IAA Journal of Biological Sciences 13(1):37-43, 2025
©IAAJOURNALS
https://doi.org/10.59298/IAAJB/2025/1313743
www.iaajournals.org
ISSN:2636-7254
IAAJB131

Addressing Bias in AI Algorithms for Health Applications

Kansiime Agnes

Department of Clinical Medicine and Dentistry Kampala International University Uganda
agnes.kansiime.2974@studwc.kiu.ac.ug

ABSTRACT

Artificial Intelligence (AI) has transformed healthcare by enhancing diagnostic accuracy, treatment personalization, and health service efficiency. However, mounting evidence reveals that AI systems can perpetuate or even amplify existing disparities related to race, gender, socioeconomic status, and geographic location. Biases often originate from imbalanced training datasets, flawed algorithm design, and unequal data collection practices. These biases have led to misdiagnoses, unequal resource allocation, and inadequate treatment recommendations, disproportionately affecting marginalized communities. This review explores the roots of algorithmic bias in healthcare AI, analyzing real-world examples such as COVID-19 triage systems and diagnostic tools that underperform in minority populations. It also examines mitigation strategies, including bias-aware data collection, algorithm design techniques, regulatory frameworks, and stakeholder engagement. Successful case studies and future research directions are presented, emphasizing fairness, transparency, and trust in computational medicine. Establishing robust, bias-resilient AI frameworks is critical to achieving equitable health outcomes and reinforcing the ethical foundations of digital health.

Keywords: AI bias, health equity, algorithmic fairness, medical AI, healthcare disparities, machine learning, ethical AI, computational medicine.

INTRODUCTION

Racial and ethnic inequities in COVID-19 mortality have been highlighted in the USA, showing that Black and Hispanic patients are more likely to die from the virus and possess higher comorbidity scores than white patients. The advent of AI/machine learning in healthcare has raised concerns about perpetuating existing biases. A deep learning system recommending COVID therapies misallocated treatments due to biases in training data, with deeper neural networks displaying even stronger biases. These biases were heritable within the networks. Gender, age, and ethnic biases were noted in AI systems for medical imaging used in surgery. Generalized adversarial networks have been utilized to create bias-protected datasets, while network ensembles have produced fairer predictions. Pre- and post-processing methods for enhancing machine learning fairness have been examined. Disparities in healthcare mirror socioeconomic inequities and cultural stereotypes, influenced by various factors including socioeconomic status, insurance, education, language, age, gender, sexual identity/orientation, and body mass index. The World Health Organization has urged global action on health inequities, emphasizing the need for health equity, ensuring fair opportunities for everyone to achieve their health potential. Reports in the early 2000s documented healthcare disparities across different racial and ethnic groups. Current trends indicate that non-Hispanic Black women experience significantly higher pregnancy-related mortality rates compared to other racial groups, driven by factors such as access to care, economic stability, and education [1, 2].

Understanding Bias in AI

Numerous instances in healthcare highlight the need for guidelines concerning bias tied to race, gender, socioeconomic status, and geography. Analysis of a large dataset designed to identify patients with complex health needs revealed that black patients were sicker than white patients at similar risk scores. However, due to a focus on total healthcare costs, the dataset failed to acknowledge this disparity, resulting in racial bias. Gender bias is also evident in medical imaging datasets used for AI in diagnosis.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Furthermore, AI skin cancer detection systems perform inconsistently across diverse backgrounds, and a diabetic retinopathy detection system misclassifies patients from lower economic statuses. A study of a clinical decision support system for sepsis in COVID-19 patients indicated that poorer individuals receive fewer necessary medications and tests, potentially skewing the system's recommendations. Moreover, biases found in healthcare algorithms resemble those in general algorithms, affecting access to jobs, housing, and loans. For instance, an algorithm trained on resumes favored white male candidates and used irrelevant patterns, creating gender bias. Other algorithms have shown favoritism towards wealthier and better-educated individuals, compounding disparities for underprivileged groups. Additionally, bias in AI algorithms on social media can lead to discriminatory ad targeting based on ethnicity and education [3, 4].

Impact of Bias in Health Applications

Machine learning algorithms have shown promising performance in healthcare problems; however, they may lead to unintended bias when making decisions involving ethnic minorities. For example, associations between Framingham risk factors and cardiovascular events differ across ethnic groups. Model-induced differences in how variables relate to outcomes also happen in prediction tasks, including breast cancer and acute kidney injury. Similar results are observed in treatment applications; though recommendations of colonoscopy screening depend on intervention assignment, they did not exceed within-group thresholds. These model-induced biases may disproportionately impact minorities. A more severe accuracy drop of a commercialized image-based model has been demonstrated for African-Americans in the skin disease detection problem, as well as for female patients in the heart disease prediction problem. Undiagnosed silent hypoxemia, a common cause of exacerbation due to COVID-19, was found to occur three times more frequently in Black compared to White patients due to the fact that dark skin responds differently to the light wavelengths in a SpO2 monitoring device. Various health inequities appear more severe than general disparities in the model deployment field of computational medicine. Despite their widespread use, no comprehensive work documents the sources and quantification methods of bias that computational medicine may inherit from preceding application fields. In recent years, many researchers have scrutinized the healthcare deployment of predictive models in terms of unintended bias; nevertheless, none have examined the extent to which proposed methods are applicable to data-driven computational medicine. After collecting and analyzing literature across health care and other fields, it is found that these works fall into the three categories of sources, metrics, and mitigation. A tutorial is also provided on how to use these works in computational medicine with examples in prediction tasks. Due to such documentation, practitioners will be better equipped to identify unintended biases as their algorithms migrate to new problems and application fields, which will in turn motivate further development of quantification approaches for new settings of bias [5, 6].

Regulatory Frameworks

The deployment of digital health technologies and algorithms into healthcare settings cannot remain unregulated. Congress is calling for investigations into the biasing of algorithms used in the context of COVID-19 and related health services and assessments. These calls are being made in part due to concerns that biased algorithms contribute to the unequal impacts of COVID-19 on marginalized communities. This algorithm was found to directly contribute to inequitable access during the pandemic, particularly for locally-operating health services and Black communities. The Centers for Medicare and Medicaid Services (CMS) audit of Face Value, an AI algorithm that assigns a score based on race and ethnicity, was another due of discrimination against communities of color. Analogously to the analysis of algorithms deployed in healthcare, the introduction of regulation and oversight approaches to mitigate bias is imperative. Legislative and policy initiatives aimed at identifying and preventing discriminatory biases in the development and deployment of AI algorithms are essential elements of a comprehensive bias impact assessment approach. These regulatory initiatives may take various forms, from broader general-purpose algorithm disclosure bills that could be adapted to the health domain to industry sector-specific transparency audit statutes. Regulatory frameworks that directly apply to the biases of AI algorithms deployed in the healthcare sector would also mitigate concerns regarding biased algorithms used in healthcare settings. Congress's recent calls for investigations into the biasing of algorithms used in the context of COVID-19 and related health services and assessments implemented by AI developers and organizations included in-depth examinations of the aforementioned algorithm disclosure statutes. Examples include the Algorithmic Accountability Act, the Justice Data Accountability Act, and the Blueprint for a Federal Data Strategy. Other regulation efforts taking shape in various US states, such as

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

New York City, New York State, and Washington State, algorithm assessments, may also address the issue of AI bias monitoring and auditing [7, 8].

Data Collection and Management

Accumulation of digital health data records of patients with various diseases allows the development of Artificial Intelligence (AI) algorithms for a wide range of applications in health, addressing clinical and operational issues of hospitals and care pathways. However, AI algorithms can be biased since they are trained on retrospective data collection of patients, leading to health inequalities in healthcare. Health inequalities can arise from multiple sources, such as data imbalance or population shift between training and evaluation datasets. There is a need to introduce data standards for data collection and management in hospitals to provide a structured approach for the design of expertise-compatible and norm-compliant protocols for data collection. Furthermore, specific guidelines for data monitoring should be defined to assess how the data records respect the standards, to propagate modified standards in a structured manner, and to help hospitals in updating their data management processes. These monitoring instruments should be low-weight software modules integrated in the PDMS to access a wide set of metrics and eventually (i) quantify and monitor the level of conformance of the data records to a standard and (ii) notify about any deviation from it. Standards regarding clinical vocabulary classifications, data types, structures and formats, metadata, ontologies, and knowledge representation for free text fields should be defined based on (meta-) data management requirements. They can range between general standards suitable for most disease datasets to specific standards only applicable to a particular disease [9, 10].

Algorithm Design and Development

There are numerous examples in healthcare that warrant the establishment of these guidelines, including bias related to race, gender, and socioeconomic status, impacting millions of lives. The inequities detected in healthcare-related algorithms mirror the biases observed in general-purpose algorithms, such as a digital tool trained on resumes that consisted primarily of white male candidates, resulting in gender bias. Healthcare disparities continue to exist in medicine as a reflection of historical and current socioeconomic inequities and group biases from the perpetuation of cultural stereotypes. Though traditionally viewed through the lens of race and ethnicity, healthcare disparities encompass a wide range of dimensions, including socioeconomic status, insurance status, education status, language, age, gender, sexual identity/orientation, and body mass index (BMI). These disparities encompass all 5 domains of the social determinants of health. Healthcare disparities began to become more widely recognized in the early 2000s, with reports documenting disparities in tobacco use and access to mental health care by different racial and ethnic groups. An example can be seen in maternal morbidity, where trends in pregnancy-related mortality in the US stratified by race/ethnicity showed significantly higher pregnancy-related deaths amongst non-Hispanic Black women due to disparate healthcare access and poor economic stability. The calculation of estimated glomerular filtration rate as a diagnostic tool for chronic kidney disease has led to an overestimation of kidney function in Black patients, directly affecting their standard of care. Understanding the sources of these disparities would guide public policy on developing new clinical criteria for early detection of underserved patients and regulating the current development of machine learning algorithms trained with biased data [11, 12].

Ethical Considerations

Healthcare disparities persist in medicine, stemming from historical and current socioeconomic inequities and biases. These disparities cover various aspects such as socioeconomic status, insurance access, education, language, age, gender, sexual orientation, and body mass index (BMI), aligning with the social determinants of health outlined by the US Department of Health and Human Services. Recognition of these disparities increased in the early 2000s, highlighted by reports on tobacco use and access to mental health care among different racial and ethnic groups. For instance, pregnancy-related mortality rates in the US reveal significantly higher instances among non-Hispanic Black women, attributed to inadequate healthcare access and economic instability. Additionally, the practice of calculating estimated glomerular filtration rate for chronic kidney disease has resulted in overestimating kidney function in Black patients when race is included. Examining these disparities can inform public policies aimed at establishing new clinical criteria for detecting underserved groups and improving machine learning algorithms to prevent biases in the data they utilize. There are substantial ethical concerns regarding the impact of AI on ethnic minority groups and underrepresented communities, as evidenced by audit studies indicating that AI may

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

reveal spurious causal relationships linked to identity status, risking privacy and misdiagnosis. In recent years, there has been considerable interest in machine learning applications within healthcare; however, there are rising concerns about biases in decisions affecting ethnic minorities. Differences in cardiovascular events correlate significantly with racial group dynamics, and undiagnosed silent hypoxemia, identified via pulse oximetry, occurs more frequently in Black individuals due to variations in skin response to light wavelengths. It is essential to recognize and address these biases and disparities within computational medicine. This review encapsulates existing research on bias sources, quantification methods, and strategies for mitigation in computational medicine [13, 14].

Stakeholder Engagement

Different stakeholders, including machine learning researchers, software engineers, and electronics engineers, need to cooperate to develop goals, methods, and training data for the creation of AI algorithms. A long-term effort to educate stakeholders will also be necessary to ensure a broad consensus on the importance of diversity evaluation. Recent investments have been made to develop large, diverse demographic representation datasets. Ownership of this data must be limited to prevent its use by competitors. However, to facilitate a proper evaluation of AI algorithms, public datasets are needed to scope for adversarial use by underserved groups. For this reason, a multi-stakeholder alliance might be necessary between algorithm developers, different governmental and non-profit organizations, and medical societies worldwide. All stakeholders are obligated to provide not just the net outputs of pre-trained AI engines but also the source codes to competitive AI engines and to improve explainability of their pre-trained tasks and datasets. Cooperating with experts in interpretability and explainability will allow greater insight into AI decision-making processes when deployed. The academic community researching fairness, accountability, transparency, explainability, and robustness of AI can also assist the medical AI community regarding these types of issues concerning AI accountability. A scientific community can be established via participatory stakeholder initiatives with a pool of diverse technical talent. Local events may also be held to ensure proper geographical representation and talent discovery and to provide easily accessible educational resources. From this pool of experts, national groups can be formed and eventually linked to international organizations. This would drive increasing awareness of diversity challenges in medical imaging AI engines. While differences in levels of early-stage medical imaging AI research resources across countries may seem like insurmountable barriers, this challenge is approached with optimism. Patience and perseverance are ultimately required to impart understanding of how to bridge these differences [15, 16].

Future Directions

This review paper presents a comprehensive bibliographic survey of AI-fairness and bias-related literature in medicine and health. A systematic analysis of 24 primary studies on methods to address bias and fairness in AI models is performed. Given the variety of methods, a 4-category taxonomy is established. Larger-scale datasets covering five different sub-domains are collected for understanding research trends. Bias evaluation metrics and a proof-of-concept pipeline to facilitate future research are also provided. It is hoped that this thorough review will lead to novel ideas and inspire researchers to advance the development of fair and unbiased AI methods for biomedicine. The high quality of this review paper reflects both its relevance and necessity in the designated area. It covers the context, extensive datasets, bias metrics, and methodologies for addressing bias in the field of AI fairness and bias in health and biomedicine. The investigation ranges from recent methods to systematic analysis, which is rare in this relatively new research area. Overall, it attempts to provide readers with a comprehensive and clear view, enabling them to understand the whole picture of this enormous research area within biomedicine. Tackling bias in AI health applications is a burgeoning research area, and methods or applications are emerging across diverse health applications. On one hand, understanding existing methods and their drawbacks is essential for the design and enhancement of new ones. On the other hand, endeavors on systematic surveys and comprehensive methods are important for research trends and comparisons. In addition to augmenting fairness in healthcare, it is equally important to enhance trustworthiness, interpretability, and safety [17, 18].

Case Studies of Successful Mitigation

Since 2021, a growing number of AI-generated health technology applications have been subjected to rigorous evaluation of their potential biases using external test datasets, often collected from institutions with different population demographics from those included in the development datasets. This systematic

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

screening of hundreds of studies from multiple health domains has provided some practical examples of the various aforementioned biases identified in AI algorithms and their associated mitigation approaches. Running a test dataset through an AI system may change the prediction prices, but the quality of output will stay consistent, which is the defining characteristic of software. This section reviews how researchers have successfully released new test datasets to reveal the limitations of AI-generated health technology and the corresponding mitigation approaches, such as model re-training and self-training, to narrow the gap across the algorithm. It illustrates how to initiate the development of fair AI-generated health technologies based on real-world cases, showing how validation datasets or independent datasets can be leveraged. Early work using the MIMIC-III dataset highlighted the racial bias in AI systems trained primarily on data from white patients, being less accurate for data from Black patients. These unavoidable biases from the data domain lead to a misalignment of expertise as testing datasets enrich knowledge from a different patient cohort, which raises broad concerns about the validity and trustworthiness of AI-generated health systems. It's developed black-box models, SCR and BLADE trained models using the development datasets from the triaging systems currently used at the Stanford hospital, but this use case would ironically not cover up personalization challenges with input from different hospitals trained on clinical records with distinct data distributions [19, 20].

Best Practices for Implementation

The aforementioned approaches can help mitigate undesired bias in AI workflows. However, these approaches can be complex and costly to implement. Therefore, before using these approaches, given a machine learning workflow, it is essential to identify which type of bias is present or more likely to be present. In this work, five major types of sources causing bias and unfairness in AI systems in general were identified. These are (1) a biased training population distribution, (2) the absence of features characterizing a certain group, (3) an insufficient amount of data, (4) training on incorrect labels, and (5) misapplication of a predictor to certain groups. It is important to underscore that accurately pinpointing the specific source of this bias is crucial, as it informs the choice of the most effective approach to counteract and diminish these biases. Population bias is one specific case, which was discussed above. A model that is exclusively trained on data collected from a hospital predominantly serving white patients may not perform as effectively for other races. In such cases, the use of adversarial learning to forget the demographic attributes from the model could help it focus more on the disease characteristics and thus promote fairness. Some populations may not be represented in the training population. Insufficient data, which is a potential cause of representation bias, is particularly acute in such domains as biomedicine. However, in such domains, the data annotation often necessitates the involvement of a highly skilled clinician or physician. Distributional techniques could help solve such issues by generating controlled synthetic data to enhance and diversify the dataset or facilitate the collaboration between different centers by utilizing the data under federated learning protocols. This is not to imply that a certain technique could solely deal with a specific source of bias. Instead, each approach has its own merits and limitations. Furthermore, in practice, multiple sources of bias could happen at the same time [21-25].

International Perspectives

Health-related AI is rapidly infiltrating medicine, raising significant concerns over biased algorithms. This survey outlines practical AI bias mitigation strategies in biomedicine, highlighting strengths, weaknesses, and factors for incorporation. Despite extensive research in criminal justice and finance, biases in AI diagnostic and therapeutic systems for health are just now garnering essential attention. Issues range from algorithmic output bias to inherent data bias, leading to questions about how models treat individuals and groups, and whether health impacts on social characteristics are justified. Disproportionate algorithmic errors towards certain subgroups threaten the fairness of health systems. This scrutiny has prompted calls for re-evaluating algorithmic predictions, raising bioethical concerns about a potentially unregulated replacement of empathetic physicians with automated systems. Nonetheless, the wide application of programmatic AI offers chances for systematic reassessment of health aims. With the gap between growing interdisciplinary interests and the lack of relevant surveys, practical insights into debiasing methods are crucial. This work deviates from the typical focus on potential, making strides towards practical solutions by identifying and categorizing methods to combat algorithmic unfairness in health-related AI. While covering text, signal-based, and image-based applications, it remains non-exhaustive. By educating technologists, regulators, ethicists, and stakeholders about AI injustices, this work aims to address grievances that could hinder the acceptance and integration of AI in health systems [26-29].

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

CONCLUSION

Addressing bias in AI algorithms used for healthcare applications is not merely a technical challenge but a moral imperative. Historical and systemic inequities embedded within healthcare data and digital tools threaten to exacerbate disparities if not carefully mitigated. Evidence from clinical AI applications reveals significant performance gaps across racial, gender, and socioeconomic lines, underscoring the urgent need for comprehensive strategies that span data governance, algorithm design, and ethical oversight. Regulatory interventions such as algorithmic audits and bias impact assessments are necessary to ensure accountability, while stakeholder engagement fosters inclusivity and transparency. Case studies of bias mitigation affirm the potential of retraining, diverse datasets, and interpretable models to reduce harm. Moving forward, interdisciplinary collaboration, public-private partnerships, and ongoing research are vital to creating AI systems that are fair, trustworthy, and capable of delivering equitable healthcare benefits to all communities. Only through sustained efforts in fairness-aware design and deployment can we align technological advancement with the principles of justice and human dignity in medicine.

REFERENCES

1. Shiels MS, Haque AT, Haozous EA, Albert PS, Almeida JS, García-Closas M, Nápoles AM, Pérez-Stable EJ, Freedman ND, Berrington de González A. Racial and ethnic disparities in excess deaths during the COVID-19 pandemic, March to December 2020. Annals of internal medicine. 2021 Dec;174(12):1693-9. acpjournals.org
2. Barnato AE, Johnson GR, Birkmeyer JD, Skinner JS, O'Malley AJ, Birkmeyer NJ. Advance care planning and treatment intensity before death among black, hispanic, and white patients hospitalized with COVID-19. Journal of General Internal Medicine. 2022 Jun;37(8):1996-2002. springer.com
3. Howard J. Algorithms and the future of work. American Journal of Industrial Medicine. 2022 Dec;65(12):943-52.
4. Kadiresan A, Baweja Y, Ogbanufe O. Bias in AI-based decision-making. InBridging human intelligence and artificial intelligence 2022 Feb 24 (pp. 275-285). Cham: Springer International Publishing. [HTML]
5. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Medical preparedness for bioterrorism and chemical warfare: A public health integration review. Medicine. 2025 May 2;104(18):e42289.
6. Parr NJ, Beech EH, Young S, Valley TS. Racial and ethnic disparities in occult hypoxemia prevalence and clinical outcomes among hospitalized patients: a systematic review and meta-analysis. Journal of General Internal Medicine. 2024 Oct;39(13):2543-53. springer.com
7. Marrone E, Cafaro J, Klein J. The Silent Saboteur: Teaching the Clinical Implications of Occult Hypoxemia & Social Determinants of Health via a Pulmonary Embolism Case. Journal of Education & Teaching in Emergency Medicine. 2025 Apr 30;10(2):O1. nih.gov
8. Jackson MC. Artificial intelligence & algorithmic bias: the issues with technology reflecting history & humans. J. Bus. & Tech. L.. 2021;16:299.
9. Ugwu CN, Ugwu OP, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI, Uti DE. Sustainable development goals (SDGs) and resilient healthcare systems: Addressing medicine and public health challenges in conflict zones. Medicine. 2025 Feb 14;104(7):e41535.
10. Siddique SM, Tipton K, Leas B, Jepson C, Aysola J, Cohen JB, Flores E, Harhay MO, Schmidt H, Weissman GE, Fricke J. The impact of health care algorithms on racial and ethnic disparities: a systematic review. Annals of Internal Medicine. 2024 Apr;177(4):484-96. acpjournals.org
11. Seyyed-Kalantari L, Zhang H, McDermott MB, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nature medicine. 2021 Dec;27(12):2176-82. nature.com
12. Flores L, Kim S, Young SD. Addressing bias in artificial intelligence for public health surveillance. Journal of Medical Ethics. 2024. [HTML]
13. Edyedu I, Ugwu OP, Ugwu CN, Alum EU, Eze VH, Basajja M, Ugwu JN, Ogenyi FC, Ejemot-Nwadiaro RI, Okon MB, Egba SI. The role of pharmacological interventions in managing

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

urological complications during pregnancy and childbirth: A review. Medicine. 2025 Feb 14;104(7):e41381.
14. Vinothini C, Suja Rose RS, Saravanabavan V. Patient's Satisfaction with Primary Healthcare Services and Its Link to Socio-Economic Conditions in Madurai District. International Journal of Scientific Research and Engineering Development. 2025;8(2):248-53. researchgate.net
15. Emon MM, Nipa MN. Exploring the gender dimension in entrepreneurship development: A systematic literature review in the context of Bangladesh. Westcliff International Journal of Applied Research. 2024;8(1):10-47670. ssrn.com
16. Chen RJ, Chen TY, Lipkova J, Wang JJ, Williamson DF, Lu MY, Sahai S, Mahmood F. Algorithm fairness in ai for medicine and healthcare. arXiv preprint arXiv:2110.00603. 2021 Oct 1.
17. Nyamboga TO, Ugwu OP, Ugwu JN, Alum EU, Eze VH, Ugwu CN, Ogenyi FC, Okon MB, Ejemot-Nwadiaro RI. Biotechnological innovations in soil health management: a systematic review of integrating microbiome engineering, bioinformatics, and sustainable practices. Cogent Food & Agriculture. 2025 Dec 31;11(1):2519811.
18. Xu J, Xiao Y, Wang WH, Ning Y, Shenkman EA, Bian J, Wang F. Algorithmic fairness in computational medicine. EBioMedicine. 2022 Oct 1;84.
19. Eweje G, Sajjad A, Nath SD, Kobayashi K. Multi-stakeholder partnerships: A catalyst to achieve sustainable development goals. Marketing Intelligence & Planning. 2021 Mar 8;39(2):186-212. academia.edu
20. Ugwu OP, Alum EU, Ugwu JN, Eze VH, Ugwu CN, Ogenyi FC, Okon MB. Harnessing technology for infectious disease response in conflict zones: Challenges, innovations, and policy implications. Medicine. 2024 Jul 12;103(28):e38834.
21. Civelek U, Eren PE, Gökalp MO. Increasing the collaboration of data science stakeholders with a knowledge management system. Business Process Management Journal. 2024 Oct 30.
22. Petersen JM, Ranker LR, Barnard-Mayers R, MacLehose RF, Fox MP. A systematic review of quantitative bias analysis applied to epidemiological research. International journal of epidemiology. 2021 Oct 1;50(5):1708-30. [HTML]
23. Page MJ, Sterne JA, Higgins JP, Egger M. Investigating and dealing with publication bias and other reporting biases in meta-analyses of health research: A review. Research synthesis methods. 2021 Mar;12(2):248-59. wiley.com
24. Meng C, Trinh L, Xu N, Liu Y. Mimic-if: Interpretability and fairness evaluation of deep learning models on mimic-iv dataset. arXiv preprint arXiv:2102.06761. 2021. [PDF]
25. Bear Don't Walk OJ, Pichon A, Reyes Nieva H, Sun T, Li J, Joseph J, Kinberg S, Richter LR, Crusco S, Kulas K, Ahmed SA. Contextualized race and ethnicity annotations for clinical text from MIMIC-III. Scientific Data. 2024 Dec 5;11(1):1-2.
26. Ferrara E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. Sci. 2023 Dec 26;6(1):3.
27. Varona D, Suárez JL. Discrimination, bias, fairness, and trustworthy AI. Applied Sciences. 2022 Jun 8;12(12):5826.
28. Sikstrom L, Maslej MM, Hui K, Findlay Z, Buchman DZ, Hill SL. Conceptualising fairness: three pillars for medical algorithms and health equity. BMJ health & care informatics. 2022 Jan 10;29(1):e100459. nih.gov
29. Kaur D, Uslu S, Rittichier KJ, Durresi A. Trustworthy artificial intelligence: a review. ACM computing surveys (CSUR). 2022 Jan 18;55(2):1-38.

CITE AS: Kansiime Agnes (2025). Addressing Bias in AI Algorithms for Health Applications. IAA Journal of Biological Sciences 13(1):37-43. https://doi.org/10.59298/IAAJB/2025/1313743

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.