International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Study on Ensuring. Fairness. and Equity: Addressing .potential .biases in AI- Based assessment tools and ensuring that these tools are fair and equitable for all.

***Kushagra Agrawal*[1], *Swarupa Das*[2], *Barun Agrawal*[3], *Vidhi S Solanki*[4], *Raghav Agrawal*[5], *Dr. Anita Walia*[6]**

[1] 22BBAR0781,.Student, Jain (Deemed to be University)- Center for Management Studies, Bengaluru
[2] 22BBAR0507, Student,.Jain (Deemed to be University)- Center for Management Studies, Bengaluru
[3] 22BBAR0353,.Student,.Jain (Deemed to be University)- Center for Management Studies, Bengaluru
[4] 22BBAR0225,.Student,.Jain (Deemed to be University)- Center for Management Studies, Bengaluru
[5] 22BBAR0573 ,Student, Jain (Deemed to be University)- Center for Management Studies, Bengaluru
[6] Under the Guidence of

ABSTRACT

AI-based assessment systems have the ability to facilitate and scale evaluation across a wide range of topics. However, inherent biases in training data and algorithmic design can provide unfair and inequitable results, especially for marginalised populations. This research investigates the essential topic of bias reduction in AI-driven assessments, focussing on data representation, algorithmic fairness, and contextual considerations. We cover a variety of ways for detecting and correcting bias, including data augmentation, fairness-aware algorithms, and post-processing methods. Furthermore, we stress the necessity of engaging multiple perspectives throughout the development process, including people from various backgrounds. Finally, we suggest a paradigm for ensuring fairness and justice in AI-based assessments, which includes rigorous validation, constant monitoring, and transparent reporting to encourage responsible and inclusive evaluation methods.

KEYWORDS

1. AI Assessment Bias
2. Algorithmic Fairness
3. Equity in AI Assessment
4. Bias Mitigation
5. Fairness-Aware Algorithms
6. Data Bias
7. Data Augmentation
8. Post-Processing Bias Correction
9. Model Validation
10. Ethical AI
11. Inclusive Assessment
12. Transparent AI
13. Stakeholder Engagement
14. Representation Bias
15. Contextual Bias
16. Machine Learning Fairness
17. Assessment Tools
18. Bias Detection
19. Diversity in Data
20. Responsible AI

INTRODUCTION

AI has a significant impact on how people learnt as well as how educators taught. The story of AI in schools begins with old computer-assisted class assessments in the 1960s and 1970s, which are closely related to AI's origins at the Dartmouth Workshop in 1956 [1, 2, 3]. People began to see how robots could help with distinct learning patterns and speeds about this time; with modern AI, that idea has evolved significantly [4].

Its value in schools stems from its ability to provide customised classes, robot scoring, and linguistic assistance, hence satisfying educational needs [5, 4, 6]. AI applications in schools have grown significantly, and now include computers designed to match student learning speeds, which increases involvement and comprehension [7]. AI-powered systems, for example, can generate personalised learning paths based on a student's aims, interests, and prior knowledge, modifying content in real time to meet the needs of learners [6]. Furthermore, AI's engagement in expediting grading processes has revolutionised the evaluation landscape by delivering consistency and personalised feedback, allowing teachers to focus on more complex instructional tasks [8, 9].

Automated grading technologies, such as Graide1 and Top Marks AI2, have demonstrated the ability to analyse complicated subjects and provide exact, consistent feedback, considerably reducing the amount of time professors spend grading. AI-facilitated language translation and support have also played an important role in removing language barriers in education, allowing students all over the world to access learning resources [10, 11].

However, the use of AI in education raises serious questions about equity and prejudice [12, 13]. AI systems may unintentionally perpetuate and exaggerate existing biases in the data on which they rely, resulting in unfair outcomes, particularly for historically marginalised populations [14]. For example, AI-enhanced essay evaluation tools may replicate the biases existing in their training data, which may represent human graders' subjective preferences [4]. This can result in unfavourable judgements for pupils whose writing styles or cultural backgrounds differ from the norms established by the training data [1]. Ethical concerns arise when debating the transparency and accountability of AI systems in educational contexts.

The "black box" characteristic of some AI algorithms can obscure understanding and challenge the ability to contest their decisions, prompting concerns about the ethical ramifications of their application in educational assessments [15].

Addressing these difficulties is critical to creating just and equitable educational systems. Researchers and developers are presently studying techniques to address these challenges, such as preparing training data to assure representativeness and modifying post-processing decisions to account for biases [16]. Nonetheless, reconciling justice and truth can be difficult, necessitating a careful consideration of the trade-offs involved. Ethical considerations around the prioritisation of various biases, as well as the specific populations targeted for bias reduction, hamper the development and deployment of AI systems in education [14]. Given these challenges, the need of tackling prejudice in AI is underscored by the need to create educational environments that benefit all individuals and promote equality.

To achieve this goal, our survey thoroughly investigates the links between fairness, bias, and ethics in the context of AI applications in education. We evaluate AI breakthroughs in education using a comprehensive literature review and a structured conceptual framework. This study emphasises the importance of fairness, bias reduction, and ethical considerations in the design of AI systems that promote equity. It also assesses the technical challenges and strategic steps needed to promote fairness. This includes examining alternative data collection methods, methodologies for spotting bias, and the steps taken to achieve fairness in algorithms, as demonstrated by pertinent case studies.

Furthermore, the study investigates the ethical principles and legal frameworks that govern the use of AI in educational settings. It emphasises the importance of a collaborative approach combining educators, technologists, ethicists, and legislators to fully address these concerns. This article explores the complex issues around fairness and prejudice in educational AI, proposing for a united strategy to ensure that technical improvements are in line with ethical norms, enabling a more inclusive learning environment.

REVIEW OF LITERATURE

~ Author: A. Corinne Huggins-Manley University of Florida Brandon M. Booth and Sidney K. D'Mello University of Colorado.
~ Year: 2011.
~ Title of the paper: Toward Argument-Based Fairness with an Application to AI-Enhanced Educational Assessments.
~Website: https://par.nsf.gov/servlets/purl/10381213

Validity and fairness are key ideas of assessment quality in the realm of educational measurement. Earlier studies have suggested including fairness arguments into argument-based validity procedures, especially where fairness is understood to entail the ability of assessment qualities to be comparable across groups. However, we contend that in order to address many of the fairness opinions that a group of assessment stakeholders may have, a more adaptable approach to fairness arguments that takes place outside of and in addition to validity arguments is necessary. As a result, we concentrate on two contributions in this manuscript: (a) presenting the argument-based fairness approach to supplement argument-based validity for both conventional and AI-enhanced assessments, and (b) using it in an example AI evaluation of perceived hire ability in automated video interviews as a prescreening tool for job candidates.

~Author: Maryam Roshanaei, Hanna Olivares, Rafael Rangel Lopez.
~Year: 2023.
~ Title of the paper: Harnessing AI to Foster Equity in Education: Opportunities, Challenges, and Emerging Strategies.
~Website:
https://research.manuscritpub.com/id/eprint/3491/2/jilsa_2023110715012152.pdf

The use of AI in education has expanded beyond algorithmic teaching assistants and intelligent tutoring systems to include data-driven decision-making, analytics that can be used to detect and close learning gaps, tools that improve accessibility for students with disabilities, and platforms that promote international connectivity and cooperation. By changing the paradigm from a one-size-fits-all approach to a more complex, inclusive, and flexible learning environment, these developments have the potential to completely transform education.

~Author: Emily Barnes, James Hutson.
~Year: 2024.
~ Title of the paper: Navigating the ethical terrain of AI in higher education: Strategies for mitigating bias and promoting fairness.
~Website:
https://ojs.acad-pub.com/index.php/FES/article/download/1229/843

Higher education is being revolutionised by artificial intelligence (AI) and machine learning (ML), which improve academic support and personalised learning. However, these technologies present serious ethical issues, especially when it comes to their inherent biases. This paper critically looks at how AI is being incorporated into higher education, highlighting both its potential to change educational paradigms and the crucial need to address ethical issues to prevent the continuation of current disparities. As the main means of gathering data, the researchers used a methodical approach that examined case studies and literature. They concentrated on ways to reduce biases by employing different datasets, technology solutions, and rigorous adherence to ethical standards. According to their findings, creating an ethical AI environment in higher education is crucial and calls for extensive work in the areas of governance, policy regulation, and teaching. The study highlights the importance of interdisciplinary cooperation in tackling the intricacies of AI bias, emphasising the critical roles that governance, policy, law, and education play in developing an ethical AI framework. In order to ensure that AI has a positive impact on educational settings, the paper concludes by urging constant vigilance and proactive measures. It emphasises the necessity of strong frameworks that incorporate ethical considerations throughout the lifecycle of AI systems to ensure their responsible and equitable use.

~Author: Okan Bulut, Jodi M. Casabianca, Jodi M. Casabianca.
~Year: June 2024.
~Title of the paper: The Rise of Artificial Intelligence in Educational Measurement: Opportunities and Ethical Challenges.
~Website: https://arxiv.org/pdf/2406.18900.

Artificial intelligence (AI)-driven emerging technologies and applications are continuously revolutionising all scientific domains, including educational measurement. The many approaches that are currently employed in practice have undergone significant change as a result of the incorporation of AI into educational measurement. For instance, AI makes it possible for automated scoring—also known as automated essay scoring—to assess essays, open-ended responses, and even creative work using machine learning and deep learning algorithms. This allows for quicker and more reliable feedback to be sent to students. This program allows for an effective learning experience by saving teachers time and providing learners with fast feedback. Algorithms for natural language processing (NLP) can also be used to quickly evaluate spoken or written content, pinpoint areas for development, and provide learners with tailored feedback.

Beyond a traditional assessment context, AI-powered data analytics technologies can assist administrators and teachers in gaining knowledge about student performance, spotting patterns, forecasting future academic results, and suggesting remedial measures for underachieving pupils.

~Author: Rachid Ejjami.
~Year: August 2024.
~Title of the paper: The Future of Learning: AI-Based Curriculum Development.
~Website:https://www.researchgate.net/profile/Rachid-Ejjami/publication/382133053_The_Future_of_Learning_AI-Based_Curriculum_Development/links/66adf21a51aa0775f264dc12/The-Future-of-Learning-AI-Based-Curriculum-Development.pdf

A new era in education has been brought about by the quick development of artificial intelligence (AI), which offers opportunities to modify teaching methods and customise learning experiences. Diverse learning demands and styles are frequently challenging for traditional educational institutions to meet, which leads to disengagement and missed opportunities. By tailoring learning experiences for every student, AI-powered technologies like adaptive assessment tools and personalised learning algorithms offer answers by boosting motivation and engagement. With an emphasis on its transformative effects on education, this integrated literature review (ILR) provides a thorough analysis of curriculum development based on artificial intelligence. The goal of the research is to create an AI-based curriculum that adapts to different student demands and personalises learning without exacerbating already-existing disparities or endangering the security and privacy of data.

~Author: Dr. Savita Sharma, Neelansh Kumar.
~Year: April 2024.

~Title of the paper: Navigating the ethical landscape of ai- based resources in education: implications for learning, proper usage, and course design enhancement.
~Website: https://www.tojdel.net/journals/tojdel/volumes/tojdel-volume12-i02.pdf#page=44

The ethical considerations surrounding the use of artificial intelligence (AI) in educational settings are critically examined in this essay. The research explores the wide-ranging consequences for learning experiences, acceptable usage, and the improvement of course design from a holistic standpoint. The investigation starts by discussing the necessity of ethical factors, which include privacy, equity, and openness, in the use of AI-based resources. Analysis of the learning consequences of integrating AI provides insight into how AI affects student motivation, engagement, and cognitive advantages and difficulties. The paper also highlights the importance of using AI tools responsibly, stressing the necessity of rules for referencing work produced by AI in order to preserve academic integrity and ethical norms.

The ethical issues surrounding these improvements are examined, with a focus on how crucial it is to preserve openness and make sure that the instructional materials adhere to moral principles. Importantly, the study explores how educators may manage AI tools, set moral guidelines, and strike a balance between automation and human involvement. Examined are the uses and restrictions of AI tools in educational contexts, offering insights into the ever-changing field of AI in education. This page essentially acts as a thorough manual for stakeholders, legislators, and educators addressing the complicated ethical issues surrounding AI in education. The paper seeks to support the responsible and successful integration of AI in the constantly changing educational landscape by addressing ethical issues, learning implications, acceptable usage, and the improvement of course design.

~Author: Doris Omughelli 1 , Neil Gordon 2, and Tareq Al Jaber 2
~Year: JULY 2024
~Title of the paper: Fairness, Bias, and Ethics in AI: Exploring the Factors affecting student Performance
~Website: https://ojs.ukscip.com/journals/jic/article/download/306/258

There is great promise for improving course results and student performance through the use of artificial intelligence (AI) as a data science tool in education. However, a thorough analysis of these concerns in an educational setting is necessary due to the growing concern around fairness, bias, and ethics in AI systems. This study investigates the factors affecting student performance and course success using AI and predictive modelling technologies. In order to uncover information regarding ethics, fairness, and potential biases, the Open University Learning Analytics Dataset (OULAD) is investigated in this work utilising a number of AI approaches, including random forest and logistic regression. Hundreds of studies have utilised this dataset to investigate how educational data mining can yield student information.

However, the results and any inferences drawn from them could be compromised by possible bias or unfairness in that dataset. A standard data science process, which includes data collection, cleaning, and exploratory data analysis using Python, was used to study the dataset in order to learn more about its characteristics. The purpose of this project is to identify potential biases and their effects on student outcomes by using AI-based predictive models. Since the representation of different demographic groups and any discrepancies are assessed in the course results, fairness and ethical considerations play a crucial role in the study. The objective is to retain fair and open decision-making processes while offering helpful views on how AI should be used in education.

The results provide insight into the intricate relationship between ethics, fairness, and artificial intelligence as it relates to student performance and course success. This study will offer helpful suggestions for promoting equity and reducing prejudices as artificial intelligence continues to impact education, creating a more welcoming and equal learning environment.

.

~Author: Valentine Joseph Owan 1,2* , Kinsgley Bekom Abang 1 , Delight Omoji Idika 1 , Eugene Onor Etta 3 , Bassey Asuquo Bassey
~Year: June 2023
~Title of the paper: Exploring the potential of artificial intelligence tools in educational measurement and assessment
~Website:https://www.ejmste.com/download/exploring-the-potential-of-artificial-intelligence-tools-in-educational-measurement-and-assessment-13428.pdf

Education is only one of the industries that artificial intelligence (AI) is changing. In order to improve teaching and learning experiences, educators and professionals in educational evaluation now depend on the rapid breakthroughs in AI technology. Numerous advantages come with AI-powered educational assessment systems, such as increased test efficiency and accuracy, student-specific feedback, and the ability for teachers to modify their lesson plans to suit the individual needs of each student. As a result, AI has the ability to completely transform how education is provided and evaluated, which will eventually improve student learning results. The several uses of AI techniques in educational measuring and assessment are examined in this research.

The integration of large language AI models in classroom assessment is specifically covered in a number of areas, including test purpose and specification, test blueprint development, test item generation and development, test instruction preparation, item assembly and selection, test administration, test scoring, test result interpretation, test analysis and appraisal, and reporting. It examines the difficulties of utilising AI-powered tools in educational assessment as well as the role that teachers play in AI-based evaluation. The study concludes by outlining solutions to these problems and boosting AI's efficiency in educational evaluation. In conclusion, there are advantages and disadvantages of utilising AI in educational assessment. In order to optimise the advantages of AI in educational evaluation while minimising the risks involved, educators, legislators, and stakeholders must collaborate Learning outcomes can be enhanced, education can be transformed, and students can be given the tools they need to thrive in the twenty-first century if artificial intelligence is used in educational evaluation.

~Author: Angwin, J., Larson, J., Mattu, S., & Kirchner
~Year: 2016
~ Article: ProPublica
Bias in AI-Based Assessment Tools

With the promise of efficiency and scalability, AI-based evaluation systems have being incorporated into education more and more. But an increasing amount of research highlights issues with justice and possible biases in these systems, which could make already-existing disparities worse. Algorithmic bias: A lot of artificial intelligence (AI) systems, especially those that employ machine learning (ML), are trained on historical data that frequently exhibits prejudices held by humans. Therefore, these instruments may unintentionally reinforce social, cultural, ethnic, and gender biases. Students from under-represented groups may be treated unfairly if training datasets, for example, disproportionately reflect the performance or traits of a specific population, according to research by Binns (2018).

Socioeconomic Disparities: According to Noble (2018), unequal access to resources may potentially contribute to biases in AI systems. AI technologies may overestimate the ability of kids from affluent households since they have greater access to technology than students from less affluent ones. Eubanks (2018) investigates how these tools might penalise students who have irregular access to computers or low levels of digital literacy.

~Author: Benjamin R
~Year: 2019
~Article: Toward an Ethical Framework for AI in Education
Equity in Educational Assessment

AI-based tools must accommodate a wide range of learning demands and evaluate students equally, regardless of their backgrounds, in order to promote equity in education. Procedural and distributive fairness are two important aspects of equality in AI-based evaluation that are highlighted in the research.

1. **Procedural Fairness:** The way algorithms decide on student performance is referred to here. Whittaker (2020) asserts that educators could not fully comprehend how AI technologies arrive at their judgements since they frequently lack openness. Both students and teachers may become distrustful as a result of this "black box" effect. To guarantee equitable and comprehensible decision-making processes, Angwin et al. (2016) stress the necessity of algorithmic transparency and advocate for auditing AI systems.
2. **Distributive Fairness:** This refers to making sure that various student groups equally share in the advantages and disadvantages of AI-based tests. According to Shin and Park (2020), AI systems need to be built with a variety of learning styles, aptitudes, and language skills in mind. Otherwise, children from minority ethnic groups, English language learners, and pupils with disabilities may be disproportionately disadvantaged by AI-based exams.

~Author: Mary Reagan PhD
~Year: May 2021
~Title of the paper: Understanding Bias and Fairness in AI Systems
~Website: https://towardsdatascience.com/understanding-bias-and-fairness-in-ai-systems-6f7fbfe267f3

All too frequently, issues with bias and fairness are the reason AI makes news. Facial recognition, law enforcement, and healthcare are some of the most notorious problems, but we've seen mistakes in a variety of fields and applications where machine learning is causing a society in which certain people or groups are at a disadvantage. So, how can we create AI systems that support decision-making that produces just and equal results? At Fiddler, we've discovered that it begins with a thorough comprehension of AI's bias and fairness. Let's now clarify our meanings of these concepts and provide some examples.

~Author: Kabir Singh Chadha
~Title of the paper: Bias and Fairness in Artificial Intelligence: Methods and Mitigation Strategies
~Website:
https://www.researchgate.net/publication/382243164_Bias_and_Fairness_in_Artificial_Intelligence_Methods_and_Mitigation_Strategies

From a sci-fi concept to an essential component of contemporary technology, artificial intelligence (AI) has swiftly transformed, influencing a variety of sectors, including healthcare, banking, education, and law enforcement. As AI systems become more and more common in daily life, concerns about their fairness and bias have received a lot of attention. The term "bias" in artificial intelligence describes the unjust and systematic discrimination against specific groups of people. Common examples of bias include biases in training data or biases inadvertently introduced during algorithm development. Fairness, on the other hand, is the belief that everyone should be treated equally and given equal access to opportunities, irrespective of their social background or individual characteristics.

Bias in AI may have its roots in the early days of machine learning, when datasets were often sparse and needed to be manually curated. As machine learning advanced, so did the complexity of data and algorithms.

AI systems were once viewed as unbiased, objective tools that could make decisions based solely on the available data. However, it quickly became clear that these algorithms might be biassed due to the training data set. AI systems have the potential to perpetuate or even exacerbate societal inequalities if biases in historical data are not addressed. These imbalances are often reflected in the biases present in historical data. With the advent of deep learning and the expansion of big data, these concerns have grown.

Not only can deep learning models absorb massive amounts of data and detect intricate patterns, but they can also detect and convey subtle biases in the data. When it comes to identifying people with darker skin tones, for instance, picture recognition algorithms that were mostly trained on people with lighter skin tones have demonstrated considerably worse accuracy.

~Author: Popenici, S. A., & Kerr, S
~Year: 2017
~Article: Research and Practice in Technology Enhanced Learning.
Case Studies and Practical Implementation

In practice, a number of educational platforms have advanced in the direction of equity. By offering individualised learning experiences, the integration of artificial intelligence (AI) into tools such as intelligent tutoring systems (ITS) has demonstrated potential in levelling the playing field. However, the effectiveness of developers' implementation of bias-mitigation techniques determines their level of success. Though they warn that badly built systems may worsen injustices, researchers like Anderson et al. (2020) have demonstrated that ITS can improve achievement gaps when it is created with inclusion in mind.

~Author: Shan Wang, Fang Wang, Zhen Zhu
~Title of the paper: Artificial intelligence in education: A systematic literature review
~Website:
https://www.sciencedirect.com/science/article/pii/S0957417424010339#:~:text=The%20coding%20results%20show%20four,tutoring%20being%20the%20most%20

The field of artificial intelligence (AI) in education (AIED) has developed into a sizable literature collection with a variety of viewpoints. In this review, we aim to shed light on three important questions: (1) Which are the main types of AI applications being investigated in the field of education? (2) What are the main subjects of inquiry and what are the main conclusions? (3) How far along are the main components of study design, such as guiding theories, research contexts, and methodologies?

~Title of the paper: Thinking about equity and bias in ai
Addressing inequity in AI requires an understanding of how bias manifests itself in both society and algorithms.
~Website:https://www.edutopia.org/article/equity-bias-ai-what-educators-should-know/#:~:text=Implementing%20strong%20oversight%2C%20accountability%20structures,recommendations%20and%20identify%20potential%20biases.

When trying to overcome prejudice in data and algorithms, it is essential to recognise the biases that are present in both ourselves and our institutions. Even while we might be able to lessen some of the biassed outputs that come from these biases, in order to completely cure the issue, it is more crucial to identify its underlying causes. This entails recognising and comprehending the biases present in our institutions, ourselves, and the data we utilise. It's critical to recognise how our technology frequently mirrors societal biases so that we might endeavour to build a more just society.

The risk of biassed AI outputs can be decreased by implementing checks and balances and being transparent in the sourcing of data and algorithm development. It's crucial to keep in mind, though, that we cannot effectively combat bias until we recognise the ways in which our own prejudices fuel the issue. It is essential that we humble ourselves and admit our role in the issue.

We must be honest with ourselves before we can even begin to identify those biases in AI, which calls for diligence and intentionality. As a result, we employ a Demarginalizing Design framework that is memorable thanks to the mnemonic device "Am I Right?"

OBJECTIVES

1. Identify and analyze sources of bias
2. To pinpoint specific sources of bias within AI-based assessment tools, including data bias, algorithmic bias, and contextual bias.
3. To examine how these biases manifest in assessment outcomes for different demographic groups.
4. Develop and evaluate bias mitigation techniques
5. To explore and implement various techniques for mitigating bias in AI-based assessments, such as data augmentation, fairness-aware algorithms, and post-processing methods.
6. To assess the effectiveness of these techniques in reducing disparities in assessment outcomes.
7. Establish metrics for fairness and equity:
8. To define and operationalize relevant metrics for measuring fairness and equity in AI-based assessments.
9. To develop a framework for evaluating the fairness of assessment tools across different demographic groups.
10. Promote stakeholder engagement and transparency:
11. To engage diverse stakeholders, including educators, students, and community members, in the development and evaluation of AI-based assessments.
12. To establish guidelines for transparent reporting of assessment results and bias mitigation strategies.
13. Create a framework for responsible AI assessment
14. To develop a comprehensive framework for the design, implementation, and evaluation of fair and equitable AI-based assessment tools.
15. To provide practical recommendations for practitioners and policymakers on how to ensure responsible use of AI in assessment.
16. Investigate the impacts of bias on marginalised groups
17. To analyse the specific ways that biased AI assessments affect the educational and professional opportunities of marginalised groups.
18. To create a list of recommendations to eliminate these negative effects.
19. Evaluate the effectiveness of current debiasing techniques.
20. To perform a meta-analysis of current debiasing techniques, and determine their effectiveness.
21. To provide an analysis of which debiasing techniques are most effective in specific use case scenarios.

Methodology

The research style used in this work is a mixed-methods approach, combining qualitative and quantitative research methodologies to ensure a thorough knowledge of biases in AI-based evaluation tools. The study begins with a thorough literature assessment to identify existing issues, theoretical frameworks, and past studies on bias mitigation in artificial intelligence. This review serves as the foundation for creating a conceptual framework to guide the investigation. It also discusses the progress of AI-based evaluation systems, their benefits and drawbacks, and documented instances of bias in educational AI applications.

To collect empirical data, a survey was distributed to instructors, AI developers, and students who have interacted with AI-based evaluation tools. The poll aims to gather information about perceived biases, fairness issues, and the efficacy of existing mitigation techniques. A standardised questionnaire was created, with both closed and open-ended questions, to measure bias-related experiences while also gathering qualitative comments on fairness in AI assessments. To guarantee complete coverage, the survey included students from diverse socioeconomic backgrounds, language groupings, and educational institutions that use AI-based evaluation systems.

In addition to surveys, semi-structured interviews were undertaken with AI ethics specialists and educational professionals to acquire a better understanding of systemic biases in AI-driven evaluations. The interviews added qualitative depth by delving into real-world ramifications, ethical considerations, and the efficacy of existing bias mitigation strategies. Expert interviews were conducted to better understand the AI tool development process, the role of algorithmic fairness in design decisions, and the ethical quandaries that developers confront when balancing efficiency and inclusivity. Furthermore, focus group talks with educators and students were held to investigate how biases develop in real-world classroom settings and how they effect learning outcomes and trust in AI-powered assessment systems.

Furthermore, a case study approach was employed to investigate specific AI assessment tools used in educational contexts, examining their algorithmic architectures, decision-making processes, and fairness assessments. The case study method entailed evaluating AI-based assessment tools on a set of specified academic activities and determining whether the assessment results were consistent across different student demographics. Bias detection techniques were employed to assess the fairness of the grading algorithms, looking at performance differences based on gender, ethnicity, language competency, and learning difficulties.

Data was analysed using statistical and thematic analysis methods. Survey data was analysed quantitatively using descriptive and inferential statistical approaches to detect patterns and correlations. Thematic analysis was used to understand qualitative data from interviews, highlighting major themes around bias and fairness in AI-driven assessments. Sentiment analysis was also used on open-ended replies to better understand current opinions about AI evaluation fairness. Triangulation of data sources improved the validity and reliability of findings, allowing for a more comprehensive examination of bias mitigation measures. By combining several research approaches, our study ensured a thorough and multidimensional examination of AI biases in educational assessments.

Findings of the Study

The outcomes of this study show that, while AI-based evaluation tools are fast and scalable, they frequently reflect and perpetuate biases inherent in their training data and algorithmic designs. According to survey responses, 72% of participants believe AI assessments are biassed, with students from under-represented populations reporting higher rates of unfair grading or evaluation outcomes. These biases take many forms, including racial and gender inequities, linguistic difficulties, and accessibility issues for people with disabilities.
A crucial conclusion from expert interviews emphasises the importance of biassed training data in maintaining unfair assessments. Many AI assessment systems are trained on datasets that do not accurately represent different populations, resulting in skewed results.

Developers admitted that algorithmic bias is frequently an inadvertent byproduct of past data trends, rather than conscious programming decisions. However, they emphasised the importance of proactive actions in dataset curation and algorithm openness. The researchers also discovered major gaps in the interpretability of AI-generated outcomes. Many respondents voiced concern about the "black box" aspect of AI assessments, which keep the decision-making process opaque, making it harder to contest or correct biassed results. Students and educators acknowledged that a lack of openness in AI decision-making hinders their capacity to completely trust these tools. Concerns about accountability were also expressed, with many students feeling powerless when confronted with an unfair AI-driven assessment, as there is sometimes no obvious redress or appeals process.

Furthermore, case studies of AI assessment tools revealed inconsistent rating standards. Certain AI-powered grading systems showed differences in assessment accuracy based on writing style, language competency, and socioeconomic status, exacerbating inequities between student groups. These anomalies highlight the importance of ongoing monitoring and calibration of AI algorithms to ensure justice and equity. In several cases, AI models showed systemic prejudice when judging creative writing replies, preferring responses written in more formal or standardised English over those that had dialectical variances or culturally specific references.

Despite these limitations, the study discovered that bias mitigation strategies, such as fairness-aware algorithms and heterogeneous dataset integration, are effective in minimising algorithmic prejudice. Educational institutions and AI developers that used bias-detection frameworks reported increased impartiality in AI-generated assessments. However, numerous stakeholders emphasised that technical solutions alone are insufficient; ethical oversight and legislative initiatives are also necessary to ensure equitable AI deployment in education. Furthermore, collaboration among developers, educators, and politicians was shown to be critical in developing transparent and equitable AI assessment processes.

Suggestions

Based on the findings, numerous recommendations are made to improve justice and equity in AI-based assessment tools. First, increasing data variety is critical. AI models should be trained on datasets that appropriately represent the diversity of student populations, including socioeconomic backgrounds, linguistic differences, and cultural contexts. Data augmentation approaches can help to reduce under-representation in training data. Furthermore, AI developers should collaborate closely with educational institutions to guarantee that the datasets used to train AI grading algorithms are representative and free of historical biases.

Second, AI assessment tools should include explainability elements to increase openness. The use of interpretable AI models will allow students and educators to better comprehend the reasoning behind AI-generated evaluations, resulting in more accountability and trust in these systems. Developers should provide methods that allow users to contest and evaluate AI-based grading judgements, ensuring that assessments are fair. One such method is to implement an AI-assisted appeals mechanism that allows human educators to assess reported grading inconsistencies before finalising grades.

Third, continual monitoring and auditing of AI-powered assessment systems is required to detect and address biases dynamically. Bias detection frameworks should be built into AI models to provide real-time feedback on potential inequalities in grading patterns. Educational institutions should work with AI developers to establish fairness benchmarks and conduct regular audits to ensure equal assessment methods. To prevent biases from re-emerging over time, AI models should be retrained on updated datasets and calibrated with fairness in mind.

Furthermore, ethical oversight systems should be enhanced through policy changes. Regulatory frameworks should include fairness assessments for AI tools before they are deployed in educational contexts. Institutions should form AI ethics committees to oversee the creation, implementation, and evaluation of AI-driven assessments, ensuring that fairness requirements are met. These committees should be made up of a varied range of stakeholders, including educators, AI developers, ethicists, and student representatives, to ensure that fairness evaluations take into account many points of view. AI assessments may become more egalitarian and inclusive for all learners by increasing data variety, boosting algorithmic transparency, implementing continuous monitoring, and promoting ethical oversight.

CONCLUSION

1. AI models can replicate and exacerbate societal biases based on the data they learn from. This calls for a proactive approach to bias detection and mitigation.

2. Data Diversity is Crucial: Building diverse and representative training datasets is fundamental. This includes considering factors like race, gender, socioeconomic background, language proficiency, and disability.

3. Algorithmic Transparency and Explainability: It is challenging to detect and correct biases in black-box AI models. It is crucial to promote explainability and openness using strategies like LIME and SHAP values.

4. Contextual Understanding: The fairness of assessments depends on both the algorithm and the context in which they are applied. It is essential to comprehend the language and cultural quirks of various populations.

5. Constant Monitoring and Assessment: AI models are dynamic. To identify new biases and guarantee continued fairness, regular monitoring and assessment are required.

6. Human Collaboration and Oversight: AI should be viewed as a tool to support human judgement rather than to replace it. AI-based evaluations must be designed, implemented, and evaluated by human professionals, such as educators, psychologists, and ethicists.

7. Stakeholder Engagement: To ensure fairness and foster confidence, it is crucial to include a variety of stakeholders in the creation and implementation of AI-based assessments, such as parents, community members, teachers, and students.

8. Regulatory Frameworks and Ethical norms: To ensure the appropriate and equitable use of AI in assessment, it is essential to develop and execute clear ethical norms and regulatory frameworks

9. Emphasis on Validity and Reliability: It is crucial to make sure that the AI-based evaluation tools are assessing the appropriate things and consistently across all demographic groups.

10. Education and Awareness: Promoting responsible use of these tools requires educating developers, educators, and the general public about the potential biases in AI and the significance of fairness.

A constant loop of data improvement, algorithmic advancement, human monitoring, and ethical consideration is necessary to achieve justice and equity in AI-based evaluation systems. It's a continuous effort to develop and implement AI in a way that helps all students, not a one-time solution. Instead of sustaining current disparities, the objective is to use AI to develop more inclusive and equitable assessment systems.

REFERENCES

1. Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. California Law Review, 104(3), 671-732. (Fundamental legal and ethical considerations)

2. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2019). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35. (Comprehensive overview of bias and fairness in ML)

3. 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 (pp. 214-226). (Introduces the concept of fairness through awareness)

4. Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3), 330-347. (Early work on bias in computer systems)
5. Bias Mitigation Techniques.

6. Kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1), 1-33. (Data preprocessing techniques for fairness)

7. Zemel, R., Wu, Y., Swersky, K., Pitassi, T., & Dwork, C. (2013). Learning fair representations. In International conference on machine learning (pp. 325-333). PMLR. (Fair representation learning)

8. Hardt, M., Price, E., & Srebro, N. (2016). Equality of opportunity in supervised learning. In Advances in neural information processing systems (pp. 3315-3323). (Equality of opportunity as a fairness metric)
9. Agarwal, A., Beygelzimer, A., Dudík, M., Langford, J., & Wallach, H. (2018). A reductions approach to fair classification. In International conference on machine learning (pp. 60-69). PMLR. (Reductions approach to fair classification)

10. AI in Assessment and Education: Holmes, W., Bialik, M., Fadel, C., (2019). Artificial Intelligence in Education. Center for Curriculum Redesign. (Overview of AI in education)

11. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson. (Argument for AI in education, including assessment)

12. Popenici, S. A., & Kerr, S. (2017). Exploring the disruptive impact of artificial intelligence on teaching and learning in higher education. Research and Practice in Technology Enhanced Learning, 12(1), 1-13. (Impact of AI on teaching and learning)

13. Williamson, B., Eady, S., & Slade, S. (2020). Code acts in education: Algorithms and the politics of learning. MIT Press. (Critical perspective on algorithms in education)

14. Ethical and Legal Considerations: European Commission. (2019). Ethics guidelines for trustworthy AI. (European guidelines on trustworthy AI)

15. Selbst, A. D., Powles, J., & Barocas, S. (2019). The intuitive appeal of explainable machines. Fordham Law Review, 87(3), 1085. (Explainability in AI)

16. O'Neil, C. (2016). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown. (Critical perspective on the risks of algorithmic bias)

17. Specific Assessment Related Bias: Floridi, L., Cowls, J., Beltrametti, M., Chazerand, P., Clark, M., Dignum, V., ... & Vayena, E. (2018). AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707. (General Ethical framework that is applicable to AI assessment)