ads," Social Science Research Network (SSRN), p. 40, 12 03 2018.