Bias Beyond Boundaries: Human vs. Algorithmic Influence in Decision-Making

YAGNESH CHALLAGUNDLA [1*],
ASLESHA SANJANA KODAVALI [2*], SANTHI DAGGUBATI [3*],
SRIKAR GOWRISHETTY [4*]
and VENKATA JYOTHI PRIYA MULAKA [5*]

[1,2,3,4,5*]Herbert Wertheim College of Engineering, University of Florida, Gainesville, 32611, Florida, USA.

*Corresponding author(s). E-mail(s): yagneshnaidu1234@gmail.com; sanjanakodavali10@gmail.com; Santhic08@gmail.com; srikargowrishetty03@gmail.com; mvjpriya@gmail.com;

Abstract

Our study analyzes the cultural, technological, and ethical issues brought about by biases in AI and algorithmic technology from multiple applications. It brings to the fore the common problem of overt and covert biases, including Own Race Bias (ORB) in facial recognition as well as the effect of implicit racial bias on neural activity. Algorithmic biases, and in particular those embedded in the medical and the criminal justice systems, are systematically evaluated, pointing to the effect that such biases have on equity and decision-making. Healthcare policies operated with narrow healthcare algorithms tend to under-diagnose and overuse some treatments while predictably policing instruments exacerbate already entrenched social injustices. It stresses the need for transparency and clear definition of responsibility for the dataset curation, enhancing model development practices. Groundbreaking tools like DermaSensor an AI skin cancer screening tool demonstrate both the promise and the ethical dangers of AI in Medicine. Extremalists conclude that there is an urgent goal to build equitable AI to promote civil rights, diversity and the well-being of communities.

Keywords: Algorithmic Bias, Artificial Intelligence Ethics, Racial Disparities in AI, Fairness in Machine Learning, Healthcare Algorithms, Predictive Policing Bias, Ethical AI Frameworks, Inclusive Data Representation.

1 INTRODUCTION

Artificial intelligence (AI) has grown to be a novel enhancer of change in all fields. However, its use in critical areas like medicine, criminal justice, and employment, raised notable ethical and technical challenges. Among these concerns, the concept of bias – whether implicit, racial, or algorithmic – emerges as a key concern having a potential to deepen inequalities and sustain stereotypes. Bias in artificial intelligence is not destructive in nature but is a product of historic tendency and structural discrimination features of the data set. For example, Own Race Bias (ORB) presents a difficulty in identifying faces of other races in the human face recognition system which is said to be influenced by the individual's or society's mindset. These factors can have adverse effects especially in judicial systems where poor identification leads to injustices. In medicine, AI models often have a bias due to unrepresentative training datasets. Tools that are meant for treatment allocation or diagnostic determination may favor some sections of the population while disadvantaging others in health care. As an example, while AI technologies such as DermaSensor, an AI tool for detection of skin cancer, holds promise for enhancing effectiveness and accessibility, its effectiveness for patients with dark skin is limited due to a lack of them in the training data set. In the same vein, algorithms such as PredPol, which are predictive in the justice system, could exacerbate institutional bias as they particularly target communities that have been disadvantaged in past policing, which is derived from biased arrest history data.

Algorithmic bias in self-service hiring refers to the use of self-service hiring systems that in the end seems to favour candidates who are similar in language or culture to those in the training dataset. These inconsistencies in hiring practices tend to disfavor some sections of the population. Such biases are best reduced through a multi-pronged approach which includes, different data sourcing, fairness constraints, and explainable solutions. This paper brings together foundational materials from different papers and field works to better explain algorithm bias. In addition, the paper examines the impact of responsible artificial intelligence through its origins, its effects, and the relationship to social justice principles in advancing the conversation surrounding social justice in technology.

2 LITERATURE WORK

Barocas and Selbst investigate in their article the aspects of big data which are not so attractive arguing that biased data and the criteria still constitute ' a disability' especially in areas such as hiring trends, lending practices, and even the criminal justice systems [1]. Kroll et al. have called for the need of explaining algorithms, arguing for ways of ensuring fairness and transparency of all decisions made in the course of applying machine learning models [1]. Kleinberg and his colleagues (2018) addressed bias in algorithms, particularly the biases present in the datasets used which worsen the existing biases within society particularly in regard to automated decision making technologies [2]. Bagaric et al. (2022) argue that AI-based, transparent and fair models should be enforced within the criminal justice system in response to the systemic biases that are present [3]. Ayling and Chapman (2022) analyze the effectiveness of artificial intelligence ethics tools arguing that they are not enough and further more broad approaches must be formulated [5]. De Cremer and Kasparov (2022) argue the ethical AI paradox stating that with the advancement of technology, moral decision making will require more humanity in the loop [6]. Frye (2022) raises questions about ethics in the use of AI in the production of academic papers wondering if the application constitutes self-plagiarism and undermines academic honesty [7]. Hagendorff (2022) points out the shortcomings in AI ethics frameworks stressing overlooked issues like social class and poverty, and unwanted side effects of algorithms [8]. Huang et al.

The challenges of incorporating technology into moral and ethical arguments in the creation of artificial intelligence are discussed by Makara et al. (2022). The in depth analysis of the discussion on ethical dissection of AI is said to have been highlighted by Makatov (2022) where he proposed a lens through which AI ethics may be examined from the stakeholder viewpoint. The research gap about Chinese artificial intelligence in the Western market has been added to by the study done by Nakonechny and Gromova (2022). Robinson (2022) argues that there is a deeper socio-political—or even revolutionary—point that Walker and his comrades were putting forth in their works. During the evaluation of the impact of globalization on Chinese politics, Wu (2020) was able to highlight the effect of globalization on Chinese economic development as he suggested that chinos opportunities to benefit from trade and foreign direct investment led to suffering from capital and technology starvation. Wu (2020) further highlighted the imposition of sanctions on Chinese companies specifically around technology products that were deemed to have been likely to pose a national security threat to the US and is seen as an emerging trend in US foreign policy. Wu believed that the American pursuit of an aggressive strategy would most likely lead China into confrontation, followed by waves of military confrontation that would result in genocide as elaborated to by Wu and about few Uebermenschen Success and their Muslim counterparts.
While (2021) examine the digital divide in the implementation of AI technologies around the world arguing that the disregard for AIs's full potential makes it possible for socioeconomic injustices to deepen further, [16] In Green (2020) discusses a total of sixteen general obstacles and possibilities in AI ethics including amongst others safety and privacy of technology, unemployment and inequality [17]. The Center for Equity, Gender, and Leadership (2022) talks of incorporating inequalities to even out by improving datasets and defining fairness constraints in AI systems [18].

3 CASE STUDY

Case Study on DermaSensor: One Way Never Works. Algorithms Can Be Biased at Times Too.

3.1 Preface

The DermaSensor device is a remarkable AI tool that enables primary care physicians (PCPs) to assist in skin cancer diagnosis by analyzing skin lesions. It uses ESS and

AI to estimate the spectral signatures of lesions and rate them as high or low for cancerous potential. Despite being developed with artificial intelligence, the gadget raised concerns during development and clinical trials regarding algorithmic bias, particularly in relation to different skin types. It became clear that there was an interplay between human and computer biases.

3.2 Human Bias in The Field of Dermatology

Human biases have always affected dermatology, especially in diagnosing skin disorders in non-Caucasian skin types. Dermatologists often rely on visual assessments, requiring extensive experience to differentiate benign from malignant lesions. However, due to bias and lack of training in recognizing skin conditions on dark skin, early-stage lesions, especially in African patients, may be missed. These biases lead to incorrect diagnoses, delayed treatments, and missed opportunities for recovery, particularly in minority races.

3.3 Algorithmic Bias in DermaSensor

In contrast to human bias, **algorithmic bias** arises from the data used to train AI algorithms. DermaSensor was predominantly trained with a cohort of Fitzpatrick Skin Types I-III patients, which impaired its ability to recognize lesions in Fitzpatrick Types V-VI. The lack of diverse representation in the training data caused the algorithm to favor lighter skin tones, weakening its ability to detect skin cancer in patients with darker skin types.

3.4 Influence of Bias on Clinical Outcomes

In clinical trials validating DermaSensor, significant differences in diagnostic accuracy were observed. In the **DERM-SUCCESS trial**, nearly all participants were White (97.1%), with only 12.7% having darker skin. This skewed representation led to poorer diagnostic performance for individuals with skin types V and VI. This highlights the ethical concerns of applying AI methods without representative sampling, worsening trends of inequity in healthcare and obstructing equality of care.

3.5 Mitigating Algorithmic Bias

To address biases, it is crucial to expand training datasets and ensure that AI tools like DermaSensor are tested across a broader range of target users. Ethical design entails that AI systems are built with diverse data, undergo extensive validation, and are subject to post-implementation reporting. Furthermore, regulatory bodies like the FDA should implement guidelines ensuring AI devices meet the needs of all racial and ethnic groups, not just one.

3.6 Final Assessment

While DermaSensor has potential for assisting PCPs in skin cancer diagnosis, the trials showed issues of algorithmic equity. Biases can be reduced, but this requires

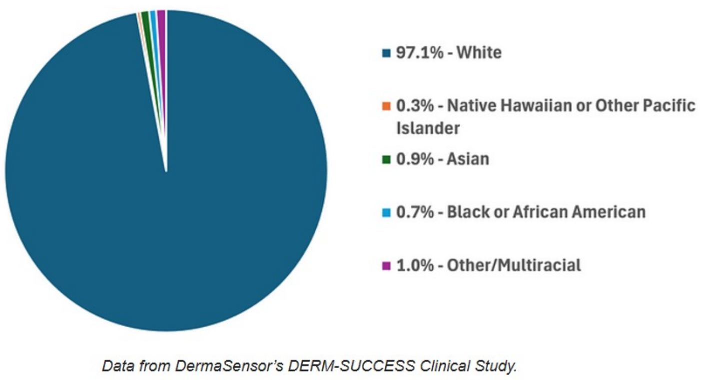

Fig. 1 A Visual Representation of Patient Demographics. (Ref. taken from Dermosensor website)

a multi-dimensional approach focusing on data diversity, better AI model training, and a commitment to equity in healthcare development. Addressing both human and computational biases is essential for creating more equitable and sustainable healthcare solutions.

4 Results

The investigation focused on various forms of algorithmic bias and its effects, Kasy (2020) focused on the algorithmic bias that is profit motivated aspects, asserting that in most cases algorithms exacerbate racial inclusions by giving advantages to some groups and in most cases the economic ones. He suggests that future algorithms should focus on charity and fairness so as to ensure that all groups including the minority groups in terms of ethnicity are not sidelined [5]. Jain et al. (2020) stressed the occurrence of of racial and ethnic bias through the use of algorithms in health care systems which influenced some of the processes in making decisions. They showed that most algorithms were based on biased data and therefore the outcome delivered was equitable but most of the minority populations did not benefit from the outcome. Lack of transparency in the application of these algorithms leads the patients and the health care professionals to be unaware of these biases in the provision of services with the result that disparities in the care provided are set in motion [6].
According to their research, Intahchomphoo and Gundersen (2020) reviewed four races issues touch points where AI penetrates Racial Disparities, that is: Restriction of Specific Racial Communities Opportunities, Identification of Racism, Health Disparities Amongst Racial Groups, and Racial Myths on Facial Recognition. Such conclusion points out that there is a need to create, more so in northern countries, legislative frameworks ready to deal with such inequalities. Huq (2019) raised similar issues in the context of the criminal justice system, and in particular, pointed out that assessments of risk and predictive policing as algorithms aggravate racial inequality. He emphasized that available legal mechanisms are indeed insufficient to rectify these biases, making it necessary to start thinking of the long-term consequences of the algorithms for ethnic minorities.

Table 1 Performance of various imaging technologies for skin cancer detection.

Device	Imaging Technology	Indication	n (Lesions n)	Sensitivity (%)
Melafind[11]	Multispectral	Disease: Melanoma	1383 (1631)	98.3
NeviSense[22]	Electrical impedance	Disease: Melanoma	1951 (2416)	97.0
DermaSensor[13]	Elastic scattering	Squamous & basal cell carcinoma	1005 (1579)	96.6

However, there might be potential biases caused by algorithmic decision making such as in the case of AI systems used for recruitment, clients who are honored in the systems are those who meet the requirements of the language and culture, these situations aggravate the inequalities in the recruitment processes, thus denying deserving candidates with other qualifications [9]. Also bias within the health care section has been demonstrated to undermine the ability for accurate diagnosis applicable to economically disadvantaged patients or individuals comming from rural communities and who are obese, suffering from mental health problems, as the researchers assert, because of a limited array of data employed in the training as well as implicit bias of the health care providers [10].
The ethical concerns regarding Ai integration include medical devices such as DermaSensor which is used to enhance the chances of identifying skin cancer. While this could potentially improve the quality of clinical services in poor areas, its training algorithms have been criticized for being biased, lacking in data for patients with darker skin tones which may lead to a poor level of diagnosis for these patients Ho(cited) 11. This raises the issue of the need to bring diverse groups of people in testing datasets so that the AI algorithms developed are equitable across all populations. Also, the use of DermaSensor has the potential to deepen healthcare disparities if its use is not carefully managed and suitably modified to address the needs of the disadvantaged groups. To avoid further contributing to bias and disparity, ethical principles around justice, responsibility, and openness should be embedded in all stages of AI solutions development and deployment.

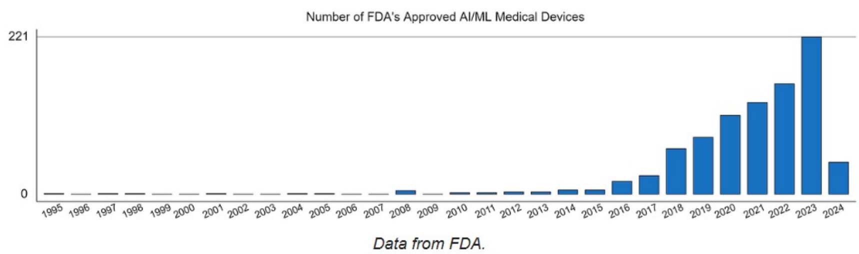

Fig. 2 From Concept to Clinic: Tracking the Growth of FDA-Approved AI/ML Medical Devices (Ref. taken from website)

5 Conclusion

The AI system in question, exemplified by DermaSensor for skin cancer diagnosis, is primarily intended for primary care physicians (PCPs), particularly in underprivileged regions with restricted access to dermatologists. It addresses the want for precise, early-stage skin cancer identification by non-specialists, enhancing diagnostic proficiency in areas lacking dermatological competence. The system assimilates into a broader cybernetic framework by linking data from patient assessments to a database of identified lesions, employing Elastic Scattering Spectroscopy (ESS) and artificial intelligence to categorize lesions. This improves PCPs' decision-making, enabling expedited referrals for high-risk situations. The ethical theories informing the system encompass justice (ensuring equal healthcare access), beneficence (striving to enhance health outcomes), and autonomy (enabling physicians to make informed decisions).
The system necessitates ongoing training with varied datasets and must be subjected to stringent testing to evaluate performance across diverse populations, ensuring equity and minimizing bias. This embodies the principle of equality, aligning the system with the overarching objectives of delivering equitable healthcare. Ensuring safety requires compliance with regulatory requirements established by organizations such as the FDA, performing post-market surveillance, and executing real-time changes informed by empirical data. Possible unintended outcomes may be excessive dependence on the AI system, diminishing clinicians' proficiency in identifying skin problems or leading to misdiagnoses among underrepresented populations.

The accountability for decisions rendered by the system rests with both the developers and primary care clinicians. Developers must provide transparency in the AI's operations, whilst physicians are required to analyze outcomes critically. Safeguards must encompass transparency in data utilization, ensuring that the system does not exacerbate current healthcare disparities. These safeguards may encompass routine audits, feedback mechanisms, and regulatory supervision. The socio-economic context entails tackling healthcare disparities, especially in rural regions and groups with restricted access to dermatological treatments. The ethical principles for regulatory frameworks must emphasize inclusive data gathering, equitable access, and continuous monitoring to guarantee that technology serves all populations, especially marginalized ones.
This holistic approach guarantees the system's safety, fairness, and ethical standards, fostering justice in healthcare outcomes among varied populations.

References

[1] Barocas, S., Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, **104**, 671-727.

[2] Kroll, J. A., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., Yu, H. (2016). Accountable algorithms. *University of Pennsylvania Law Review*, **165**, 633-706.

[3] Kleinberg, J., Ludwig, J., Mullainathan, S., Sunstein, C. R. (2018). Discrimination in the Age of Algorithms. *Journal of Legal Analysis*, **10**, 113-174. doi:10.1093/jla/laz001

[4] Bagaric, M., Svilar, J., Bull, M., Hunter, D., Stobbs, N. (2022). The solution to the pervasive bias and discrimination in the criminal justice system: Transparent and fair artificial intelligence. *American Criminal Law Review*, **59**, 95-98.

[5] Ayling, J., Chapman, A. (2022). Putting AI ethics to work: Are the tools fit for purpose? *AI and Ethics*, **2**(3), 405-429. doi:10.1007/s43681-021-00092-7

[6] De Cremer, D., Kasparov, G. (2022). The ethical AI paradox: Why better technology needs more and not less human responsibility. *AI and Ethics*, **2**(1), 1-4. doi:10.1007/s43681-021-00085-6

[7] Frye, B. L. (2022). Should using an AI text generator to produce academic writing be plagiarism? *SSRN Scholarly Paper*, 2022. doi:10.2139/ssrn.4292283

[8] Hagendorff, T. (2022). Blind spots in AI ethics. *AI and Ethics*, **2**(4), 851-867. doi:10.1007/s43681-021-00073-w

[9] Huang, C., Zhang, Z., Mao, B., Yao, X. (2022). An overview of artificial intelligence ethics. *IEEE Transactions on Artificial Intelligence*, 1-21. doi:10.1109/TAI.2022.3178901

[10] Huston, J. (2022). Artificial intelligence as a content creator in the publishing industry. *Book Publishing Final Research Paper*, 2022. pdxscholar.library.pdx.edu

[11] Ibáñez, J. C., Olmeda, M. V. (2022). Operationalising AI ethics: How are companies bridging the gap between practice and principles? *AI SOCIETY*, **37**(4), 1663-1687. doi:10.1007/s00146-021-01258-6

[12] Kieslich, K., Keller, B., Starke, C. (2022). Artificial intelligence ethics by design: Evaluating public perception on the importance of ethical design principles of artificial intelligence. *Big Data Society*, **9**(1), doi:10.1177/20539517221092956

[13] Li, Y., Gu, J., Wang, L. (2020). Research on artificial intelligence ethics in the field of art design. *Journal of Physics: Conference Series*, **1673**(1), 012052. doi:10.1088/1742-6596/1673/1/012052

[14] McLennan, S., Fiske, A., Tigard, D., Müller, R., Haddadin, S., Buyx, A. (2022). Embedded ethics: A proposal for integrating ethics into the development of medical AI. *BMC Medical Ethics*, **23**(1), 6. doi:10.1186/s12910-021-00735-5

[15] Ashok, M., Madan, R., Joha, A., Sivarajah, U. (2022). Ethical framework for artificial intelligence and digital technologies. *International Journal of Information Management*, **62**, 102433. doi:10.1016/j.ijinfomgt.2021.102433

[16] Carter, L., Liu, D., Cantrell, C. (2021). Exploring the intersection of the digital divide and artificial intelligence. *AIS Transactions on Human-Computer Interaction*, **12**(1).

[17] Green, B. P. (2020). Artificial intelligence and ethics: Sixteen challenges and opportunities. *Markkula Center for Applied Ethics*, Santa Clara University. dataethicsrepository.iaa.ncsu.edu

[18] Center for Equity, Gender, Leadership. (2022). Mitigating bias in artificial intelligence. *Berkeley Haas*, 2022. haas.berkeley.edu