ISSN: 2977-0041

Review Article

Journal of Material Sciences and Engineering Technology

Adapting Fair Teaching and Assessment Methods for Fairness in the Age of AI Chabot's and LLMs

Basheer Riskhan*, Abdullah Hakeem Mohamad and Maria Iqbal

School of Computing & Informatics, Albukhary International University Alor Setar Kedah 05200 Malaysia

Corresponding author
Basheer Riskhan, School of Computing & Informatics, Albukhary International University Alor Setar Kedah 05200 Malaysia.

Received: October 13, 2025; **Accepted:** October 21, 2025; **Published:** October 27, 2025

ABSTRACT
This research explores how educators are adapting their teaching and assessment strategies in response to the rise of Large Language Models (LLMs) like ChatGPT. While institutions are increasingly aware of these tools, there is limited understanding of how fairness and academic integrity are maintained when students integrate LLMs into their learning. This study uses a qualitative approach, analyzing responses from educators across different institutions from various countries to understand the practical challenges, strategies, and perceptions surrounding AI in education. Key themes for this paper include LLM detection, redesigning assessments, fairness concerns, institutional support, and the dual role of AI in both enhancing and undermining student learning. The findings suggest that while many educators support responsible use of LLMs, many gaps in policy and support systems persist. Our study highlights certain portions of the ongoing debate on fair educational assessments in the age of AI, and calls for stronger institutional guidance, faculty training, and pedagogical adaptation.

Introduction

In recent years, the rise of generative AI tools such as ChatGPT has sparked widespread debate in the education sector; Shahzad et al., 2025; Isaak et al., 2024; Abd-Alrazaq et al., 2023; Beak et al., 2024; Lin et al., 2023; Vargas-Murillo et al., 2023; Anders, 2023; Bin-Nash wan et al., 2023) [1]. Large Language Models (LLMs), trained on massive datasets, can now generate high-quality text, answer complex questions, and even complete assignments [1]. As these tools become easily accessible to students, they are quickly reshaping how learning, teaching, and assessments are approached in academic environments; Isaak et al., 2024; [1,2].

This transformation presents both opportunities and challenges (Abd-Alrazaq et al., 2023; Shahzad et al., 2025) [3]. On one hand, LLMs can support students in brainstorming, clarifying concepts, and improving their writing, offering enhanced personalization, increased access to information, efficient feedback mechanisms, and 24/7 accessibility; Madasamy et al., 2022; Seetharaman, 2023) [1,2]. On the other hand, they raise concerns about academic integrity, originality, and fairness especially when students use them during assessments (Anders, 2023; Bin-Nash wan et al., 2023; Khalil & Er, 2023; Kiesler & Schiffner, 2023) [1]. Educators are left questioning how to assess students' true understanding in an age where answers can be AI-generated within seconds (Jamil, 2023) [1].

In this context, fairness refers to ensuring that all students are assessed on their own understanding while academic integrity is about upholding honesty, accountability, and ethical conduct in learning (Green, 2025; Bin-Nashwan et al., 2023) [4]. The use of LLMs blurs these boundaries, especially when institutional policies and teaching methods have not yet adapted to this technological shift (Bin-Nashwan et al., 2023) [1].

The aim of this study is to explore how educators are responding to these challenges. Specifically, it investigates the lived experiences of academic staff in adapting their teaching and assessment strategies to maintain fairness and academic integrity in classrooms where AI tools are readily available to students [2,4].

While educational institutions have rapidly embraced tools like ChatGPT and other Large Language Models (LLMs), there remains limited research on how teaching and assessment practices are evolving in response [1]. Most existing literature

Citation: Basheer Riskhan, Abdullah Hakeem Mohamad, Maria Iqbal. Adapting Fair Teaching and Assessment Methods for Fairness in the Age of AI Chabot's and LLMs. J Mat Sci Eng Technol. 2025. 3(4): 1-8. DOI: doi.org/10.61440/JMSET.2025.v3.80

focuses on the ethical implications of AI use, student behavior, or institutional policies, but rarely addresses how educators themselves are adapting in real-time to maintain fairness and academic integrity. This lack of guidance leaves educators with few practical strategies. As a result, many may resort to either outright prohibiting LLMs or allowing unchecked use both of which risk undermining the validity of assessments and learning outcomes. A complete ban can stifle technological engagement and limit learning opportunities, while unrestricted use may compromise originality, understanding, and fair evaluation. Given the growing presence of AI in education, it is critical to understand how teaching and assessment methods are being adapted on the ground. This study addresses that gap by exploring how educators perceive and manage the challenges posed by LLMs, and how they are working to ensure fairness in assessments amidst rapid technological change.

Literature Review

An Experiment with LLMs as Database Design Tutors Persistent Equity and Fairness Challenges in Online Learning

Jamil critically investigates the fairness and equity implications of using large language models (LLMs) like ChatGPT, Gemini, and CoPilot as intelligent tutors for database design education, particularly in functional dependency theory and normalization tasks [1]. Through extensive experiments, the study reveals that earlier LLM versions (e.g., ChatGPT 3.5) often provide factually incorrect solutions and misleading derivations what the author terms "ignorant bias" which can disadvantage learners who lack the expertise to detect such errors. Even advanced versions like ChatGPT 4o1 showed occasional flaws, raising concerns about equitable learning outcomes when access to accurate AI tools depends on factors like subscription level, technical expertise, or socioeconomic status. The paper emphasizes that while LLMs can support scalable digital learning, their limitations in reasoning and inconsistency across versions may inadvertently widen the digital divide, especially in self-paced or under-supervised educational settings. Jamil proposes the development of integrated intelligent tutoring systems (ITS), such as NoDD, which combine the explainability of LLMs with the reliability of structured algorithmic approaches, as a potential remedy to these persistent equity challenges [1].

Debiasing Education Algorithms

Idowu conducts a systematic literature review (SLR) to assess the state of fairness in educational algorithms, analyzing 12 peer-reviewed studies across domains such as dropout prediction, forum post classification, and recommender systems [5]. The study identifies a wide variety of fairness metrics used in practice including ABROCA, group performance disparity, TPR/FPR, and counterfactual fairness and emphasizes that no single metric suits all contexts, reinforcing the importance of aligning metrics with specific educational goals. Key bias mitigation strategies examined include class balancing, adjusted sample weights, adversarial learning, vector projection-based debiasing, and model-specific approaches like MCCM. Notably, the review critiques the dominant focus on gender and race as sensitive attributes, urging future work to consider broader factors like disability, socioeconomic status, and native language. Another major insight is the importance of assessing data and feature bias before model-level fairness, as discriminatory patterns in input data often carry through to algorithmic outcomes. Surprisingly, many of the reviewed studies found no strict trade-off between fairness and accuracy, suggesting that performance can be improved alongside fairness under the right strategies. Idowu concludes by recommending context-sensitive fairness strategies, expanded demographic focus, and stronger alignment between algorithmic fairness and human perception to ensure more equitable educational AI systems.

Educational Data Mining and Predictive Modeling in the Age of Artificial Intelligence: An In-Depth Analysis of Research Dynamics

López-Meneses et al. present a comprehensive bibliometric and systematic review of 793 Scopus-indexed articles (2000–2024) to map the evolution, themes, and challenges of educational data mining (EDM) and predictive modeling (PM) in the AI era. The study identifies three major phases in research output initial growth (2000–2013), consolidation (2014–2019), and a boom in maturity (2020–2024), driven by increased data access, AI advances, and the COVID19 shift to digital education [2]. The authors highlight how EDM and PM have been used to predict student performance, personalize learning, and identify at-risk students through techniques like decision trees, SVMs, and deep learning. Key applications include early interventions, adaptive content delivery, and performance forecasting. The paper emphasizes ethical and technical challenges such as data privacy, algorithmic bias, and interoperability across educational systems. It also outlines emerging research areas like hierarchical active learning, vulnerability detection, and collaborative AI consortia. A key insight is the potential of AI to democratize access to quality education if implemented equitably, ethically, and transparently. The study concludes that future research must prioritize explainability, personalization, and inclusion, ensuring AI in education reduces rather than reinforces systemic inequalities.

Escaping the Impossibility of Fairness from Formal

Green analyzes the dominant “formal algorithmic fairness” paradigm built on mathematical models like separation and sufficiency for failing to account for systemic inequalities that shape decision-making contexts, especially in public policy domains like education, criminal justice, and welfare [6]. He introduces the concept of substantive algorithmic fairness, grounded in legal and philosophical theories of substantive equality, which shifts the focus from isolated decision points to broader relational and structural inequalities. Green argues that the so-called “impossibility of fairness” the proven mathematical incompatibility of fairness definitions is a direct result of formalism’s narrow scope, which cannot capture the social context or long-term impact of algorithmic decisions. Through examples like the COMPAS pretrial risk assessment tool, he shows that even perfectly accurate models can perpetuate injustice if they ignore historical disadvantages. Substantive fairness, by contrast, encourages algorithmic reform that (1) reduces upstream disparities, (2) lowers downstream harms, and (3) uses algorithms only if they can support broader social reforms. Green proposes a three-step framework to implement this method, emphasizing diagnosis of inequality, design of structural interventions, and careful consideration of whether algorithms should be used at all. This approach repositions fairness not as a fixed technical goal, but as an evolving normative practice tied to justice and political struggle [6].

Threats and Opportunities of Students' Use Of AI-Integrated Technology (ChatGPT) in Online Higher Education: Saudi Arabian Educational Technologists' Perspectives

Mihmas Mesfer Aldawsari & Rashed Ibrahim Almohish provide an in-depth qualitative investigation into the perceived threats and opportunities associated with students' use of AI integrated technologies specifically ChatGPT in online higher education, based on interviews with 20 Saudi Arabian educational technologists [7]. The study identifies ten key opportunities, including enhanced personalization, increased access to information, efficient feedback, 24/7 accessibility, global collaboration, interactive learning, cost-effectiveness, remote learning support, data-driven insights, and better preparation for future job markets. Simultaneously, it surfaces eight perceived threats: privacy concerns, variable content quality, digital dependence, teacher redundancy, unequal technological access, social isolation, ethical dilemmas (e.g., algorithmic bias), and human-AI interaction challenges. The authors emphasize that while AI tools like ChatGPT can democratize education and enhance flexibility, they may also exacerbate inequality, weaken critical thinking, and undermine educator roles if not ethically and pedagogically regulated. The study advocates for balanced integration, proactive policy development, and transparent practices to ensure that AI enhances rather than undermines educational equity and outcomes.

A Comprehensive Survey on Bias Mitigation in Machine Learning: Techniques, Challenges, and Future Directions

Siddique et al. discusses the challenges and strategies in mitigating bias in machine learning, particularly within high-stakes domains like education, healthcare, and hiring [8]. It categorizes biases into selection, confirmation, algorithmic, and data label biases, and reviews methods such as pre-processing, in-processing, and post-processing for mitigation. The authors highlight the trade-off between fairness and model accuracy, noting that no single method performs best in all scenarios. For instance, while some techniques improve fairness metrics like equalized odds or demographic parity, they often reduce predictive accuracy. The survey also discusses fairness constraints such as demographic parity and equalized odds, emphasizing their implementation through objective function modifications or post-hoc adjustments. Furthermore, it introduces intersectional fairness, fairness in deep learning, and privacy-preserving ML as emerging areas. Future directions include standardizing evaluation metrics, improving transparency, and contextualizing fairness frameworks. The use of tools like VOSviewer for bibliometric mapping reveals research trends and gaps, while the paper calls for multidimensional and interdisciplinary collaboration to ensure ML systems are fair, inclusive, and trustworthy.

Are Algorithms Biased in Education? Exploring Racial Bias in Predicting Community College Student Success by Bird, Castleman, and Song (2025)

Bird et al. present an in-depth investigation into algorithmic bias in higher education by analyzing two random forest models one predicting course completion and the other degree completion using student-level data from the Virginia Community College System (VCCS). The authors find significant evidence of calibration bias, where Black students with the same predicted risk scores as White students have lower actual success rates, meaning they would receive fewer resources under typical "at-risk" targeting systems [9]. The study also uncovers accuracy bias, with slightly lower c-statistics for Black students (e.g., −3.01% for course completion), highlighting reduced model reliability for minority groups. Surprisingly, they observe that including racial predictors or creating race-specific models reduces bias in course prediction but worsens it for degree prediction emphasizing the contextual complexity of fairness interventions. The authors further rule out underrepresentation and differential sorting as primary bias sources, instead attributing the problem to Black students' shorter enrollment histories and data limitations in existing administrative records. They conclude that predictive systems may unintentionally perpetuate inequity and urge institutions to collect richer data and demand transparency from vendors to ensure equitable allocation of academic support.

Algorithmic Bias in Educational Systems: Examining the Impact of AI-Driven Decision Making in Modern Education" by Boateng and Boateng [10]

Boateng & Boateng provide a critical overview of how algorithmic bias manifests across various educational systems, focusing on its implications for equity in admissions, assessment, and learning management [10]. They outline that biases can originate at multiple stages data collection, model design, and institutional implementation often disadvantaging marginalized groups by reinforcing historical inequalities. The paper identifies major sources of bias, including proxy discrimination, feature selection bias, and algorithmic discrimination, and emphasizes how these issues disproportionately affect students by race, gender, socioeconomic status, and disability. Through a detailed review of current studies, the authors demonstrate that AI systems in education risk amplifying disparities when built on incomplete or biased data, especially in high stakes domains like admissions and grading. They propose mitigation strategies such as algorithmic auditing, fairness metrics (e.g., ABROCA), diverse developer teams, and policy reforms grounded in ethical AI principles. The study concludes that addressing algorithmic bias requires not only technical solutions but also robust institutional accountability and interdisciplinary collaboration to build fair, transparent, and inclusive educational technologies.

How Do the Existing Fairness Metrics and Unfairness Mitigation Algorithms Contribute to Ethical Learning Analytics? by Deho et al. [11]

Deho et al. conduct a comprehensive evaluation of fairness metrics and unfairness mitigation algorithms within the context of learning analytics (LA), using a dropout prediction task as a case study [11]. The study compares eight widely cited mitigation algorithms across the entire ML pipeline pre-processing, in processing, and post-processing on both educational and benchmark datasets. Notably, the authors find that data bias does not always lead to predictive bias, and that fairness-enhancing techniques may sometimes improve utility rather than compromise it. Algorithms like Learning Fair Representations (LFR) and Disparate Impact Remover (DIR) occasionally produced more accurate models than the original biased data, suggesting the potential of debiased data to be "richer." However, they also observe that no single algorithm performs best across all fairness metrics, and that performance can vary between models even when trained on the same fair data. Through experimental analysis, the study shows how careful hyperparameter tuning is key to achieving balance between fairness and predictive performance, and introduces the concept of "bounded consistency" between group and individual fairness metrics. The authors conclude that while algorithmic fairness tools can support ethical LA, they must be deployed cautiously, with humans-in-the-loop and a shift in focus from equality to equity.

Fair AIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications by Chinta et al. [4]

Chinta et al. present a comprehensive survey of fairness, bias, and ethical considerations in AI-driven educational systems, examining technical, societal, and policy dimensions [4]. The paper categorizes bias in AIED into three main types: data-related (e.g., historical, measurement, and representation biases), algorithmic (e.g., learning, mapping, and confirmation biases), and user interaction biases (e.g., stereotyping, exclusion, and interaction feedback loops). Through case studies across grading, admissions, recommender systems, and curriculum design, the authors highlight real-world consequences of biased AI, such as minority students receiving lower essay scores or being excluded from advanced course recommendations. The paper reviews fairness metrics like statistical parity, equal opportunity, and individual fairness via Lipschitz constraints and introduces the Fairness-Bounded Utility (FBU) framework to visualize the trade-offs between fairness and model performance. Bias mitigation strategies are grouped into pre-processing (e.g., reweighing, SMOTE, GANs), in-processing (e.g., regularization, adversarial debiasing), and postprocessing (e.g., threshold adjustments). The study also surveys regulatory frameworks like UNESCO's and IEEE's ethical AI guidelines. Concluding, the authors call for interdisciplinary collaboration, personalized fairness approaches, and robust privacy safeguards to ensure AI supports equity rather than exacerbating educational disparities.

Methodology

Research Approach

This study uses a qualitative research approach to explore educators' experiences, perspectives, and strategies regarding the use of LLMs like ChatGPT in assessments. The aim is to understand real-world challenges and fairness concerns from those directly involved in managing academic integrity.

Participants

Participants were academic staff and educators from various institutions where students have access to LLMs. Educators with firsthand experience in dealing with LLM use during assessments were selected using purposeful sampling.

Data Collection Methods

This study primarily relied on semi-structured responses collected via a Google Form containing a consistent set of open-ended questions. The list of questions is available in Appendix A. While interviews were initially considered, participants were instead offered the form-based method due to time and access constraints. The questions covered areas such as LLM usage, fairness, teaching adaptations, institutional guidance, and educator strategies. All responses were gathered with informed consent and remained anonymous. The questions were forwarded to around 400 professionals; the amount of received responses were 121(Refer to Figure 2) whereas out of 121 responses we noticed that 73 number of responses were positive and 48 were negative. (Refer to Figure 3).

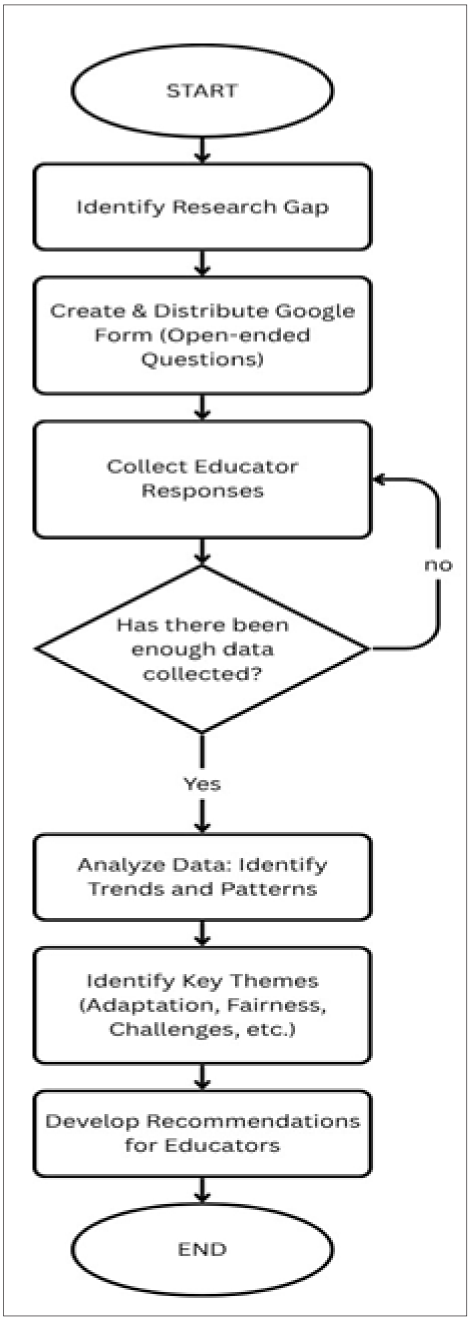

Figure 1: Explains the overall flow of the complete study

Research Procedure

The study began by identifying a relevant research gap and defining key research objectives through literature review. After selecting participants from a personal network of educators, data was collected exclusively through the google form and interview. The responses were reviewed and analyzed to identify trends, challenges, and educator responses concerning LLM usage in assessments. All the participant responses that were relevant to our study are available in Appendix B, with the participants being labelled as P1 to P8.

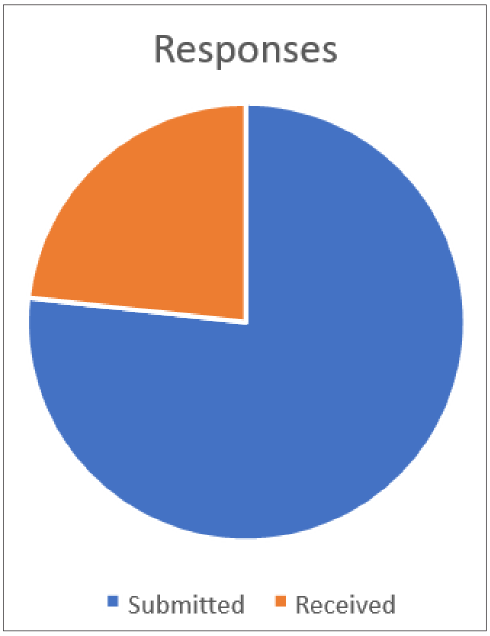

Figure 2: Presents the actual responses received from the professionals

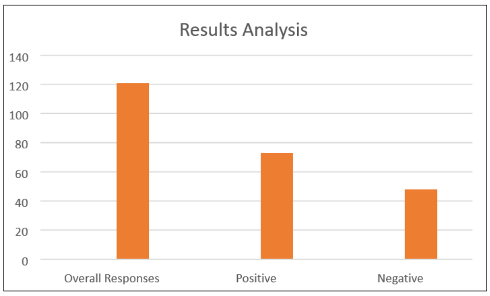

Figure 3: Show the actual result of study

Data Analysis

Responses were thematically reviewed to identify recurring patterns across the educators' perspectives. The analysis focused on common strategies, fairness concerns, adaptation methods, and institutional limitations shared by participants.

Ethical Considerations

Although an explicit digital consent form was not embedded in the Google Form, participants were clearly informed that their responses would be used for academic research. Participation was entirely voluntary, and no personally identifying data (apart from email addresses used solely for response validation) was included in the final report. All responses were anonymized (e.g., P1– P8) and treated with strict confidentiality throughout the analysis.

Expected Contribution

This study aims to offer deep qualitative insights into how educators manage fairness and validity in assessments when LLMs are in use. The findings may help inform future institutional policies and practices that adapt to evolving AI technologies in education.

Qualitative Theme Analysis

Detection of LLM Use

Most educators indicated that they are able to detect the use of AI-generated content by observing distinct patterns in students' writing. These include unusually formal language, overly polished phrasing, and vocabulary that doesn't match a student's typical style or level. For example, P3 remarked, “ChatGPT is known for using certain words and jargon... no student can write in this manner,” pointing to clear linguistic indicators. Similarly, P7 noted that abrupt shifts in tone or sentence structure often signal AI involvement. Other participants relied on AI detection tools like plagiarism checkers (P1, P6), while some (like P4) made inferences based on inconsistencies during classroom discussions, suggesting an intuitive, experience-based form of detection.

Assessment Redesign and Pedagogical Strategies

In response to the increasing use of LLMs, several educators have begun rethinking how they assess students. P1 and P3, for instance, shifted toward in-person assessments that require Realtime application of knowledge, reducing opportunities for external AI assistance. P6 introduced a hybrid model where students submit AI-generated content but are also required to defend their work orally. This method encourages students to engage more deeply with their submissions. Meanwhile, P8 transitioned to problem-based learning and case study approaches, promoting creativity, critical thinking, and authentic engagement that AI tools cannot easily replicate.

Academic Integrity and Fairness

Most participants acknowledged that unregulated use of LLMs could undermine academic integrity and fairness. P1 and P3 expressed concern that AI-generated work might be submitted without genuine understanding, diminishing the purpose of assessment and fostering surfacelevel learning. However, not all responses were critical P7 and P8 offered more nuanced views, arguing that LLMs are not inherently unethical but require proper contextualization. P7 compared LLMs to calculators, highlighting that their ethical use depends on how they’re integrated into teaching. This perspective suggests that fairness is less about the tool itself and more about the framework surrounding its use.

Institutional Response

Institutional support appeared to vary greatly across participants. While some educators (P6, P7, and P8) reported that their institutions had started providing training sessions, AI detection tools, or written guidelines, others (P1, P4, and P5) felt that no concrete support had been offered. This inconsistency created uncertainty and placed the burden of response squarely on individual educators. P3 shared that their institution implemented an “AI Assessment Scale,” hinting at a more structured approach but without elaborating on its success or adoption. Overall, these findings indicate that institutional readiness is still uneven, affecting educators’ ability to respond effectively.

Instructor-Level Strategies and Challenges

Despite limited guidance, many educators demonstrated proactive strategies at a personal level. P7, for example, engaged students in post-submission conversations to confirm whether they truly understood the material. P5 encouraged students to use AI as a learning tool but insisted on paraphrasing and critical reflection. P4 admitted to having no structured plan yet, showing the gap that can exist even among motivated instructors. P8 also expressed that their current strategies were still being tested, suggesting that educators are experimenting but haven't yet arrived at fully reliable solutions.

Student Learning and Dependency

Participants expressed mixed views on how LLMs affect student learning. On one hand, some educators (P1, P5, P7) noticed increased curiosity and engagement among students who used AI tools to deepen understanding or explore unfamiliar topics. These tools, when used wisely, can enhance learning outcomes. On the other hand, concerns were raised about overreliance, where students begin to trust AI-generated content blindly (P3), or fail to critically assess its quality (P4). This dependency can potentially undermine independent thinking and reduce students' ability to form their own arguments or solutions.

Support Needs and Recommendations

Almost all participants voiced a strong need for better support to deal with the rise of AI tools in education. P7 clearly articulated that "educators need a combination of training, clear guidelines, and institutional support." Participants suggested targeted workshops (P4), increased access to AI tools for learning and detection (P2), and ongoing discussions about ethical usage. P8 emphasized that beyond technical resources, a cultural and pedagogical shift is needed one that embraces innovation while reinforcing accountability. The consensus was that a coordinated, well-informed approach would empower educators to adapt fairly and confidently.

Discussion

The findings of this study highlight the complex, often inconsistent ways educators are adapting to the presence of large language models (LLMs) such as ChatGPT in assessment contexts. While most prior research has focused on algorithmic fairness, bias mitigation, or the technical accuracy of AI systems in education (Iowa, 2024), this study shifts the lens toward human-level responses how educators are navigating fairness, ethics, and assessment validity in real classrooms [12,4].

Educators in this study reported varied levels of institutional guidance and support, with most describing either general integrity guidelines or a complete absence of structured policy around LLM usage. This aligns with Jamil, who emphasized the uneven rollout of AI tutoring tools and the resulting inequities in access and usage [1]. The absence of clear institutional frameworks leaves many educators to independently develop strategies ranging from redesigning assessments to incorporating AI detection tools to manage fairness and academic honesty.

There was no single consensus on whether LLMs harm or support fairness. Some respondents viewed them as a threat to academic integrity, citing risks like plagiarism or shallow engagement with content. Others framed them as learning aids, much like calculators or grammar checkers, echoing arguments by fairness is less about the tools themselves and more about their use in context. This perspective reflects a shift from formal fairness metrics toward more substantive, educator-defined interpretations of fairness, especially in fast-evolving digital classrooms [6,7].

Interestingly, several educators acknowledged the potential benefits of LLMs, including deeper student engagement, better brainstorming, and improved confidence for struggling learners. However, these advantages were usually accompanied by concerns over over-reliance, echoing Jamil's idea of "ignorant bias" Where students fail to critically engage with LLM outputs [1]. This tension mirrors broader findings in the literature about the double-edged nature of educational AI [2,13].

One of the more actionable insights from this study is the call for targeted training and clear policy development. Educators want to use these tools effectively but they need support. As echoed by effectiveness of AI in education depends not just on algorithms, but on the systems and stakeholders that surround them. Without institutional buying, educators are left to balance fairness and innovation on their own [4,11].

Finally, while fairness is a central concern, this study also revealed that fairness in the LLM era cannot be resolved through binary decisions like prohibiting or allowing AI. As noted by several participants, setting clear, contextual boundaries rather than rigid rules may be the more sustainable path forward [14-23].

Recommendations

Incorporate Oral Defenses or Presentations to Validate Understanding

One of the most effective ways to mitigate the misuse of LLMs is to require students to orally defend or present their submitted work. As found in the study, several educators (e.g., P6, P7) already use this strategy, combining AI-written content with live presentations or follow-up conversations. This method ensures that students actually understand the material and are not merely submitting AI-generated output without engagement. In practice, even a five-minute viva or informal check-in can reveal whether the student has critically interacted with the topic or relied entirely on generative tools.

Enhance Institutional Support through Training and Clear Guidelines

A recurring theme across the responses was the lack of structured institutional support. Many educators (e.g., P1, P4, P5) highlighted that they received little to no formal guidance on how to handle LLM use in academic settings. This lack of direction leads to inconsistent enforcement and personal stress for educators. Institutions should invest in regular workshops, training modules, and policy documents that not only explain what LLMs are, but also how to integrate, regulate, and assess them fairly. Clear expectations can ease the burden on individual educators and ensure a standardized approach across departments.

Redesign Assessment Strategies to Resist AI Exploitation

Educators are already trying to make their assessments more resistant to AI misuse as seen with P1, P3, and P8. However, as P3 mentioned, these efforts are still in the trial-and-error phase. A more systematic redesign of assessments is required this includes a shift toward open-book exams, project-based tasks, and in-person evaluations. These formats demand authentic engagement and personal reflection, which AI tools cannot easily replicate. Moreover, incorporating interdisciplinary or real-world problem-solving questions can make assessments more meaningful and less susceptible to being solved solely through LLMs.

Promote Ethical and Responsible Use of LLMs in the Curriculum

Prohibiting LLMs entirely is neither practical nor educationally beneficial. Instead, educators should explicitly teach students how to use AI responsibly. This means framing LLMs as tools for ideation, brainstorming, clarification, or structure not as content creators to be blindly submitted. Participants like P5 and P7 highlighted the value of embracing the tool but setting boundaries. Embedding this philosophy into classroom culture will help students develop discernment and avoid the trap of academic dishonesty.

Establish Transparent and Adaptive Policies on LLM Use

Currently, institutional policies on LLMs are vague or nonexistent, leading to a grey area where both students and staff are unsure of what is allowed. This ambiguity can damage trust and lead to inconsistent academic practices. Educational institutions should urgently create and disseminate detailed policies that outline acceptable use cases, assessment expectations, and consequences of misuse. These policies should be regularly updated in response to evolving AI capabilities and should include input from both educators and students.

Foster Reflective and Critical Learning Approaches

One of the risks identified in the findings is that students may become overly dependent on AI, potentially weakening their critical thinking and subject mastery. To counter this, educators should promote learning strategies that emphasize reflection, critique, and iteration. For instance, after using ChatGPT to draft an idea, students could be asked to critique the AI's output, compare it to academic sources, or explain their reasoning in a reflection journal. This not only enhances learning but also trains students to see LLMs as starting points rather than final answers.

Encourage Cultural and Pedagogical Shifts among Educators

Finally, the success of any adaptation strategy depends on the educators themselves. As shown by participants like P8, adapting to LLMs is not just about tools it is also about mindset. Institutions should foster a culture of adaptability, innovation, and openness to change among academic staff. Providing peer support groups, innovation grants, or communities of practice can help educators feel less isolated and more empowered to experiment with new methods and share what works.

Limitations

While the findings offer useful insight into how educators are dealing with student use of LLMs like ChatGPT, the study does have a few key limitations.

Small Sample Size

The most significant limitation is the limited sample size of eight participants; all selected through purposeful sampling. While their responses offer rich qualitative insights, there may be sampling bias due to accessibility constraints meaning only those within reach or willing to respond were included. As a result, the findings may not be generalizable across broader educational contexts. Future research could benefit from a mixed sampling strategy, incorporating educators from different institutions, disciplines, and regions (including international participants), to provide more diverse perspectives and reduce contextual limitations.

Self-Reported Data

All data collected was based on self-reported responses through an online form. This method relies on participant honesty and recall, and there is a chance of bias, overestimation, or underreporting. Additionally, because the form was asynchronous, the researcher could not ask clarifying or follow-up questions to deepen or contextualize ambiguous answers.

Lack of Student Perspectives

This study focuses solely on the views of educators. While this aligns with the research aim, it excludes the student voice, which could offer contrasting or complementary insights regarding fairness, assessment changes, and the educational value of LLMs.

Time and Access Constraints

Due to time limitations and the difficulty of securing interviews, a more detailed, mixed-method approach (such as follow-up interviews or classroom observations) was not feasible. As a result, the depth of interpretation may be constrained by the fixed structure of the Google Form used.

Evolving Nature of LLMs and Policy Landscape

The study captures educator experiences at a specific point in time during a rapidly evolving period of AI integration. As tools like ChatGPT develop and institutional policies shift, perceptions and practices may change significantly, meaning the findings might not remain representative in the long term.

Conclusion

This study explored how educators are adapting their teaching and assessment practices in response to the growing presence of Large Language Models (LLMs) like ChatGPT in higher education. By analyzing qualitative responses from eight educators across different institutions, the research highlighted key themes related to academic integrity, fairness, institutional support, student learning behaviors, and the strategies educators are currently employing or struggling to implement in real-world settings. Findings suggest that while most educators acknowledge the presence of AI-generated content, their responses vary widely. Some have redesigned assessments, embraced oral defenses, and encouraged critical AI engagement, while others remain uncertain or unsupported. Importantly, the study reveals a clear gap between the pace of AI integration and institutional readiness. There's a need for urgent, structured support to help educators navigate the pedagogical, ethical, and technical challenges posed by these tools.

This research contributes to a relatively underexplored area by foregrounding educator perspectives not just policy, not just tech but the people responsible for upholding fairness and academic standards in classrooms increasingly influenced by generative AI. The findings emphasize that responsible integration, not restriction, is the most sustainable path forward. Institutions should prioritize practical strategies, such as training staff, drafting adaptable policies, and encouraging mixed assessment formats like presentations or oral vivas alongside written work to assess real understanding. These adjustments can help reduce over-reliance on AI while preserving the learning outcomes intended by academic tasks. Future research could expand on this work by involving a larger, more diverse sample of educators, including those from nontertiary or global institutions. It could also focus on discipline-specific differences, the impact of AI-aware pedagogy on student outcomes, or longitudinal changes in educator attitudes as LLM tools become even more embedded in educational ecosystems.

The call to action is clear: support educators now with training, policies, and room to experiment or risk leaving them overwhelmed in a rapidly evolving academic landscape.

References

1. Jamil HM. An Experiment with LLMs as Database Design Tutors: Persistent Equity and Fairness Challenges in Online Learning. Education Sciences. 2025. 15.
2. López-Meneses E, Mellado-Moreno PC, Gallardo Herrerías C, Pelícano-Piris N. Educational Data Mining and Predictive Modeling in the Age of Artificial Intelligence: An InDepth Analysis of Research Dynamics. Computers. 2025. 14: 1-26.
3. Ahmad S, El-Affendi MA, Anwar MS, Iqbal R. Potential future directions in optimization of students' performance prediction system. Computational Intelligence and Neuroscience. 2022. 6864955.
4. Chinta SV, Wang Z, Yin Z, Hoang N, Gonzalez M, et al. FairAIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications. 2024. 1-47.
5. Idowu JA. Debiasing Education Algorithms. International Journal of Artificial Intelligence in Education. 2024. 34: 1510-1540.
6. Green B. Escaping the Impossibility of Fairness: From Formal to Substantive Algorithmic Fairness. Philosophy and Technology. 2022. 35.
7. Aldawsari MMM, Almohish NRI. Threats and Opportunities of Students' Use Of AI-Integrated Technology (ChatGPT) in Online Higher Education: Saudi Arabian Educational Technologists' Perspectives. International Review of Research in Open and Distributed Learning. 2024. 25: 19-36.
8. Siddique S, Haque MA, George R, Gupta KD, Gupta D, et al. Survey on Machine Learning Biases and Mitigation Techniques. Digital. 2024. 4: 1-68.
9. Bird KA, Castleman BL, Song Y. Are algorithms biased in education? Exploring racial bias in predicting community college student success. Journal of Policy Analysis and Management. 2024. 379-402.
10. Boateng O, Boateng B. Algorithmic bias in educational systems: Examining the impact of AI-driven decision making in modern education. World Journal of Advanced Research and Reviews. 2025. 25: 2012-2017.
11. Deho OB, Zhan C, Li J, Liu J, Liu L, et al. How do the existing fairness metrics and unfairness mitigation algorithms contribute to ethical learning analytics? British Journal of Educational Technology. 2022. 53: 822-843.
12. Wongvorachan T, Bulut O, Liu JX, Mazzullo E. A Comparison of Bias Mitigation Techniques for Educational Classification Tasks Using Supervised Machine Learning. Information (Switzerland). 2024. 15: 1-25.
13. Sha L, Gašević D, Chen G. Lessons from debiasing data for fair and accurate predictive modeling in education. Expert Systems with Applications. 2023. 228.
14. Cohausz L, Kappenberger J, Stuckenschmidt H. What Fairness Metrics Can Really Tell You: A Case Study in the Educational Domain. 2024. 24.
15. Harry A, Sayudin S. Role of AI in Education Abstract. Injuruty: Interdisciplinary Journal and Humanity. 2023. 2: 2963-3397.
16. Hemal SH, Khan MAR, Ahammad I, Rahman M, Khan MAS, et al. Predicting the impact of internet usage on students' academic performance using machine learning techniques in Bangladesh perspective. Social Network Analysis and Mining. 2024. 14.
17. Huang X, Gadavanij S. Power and marginalization in discourse on AI in education (AIEd): social actors' representation in China Daily (2018–2023). Humanities and Social Sciences Communications. 2025. 12: 1-3.
18. Ibarra-Vazquez G, Ramírez-Montoya MS, Terashima H. Gender prediction based on university students' complex thinking competency: An analysis from machine learning approaches. Education and Information Technologies. 2024. 29: 2721-2739.
19. Lünich M, Keller B, Marcinkowski F. Fairness of Academic Performance Prediction for the Distribution of Support Measures for Students: Differences in Perceived Fairness of Distributive Justice Norms. Technology, Knowledge and Learning. 2024. 29: 1079-1107.
20. Rashed A, Kallich A, Eltayeb M. Analyzing Fairness of Computer Vision and Natural Language Processing Models. Information (Switzerland). 2025. 16: 1-21.
21. Setiadi H, Sanjaya K, Wijayanto A, Wardhani DW, Cahyono HD. Comparative Analysis of Classification Algorithms Using Feature Selection Techniques to Predict On-Time Student Graduation. Ingenierie Des Systemes d'Information. 2024. 29: 1365-1379.
22. Xia F, Guo T, Bai X, Shatte A, Liu Z, et al. SUMMER: Bias-aware Prediction of Graduate Employment Based on Educational Big Data. ACM Transactions on Data Science. 2022. 2: 39.
23. Zong R, Zhang Y, Stinar F, Shang L, Zeng H, et al. A Crowd–AI Collaborative Approach to Address Demographic Bias for Student Performance Prediction in Online Education. Proceedings of the Eleventh AAAI Conference on Human Computation and Crowdsourcing. 2023.