IJARBM – International Journal of Applied Research in Business and Management
Vol. 05 / Issue 01, pp. 209-224, January 2024
ISSN: 2700-8983 | an Open Access Journal by Wohllebe & Ross Publishing

This paper is available online at
www.ijarbm.org

Mitigating Biases in Training Data: Technical and Legal Challenges for Sub-Saharan Africa

Alexander Oluka
olukaam@gmail.com
Europa-Universität Viadrina, Frankfurt (Oder), Germany

https://doi.org/10.51137/ijarbm.2024.5.1.10

Abstract – The study examines the challenges of mitigating biases in AI training data within Sub-Saharan Africa. A qualitative research approach with semi-structured interviews was employed to gather insights from eight participants with law, IT, and academic background. Thematic analysis was utilised to categorise the data into key themes, revealing insights into the complexities of developing fair AI technologies that reflect the socio-cultural diversity of the region. The findings emphasise the importance of incorporating local values and ethical considerations into AI development and highlight the need for enhanced collaborative efforts to establish resilient, culturally sensitive AI governance frameworks. The research contributes to the broader discourse on ethical AI deployment in diverse global contexts.

Keywords – Bias audits, Data biases, Ethical norms, Fairness, Mitigating biases, Training data.

Submitted: 2022-04-28. Revised: 2024-05-14. Accepted: 2024-05-14.

1. Introduction

Machine learning relies on training data to build models for accurate prediction and classification. However, a significant challenge in machine learning is the president of biases in training data (Gerard, 2020). Data biases can lead to unfair or inoculate prediction which may impact the performance and reliability of machine learning models. Training data bias refers to systematic errors in the data set resulting in unfair, partial, or skewed outcomes when algorithms are trained on it. These errors and predispositions can disadvantage particular groups based on characteristics like gender, race, social, economic status, geography and other distinguishing factors. Errors and predispositions result from a variety of factors, which may include data that reflects historical disparities, subjective labelling techniques, the under or overrepresentation of specific groups, and the inclusion of social or cultural standards that are not always applicable. Dai et al. (2023) suggest data argumentation, which involves expanding the training data to capture data invariance and increase the sample size as an appropriate approach to mitigate biases.

Training data is the presence of inherent biases that can be encoded in embeddings during self-supervised training (Orr et al., 2021). Biases in training data can affect the performance of machine learning models in scenarios where the embeddings do not accurately represent or under-presented elements in the data. Additionally, class imbalance in the dataset can impact the performance of the natural networks for classification tasks (Nolte et al., 2018). When certain classes are over-represented or under-represented in the training data, it can lead to degraded classification performance. Moreover, the quality and quantity of cleaning data play a role in the effectiveness of machine learning models. Qu et al. (2020) argue that limited training data can introduce challenges in model training inference, leading to disparities in and limited performance.

Types and sources biases in machine learning models

Baises in machine learning models can arise from various sources, leading to challenges in model performance and fairness. The common type of bias is selection bias, where the training data is not representative of the entire population, resulting in a skewed prediction (Singh & Sinha, 2022). In addition, algorithmic bias occurs when the machine learning algorithm itself introduces discriminatory patterns based on the data it is trained on the source of bias (Maeda, 2018). These biases can perpetuate existing inequalities and lead to unfair outcomes in decision-making processes.

Moreover, biases in machine learning models can also stem from the quality and quantity of training data used to train models. For instance, biases can be introduced through labelling, where human annotators may inadvertently input their own biases into the training data (Kauwe et al., 2020). Biases can emerge from imbalances in the distribution of classes within the dataset leading to challenges in accurately representing all classes during model training (Shang & Wang, 2016; Wang & Deng, 2020). Addressing these biases is critical to ensure that machine learning models make fair and unbiased predictions across different demographic groups. Furthermore, bias in machine learning models can also be exacerbated by the lack of diversity in the training data, where models trained on homogeneous datasets may fail to generalise well to diverse populations, leading to performance disparities across different groups (Shi et al., 2018). Additionally, bias can be introduced through a feature selection process where certain features may be overemphasised or underrepresented, impacting the model prediction predictive capabilities (Budiman, 2016).

2. Literature review

2.1 Technical Challenges in Mitigating Biases

Detecting biases in datasets is essential for ensuring the fairness and accuracy of machine learning models. The primary obstacle stems from the inherent trade-offs between bias mitigation and model performance. Wang et al. (2019) note that attempts to correct biases, such as through re-sampling or algorithmic fairness approaches, often lead to a decrease in the model's predictive accuracy. Correcting biases in training data may increase error rates for certain subgroups (Thompson et al., 2021). Thus, the deployment of unbiased AI systems in real-world applications may be hampered by the need to achieve an optimal balance between the ethical imperative of fairness and the technical objective of accuracy. The representation of minority groups in datasets is often insufficient, leading to models performing suboptimal for these groups compared to their majority counterparts (Gianfrancesco et al., 2018). This underrepresentation exacerbates existing societal biases, as the AI systems are trained on data that do not accurately reflect the diversity of the global population.

Biases in training data evolve as societal norms and values change, making continuous monitoring and updating of AI models essential for maintaining fairness over time (Chakraborty et al., 2020). Complex machine learning models often operate as "black boxes," where the decision-making processes are not readily understandable by humans. The lack of transparency complicates efforts to detect biases, as it obscures the causal pathways through which biased decisions are made (Belkacemi et al., 2021). Issues of interpretability and transparency hinder the identification of biases within AI models and training datasets. While technical solutions are essential, their effectiveness is often contingent upon supportive legal and regulatory frameworks that promote fairness and accountability in AI systems (Chakraborty et al., 2020). The regulatory landscape presents additional complexities in the quest to mitigate biases in AI.

2.2 Challenges in implementing effective legal measures to combat biases in AI

Implementing effective legal measures to combat biases in AI presents challenges due to the dynamic and complex nature of AI technologies and the data on which they operate. Artificial intelligence systems can manifest biases in various ways, influenced by skewed training data, algorithmic design, or the objectives set by developers, which may inadvertently reinforce existing social inequalities (Barocas & Selbst, 2016). Given the sensitivity and complexity of biases, legal frameworks require a comprehensive understanding of technological processes to effectively mitigate these issues. However, the rapid pace of AI development often surpasses the slower, deliberative processes involved in legislative and regulatory framework development, leading to a lag in responsive legal measures (Cath et al., 2018).

Another challenge is the transnational nature of data and AI technologies, which complicates jurisdictional authority and the enforceability of legal measures. Artificial intelligence systems and the data they use can cross geographical and jurisdictional boundaries, making it difficult to apply national laws effectively. This is particularly relevant in the context of multinational corporations that operate across different legal regimes, potentially exploiting these gaps to avoid stringent compliance (Koops, 2014; Yeung, 2017). The need for international cooperation and harmonisation of laws is evident, yet achieving consensus among diverse legal systems and cultural norms is inherently challenging. Furthermore, the enforcement of laws against biases in AI requires robust monitoring and reporting mechanisms, which are often lacking or underdeveloped, especially in jurisdictions with limited technological infrastructure (Yeung, 2017).

Furthermore, the challenge of ensuring that legal measures are both effective and adaptive is compounded by the inherent opacity of AI algorithms. The 'black box' nature of many AI systems, where the decision-making processes are not transparent, poses a significant barrier to assessing and addressing biases (Pasquale, 2015). This lack of transparency not only makes it difficult for regulators to pinpoint the source of biases but also hinders efforts to enforce accountability and remedial actions. Consequently, there is an increasing call for incorporating explainability and transparency requirements into legal frameworks governing AI. However, achieving this balance without stifling innovation requires careful consideration, as overly prescriptive regulations may limit the development of AI technologies that could offer substantial societal benefits.

2.3 Legal frameworks and policies in Africa concerning data protection and AI

The regulatory environment in many African countries remains in its infancy despite the rapid proliferation of AI technologies and the attendant need for robust data governance mechanisms. A notable exception is the African Union's Convention on Cyber Security and Personal Data Protection, adopted in

2014, which seeks to establish a comprehensive legal framework for cybersecurity and data protection across the continent (African Union, 2014). However, its implementation has been uneven, with only a handful of member states ratifying the convention. This reflects a broader trend of fragmented and inconsistent legal landscapes, wherein a unified approach to data protection and privacy is still emerging. Countries such as South Africa, Kenya, and Nigeria have made strides in enacting national data protection laws, yet the degree to which these laws are enforced and their effectiveness in regulating AI applications varies widely (Kshetri, 2019; Ademuyiwa & Adeniran, 2020). Moreover, the existing legal frameworks in many African countries often lack specific provisions addressing the ethical development, deployment, and use of AI systems.

South Africa, for instance, has enacted the Protection of Personal Information Act (POPIA) in 2013. While the Act is not explicitly designed to regulate AI, POPIA sets a benchmark for data privacy and protection, which indirectly influences AI practices by controlling how personal data can be collected, processed, and stored. In Kenya, the Data Protection Act of 2019 represents a step forward in aligning the country with international data protection standards, mirroring principles in the European Union's General Data Protection Regulation (GDPR). Although the Act primarily focuses on data protection, it provides a legal framework that addresses concerns related to data privacy and the ethical use of AI. Nigeria's National Digital Economy Policy and Strategy (2020-2030) demonstrates the country's commitment to harnessing digital technologies to drive economic growth. The policy highlights the importance of developing legal and regulatory frameworks that support the digital economy while ensuring data protection, privacy, and cybersecurity. Even though it is not an AI regulation per se, the strategy acknowledges the critical role of AI in achieving the country's digital economy objectives, signalling a move towards more comprehensive AI governance frameworks.

2.4 The state of AI and data science infrastructure in Africa

Several African countries are making notable strides in building their AI and data science capabilities. Initiatives like the African Institute for Mathematical Sciences (AIMS) are offering advanced training in mathematical sciences, including data science and machine learning, across multiple African countries. These educational programs are instrumental in developing local talent to drive AI innovation (Nakatumba-Nabende et al., 2023). Additionally, tech hubs and innovation centres across the continent, such as iHub in Nairobi and CcHub in Lagos, provide vital support for startups and researchers in AI and data science (Ehimuan et al., 2024). These initiatives foster a culture of innovation and collaboration in sub-Saharan Africa.

In addition, the African Centre of Excellence in Data Science in Rwanda, the AI & Data Science Research Group at Makerere University in Uganda, Data Science Africa, and the Deep Learning Indaba are structures and training programs created to stimulate research and capacity development in AI. The availability of large and diverse datasets for training AI models remains a significant challenge. Many African countries lack comprehensive data collection and management systems, which impedes the development of AI applications tailored to local needs. Moreover, issues related to data privacy and protection are of concern, given the nascent stage of regulatory frameworks in many African nations. However, the potential for AI and data science to drive socioeconomic development in Africa is immense.

2.5 The availability, quality, and representativeness of Sub-Saharan African data

The data landscape in Sub-Saharan Africa is marked by significant challenges related to availability, quality, and representativeness, impacting the development and application of AI technologies across the continent. There is a scarcity of accessible and reliable data sets that accurately reflect the diverse demographics and contexts within African nations. The scarcity is partly due to limited digital infrastructure and the lack of comprehensive data collection and management systems in many countries (Aker & Mbiti, 2010; Alzubaidi et al., 2023). Furthermore, the digital divide exacerbates these challenges, as a significant portion of the population in various African countries lacks access to digital technologies. The digital divide leads to gaps in the collected data, thus affecting the quality and completeness of datasets. These factors contribute to developing AI systems that may not be optimised for local contexts, potentially leading to biased outcomes and inefficiencies (Mittelstadt et al., 2019).

Moreover, the representativeness of data is a critical concern, with existing datasets often failing to capture the full scale of linguistic, cultural, and socio-economic diversity present within the continent. This lack of representativeness can lead to AI models that perform poorly when deployed in different African settings, thereby limiting their effectiveness and applicability. For instance, AI applications in healthcare or agriculture developed using non-representative datasets might not account for local disease patterns or crop varieties, diminishing their utility. Efforts to address these issues involve enhancing data collection methodologies to ensure broader inclusivity and employing advanced machine learning techniques to compensate for data imbalances. Nevertheless, the path forward requires a concerted effort from governments, the private sector, and international partners to build robust digital ecosystems to generate high-quality, representative data for AI and other technological applications in Africa (Ehimuan et al., 2024).

3. Methodology

The research methodology for this study utilised a qualitative approach, focusing on semi-structured interviews with eight experts in AI development, legal expertise, and academia. These participants were selected for their deep knowledge and active involvement in AI. The interviews aimed to gather diverse insights on the challenges and strategies for addressing biases in AI training data across different sectors. Data from the interviews were analysed using thematic analysis, which allowed for the identification and interpretation of key themes from the discussions. The process involved data familiarisation, initial coding, searching for themes, reviewing themes, and finally defining and naming the themes. This structured approach ensured that the findings were based on the actual data provided by the participants, facilitating a grounded understanding of the complex issues surrounding AI bias mitigation in Sub-Saharan Africa.

4. Results and discussion

Word cloud (Fig:1) provides a visual representation of the frequency of words mentioned, with words like biases, fairness, ethical, legal, standards and jurisdictions being prominent, reflecting the participant’s views on the technical and legal challenges in mitigating biases in AI training data.

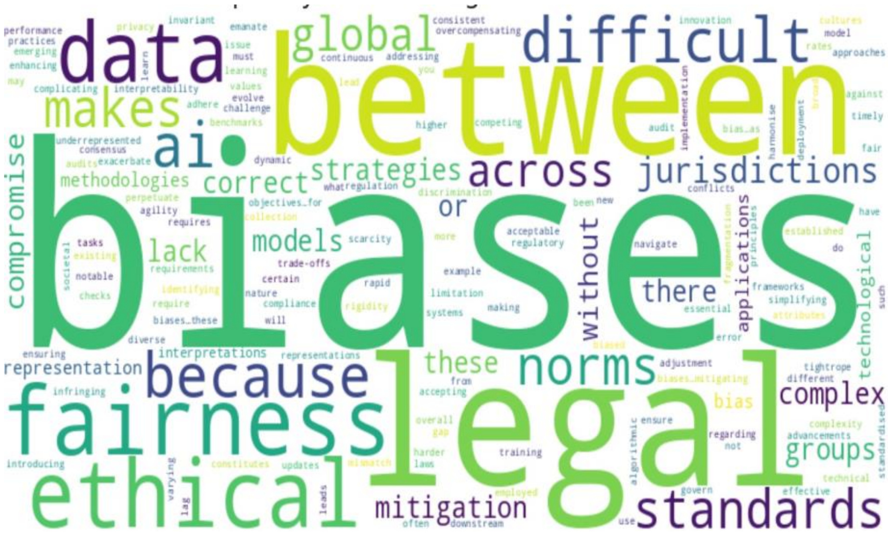

Figure 1: Word cloud

4.1 Technical challenges in mitigating biases

4.1.1 Data scarcity

The challenge of mitigating biases in AI models is pronounced in regions like Sub-Saharan Africa, where data scarcity for underrepresented groups is a significant issue. The scarcity stems from the limited digital infrastructure and the lack of diverse data collection initiatives, leading to datasets that do not fully capture the demographic and cultural diversity of the region. Efforts to correct these biases by augmenting datasets or modifying algorithms must be carefully managed to avoid introducing new biases. Correcting biases in training data may be complicated by the dynamic nature of societal norms and values across different African societies. As these norms and values evolve, continuous adjustment and localisation of mitigation strategies become crucial, necessitating a tailored approach that considers the unique socio-cultural context of Sub-Saharan Africa (Ogbonnaya-Ogburu et al., 2020).

"There is the issue of data scarcity for certain groups, making it difficult to correct biases without overcompensating and introducing new biases because of the dynamic nature of bias…as societal norms and values evolve, it will require continuous adjustment of mitigation approaches" (#5).

Adapting bias mitigation strategies to the evolving societal norms in Sub-Saharan Africa requires flexible models. This requires a proactive engagement with local communities to understand the changing landscape of fairness and representation and to develop AI systems responsive to these shifts. Such engagement can inform the development of adaptive models capable of self-assessment and adjustment in response to identified biases of the region. However, operationalising these adaptive, context-aware models within the technological and infrastructural constraints in many parts of Sub-Saharan Africa poses a complexity regarding local knowledge and resources.

4.1.2 Model complexity and interpretability

The trade-off between model complexity and interpretability presents a challenge when auditing for biases in artificial intelligence. The challenge is acutely felt in Sub-Saharan Africa, where the diversity of languages, cultures, and socio-economic conditions necessitates AI models that are both sophisticated and interpretable. While complex models may capture the intricacies of diverse datasets more accurately, they become less transparent, making it difficult to identify and eliminate biases (Raji & Buolamwini, 2019). The complexity impedes efforts to ensure that AI systems are fair and equitable across all demographics in regions like Sub-Saharan Africa, where digital technologies have the potential to impact societal development.

"A limitation may emanate from a compromise between model complexity and interpretability because more complex models are harder to audit for biases…mitigating these biases often requires simplifying models or accepting higher error rates for certain groups, which can compromise overall performance" (#2).

Mitigating biases in training data in complex models may involve a trade-off either by simplifying the models to enhance interpretability or by accepting higher error rates for certain groups to maintain model complexity. In Sub-Saharan Africa, simplified models may not adequately capture the rich, diverse datasets representative of the region's population, leading to biased outcomes that can exacerbate existing inequalities. However, accepting higher error rates for marginalised groups can further entrench disparities in access to services and opportunities facilitated by AI technologies (Mohamed et al., 2020).

4.1.3 Lack of standardised benchmarks on fairness

The lack of established benchmarks defining acceptable fairness hinders initiatives to achieve fairness in machine learning (ML) and artificial intelligence (AI) systems. The absence of benchmarks reflects a broader challenge because of the complex nature of fairness, which varies across contexts, cultures, and legal systems. Dwork et al. (2012) and Nakao et al. (2022) argue that fairness cannot be distilled into a single metric or definition but rather should be understood as a spectrum of considerations that balance societal values, individual rights, and the technical constraints of AI systems. The rapid pace of technological advancement leads to a scenario where practitioners often rely on ad-hoc measures to assess fairness (Barocas et al., 2023).

"There is a lack of standardised benchmarks for what constitutes acceptable fairness and the complex trade-offs between competing legal and ethical objectives…for example, the ethical collection and use of diverse data in which you must navigate the tightrope of enhancing representation without infringing on privacy" (#3).

The ethical collection and utilisation of data to enhance representation without violating privacy rights reveals the intricate balance required to develop equitable AI systems. Efforts to improve the diversity of datasets are a decisive step toward mitigating biases but must be weighed against the potential risks to privacy and consent. This is true for marginalised groups that may be disproportionately impacted by data misuse. Kasy and Abebe (2021) note that while inclusive data practices are essential for fairness, they must not infringe upon individual rights to privacy. The trade-off is exacerbated by varying global data protection and privacy standards, which set stringent data handling guidelines that may inadvertently constrain efforts to collect diverse datasets (Veale & Binns, 2017). The development of universally accepted benchmarks for fairness in AI is a technical, ethical, and philosophical endeavour that requires broad consensus among stakeholders.

4.2 Technical approaches in mitigating bias in training data

4.2.1 Bias audits and compliance checks

Bias audits and compliance checks against established fairness standards represent pivotal methodologies in mitigating biases in AI systems. The deployment of AI technologies must be carefully audited to ensure they do not exacerbate existing inequalities or introduce new forms of discrimination (Raji & Buolamwini, 2019). By implementing rigorous bias audits, organisations can systematically assess the fairness of their AI systems, identifying and addressing potential biases that may disproportionately affect marginalised or underrepresented groups.

"Bias audits and compliance checks against established fairness standards have been effective technical methodologies for identifying and addressing biases…these practices are essential for ensuring AI systems adhere to ethical norms and legal requirements regarding discrimination and fairness" (#8).

Compliance checks against established fairness standards ensure that AI systems meet ethical norms and legal requirements. This is fundamental in fostering trust and acceptance of AI technologies within diverse communities across Sub-Saharan Africa. However, the effectiveness of bias audits and compliance checks hinges on the availability and applicability of fairness standards that resonate with the cultural and socio-economic realities of Sub-

Saharan Africa. Current standards, often developed in Western contexts, may fail to capture fairness and discrimination encountered in the region (Mohamed et al., 2020).

4.2.2 Algorithmic fairness approaches

Algorithmic fairness approaches, like fair representation learning, have emerged as potent methodologies for addressing biases within AI systems by learning data representations invariant to biased attributes. The technique is pertinent in Sub-Saharan Africa, where diverse socio-cultural dynamics necessitate AI models that fairly represent the region's myriad communities. These approaches aim to provide more equal outcomes across multiple AI applications, such as healthcare diagnostics and financial services, by abstracting features in a way that reduces the influence of biased attributes (Zemel et al., 2013; Starke et al., 2022). The effectiveness of such methodologies in Sub-Saharan Africa hinges on their ability to encapsulate the diverse understanding of fairness within the region.

"Algorithmic fairness approaches, such as fair representation learning, can be employed to correct biases in training data because these methodologies learn representations of the data invariant to the biased attributes to ensure that the downstream tasks do not perpetuate or exacerbate biases" (#1).

Advancing algorithmic fairness in Sub-Saharan Africa requires continuous innovation and research tailored to the region's specific needs and challenges. However, the scarcity of localised data that accurately reflects the continent's diverse populations is necessary for training algorithms that are truly representative and unbiased (Sambasivan et al., 2021). Moreover, the complexity inherent in discerning and quantifying biased attributes within these datasets necessitates a deep understanding of the local socio-cultural structure.

4.3 Legal challenges in mitigating biases

4.3.1 Rapid technological advancements

The inconsistency between the rapid pace of technological advancements in AI and the slower evolution of regulatory frameworks is a global challenge. Sub-Saharan Africa faces unique challenges in crafting and updating regulations that can keep pace with AI advancements because of the diverse socio-economic landscapes and varying levels of technological adoption. The agility of technological innovation often surpasses the capacity of existing legal systems to provide timely and effective oversight. The lag in regulatory implementation can lead to a regulatory vacuum where emerging AI applications operate without comprehensive guidance. This scenario can result in ethical dilemmas and governance issues that may hinder the potential benefits of AI technologies or exacerbate existing inequalities (Karimi et al., 2018). The need for regulatory frameworks that are both flexible and robust enough to adapt to new developments in AI is critical to ensuring these technologies are deployed responsibly and equitably across the continent.

"The lag between rapid technological advancements and regulatory updates makes it difficult for existing laws to govern emerging AI applications.

The gap between the agility of technological innovation and the rigidity of legal frameworks makes timely regulation difficult" (#6).

The disparity in legal and institutional capacities across Sub-Saharan Africa exacerbates the lag in regulation updates compared to technological developments. While some countries may possess the infrastructural and regulatory foundation to keep pace with AI developments, others may lack the necessary resources or expertise, leading to disparities in governance and oversight of AI technologies. The inconsistency can impede the harmonisation of AI regulations across the region, potentially creating barriers to cross-border technological collaboration and innovation. Moreover, the absence of region-wide standards for AI ethics and governance may leave vulnerable populations at risk of harm from unregulated AI applications (Ntoutsi et al., 2020).

4.3.2 Lack of global consensus on ethical norms

The lack of global consensus on ethical norms impedes harmonising legal standards across jurisdictions, especially in AI. The challenge is acutely felt in Sub-Saharan Africa because the region is characterised by a rich tapestry of cultures, languages, and legal systems. As AI technologies continue to permeate various aspects of society, the diversity within the region makes establishing uniform ethical guidelines and legal frameworks challenging (Jobin et al., 2019). The disparity between broad ethical principles and their legal interpretations across different countries exacerbates the difficulty of implementing AI governance that is effective and culturally sensitive. This dissonance hampers the development of AI technologies that are equitable and just but also restricts the ability of nations within the region to collaborate on AI initiatives and share technological advancements.

"A notable challenge is the lack of global consensus on ethical norms, which makes it difficult to harmonise legal standards across jurisdictions. This leads to a mismatch between the broad ethical principles and the legal interpretations" (#4).

The absence of a global consensus on ethical norms for AI is complicated by the rapid pace of technological innovation which often surpasses the ability of regulatory bodies to adapt. In Sub-Saharan Africa, where regulatory capacities and digital infrastructures may vary significantly across countries, aligning emerging AI applications with existing legal norms becomes an even more daunting task (Floridi & Cowls, 2022). This misalignment can lead to situations where AI technologies are deployed without adequate oversight, potentially resulting in ethical breaches and societal harm. Moreover, the varied interpretations of ethical principles across the region necessitate a flexible approach that can accommodate diverse values and ethical considerations while ensuring the responsible development and deployment of AI systems.

4.3.3 Interpretations of fairness

The diversity of cultural norms and legal systems across Sub-Saharan Africa and the wider world presents a significant challenge to the global deployment of AI technologies. The concept of fairness, a cornerstone in the ethical development and application of AI, is subject to varying interpretations depending on cultural and jurisdictional contexts. This variability complicates the creation of universally applicable AI systems, as what is considered fair in one region may not hold the same meaning in another. For instance, a study conducted by Mhlambi (2020) highlighted how Western-centric AI ethics frameworks may not adequately address or respect the values and societal norms prevalent in Sub-Saharan Africa.

"The varying interpretations of fairness across different cultures and jurisdictions can lead to conflicts in global AI applications because of the fragmentation of legal standards across jurisdictions, complicating the global deployment of AI and the consistent implementation of bias mitigation strategies" (#7).

The fragmented legal framework poses a challenge for multinational corporations and global AI initiatives seeking to deploy technologies across Sub-Saharan Africa. The lack of harmonised legal standards means that AI developers must navigate a patchwork of regulations, which can hinder the efficient and equitable distribution of AI technologies. Moreover, the attempt to implement bias mitigation strategies that align with diverse legal and cultural standards can lead to inconsistent applications of AI. Some bias mitigation strategies can exacerbate existing inequalities or introduce new forms of bias (Benjamin, 2019). The resulting conflicts undermine the trust in and the efficacy of AI systems but also pose a risk to the societal acceptance of these technologies, especially in regions with a deep mistrust of systems perceived as foreign or neo-colonial.

5. Conclusion and recommendations

The study reveals the challenges of mitigating biases in AI training data within Sub-Saharan Africa, where technological advancement intersects with diverse socio-cultural diversity. The pervasive issues of data scarcity, model complexity versus interpretability, and the absence of standardised benchmarks for fairness pose significant obstacles. These technical challenges are further compounded by the evolving societal norms and the sticky regulatory frameworks, which struggle to keep pace with rapid technological advancements. The findings emphasise the need for ongoing adaptation of bias mitigation strategies sensitive to the local context and responsive to global technological trends.

Bias mitigation strategies should include enhancing collaborative efforts among governments and all the stakeholders to improve the diversity and accuracy of data collection, reflecting the true demographic and cultural diversity of Sub-Saharan Africa. Moreover, there is an urgent need to establish localised fairness benchmarks developed in consultation with various stakeholders, including local communities, to ensure they embody the region's diverse ethical and cultural standards. Additionally, these efforts should be supported by significant investments in strengthening the regulatory frameworks within the region, coupled with robust capacity-building initiatives that empower local policymakers, AI developers, and regulatory bodies with the necessary expertise to navigate the complexities of AI deployment. These initiatives will be necessary for ensuring that AI technologies not only advance technological progress but also reinforce the ethical, legal, and social frameworks that support equitable development across Sub-Saharan Africa.

6. References

Ademuyiwa, I., & Adeniran, A. (2020). Assessing Data Protection and Privacy in Africa. In *Assessing Digitalization and Data Governance Issues in Africa* (4–6). Centre for International Governance Innovation. http://www.jstor.org/stable/resrep25330.7

Aker, J. C., & Mbiti, I. M. (2010). Mobile phones and economic development in Africa. *Journal of economic Perspectives*, *24*(3), 207-232. DOI: 10.1257/jep.24.3.207

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A.S., Al-dabbagh, B.S.N., Fadhel, M.A., Manoufali, M., Zhang, J., Al-Timemy, A.H. and Duan, Y., 2023. A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. *Journal of Big Data*, *10*(1), 46. DOI: 10.1186/s40537-023-00727-2

Barocas, S., Hardt, M., & Narayanan, A. (2023). *Fairness and machine learning: Limitations and opportunities*. MIT Press.

Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *Calif. L. Rev.*, *104*, 671. DOI: 10.2139/ssrn.2477899

Belkacemi, Z., Gkeka, P., Lelièvre, T., & Stoltz, G. (2021). Chasing collective variables using autoencoders and biased trajectories. *Journal of chemical theory and computation*, *18*(1), 59-78. https://doi.org/10.48550/arxiv.2104.11061

Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. John Wiley & Sons.

Budiman, A. (2016). Distributed averaging cnn-elm for big data. https://doi.org/10.48550/arxiv.1610.02373

Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the 'good society': the US, EU, and UK approach. *Science and engineering ethics*, *24*, 505-528. http://dx.doi.org/10.2139/ssrn.2906249

Chakraborty, J., Majumder, S., Yu, Z., & Menzies, T. (2020). Fairway: a way to build fair ml software. In *Proceedings of the 28th ACM joint meeting on European software engineering conference and symposium on the foundations of software engineering* (654-665). https://doi.org/10.1145/3368089.3409697

Dai, H., Liu, Z., Liao, W., Huang, X., Cao, Y., Wu, Z., Zhao, L., Xu, S., Liu, W., Liu, N. & Li, S. (2023). Auggpt: leveraging chatgpt for text data augmentation. https://doi.org/10.48550/arxiv.2302.13007

Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012). Fairness through Awareness. *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, 214-226. DOI: 10.48550/arXiv.1104.3913

Ehimuan, B., Anyanwu, A., Olorunsogo, T., Akindote, O. J., Abrahams, T. O., & Reis, O. (2024). Digital inclusion initiatives: Bridging the connectivity gap in Africa and the USA–A review. *International Journal of Science and Research Archive*, *11*(1), 488-501. DOI: 10.30574/ijsra.2024.11.1.0061

Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545. DOI: 10.2139/ssrn.3831321

Gerard, C. (2021). Bias in machine learning. *Practical Machine Learning in JavaScript: TensorFlow. js for Web Developers*, 305-316. https://doi.org/10.1007/978-1-4842-6418-8_7

Gianfrancesco, M., Tamang, S., Yazdany, J., & Schmajuk, G. (2018). Potential biases in machine learning algorithms using electronic health record data. *Jama Internal Medicine*, 178(11), 1544. https://doi.org/10.1001/jamainternmed.2018.3763

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1, 389–399. DOI: 10.1038/s42256-019-0088-2

Karimi, F., Génois, M., Wagner, C., Singer, P., & Strohmaier, M. (2018). Homophily influences ranking of minorities in social networks. *Scientific reports*, *8*(1), 11077. DOI: 10.1038/s41598-018-29405-7

Kasy, M., & Abebe, R. (2021). Fairness, Equality, and Power in Algorithmic Decision-Making. *Foundations and Trends® in Econometrics*, 14(1-2), 1-144. DOI: 10.1145/3442188.3445919

Kauwe, S., Welker, T., & Sparks, T. (2020). Extracting knowledge from dft: experimental band gap predictions through ensemble learning. *Integrating Materials and Manufacturing Innovation*, *9*(3), 213-220. https://doi.org/10.1007/s40192-020-00178-0

Koops, B. J. (2014). The trouble with European data protection law. *International Data Privacy Law*, 4(4), 250-261. DOI: 10.1093/idpl/ipu023

Kshetri, N. (2019) Cybercrime and Cybersecurity in Africa, *Journal of Global Information Technology Management,* *22*(2), 77-81. DOI: 10.1080/1097198X.2019.1603527

Maeda, T. (2018). Technical note: how to rationally compare the performances of different machine learning models?. *PeerJ Preprints* 6:e26714v1 https://doi.org/10.7287/peerj.preprints.26714

Mhlambi, S. (2020). From rationality to relationality: ubuntu as an ethical and human rights framework for artificial intelligence governance. *Carr Center for Human Rights Policy Discussion Paper Series*, *9*, 31.

Mittelstadt, B., Russell, C., & Wachter, S. (2019). Explaining explanations in AI. In *Proceedings of the conference on fairness, accountability, and transparency* (279-288). DOI: 10.1145/3287560.3287574

Mohamed, S., Png, M. T., & Isaac, W. (2020). Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33, 659-684. DOI: 10.1007/s13347-020-00405-8

Nakao, Y., Stumpf, S., Ahmed, S., Naseer, A., & Strappelli, L. (2022). Toward involving end-users in interactive human-in-the-loop AI fairness. *ACM*

Transactions on Interactive Intelligent Systems (TiiS), *12*(3), 1-30. https://doi.org/10.48550/arXiv.2204.10464

Nakatumba-Nabende, J., Suuna, C., & Bainomugisha, E. (2023). AI Ethics in Higher Education: Research Experiences from Practical Development and Deployment of AI Systems. In *AI Ethics in Higher Education: Insights from Africa and Beyond* (pp. 39-55). Cham: Springer International Publishing. DOI: 10.1007/978-3-031-23035-6_4

Nolte, M., Kister, N., & Maurer, M. (2018). Assessment of deep convolutional neural networks for road surface classification. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC).* https://doi.org/10.1109/itsc.2018.8569396

Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., & Kompatsiaris, I. (2020). Bias in data-driven artificial intelligence systems—An introductory survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1356. DOI: 10.1002/widm.1356

Ogbonnaya-Ogburu, I. F., Smith, A. D. R., To, A., & Toyama, K. (2020). Critical race theory for HCI. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1-16. https://doi.org/10.1145/3313831.3376392

Orr, L., Sanyal, A., Ling, X., Goel, K., & Leszczynski, M. (2021). Managing ml pipelines: feature stores and the coming wave of embedding ecosystems.. https://doi.org/10.48550/arxiv.2108.05053

Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Harvard University Press.

Qu, Y., Ding, Y., Liu, J., Liu, K., Ren, R., Zhao, W.X., Dong, D., Wu, H. and Wang, H. (2020). Rocketqa: an optimized training approach to dense passage retrieval for open-domain question answering. https://doi.org/10.48550/arxiv.2010.08191

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *AIES '19: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 429-435. https://doi.org/10.1145/3571151

Sambasivan, N., Kapania, S., Highfill, H., Akrong, D., Paritosh, P., & Aroyo, L. M. (2021). Everyone wants to do the model work, not the data work: Data Cascades in High-Stakes AI. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-15. https://doi.org/10.1145/3411764.3445518

Shang, B., & Wang, K. (2016). A data flow model to solve the data distribution changing problem in machine learning. In *Itm Web of Conferences*, 7, 05012. https://doi.org/10.1051/itmconf/20160705012

Shi, Y., Sagduyu, Y., Davaslioglu, K., & Li, J. (2018). Active deep learning attacks under strict rate limitations for online api calls. In *2018 IEEE international symposium on technologies for homeland security (HST)* (1-6). IEEE. https://doi.org/10.1109/ths.2018.8574124

Singh, B. K., & Sinha, G. R. (2022). *Machine Learning in Healthcare: Fundamentals and Recent Applications*. CRC Press. (107-133). https://doi.org/10.1201/9781003097808-7

Starke, C., Baleis, J., Keller, B., & Marcinkowski, F. (2022). Fairness perceptions of algorithmic decision-making: A systematic review of the empirical literature. *Big Data & Society*, *9*(2), 20539517221115189. https://doi.org/10.1177/20539517221115189

Thompson, H.M., Sharma, B., Bhalla, S., Boley, R., McCluskey, C., Dligach, D., Churpek, M.M., Karnik, N.S. & Afshar, M. (2021). Bias and fairness assessment of a natural language processing opioid misuse classifier: detection and mitigation of electronic health record data disadvantages across racial subgroups. *Journal of the American Medical Informatics Association*, 28(11), 2393-2403. https://doi.org/10.1093/jamia/ocab148

Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2), 1-17. https://doi.org/10.1177/2053951717743530

Wang, M., & Deng, W. (2020). Mitigating Bias in Face Recognition Using Skewness-Aware Reinforcement Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (9319-9328). IEEE. https://doi.org/10.48550/arXiv.1911.10692

Wang, T., Zhao, J., Yatskar, M., Chang, K., & Ordóñez, V. (2019). Balanced datasets are not enough: estimating and mitigating gender bias in deep image representations. In *Proceedings of the IEEE/CVF international conference on computer vision* (5310-5319). https://doi.org/10.1109/iccv.2019.00541

Yeung, K. (2017). Hypernudge: Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118 –136. DOI: 10.1080/1369118X.2016.1186713

Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). Learning Fair Representations. *Proceedings of the 30th International Conference on Machine Learning*, 28, 325-333.