NLA – Delta Library Journal
Vol. 13, June, 2025. pp. 1 – 12 (Special Edition)

Bias and Fairness in AI-Driven Library Services in Ensuring Equity in Adult and Non-Formal Education

Habiba Muhammad

Department of Adult and Non-Formal Education Shehu Shagari College of Education, Sokoto

bibahmuhd@gmail.com

Received: 19, November 2024 **Accepted:** 30, May 2025 **Published:** 30, June 2025

Abstract
The integration of Artificial Intelligence (AI) in library services holds great potential to enhance access to educational resources, especially for adult and non-formal education. However, despite these advancements, concerns have arisen about the bias and fairness embedded in AI systems. As libraries increasingly rely on AI-technology for educational advancement, the risk of perpetuating social and educational inequalities increases. Bias in AI can result from historical data, algorithmic design, or unintended consequences of system deployment, which can disproportionately affect marginalized groups like the adult and non-formal learners. This paper explores the role of AI systems in adult and non-formal education. The study also examines the impacts of bias and fairness in AI-driven library services on adult and non-formal education. Drawing on the current works and best practices, this paper proposes strategies for mitigating bias and promoting fairness in AI-driven library services in relation to adult and non-formal education.
Keywords: AI, bias, fairness, equity, library services, adult and non-formal education.

Introduction

Artificial Intelligence (AI) technology is rapidly being incorporated into library services, particularly in the area of adult and non-formal education. However, the issue of bias and fairness have emerged as crucial concerns (Crawford, 2021). Libraries are evolving from traditional information axis to dynamic platforms that leverage on AI to personalize learning experiences, recommend resources, and facilitate access to educational materials. As much as AI holds great prospective for enhancing equity and inclusion, the risks of perpetuating bias and inequities remain important (Eubanks, 2018).

Studies have identified biases against certain groups in AI systems, such as the facial recognition systems studied by Buolamwini and Gebru (2018) and hiring algorithms examined by Dastin (2018). According to Eubanks (2018); Kleinberg et al. (2018), these biases can perpetuate systemic discrimination and inequality, with negative effects on individuals and communities in areas like hiring, lending and criminal justice. However, various mitigation strategies have been proposed, such as improving data quality (Asan et al., 2020) and designing explicitly fair algorithms (Berk et al., 2018; Yan et al., 2020).

The growing application of AI technologies has intensified debate around bias and fairness as their potential negative effect on individual become more evident. This study thus examines bias and fairness in AI-driven library services with the sole purpose of ensuring equity in adult and non-formal education.

Statement of the Problem

Artificial Intelligence (AI) has been a transformative force, reshaping the landscape of learning and teaching methodologies. According to O'Neil (2016), AI applications in education have grown speedily, with systems designed to adapt to the learning pace of students, thereby enhancing engagement and understanding. Language translation and support facilitated by AI have also significantly contributed to breaking down language barriers in education, making learning materials accessible to the global audience (Yan et al., 2020).

Despite the benefits, the deployment of AI in education raises critical issues of fairness and bias (Caliskan et al., 2017). AI systems can inadvertently perpetuate and amplify existing biases present in the data they are trained on, leading to unfair outcomes, particularly for marginalized groups like the adult and non-formal learners (Ferrara, 2023). As Artificial Intelligence (AI) is being increasingly integrated into library services, particularly in the realm of adult and non-formal education, questions of bias and fairness have emerged as critical concerns. Addressing these issues is crucial for creating fair and equitable educational systems, thus the need to address the rising issue of bias and fairness in AI system.

Objective of the Study

The main objective of this study is to examine bias and fairness in AI-driven library services in a bid to ensure equity in adult and non-formal education. However, the specific objectives are to:

1. Identify the types of bias and fairness in AI system in adult and non-formal education.
2. Determine the role of AI system in adult and non-formal education.
3. Evaluate the effect of bias and fairness in AI system in adult and non-formal education.

Literature Review

The review of related literature covers the followings areas:

Concept of Bias and Fairness

Bias is defined as a systematic error in decision-making processes that result in unfair outcomes (Ferrara, 2024). In the context of AI, bias can arise from various sources, including data collection, algorithm design, and human interpretation. For example, machine learning models, which are a type of AI system, can learn and replicate patterns of bias present in the data used to train them, resulting in unfair or discriminatory outcomes. The Merriam-Webster Dictionary

(2024) defined bias as an inclination of temperament or outlook, especially a personal and sometimes unreasoned judgment. On a general term, bias is a prejudice or predisposition towards one side of an issue, person or group, often in a way considered to be unfair. In machine learning, bias can refer to both algorithmic bias, when models systematically produce prejudiced results and bias in model range such as using incorrect assumptions in a model (Barocas et al., 2019). Example is a facial recognition algorithm that performs poorly on dark skinned individual due to biased training data. In the context of library services, these biases could manifest in the way resources are recommended, the way information is categorized, or even in the way access to services is provided (Ferrara, 2024).

On the other hand, fairness in AI, according to Ferrara (2024), refers to the ability of an AI system to treat all individuals equitably, ensuring that no particular group is disadvantaged or discriminated against. AI fairness in the educational context refers to the principle of ensuring that the use of AI systems and algorithms in educational settings does not lead to unfair or biased outcomes for students (Ferrara, 2024). Fairness can also be defined as treating people in a way that is free from bias, favouritism or injustice.

In a legal and ethical context, fairness is the equitable and impartial treatment of all individuals under the law or moral, ensuring that no one receives undue advantage or harm due to bias (Rawis, 2021). Educational context, fairness means equal access to learning opportunities and non-discriminatory assessments for students of different backgrounds, abilities and needs (Tierney and Simon, 2020). In data science or machine learning, fairness refers to the absence of bias or discrimination in automated decisions, ensuring outcomes do not systematically disadvantage certain groups (Barocas at al., 2019). Achieving fairness requires careful design, continuous monitoring, and assessment of AI systems to ensure they meet ethical and equitable standards.

There are two main fairness notions in education: individual fairness and group fairness. Individual fairness in education refers to the principle that students with similar academic abilities and educational backgrounds should receive similar treatment, regardless of their demographic characteristics. This fairness notion ensures that educational opportunities, resources, and outcomes are distributed equitably among students, rather than being influenced by factors such as race, gender, or socioeconomic status (Fleisher, 2020; Heidari et al., 2019). For example, students who demonstrate comparable academic performance and have similar potential should have access to comparable educational resources, experiences, support, and outcomes, without any discrimination or bias (Gillen et al., 2018). Implementing it in education can involve techniques like personalized learning, adaptive assessment, etc., to address the unique needs and circumstances of each student (Hu & Rangwala, 2020).

Group fairness in education on the other hand, refers to the principle that groups of students should be treated fairly, regardless of their individual characteristics. This approach aims to minimize disparities in educational opportunities, resources, and outcomes based on group membership, such as race, gender, or socioeconomic status (Hu & Rangwala, 2020). Group fairness measures in educational data mining and student prediction models focus on ensuring that the model's predictions or decisions do not systematically favor or disadvantage any particular group. It is essential to ensure that these applications avoid inequitable outcomes and negative impacts on learning for certain groups of students (Hu & Rangwala, 2020).

It is crucial to understand the specific types of biases that these systems can exhibit. These biases significantly affect the fairness and accuracy of AI applications in educational settings. Below are the potential biases categorized into three types: Data-related Bias, Algorithmic Bias, and User-Interaction Bias.

Data-Related Biases

The biases observed in AI systems derive from fundamental flaws within the datasets utilized for training and evaluation (Baker & Yacef, 2019). These biases reflect societal imbalances and inaccuracies inherent in data gathering and representation methodologies. As a result, these biases have the potential to result in AI systems that exhibit unequal performance among various student populations, hence intensifying pre-existing imbalances within educational environments.

Algorithmic-Related Biases

Addressing algorithmic-related biases in education involves a deep dive into how the design and implementation of algorithms can inadvertently lead to unfair outcomes. This exploration is critical, given the increasing reliance on AI to personalize learning, assess student performance, and even make administrative decisions in educational settings.

User-Interaction Related Biases

User Interaction-Related Biases in educational AI systems encompass biases that emerge from interactions between users (such as students and educators) and AI technologies. These biases can significantly influence the behavior and evolution of AI systems over time, potentially leading to unfair or discriminatory outcomes.

The Role of AI in Adult and Non-Formal Education

Artificial Intelligence (AI) has the potential to transform adult and non-formal education by providing tailored, efficient, and accessible learning opportunities. The integration of AI into these fields addresses specific challenges and opens up new rooms for learners of all ages and background (Rehm & Siegel, 2020). According to them, adult and non-formal education programme often serve diverse learners who come from varied socio-economic, cultural, and educational backgrounds. AI technologies, such as machine learning, natural language processing, and data analytics, are becoming essential in enhancing learning experiences, personalizing educational settings, and supporting lifelong learning (Rehm & Siegel, 2020).

In addition, Rehm and Siegel (2020) opined that AI-driven library services can play a transformative role in supporting these learners by offering: (1) Personalized Learning Paths: AI can analyze a learner's history, preferences, and goals to suggest tailored resources helping learners access information relevant to their specific needs, (2) Adaptive Learning Tools. AI can power adaptive learning systems that adjust the level of content based on a learner's progress, ensuring that content remains challenging but not overwhelming, (3) Automatic Resource Recommendations. By analyzing patterns in reading behavior or preferences, AI can recommend materials that learners might otherwise miss, exposing them to diverse viewpoints and content, (4) Access to Educational Support. AI-powered chatbots and virtual assistants can offer 24/7 support to adult learners, answering questions about library resources, providing assistance with search techniques, or explaining how to use library databases and (5) Language Support and Translation. AI-driven translation tools can break down language barriers, enabling access to resources in multiple languages, thus helping non-native speakers to engage in learning.

Effects of bias and fairness in AI system

The rapid advancement of Artificial Intelligence (AI) has brought numerous benefits, but it also comes with potential risks and challenges. One of the key concerns is the negative effects of bias in AI on individuals and society (Sweeney, 2013). According to Sweeney (2013), bias in AI can perpetuate and even amplify existing inequalities, leading to discrimination against marginalized groups and limiting their access to essential services.

To ensure that AI systems are fair and equitable to serve the needs of all users, it is crucial to identify and lessen bias in AI. The use of biased AI has numerous ethical implications, including the potential for discrimination, undermining public trust in technology, and limiting human agency and autonomy (Sweeney, 2013).

Fairness in AI has also received a lot of attention in both academic and industry circles. At its core, fairness in AI refers to the absence of bias or discrimination in AI systems, which can be challenging to achieve due to the different types of bias that can arise in these systems (Sweeney, 2013). While fairness and bias are closely related concepts, they differ in important ways, including that fairness is inherently a deliberate and intentional goal, while bias can be unintentional.

According to Mehrabi, et al. (2021), the effects of bias and fairness in AI systems are significant and wide-ranging, especially as these systems are increasingly integrated into high-impact areas such as healthcare, finance, education, hiring, policing, and social services. Below is a structured overview of these effects:

Negative Effects of Bias in AI Systems

1. Discrimination and Inequality: AI systems trained on biased data can reinforce or even amplify existing social inequalities.
2. Loss of Trust and Credibility: When users perceive AI systems as unfair or biased, public trust in both the technology and the institutions using it declines.
3. Legal and Ethical Consequences: Biased AI can result in unlawful discrimination, violating human rights or anti-discrimination laws. Organizations may face legal challenges, regulatory fines, or reputational damage.

4. Exclusion of Marginalized Groups: Certain groups may be underserved or misrepresented in datasets (e.g., racial minorities, people with disabilities), leading to poor outcomes or denial of services.
5. Feedback Loops: Biased outcomes can perpetuate further bias by reinforcing flawed patterns in future training data.

Positive Effects of Fairness in AI Systems

1. Promotes Equity and Inclusion: Fair AI systems ensure that all users, regardless of gender, race, age, or socioeconomic background, receive equitable treatment and outcomes.
2. Improved Decision-Making: Removing bias improves the accuracy and reliability of AI outputs, leading to better decisions in areas like loan approval, medical diagnosis, and job recruitment.
3. Enhances Trust and Adoption: Fair systems are more likely to gain user trust, leading to wider and more confident adoption of AI technologies.
4. Regulatory Compliance: Ensuring fairness helps organizations meet legal and ethical standards, reducing risk and promoting accountability.
5. Innovation and Diversity: Fair systems are more likely to consider diverse perspectives and needs, leading to more innovative and broadly useful solutions.

Mitigation strategies for bias in AI system

Mitigating bias in AI systems requires a multifaceted approach, focusing on data, model development and deployment. Strategies include ensuring diverse and representative data, employing explainable AI techniques, and implementing human-in-the-loop systems for oversight. Mitigating bias in AI systems is essential for ensuring fairness, accountability, and trust (Raji & Buolamwini, 2019).

Buolamwini and Gebru (2018) opined that various approaches have been proposed to mitigate bias in AI system by researchers, including dataset augmentation, bias-aware algorithms, and user feedback mechanisms. Dataset augmentation involves adding more diverse data to training datasets to increase representativeness and reduce bias. Bias-aware algorithms involve designing algorithms that consider different types of bias and aim to minimize their impact on the system's outputs. User feedback mechanisms involve soliciting feedback from users to help identify and correct biases in the system. However, each approach has its limitations and challenges, such as the lack of diverse and representative training data, the difficulty of identifying and measuring different types of bias, and the potential trade-offs between fairness and accuracy.

As the use of Artificial Intelligence (AI) continues to grow, ensuring fairness in its decision-making is becoming increasingly important. To address this challenge, various approaches have been developed, including group fairness and individual fairness. However, these approaches are not without limitations and challenges, such as trade-offs between different types of fairness and the difficulty of defining fairness itself (Kamiran & Calders, 2012).

Despite these challenges, mitigating bias in AI-driven library services is essential for creating fair and equitable systems that benefit all individuals and society (Kamiran & Calders, 2012). It is therefore important to continue to investigate and develop these approaches to create AI systems that are more equitable and fairer for all users especially as it relates to adult and non-formal education. In addition, the key strategies used to reduce or eliminate bias at different stages of the AI development lifecycle include:

1. Data-Level Mitigation Strategies: These focus on improving the quality and balance of the dataset used to train AI models. This can be achieved through

Data Collection Diversity: Ensure data represents all relevant groups (e.g., race, gender, age). Avoid over- or under-representation of particular populations.

Data Preprocessing (Debiasing): Techniques like re-sampling (over/under-sampling) to balance classes.

Anonymization and De-identification: Remove sensitive attributes (e.g., gender, ethnicity) to reduce bias though this may not always eliminate indirect bias.

2. Algorithm-Level Mitigation Strategies: These approaches modify how the model learns from data. Example includes:

Fairness-Aware Learning Algorithms: Use models that optimize for both accuracy and fairness metrics (e.g., Equal Opportunity, Demographic Parity).

Regularization Techniques: Add fairness constraints or penalties in the loss function to reduce bias during training.

Adversarial Debiasing: Train a model while simultaneously training an adversary to detect bias. The main model learns to make accurate predictions without revealing sensitive attributes.

3. Evaluation-Level Mitigation Strategies: This involves monitoring and assessing model outputs for fairness and bias using

Fairness Metrics: like sstatistical parity, equal opportunity and equalized odds

Predictive parity: Choose the appropriate metric based on the domain and fairness goals.

Audit and Benchmarking: Regular audits with tools like IBM's AI Fairness 360, Google's What-If Tool, or Microsoft's Fair learn.

4. Post-processing Mitigation Strategies: Adjust predictions after the model is trained, without changing the model itself using the followings:

Threshold Adjustment: Modify decision thresholds for different demographic groups to achieve fairness.

Output Recalibration: Recalibrate scores or rankings to ensure more equitable treatment.

5. Organizational and Policy Strategies: Technical solutions must be supported by strong governance through the followings:

Bias Impact Assessments: Conduct structured assessments before and after deployment to identify potential sources of bias.

Interdisciplinary Teams: Involve ethicists, social scientists, legal experts, and diverse community members in AI development.

Transparency and Explainability: Use interpretable models or explanation tools to understand how decisions are made.

Methodology

This study employs survey research design for the study. A self-designed questionnaire serves as the primary source for data collection for this study. The target population includes 105 adult and non-formal learners in the metropolis. This refers to individuals who engage in educational activities outside the traditional school system, typically beyond the age of formal schooling usually 25 years and above. A sample size of 80 was obtained using Krejcie and Morgan method of sample collection. Descriptive statistics was used to analyse the data collected.

Data Presentation and Analysis

Base on the study objectives, the data obtained was presented in tables in form of frequency and percentage and the data was descriptively analysed as seen below:

Table 1: Types of Bias and Fairness

Variable	Frequency	Percentage (%)
Data-related Bias,	10	13
Algorithmic Bias	6	7
User-Interaction Bias	4	5
Individual fairness	15	19
Group fairness	5	6
None of the above	0	0
All of the above	40	50
Total	80	100

Source: Questionnaire Administered, 2024

Table 1 present the types of bias and fairness in AI system in adult and non-formal education. It can be observed that majority of the respondents (40%) indicated that all of the above listed are the types of bias and fairness in AI system in adult and non-formal education.

Table 2: Roles of AI System in Adult and Non-Formal Education

Variable	Frequency	Percentage (%)
AI transforms adult and non-formal education	12	15
AI addresses specific challenges and opens up new rooms for learners of all ages	5	6
AI can power adaptive learning systems that adjust the level of content based on a learner's progress	15	19
AI-powered chatbots and virtual assistants can offer 24/7 support to adult learners	18	22
None of the above	0	0
All of the above	30	38
Total	80	100

Source: Questionnaire Administered, 2024

Table 2 present the roles of AI system in adult and non-formal education. It can be observed that majority of the respondents (30%) indicated that all of the above listed are the roles of AI system in adult and non-formal education.

Table 3: Effect of Bias and Fairness in AI System

Variable	Frequency	Percentage (%)
Bias brings about discrimination and inequality	9	11
Bias leads to loss of trust and credibility	4	5
Bias brings about exclusion of marginalized groups	12	15
Fairness promotes equity and inclusion	10	13
Fairness enhances trust and adoption	14	17
Ensuring fairness helps meet legal and ethical standards	6	8
None of the above	0	0
All of the above	25	31
Total	80	100

Source: Questionnaire Administered, 2024

Table 3 present the effect of bias and fairness in AI system in adult and non-formal education. It can be observed that majority of the respondents (25%) indicated that all of the above listed are the effect of bias and fairness in AI system in adult and non-formal education.

Findings

The following results were obtained from the study

1. There are different types of bias and fairness in AI system
2. AI played several roles in adult and non-formal education.
3. Bias and fairness in AI system have effect on adult and non-formal learners

Discussion of the Findings

The study found that data-related bias, algorithmic bias and user-interaction bias are the different type of bias in education. This is similar to the findings of Ferrara (2023). Data bias occurs when the data used to train machine learning models are unrepresentative or incomplete, leading to biased outputs. This can happen when the data are collected from biased sources or when the data are incomplete, missing important information, or contain errors. Algorithmic bias, on the other hand, occurs when the algorithms used in machine learning models have inherent biases that are reflected in their outputs. This can happen when algorithms are based on biased assumptions or when they use biased criteria to make decisions. User bias occurs when the people using AI systems introduce their own biases or prejudices into the system, consciously or unconsciously. This can happen when users provide biased training data or when they interact with the system in ways that reflect their own biases.

Also, the study also found that different types of fairness exist including group fairness, individual fairness, procedural fairness and causal fairness. This is in agreement with the findings of Ferrara (2023). Group fairness ensures that different groups are treated equally in AI systems. Individual fairness ensures that similar individuals are treated similarly by AI systems, regardless of their group affiliation. Procedural fairness on his part involves ensuring that the process used to make decisions is fair and transparent. Lastly, causal fairness involves ensuring that the system does not propagate historical biases and inequalities.

The study further found that AI played several roles in adult and non-formal education. Artificial Intelligence (AI) has the potential to transform adult and non-formal education by providing tailored, efficient, and accessible learning opportunities. The integration of AI into education addresses specific challenges and opens up new rooms for learners of all ages and background. In addition, AI-driven library services can play a transformative role in supporting these learners by offering personalized learning, adaptive learning tools, automatic resource recommendations, access to educational support and language support and translation. This is in agreement with the finding of Rehm and Siegel (2020).

Bias in AI system has negative effect on individuals and society. Discrimination is a key concern when it comes to biased AI systems, as they can perpetuate and even amplify existing inequalities (Sweeney, 2013). Biased algorithms can lead to the underrepresentation of certain groups, such as people of colour or those from lower socioeconomic backgrounds, in credit scoring systems, making it harder for them to access loans or mortgages (Dwork et al., 2012). Furthermore, bias in AI system can perpetuate gender stereotypes and discrimination. On a societal level, the widespread use of such biased AI systems can entrench discriminatory narratives and hinder efforts toward equality and inclusivity. This is in line with the finding of Mehrabi et al. (2021).

On the part of fairness, it was found that fair AI systems ensure that all users, regardless of gender, race, age, or socioeconomic background, receive equitable treatment and outcomes. Also, fairness in AI system are more likely to gain user trust, leading to wider and more confident adoption of AI technologies. Ensuring fairness helps organizations meet legal and ethical standards, reducing risk and promoting accountability. Finally, fair systems are more likely to consider diverse perspectives and needs, leading to more innovative and broadly useful solutions. This is in accordance with the finding of Mehrabi et al. (2021).

Conclusion

In conclusion, AI-driven library services offer immense potential for enhancing equity in adult and non-formal education by providing personalized learning experiences, improving access to educational resources, and supporting diverse learners. However, to realize this potential, it is essential to address the negative effect of bias that may arise from poorly designed AI system. By adopting inclusive data practices, ensuring transparency, and prioritizing equity in both design and implementation, libraries can harness the power of AI to create more equitable and inclusive learning environments for adult and non-formal education.

Recommendations

The following recommends are put forward:

1. Research in fairness and bias in AI system should prioritize the diversification of training data and address the nuanced challenges of bias in generative models, especially those used for synthetic data creation and content generation.
2. It is imperative to develop comprehensive frameworks and guidelines for responsible AI, which includes transparent documentation of training data, model choices, and generative processes.
3. Establishment of robust ethical and legal frameworks governing AI systems is necessary in ensuring that privacy, transparency, and accountability are not afterthoughts but foundational elements of the AI development lifecycle.

References

Baker, R.S. & Yacef, K. (2019). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining* 1(1), 3–17.

Barocas, S., Hardt, M. & Narayanan, A. (2019). Fiarness and machine learning. Fairnessbook.org.

Berk, R., Heidari, H., Jabbari, S., Kearns, M. & Roth, A. (2018). Fairness in criminal justice risk assessments: The state of the art. *Sociology Methods Research,* 47, 175–210.

Buolamwini, J. & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st conference on fairness, accountability and transparency, New York, NY, USA, 23–24; 77–91.

Caliskan, A., Bryson, J.J. & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356, 183–186.

Crawford, K. (2021). There is a blind spot in AI research. *Nature*, 538, 311–313.

Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. In ethics of data and analytics; Auerbach Publications: Boca Raton, FL, USA, 296–299.

Dwork, C., Hardt, M., Pitassi, T., Reingold, O. & Zemel, R. (2012). Fairness through awareness. In proceedings of the 3rd innovations in theoretical computer science conference, Cambridge, MA, USA, 8–10.

Ferrara, E. (2023). Should ChatGPT be Biased? Challenges and risks of bias in large language models. First Monday, 28.

Ferrara, E. (2024). Fairness and bias in Artificial Intelligence: A brief survey of sources, impacts, and mitigation strategies. *Science*. 6, 3.

Fleisher, W. (2020). What's fair about individual fairness? PhilSci-Archive.

Gillen, S., Kearns, M., Roth, A. & Wu, Z.S. (2018). Fairness and efficiency in ranking from pairwise comparisons. In: Proceedings of the 2018 ACM conference on economics and computation, 329–346.

Heidari, H., Loi, M., Gummadi, K.P. & Krause, A. (2019). A moral framework for understanding fair ml through economic models of equality of opportunity. In: proceedings of the conference on fairness, accountability and transparency, 181–190.

Hu, X. & Rangwala, H. (2020). Towards individually fair grade prediction models. In: Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020).

Kamiran, F. & Calders, T. (2012). Data pre-processing techniques for classification without discrimination. *Knowledge Information System*, 33, 1–33.

Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J. & Mullainathan, S. (2018). Human decisions and machine predictions. *Quantitative Journal of Economics,* 133, 237–293.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K. & Galstyan, A. (2021). A Survey on bias and fairness in machine learning. *ACM*, 54(6), 1–35.

O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy; Broadway Books: New York, NY, USA.

Rawis, J. (2021). A theory of justice. Harvard University Press.

Rehm, M. & Siegel, M. (2020). "Artificial Intelligence and adult education." *European Journal of Education*, 55(2), 283-298.

Sweeney, L. (2013). Discrimination in online ad delivery. *Community ACM Computing Surveys*, 56, 44–54.

Tierney, R. & Simon, M. (2020). What is still unfair about standardized testing. *Educational leadership,* 62(3), 48-52.

Yan, S., Kao, H.T. & Ferrara, E. (2020). Fair class balancing: Enhancing model fairness without observing sensitive attributes. In Proceedings of the 29th ACM international conference on information & knowledge management, 1715–1724.