International Journal of Business Marketing and Management (IJBMM)
Volume 11 Issue 1 Jan-Feb 2026, P.P. 04-15
ISSN: 2456-4559
www.ijbmm.com

Managing Bias and Prejudice in AI-Driven Enterprise Systems: Implications for Governance, Risk, Controls, and Ethical Deployment in Business and Economic Environments

Emmanuel Opara[1] , Dieli Onochie Jude[2] , Gbolahan Solomon Osho[3] , Sudhir Tandon[4]
College of Business, Prairie View A&M University
Prairie View, Texas

Abstract: *In an era in which artificial intelligence (AI) increasingly influences sectors, this study focuses on identifying and mitigating biases in AI-driven systems, particularly in healthcare and other enterprise environments. By examining the origins of bias, including unrepresentative datasets and algorithmic errors, the research aims to provide a comprehensive analysis across industries employing AI for decision-making. The objectives include uncovering sector-specific biases, conducting an interdisciplinary review of these issues, and formulating strategies to minimize their impact. Recommendations will be made on best practices for enterprise systems to ensure equitable and practical AI applications.*

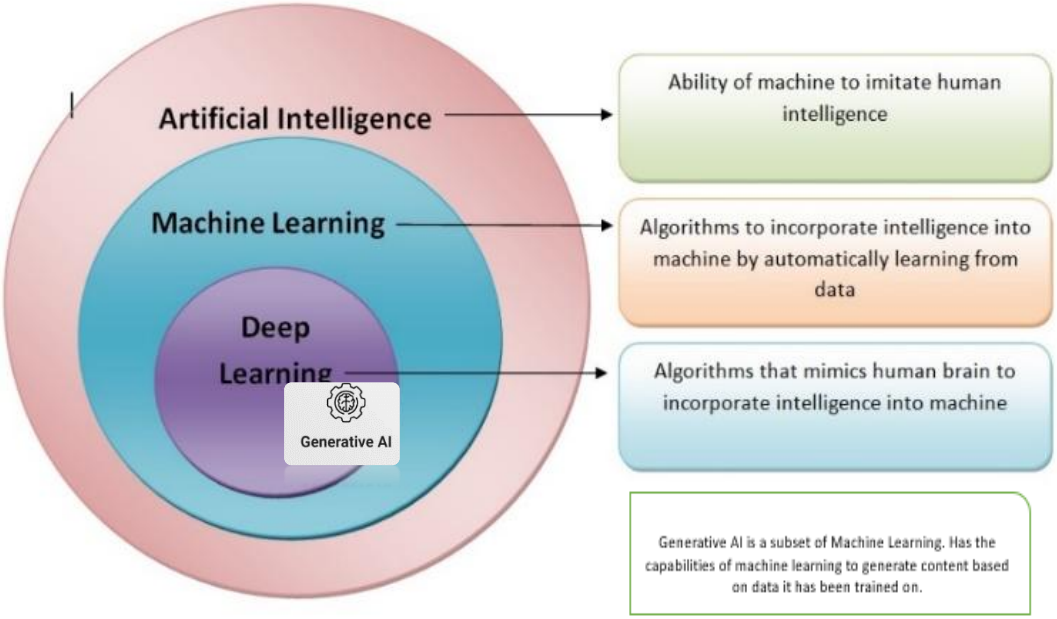
Keywords: *Artificial Intelligence Bias; Algorithmic Fairness; Enterprise Systems; Machine Learning Governance; Data Ethics; Responsible AI; Bias Mitigation Strategies.*

I. Introduction

Artificial Intelligence (AI) bias, manifesting as algorithmic or machine learning bias, reflects the systematic skewing of results by algorithms that replicate human prejudices. This phenomenon occurs when AI systems produce outcomes that inadvertently reinforce societal stereotypes, particularly against marginalized groups, echoing prejudices based on race, gender, and other societal divisions. As highlighted by the Artificial Intelligence Index Report 2023, AI bias is a critical concern when it perpetuates stereotypes, leading to discriminatory practices against specific communities. ML, as a subfield of AI, involves training machines to learn from data without being explicitly programmed. ML algorithms can find patterns and trends in data and utilize them to make predictions and decisions. As an advanced AI, ML is used to build predictive models, classify data, and identify patterns, which are indispensable tools for many AI applications. DL technology uses artificial neural networks to perform sophisticated computations on large datasets.

[1] The following graduate students collaborated with Professor Emmanuel Opara on the preparation of the initial draft of this manuscript: Barghavi Krishnan (College of Juvenile Justice and Psychology), Bernard Nyarko (College of Engineering), and Ugoaghalam Uche James (College of Engineering), Prairie View A&M University.

Figure 1. Structural Differences Between Deep Learning and Machine Learning Models



Source (46) Modified DL versus ML: What Marketers Need to Know (hubspot.com)

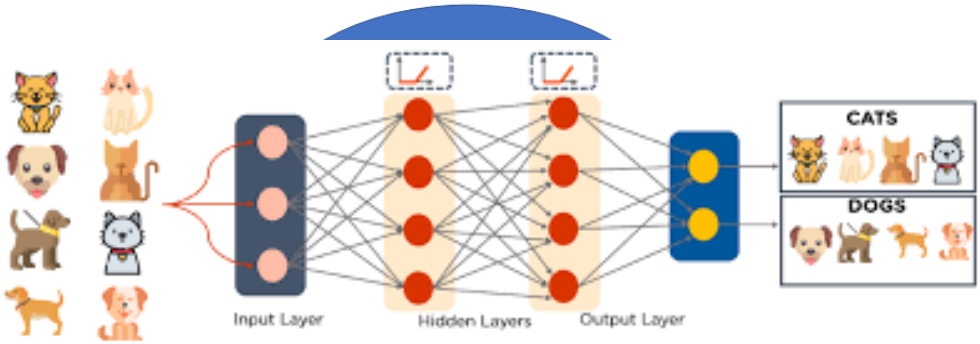
This technology leverages the structure and function of the human brain and can train machines by having them learn from examples. This technology is used by major industries such as healthcare, e-commerce, entertainment, and advertising. There are four types of machine learning algorithms used to predict analytics for enterprise systems. These are supervised, unsupervised, semi-supervised, and reinforcement machine learning algorithms. The supervised machine learning algorithm is based on accurately labeled data and oversight from a researcher. The process involves the algorithm feeding data into the system, which includes input and desired output as defined by the researcher (1). The system then learns from the relationship between the input and output, training data to build the model (4). The model maps input data to the desired output and is trained until the model reaches a high level of accuracy (9).

A researcher does not directly control unsupervised machine learning algorithms trained on unlabeled datasets. This algorithm is used to identify patterns, trends, or groupings in a dataset where these elements are unknown (39,41). The third algorithm is Semi-supervised learning, a combination of the supervised and unsupervised approaches (20). The key here is that it is used with datasets that have only a portion of the data accurately labeled.

The fourth algorithm is Reinforcement machine learning, which allows a system to learn and improve the performance of a function through a trial-and-error process. Over time, the model will learn to find the best solution to the issue under study in a specific environment. Successful actions are rewarded and reinforced through a feedback loop.

Figure 2. Supervised Machine Learning Algorithm

Source: Modified from Simplilearn.com (45)



The "Sources and Types of AI Bias" section has been expanded to detail the intricacies of AI bias, emphasizing design, contextual, and application biases. This includes how biases in AI design can stem from initial algorithm frameworks, affecting outcomes through predetermined preferences or oversights. Contextual bias reflects the environment in which the AI operates, potentially skewing data interpretation based on local norms or cultural assumptions. Application bias arises when AI tools are deployed in specific fields, such as marketing, where their use might inadvertently favor specific demographics over others. The discussion extends to implicit biases, uncovered through methods such as implicit association tests and field experiments, demonstrating that, despite its objective facade, AI can perpetuate human prejudices in its decision-making. This comprehensive view underscores the necessity for a multifaceted approach to mitigating AI biases, integrating awareness at every phase from design to deployment.

Logistic regression methods for estimating discrete outcomes from a given set of independent variables can lead to bias. This method helps researchers predict the likelihood of an event happening by fitting data to a logit function. As it indicates the probability, its output value must lie between 0 and 1. However, bias in logistic regression methods may lead to underperformance when there are numerous or non-linear decision boundaries. Since these algorithms are not flexible, the results might not capture the complex relationship in the capture mode. Bias could also occur in the classification of regression algorithms. This type of bias occurs when the algorithm attempts to label an input as two distinct classes (binary classification) or when it selects from more than two classes (multiclass classification). Bias could occur when unconstrained, individual trees overfit. Studies (3,5,8,16,34,35), among others, have shown evidence that biases could occur due to any of these:

- Invalid sample selection and training data input.
- Poor data preparation, preprocessing, and stereotypes prejudice.
- False positive outcomes and partial training data.
- Datasets overtrained by exceeding expected results.
- Measurement and selection of solid classifier training from each class.
- Inadequate Big Data classification.
- Insufficient computation time.
- Poor datasets, bad models, weak algorithms, and human error

Any of these could result in false-positive, biased reports. AI systems are only as good as the data they are fed. So, what if that dataset has its own biases? Since the technology is not stable at this time, data input errors may occur because of biased input. AI bias may also occur when the underlying prejudice in the dataset leads to race- and gender-based discrimination, and the weights used to train the algorithms result in false-positive outcomes. The consequences of the introduction of such bias in AI algorithms usually come in the form of discrimination against minorities and underrepresented members of society.

Employees of Amazon sued the corporation and filed discrimination charges on the grounds of race and gender against the giant technology company (32). Amazon Corporation's algorithm discriminated against women in its employment practices. The technology evaluated applicants based on their suitability for various positions and roles. The AI technology learned over time to identify whether someone was suitable for a position at the company by analyzing resumes from previous candidates. The effect of that was bias against women in the process. Due to the underrepresentation of women in technical roles, Amazon's AI system preferred male candidates to female candidates. The system's bias penalized resumes from female applicants, assigning lower scores.

In another case, AI underestimated the needs of Black patients in the healthcare system. AI used to predict which patients needed additional medical help in the diagnosis and evaluation process was biased. The technology analyzed patients' healthcare cost history and assumed that higher costs indicate greater healthcare need. However, this assumption was discriminatory and created a false equivalence with the actual result. The AI algorithm did not consider the myriad of ways black and white patients pay for healthcare services. Black patients received lower risk scores compared to their white counterparts and, as a result, did not qualify for additional care as compared to their white counterparts with similar needs (5, 22).

This study aims to unravel the complexities of AI bias across sectors, with a particular focus on healthcare, to understand its implications for decision-making processes. Through a detailed examination of how biases originate from unrepresentative datasets to algorithmic inaccuracies, the research seeks to provide an

interdisciplinary analysis that not only identifies these biases but also proposes methodologies for their mitigation. By incorporating examples, such as Amazon's employment algorithm discriminating against women and healthcare algorithms underestimating the needs of black patients, the introduction sets the stage for a thorough exploration of AI biases in the workplace. The study is motivated by the urgent need to develop effective strategies to counteract these biases, ensuring that AI technologies serve all segments of society equitably. It concludes with recommendations for best practices in enterprise systems, aiming to foster an environment in which AI facilitates fair and unbiased decision-making (Opara et al., 2026).

II. Literature Review

Recent investigations [1,4,14,25] into artificial intelligence (AI) have highlighted its ability to emulate human intelligence, with robots performing tasks historically assigned to humans in corporate environments. Despite these advancements, the researchers identified inherent biases within AI's operational framework in the digital economy. Their qualitative analysis underscores the widespread effects of these biases, particularly highlighting gender and racial prejudices across various sectors. This revelation underscores the urgent need to implement responsible AI practices to mitigate such risks. The literature further elaborates on the diverse nature of these biases. It stresses the crucial role of policymakers, managers, and employees in understanding and addressing the potential adverse outcomes of AI applications, especially the issue of false positives in industrial settings.

Further studies [3,6,9,24,39] delve into the specific domain of marketing tools, outlining a framework for identifying sources of algorithmic bias rooted in the micro-foundations of dynamic capability. Through engaging discussions with machine learning (ML) professionals, the research delineates three primary dimensions: design bias, contextual bias, and application bias, alongside ten subdimensions, including model, data, method, and cultural biases. This comprehensive framework aims to foster the development of dynamic algorithm management systems to reduce algorithmic bias in ML-driven marketing decisions.

The growing integration of AI into customer service, marketing, and sales technologies [7,8,14,15,17] underscores its increasing presence and anticipated expansion. AI's role in enhancing business operations and consumer customization options underscores its potential to significantly bolster competitive advantage. However, this integration has also spurred discussions about how human cognitive biases are mirrored or amplified in AI-driven sales predictions and outcomes.

Reports [29,30,31] on the adoption of facial recognition payment (FRP) services in China reveal a nuanced landscape where technological advancement meets societal resistance. Despite FRP systems offering distinct advantages over traditional payment methods, legal and privacy concerns have catalyzed a critical examination of the technology's impact on societal norms and individual privacy, highlighting resistance among various Chinese demographics.

Moreover, instances of image recognition models failing to identify individuals of color [43,44] accurately have brought to light significant ethical concerns. These failures, attributed to a lack of diversity in training datasets, not only perpetuate societal prejudices but also underline the profound social consequences of biased AI technologies. The critique extends to major tech entities such as Flickr, Hewlett-Packard, and Google, emphasizing the need for institutional measures to ensure that AI technologies are developed and deployed in ways that fairly represent and serve all sections of society.

Emerging research [32,33,34] underscores the transformative impact of artificial intelligence (AI) on societal functions and personal lifestyles, highlighting its role in enhancing decision-making through data integration and analysis. Despite AI's significant benefits, concerns are raised about its potential downsides and unintended effects. Additionally, investigations [23, 40,41] into AI's application within e-business reveal its potential to inadvertently perpetuate biases, particularly affecting minority groups, and question the fairness of algorithms used by companies like Uber, Lyft, and Via in fare calculations. These findings, corroborated by a comprehensive analysis involving ACS data, show how demographic factors influence algorithmic pricing models. Further studies [28,39,40] demonstrate algorithms' ability to detect patterns within large datasets for predictive outcomes, yet highlight the risk of inheriting biases from these datasets, complicating the perception of algorithmic neutrality. The discourse [26, 41] around AI and algorithmic systems critiques their role in reinforcing social inequalities, suggesting a nuanced view of bias as both a challenge and an opportunity for fostering equitable technological advancements.

III. Methodology

Our investigation used a systematic review of secondary data, drawing on the Scopus database to identify a wide array of scholarly articles on AI bias in enterprise systems. This approach allowed us to compile a diverse set of findings without primary data, leveraging Scopus for its extensive repository and analytical capabilities. Keywords such as "Artificial Intelligence", "artificial intelligence bias", and "Algorithm bias" were used alongside Boolean operators to refine our search, yielding 742 relevant articles. Systematic inclusion and exclusion criteria were applied to distill the data for further analysis. The items were then filtered to publications from 2019 to 2024, yielding 678 publications. We filtered by subject areas and document types as listed below, and the number of items retrieved is provided.

Table 1. Filter by Subject Area: Sort by Subject Area

Computer Science 323	Health Professions 23
Medicine 184	Environmental Science 21
Social Sciences 115	Economics, Econometrics and Finance 15
Engineering 108	Multidisciplinary 15
Business, Management and Accounting 54	Neuroscience 14
Mathematics 54	Chemistry 12
Decision Sciences 48	Materials Science 11
Arts and Humanities 43	Chemical Engineering 10
Earth and Planetary Sciences 32	Agricultural and Biological Sciences 8
Biochemistry, Genetics and Molecular Biology 26	Nursing 8
Physics and Astronomy 26	Energy 6
Psychology 26	Pharmacology, Toxicology, and Pharmaceutics 4

Table 1 provides a detailed breakdown of the subject areas represented in the scholarly literature on AI bias within enterprise systems, based on the Scopus database classification. The distribution reveals that Computer Science dominates the field with 323 indexed publications, reflecting the technical foundation of AI research and the centrality of algorithm development, data structuring, and system architecture in understanding how bias emerges. Substantial representation in Medicine (184) and the Social Sciences (115) highlights the broader societal and ethical implications of biased AI, especially in sensitive sectors such as healthcare, public administration, and social services, where decision-making algorithms can directly impact human well-being. Additional concentration in Engineering (108) and Business, Management & Accounting (54) indicates that organizations are increasingly engaging with AI at operational and strategic levels, raising concerns about responsible deployment, governance, and risk oversight within enterprise environments.

Table 2. Filter by Document Type

Article 345
Conference paper 185
Review 51
Book chapter 28
Note 23
Letter 20
Editorial 19
Conference review 2
Book 1

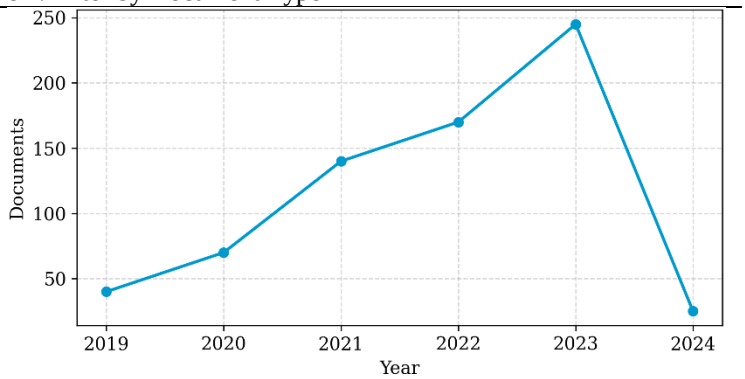


Table 2 categorizes the Scopus search results by document type, offering insight into how knowledge on AI bias in enterprise systems is created and disseminated. The data show that journal articles (345) constitute the largest share of scholarly output, indicating that peer-reviewed empirical and conceptual research remains the dominant vehicle for advancing academic understanding of AI governance, fairness, and algorithmic risk. The strong presence of conference papers (185) further underscores the field's fast-evolving nature; conferences in computer

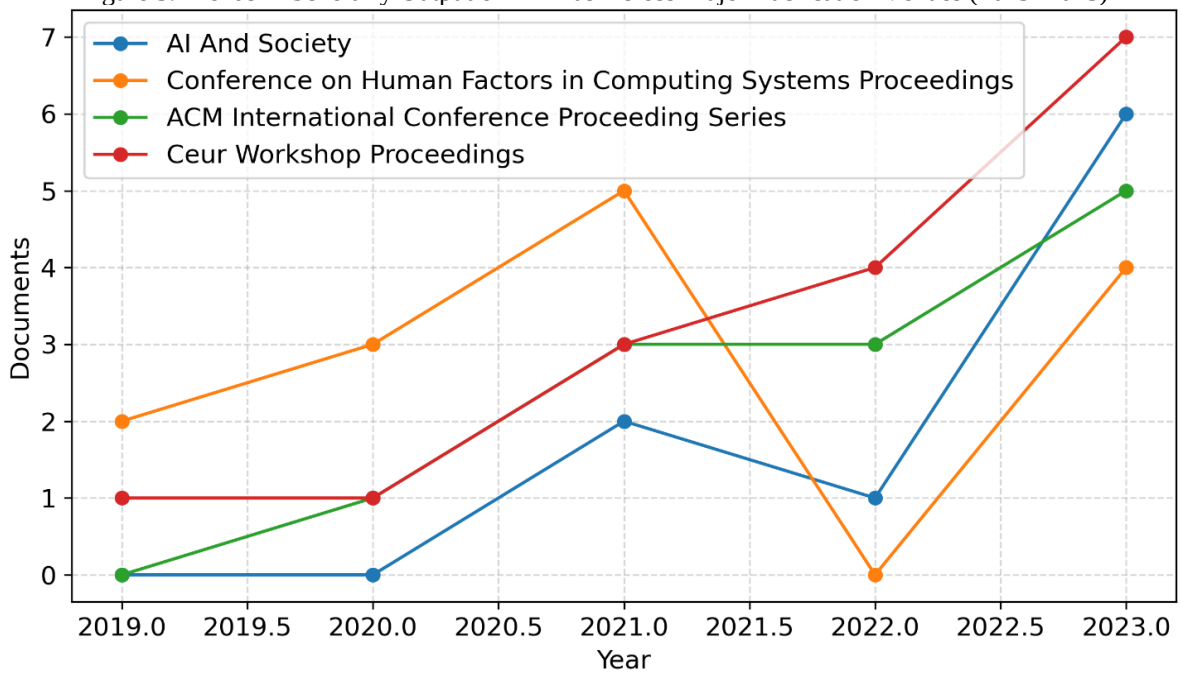
science, engineering, and information systems often serve as early platforms for presenting cutting-edge models, auditing frameworks, and technical approaches before journal publication. Meanwhile, review papers (51) demonstrate that sufficient scholarly maturity has been reached for meta-level synthesis, comparative analysis, and thematic mapping of AI bias literature across disciplines, particularly as organizations and researchers seek clearer taxonomies and mitigation frameworks.

The trend shown in Table 2 also indicates a notable increase in research output on AI bias and related topics over the six years analyzed, with publication counts rising steadily from 2019 through 2023, followed by a sharp decline in 2024. This upward trajectory reflects growing scholarly and societal attention to the ethical and technical challenges posed by algorithmic decision-making across sectors such as healthcare, finance, and public administration. The peak observed in 2023 suggests that the field reached a heightened point of academic engagement, likely influenced by policy developments, industry controversies, and heightened awareness of fairness and accountability in AI systems. The decline in 2024 is not necessarily indicative of reduced interest. Still, it may instead reflect incomplete indexing for the current year, a typical pattern in bibliometric analyses conducted before the close of a calendar year. Overall, the publication pattern demonstrates the emergence and maturation of AI bias as a significant interdisciplinary research domain.

The trend illustrated in Figure 3 shows a steady increase in publications on AI bias across major scholarly outlets from 2019 to 2023. Early publication activity in 2019 and 2020 remained relatively modest, with most venues producing fewer than three documents per year. This initial pattern likely reflects emerging scholarly awareness of algorithmic fairness, responsible AI practices, and the need to empirically evaluate AI tools in applied settings. By 2021, notable growth occurred in multiple venues, indicating the acceleration of academic interest and the expanding relevance of AI ethics, governance, and societal impacts within both technical and interdisciplinary research communities.

Publication growth continued into 2022 and 2023, with a sharper increase in output across nearly all tracked venues. Notably, CEUR Workshop Proceedings and *AI and Society* recorded substantial expansion in 2023, suggesting that conferences and interdisciplinary journals have become particularly active channels for disseminating research on AI bias. This growth may correspond with increased global attention to high-profile incidents of algorithmic discrimination, industry adoption of AI governance frameworks, and policy interventions advocating for transparency and accountability in AI systems (Tamez & Osho, 2025). The spike in conference proceedings also suggests that the field is maturing rapidly, with researchers presenting findings in early-stage venues before journal submission.

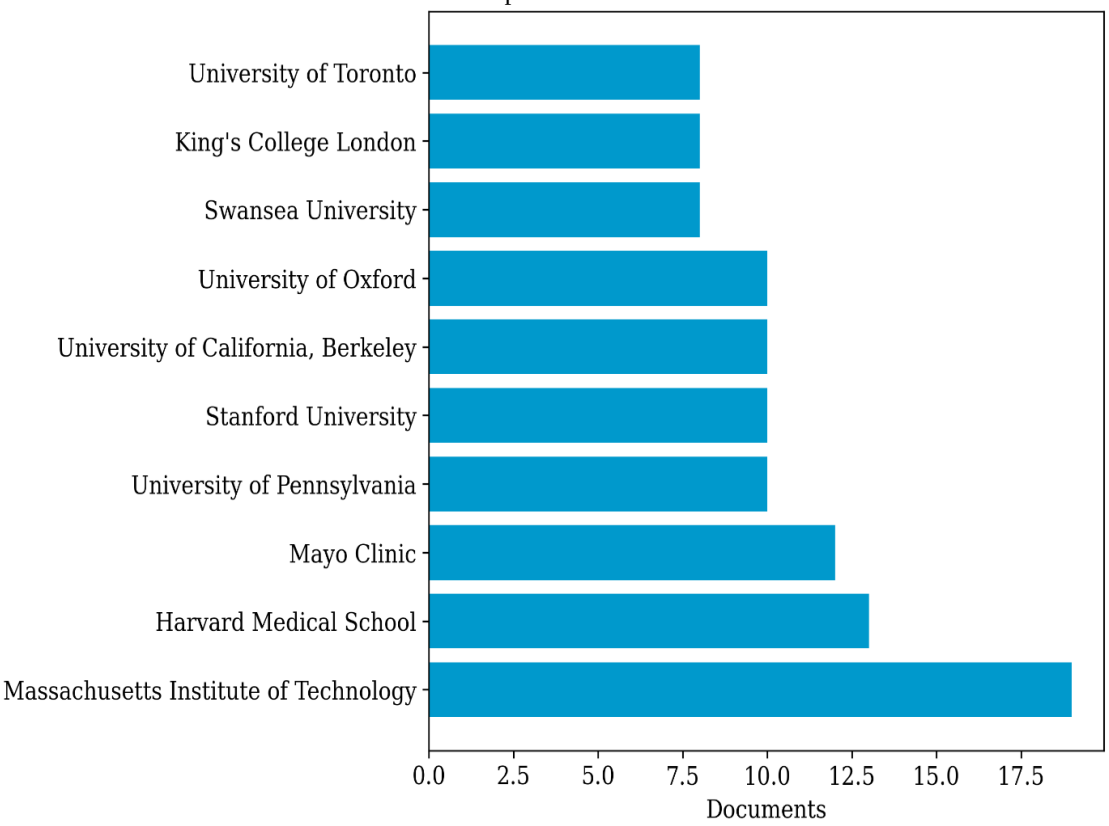
Figure 3. Trends in Scholarly Output on AI Bias Across Major Publication Venues (2019–2023)



Collectively, the distribution of publication activity underscores the dynamic and evolving nature of research on AI bias. The observed upward trajectory reflects not only heightened scholarly concern but also broader institutional and societal recognition of the risks associated with unregulated algorithmic decision-making. As research ecosystems continue to diversify, contributions from fields including computer science, human–computer interaction, ethics, and social sciences are increasingly converging. This multidisciplinary engagement supports a more comprehensive understanding of AI bias and informs the development of mitigation strategies, policy recommendations, and industry standards aligned with responsible and equitable AI deployment.

The data presented in Figure 4 suggests the institutional landscape of research productivity in the domain of artificial intelligence (AI) bias. Notably, the Massachusetts Institute of Technology emerges as the leading contributor, publishing the most on the topic. This level of engagement is consistent with the institution's broader leadership in AI, computational science, and technology policy. Harvard Medical School and Mayo Clinic also demonstrate strong publication output, suggesting that concerns regarding AI bias extend beyond computer science into applied domains such as healthcare, where algorithmic decision-making has direct implications for diagnosis, treatment, and patient equity.

Figure 4. Comparison of publication output among the top contributing institutions in AI bias research from the Scopus database.



In addition to institutions at the forefront of medical research, several academically oriented universities contribute substantially to the discourse on AI ethics and fairness. Universities such as Stanford, the University of Pennsylvania, the University of California, Berkeley, and the University of Oxford demonstrate comparable research activity, emphasizing the interdisciplinary nature of AI bias inquiry. These institutions are hubs for both technical innovation and ethical analysis, indicating that the study of bias in machine learning systems requires collaboration across computer science, information systems, bioethics, law, and the social sciences. Their sustained production of scholarly work reinforces the need to integrate responsible AI frameworks into academic research environments.

The distribution of research activity across affiliations underscores the global relevance of AI bias as a research concern. While the top contributors are predominantly U.S.-based institutions, notable representation from Swansea University, King's College London, and the University of Toronto reflects international engagement and

recognition of algorithmic fairness as a cross-border issue. The diversity of contributing institutions also suggests that AI bias is not confined to a single disciplinary perspective but is instead shaped by medical, computational, social, and regulatory contexts. Overall, the institutional distribution captured in Figure X reveals an active and expanding research community committed to addressing the ethical and practical implications of AI in society.

IV. Results and Discussion

The findings from Case Studies of Biases in AI Systems: Contrary to the assumption of AI's objectivity, it has been demonstrated to exhibit cognitive biases and to suffer from incomplete data sets, leading to skewed judgments and decision-making processes [11,12]. The essentiality of inputting clean, unbiased data into AI systems is underscored by instances of AI reflecting human prejudices [2,18]. Without intervention, these biases in AI programming could persist, highlighting the need to implement best practices within enterprise systems to mitigate such errors and foster technological excellence.

Specific Case Studies

Amazon: Despite its status as a global e-commerce leader, the Company encountered biases in its AI-driven recruitment system that favored male applicants over female applicants for STEM roles, reflecting an underrepresentation of women in these fields [19,22,42,27].

USA Healthcare Industry: AI algorithms deployed in healthcare were found to exhibit racial biases in patient care predictions, inaccurately assessing the healthcare needs of black patients compared to their white counterparts [31,5,36].

Robotic Facial Recognition: Studies have revealed biases in facial recognition algorithms, leading robots to incorrectly categorize individuals by gender and ethnicity, resulting in discriminatory classifications [16,37,13].

Canvas Learning Management System: The use of AI in educational technologies like Canvas LMS has raised concerns about the potential for bias, affecting the inclusivity and diversity of the academic content and misinterpreting student emotions [20,37,21,16].

The results of the bibliometric analysis reveal a steadily growing research interest in artificial intelligence (AI) bias across interdisciplinary publication venues over the five years examined. The distribution of documents by source shows that academic dialogues on algorithmic fairness have transitioned from isolated conference contributions to broader journal and workshop dissemination. Early years (2019–2020) show relatively modest publication counts, but beginning in 2021, a noticeable upward trajectory is observed across several major venues. CEUR Workshop Proceedings and *AI and Society* exhibited robust growth by 2023, reflecting increased engagement with the technical and socio-ethical dimensions of AI bias. The spike in workshop and conference papers suggests that the field is characterized by rapid knowledge diffusion, with emerging methods and conceptual debates often presented in early-stage formats before transitioning to journal outlets. This growth pattern aligns with previous literature, which notes that responsible AI research remains in a formative yet rapidly maturing phase, fueled by concerns about transparency, discrimination, and ethical governance.

The findings also indicate that research output on AI bias is not evenly distributed across publication types. Journal articles remain the dominant medium, demonstrating that the field has established sufficient empirical, theoretical, and methodological depth to support peer-reviewed scholarship. Conference proceedings contribute significantly to early dissemination, highlighting their role in driving technical innovation, benchmarking studies, and methodological refinement. The presence of review articles and book chapters further illustrates that the research community has entered a stage of conceptual consolidation, enabling thematic synthesis and comparative evaluations across sectors such as healthcare, finance, education, and public policy. These patterns collectively support the view that AI bias research has evolved from a niche technical concern into a broader socio-technical inquiry with increasing policy relevance.

Institutional analysis provides additional insight into the global research landscape. The affiliation data show that leading contributors to AI bias scholarship are predominantly North American institutions, with the Massachusetts Institute of Technology, Harvard Medical School, and Mayo Clinic yielding the highest document counts. The strong representation of medical and clinical institutions underscores that AI bias is increasingly recognized as a critical issue in healthcare, particularly in diagnostic modeling, risk prediction, and resource allocation. Institutions such as Stanford University, the University of Pennsylvania, and the University of California,

Berkeley also feature prominently, reflecting the intersection of machine learning innovation with legal, ethical, and societal dimensions of AI deployment. The presence of the University of Oxford, Swansea University, King's College London, and the University of Toronto highlights international engagement and cross-border recognition of AI fairness as a global research priority.

Taken together, these results suggest that AI bias research has transitioned from an emergent topic to a recognized field of inquiry with diverse academic, clinical, and policy stakeholders. The geographic concentration of leading institutions reflects both resource availability and the intensity of AI integration in advanced healthcare and technological ecosystems. At the same time, the growing diversity of publication venues indicates expanding disciplinary participation, including computer science, medicine, ethics, and the social sciences. This multidimensional engagement reinforces the need for holistic approaches to algorithmic governance that integrate technical debiasing methods with regulatory frameworks, ethical guidelines, and organizational accountability structures. Future work should continue to examine how research output translates into industry standards, clinical protocols, and public policy, as well as how participation can be broadened to include institutions from regions heavily affected by algorithmic decision-making but underrepresented in current authorship networks.

V. Conclusion

The present study examined the evolving scholarly landscape on artificial intelligence (AI) bias through a structured bibliometric review of publication trends, document types, and institutional affiliations. The findings underscore the rapid growth and interdisciplinary expansion of research devoted to understanding, measuring, and mitigating bias in AI-driven systems. While early contributions appeared primarily within computer science and technical conference venues, the broader trajectory revealed a shift toward journal publications, review studies, and cross-sector analyses involving healthcare, business, and social sciences. This shift illustrates that AI bias is no longer perceived solely as an algorithmic or data-preprocessing issue but rather as a complex sociotechnical problem with tangible implications for equity, human rights, and societal trust in automated decision-making technologies. The increasing involvement of clinical and medical research institutions further highlights the practical urgency of addressing algorithmic discrimination. As healthcare systems integrate machine learning models for diagnostic support, triage, and patient-risk prediction, the consequences of biased AI outputs become both immediate and ethically charged. Institutions such as Harvard Medical School and Mayo Clinic, identified as leading contributors in the institutional analysis, demonstrate that AI fairness has transitioned from a theoretical concern into a domain of applied biomedical policy and clinical risk management. At the same time, the continued presence of leading technical universities such as the Massachusetts Institute of Technology, Stanford University, and the University of California, Berkeley reflects the ongoing need for advances in explainable AI, debiasing algorithms, model validation, and responsible data governance.

One of the most important implications of the present review is the recognition that addressing AI bias requires sustained collaboration across disciplinary lines. Technical solutions alone, such as adversarial debiasing, fairness-aware training, or balanced dataset construction, cannot fully resolve systemic inequities embedded in the contexts in which AI systems are deployed. Ethical guidelines, regulatory frameworks, organizational governance policies, and end-user education must complement computational methods to ensure that AI systems do not reinforce existing forms of discrimination or introduce new ones. Additionally, the geographic clustering of research output in North American and European institutions raises questions regarding representational equity. Many populations that are most likely to be affected by biased AI, particularly in low-resource healthcare environments, financial inclusion contexts, and public administration systems, remain underrepresented in current research authorship and training datasets. Future work must therefore expand participation and data representation to avoid perpetuating global disparities in algorithmic decision-making.

The findings of this study also point to several promising directions for future research. First, longitudinal studies of policy adoption and standardization efforts can clarify how principles of responsible AI are operationalized in real-world organizational settings. Second, cross-sector comparative analyses may reveal how bias manifests differently across domains such as employment, finance, transportation, and healthcare. Finally, deeper engagement with civil society, ethicists, and public stakeholders will be essential for shaping AI governance structures that align with democratic values and public accountability. As AI technologies continue to influence increasingly critical aspects of daily life, questions of fairness, transparency, and social impact will remain central to both academic inquiry and policy debates.

Finally, this review highlights that AI bias research has entered a phase of accelerated growth, interdisciplinary collaboration, and practical urgency. The scholarly community, industry practitioners, healthcare institutions, and

policymakers share a collective responsibility to ensure that AI systems are designed and deployed in ways that promote fairness, protect vulnerable populations, and uphold ethical standards. By expanding the scope of research participation, strengthening methodological rigor, and aligning technological innovation with societal needs, the field is well-positioned to develop AI systems that are not only intelligent but also just, accountable, and reflectively integrated into the fabric of human decision-making.

Research Implications

Artificial intelligence's effectiveness hinges on the utilization of clean, accurate data for training algorithms, especially in supervised learning environments. Researchers bear the responsibility of employing datasets that accurately reflect society's diversity, ensuring AI systems do not favor specific outcomes due to biased data inputs. It is imperative for organizations developing AI to establish and enforce policies that prohibit the use of biased data and to adhere to industry standards throughout the development process. This includes rigorous user acceptance testing and addressing biases through continuous monitoring and review of AI models. The selection of algorithms should prioritize precision and accuracy, with ongoing assessments to prevent overfitting and ensure real-world applicability. Fairness and unbiased predictions across all demographics are crucial, requiring a comprehensive approach to data handling, including meticulous data cleaning and structuring to improve model performance. Balancing the number of epochs in model training is essential to avoid underfitting or overfitting, optimize computation time, and ensure the developed models are free of bias.

Future Research

Artificial intelligence and its benefits cannot be overstated, as it has replaced humans in performing tasks once considered impossible or repetitive. However, how to develop AI models for use has been called into question, as there are a lot of biases that emanate from its development in terms of the kind of data used in training, prejudice on the part of researchers, not asking the right questions, and the lack of policies and guidelines to guide developers under proper supervision in developing AI Algorithms. A lot of research needs to be done in these areas to change the myriad of adverse reports associated with the development of AI and the consequences of introducing biases in the development of AI. More often than not, when biases are introduced into AI, steps must be taken to rectify them so they do not negatively affect members of society and positively impact how we live.

References

[1]. Opara, E. U., Jude, D. O., Osho, G. S., Stiff, C., & Gordon, K. (2026). Generative artificial intelligence as a catalyst for enterprise transformation: Evidence from revenue growth, productivity, and industry disruption. *International Journal of Artificial Intelligence (AI) in Business and Management Research*, *3*(1).

[2]. Accenture (2021). "Responsible AI: From principles to practice"https://www.accenture.com/us-en/insights/artificial-intelligence/responsible-ai-principles-practice Retrieved from 14h August 2022.

[3]. Akter et al., 2022 S. Akter, Y. K. Dwivedi, S. Sajib, K. Biswas, R.J. Bandara, K. Michael Algorithmic bias in machine learning-based marketing models Journal of Business Research, 144 (2022), pp. 201-216, 10.1016/j.jbusres.2022.01.083 View PDFView articleView in ScopusGoogle Scholar

[4]. Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. Journal of Business Research, 144, 201–216. https://doi.org/10.1016/j.jbusres.2022.01.083

[5]. Arrieta et al., 2020 A. B. Arrieta, N. Díaz-Rodríguez, J. Del er, A. Bennetot, S. Tabik, A. Barbado, *et al.* Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI Information fusion, 58 (2020), pp. 82-115

[6]. Beattie and Johnson, 2012 G. Beattie, P. Johnson Possible unconscious bias in recruitment and promotion and the need to promote equality Perspectives: Policy and Practice in Higher Education, 16 (1) (2012), pp. 7-13, 10.1080/13603108.2011.611833 View PDF

[7]. Britt, 2020. Brit, P. (2020). "Overcoming Bias Requires an AI Reboot". Speech Technology; Medford Vol. 25, Iss. 2, (Spring 2020): 14–17.

[8]. Britt, 2021 Brit, P. (2021). Tips for battling bias in AI-based personalization https://www.destinationcrm.com/Articles/Editorial/Magazine-Features/Tips-for-Battling-Bias-in-AI-Based-Personalization-147143.aspx Retrieved from 21st February 2022.

[9]. Brown, 2021 Brow, A. (2021). "The AI-bias problem and how Fintech's should be fighting it: A deep-dive with Sam Farao",https://www.forbes.com/sites/anniebrown/2021/09/29/the-ai-bias-problem-and-how-fintechs-should-be-fighting-it-a-deep-dive-with-sam-farao/?sh=38b226492129 Retrieved from 28th February 2022.

[10]. Cao, Duan, Edwards and Dwivedi, 2021 G. Cao, Y. Duan, J.S. Edwards, Y.K. Dwivedi Understanding managers' attitudes and behavioral intentions towards using artificial intelligence for organizational decision-making Technovation, 106 (2021), Article 102312, 10.1016/j.technovation.2021.102312

[11]. Casualty Actuarial Society (2022) Casualty Actuarial Society (2022)., "Approaches to Address Racial Bias in Financial Services: Lessons for the Insurance Industry", https://www.casact.org/sites/default/files/2022-03/Research-Paper_Approaches-to-Address-Racial-Bias_0.pdf. Retrieved from 29 December 2022.

[12]. Cheng, X., Lin, X., Shen, X. L., Zarifis, A., & Mou, J. (2022). The dark sides of AI. Electronic Markets, 32(1), 11–15. https://doi.org/10.1007/s12525-022-00531-5.

[13]. Chintalapudi, N., Battineni, G., & Amenta, F. (2021). Sentimental analysis of COVID-19 tweets using deep learning models. Infectious Disease Reports, 13(2), 329–339. https://doi.org/10.3390/idr13020032.

[14]. Dicki, J. (2021). "The ethics dilemma of AI for sales" https://www.destinationcrm.com/Articles/Columns-Departments/Reality-Check/The-Ethics-Dilemma-of-AI-forSales-149181.aspx.

[15]. Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., et al., (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, Article 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002.

[16]. Fatem, F. (2020). "Three Platforms Where AI Bias Lives", https://www.forbes.com/sites/falonfatemi/2020/04/15/three-platforms-where-ai-bias-lives/?sh=68393df3b0c1 Retrieved from 28th December 2022.

[17]. Flori Needle (2023). What is AI bias? [+ Data] (hubspot.com) published June 6, 2023

[18]. Gonzales, R. M. D., & Hargreaves, C. A. (2022). How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system. International Journal of Information Management Data Insights, 2(2), Article 100130. https://doi.org/10.1016/j.jjimei.2022.100130.

[19]. Grundner, L., & Neuhofer, B. (2021). The bright and dark sides of artificial intelligence: A futures perspective on tourist destination experiences. Journal of Destination Marketing & Management, 19, Article 100511.

[20]. Hao, K. (2021). Facebook's ad algorithms are still excluding women from seeing jobs. MIT Technology Review, 21, 2022 Retrieved January.

[21]. A nalysis & Prevention, 45, 373–381. https://doi.org/10.1016/j.aap.2011.08.004. Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. Journal of the Academy of Marketing Science, 49(1), 30–50. https://doi.org/10.1007/s11747-020-00749-9/.

[22]. Jared Council (2021). "How adobe's ethics committee helps manage AI bias, A diverse selection of voices can help companies spot potential problems", https://www.wsj.com/articles/how-adobes-ethics-committee-helps-manage-ai-bias-11620261997. Retrieved from 28th February 2022.

[23]. Jayson Del Ray (2021), Amazon hit by 5 more lawsuits from employees who allege race and gender discrimination - Vox, May 19, 2021

[24]. Kar, A. K., & Dwivedi, Y. K. (2020). Theory building with big data-driven research– Moving away from the "What" towards the "Why". International Journal of Information Management, 54, Article 102205. https://doi.org/10.1016/j.ijinfomgt.2020. 102205.

[25]. Kar, A. K., & Kushwaha, A. K. (2021). Facilitators and barriers of artificial intelligence adoption in business–insights from opinions using big data analytics. Information Systems Frontiers, 1–24. https://doi.org/10.1007/s10796-021-10219-4.

[26]. Kumar, P., Hollebeek, L. D., Kar, A. K., & Kukk, J. (2022). Charting the intellectual structure of customer experience research. Marketing Intelligence & Planning ahead-of-print. https://doi.org/10.1108/MIP-05-2022-0185.

[27]. Liu, Y., Yan, W., & Hu, B. (2021). Resistance to facial recognition payment in China: The influence of privacy–related factors. Telecommunications Policy, 45(5), Article 1021155. https://doi.org/10.1016/j.telpol.2021.102155.

[28]. Lockey, S., Gillespie, N., Holm, D., & Someh, I. A. (2021). A review of trust in artificial intelligence: Challenges, vulnerabilities and future directions. http://hdl.handle.net/10125/71284. Retrieved 14 Dec 2022.

[29]. Dr.V. P.S. International Journal of Information Management Data Insights 3 (2023) 100165 Messner, W. (2022). Improving the cross-cultural functioning of deep artificial neural networks through machine enculturation. International Journal of Information Management Data Insights, 2(2), Article 100118. https://doi.org/10.1016/j.jjimei.2022.100118.

[30]. Mikalef, P., Conboy, K., Lundström, J. E., & Popovič, A. (2022). Thinking responsibly about responsible AI and 'the dark side'of AI. European Journal of Information Systems, 1–12. https://doi.org/10.1080/0960085X.2022.2026621.

[31]. Miller, Alex, & Hosanagar, Kartik (2020). Personalized discount targeting with causal machine learning. In Proceedings of the ICIS 2020 https://aisel.aisnet.org/icis2020/digital commerce/digital commerce/7.

[32]. Morande, S. (2022). Enhancing psychosomatic health using artificial intelligence-based treatment protocol: A data science-driven approach. International Journal of Information Management Data Insights, 2(2), Article 100124. https://doi.org/10.1016/j.jjimei.2022.100124

[33]. Nagwani, N. K., & Suri, J. S. (2023). An artificial intelligence framework on software bug triaging, technological evolution, and future challenges: A review. International Journal of Information Management Data Insights, 3(1), Article 100153. https://doi.org/10.1016/j.jjimei.2022.100153.

[34]. Singh, V., Konovalova, I., & Kar, A. K. (2022). When to choose ranked area integrals versus integrated gradient for explainable artificial intelligence–a comparison of algorithms. Benchmarking: An International Journal ahead-of-print. https://doi.org/10.1108/BIJ-02-2022-0112.

[35]. Teleaba, F., Popescu, S., Olaru, M., & Pitic, D. (2021). Risks of observable and unobservable biases in artificial intelligence used to predict consumer choice. Economic, 23(56), 102–119.

[36]. The Conversation (2022). "Artificial Intelligence can discriminate on the basis of race, gender and also age". https://theconversation.com/artificial- intelligencecan-discriminate-on-the-basis-of-race-and-gender-and-also-age-173617 Retrieved from 1st March 2022.

[37]. Vimalkumar, M., Gupta, A., Sharma, D., & Dwivedi, Y. (2021). Understanding the effect that task complexity has on automation potential and opacity: Implications for algorithmic fairness. AIS Transactions on Human-Computer Interaction, 13(1), 104–129.

[38]. Wagerer, J. C., & Langer, P. F. (2020). Bias and Discrimination in artificial intelligence: Emergence and Impact in E-business. In Interdisciplinary approaches to digital transformation and innovation (pp. 256–283). IGI Global. https://doi.org/10.4018/978-1-7998-1879-3.ch011.

[39]. Wigger, K. (2020). "Researchers find racial discrimination in 'dynamic pricing' algorithms used by Uber, Lyft, and others" https://venturebeat.com/2020/06/12/researchers-find-racial-discrimination-in-dynamic-pricing-algorithms-used-by-uberlyft-and-others/ Retrieved February 28, 2022.

[40]. Wong, P. H. (2020). Democratizing algorithmic fairness. Philosophy & Technology, (2), 225–244. https://doi.org/10.1007/s13347-019-00355-w

[41]. Yen, C., & Chiang, M. C. (2021). Trust me, if you can: A study on the factors that influence consumers' purchase intention triggered by chatbots based on brain image. evidence and self-reported assessments. Behavior & Information Technology, 40(11), 1177–1194.

[42]. Zajko, M. (2022). Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates. Sociology Compass, 16(3), e12962. https://doi.org/10.1111/soc4.12962

[43]. Jason Del Rey (2021) Amazon hit by 5 more lawsuits from employees who allege race and gender discrimination - Vox

[44]. Yapo A., Weiss J. Ethical implications of bias in machine learning. Proceedings of the 51st Hawaii International Conference on System Sciences | 2018, Retrieved from https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/d062bd2a-df54-48d4-b27e-76d903b9caaa/content

[45]. Tamez, V., & Osho, G. S. (2025). AI Disruption at Scale: DeepSeek's Open-Source Model and Its Macroeconomic Impact on Markets, Labor, and Global Growth. *International Journal of Business Management and Finance Research*, *8*(3), 1-11.

[46]. 44. Dai D., Li Y., Wang Y., Bao H., Wang G., Rethinking the image feature biases exhibited by deep convolutional neural network models in image recognition. November https://ietresearch.onlinelibrary.wiley.com/doi/epdf/10.1049/cit2.12097

[47]. www.simplleam.com/tutorials/deep-learning-tutorial/deep-learning-algorithm deep learning algorithm image - Google Search

[48]. Deep Learning vs. Machine Learning: What Marketers Need to Know (hubspot.com)