Research Article
2025, Vol. 01, Issue. 2, 35–47
https://doi.org/xxx

Journal of Innovations in
Social and Applied Sciences

Equity-Centered Oversight of Generative AI Systems in Education: Toward Inclusive, Transparent, and Accountable Learning Technologies

Haroon Arif [1], Aamir Raza[1]

[1] Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, USA

harif@hawk.iit.edu; araza7@hawk.iit.edu ✉;

Abstract

In the education sector, generative AI is gaining traction across applications such as intelligent tutoring, automated feedback, content generation, assessment support, and student advice. The use of these applications can be used to expand access to learning and provide personalized learning experiences. They also introduce relevant questions, including those about the potential for inaccurate or discriminatory results, opaque algorithms, privacy issues, overreliance on automated systems and unequal outcomes. This article suggests that equity-based oversight is a socio-technical approach thaed to help guide the use of generative AI in education. They are designed in keeping with the principle of fairness, inclusion, transparency, accountability and contestability, which are not an afterthought or a part of the process after they haveIt highlights the need to engage students and educators in the design of the systems, to explain why their work is important and meaningful, to provide an understanding of the outputs from AI, to include effective human review, to conduct regular auditing of AI to ensure equity, fairness, and the prevention of harmful interactions.ul interactions. The article goes beyond the technical aspects of AI's performance to emphasize the implications for educational justice, institutional accountability, and learner protection. It provides actionable insights on creating and deploying generative systems that can improve personalization without worsening social and learning inequities.

Article History

Manuscript Received
July 26, 2025

Revision Received
November 29, 2025

Final Acceptance
December 11, 2025

Keywords: Generative Artificial Intelligence; Equity-Centered Oversight; Inclusive Education; AI Governance; Responsible AI

1 Introduction

The use of GenAI in education is gaining traction, with solutions like intelligent tutoring, automated feedback, AI-generated learning content, assessment support, writing assistance, and student advising at the forefront. These systems provide great potential for customized learning, timely academic support, and access to wider learning resources. Meanwhile, the quick adoption of these creates significant issues around bias, exclusion, opacity, privacy, over-reliance, and inequities in learning. Prior studies on the use of artificial intelligence in higher education have demonstrated that AI tools are developed and tested mostly from a technical perspective, with limited consideration for the pedagogical, learner, and institutional contexts of educational technology tools and processes [1]. With the advent of large language models, these concerns become more relevant, as GenAI systems can generate convincing educational content and inaccurate, biased, culturally narrow, and pedagogically inappropriate outputs [2].

Managing GenAI in education is not merely a matter of evaluating technical performance. AI systems are now recognized as socio-technical systems shaped by data, design choices, institutional practices, cultural beliefs, and relations of power [3]. In the field of education, these factors are especially relevant, as AI systems can impact the learning process, performance evaluation, and the forms of support that can be offered by educational institutions, among other aspects, as well as how students from different socio-economic backgrounds gain access to and participate in education. This is a particular consideration for students for whom language, disability, socioeconomic status, gender, geography, and a lack of digital infrastructure are already a disadvantage.

There are emerging trends of digital systems becoming an antagonist to manipulative, exclusionary or difficult-to-challenge interaction designs as well. Hostile interaction design, as described by Foth [4] is when the combination of manipulation, friction, extraction and weak oversight diminishes the autonomy and accountability of the user of a digital system. Similar dangers can also occur in the classroom when using GenAI tools, including when they provide unclear recommendations, diminish students' ability to act, standardize judgments, and confuse teachers and students with AI-generated content. This does not mean, however, that traditional human-in-the-loop systems are still adequate, unless a way can be found to engage the human in a meaningful manner, stay in touch with the context, and review, challenge, and/or override decisions of AI systems.

Equity-focused oversight can help respond to these concerns by making fairness, inclusion, transparency, and contestability integral to governance rather than merely a moral requirement. The overall idea of this article is to suggest a socio-technical governance approach to GenAI systems in education grounded in equity. The framework examines the design, implementation, and responsible monitoring of educational GenAI systems that do not harm diverse learners, and includes concepts of responsible AI, equity, diversity, and inclusion, human-centered design, and hostile interaction design. Equitable AI-supported learning is a comparatively new field, however, and focuses on doing more to prevent harm by paying attention to accessibility and learner diversity and pedagogical justice [5]. Therefore, in education, the use of GenAI must be evaluated not only on its effectiveness and accuracy, but also on its ability to empower student agency, reduce rates of exclusion, make education contestable, and promote responsible education.

The objectives of this article are as follows:

1. To examine the major equity-related risks associated with the use of generative AI systems in educational contexts.
2. To analyze the limitations of conventional human-in-the-loop and technical performance-based approaches to AI oversight in education.
3. To propose equity-centered oversight as a socio-technical framework for governing generative AI systems across the educational AI lifecycle.
4. To identify the key dimensions of the proposed framework, including inclusive data practices, participatory design, transparency, contestability, meaningful human review, and continuous bias auditing.
5. To provide practical guidance for researchers, educators, institutions, and AI developers seeking to design and deploy generative AI systems that support diverse learners without reinforcing existing educational inequalities.

2 Literature Review

AI in Education (AIEd) is now an established interdisciplinary area that covers intelligent tutoring systems, adaptive learning environments, learning analytics, automated assessment, student modeling, predictive systems, and learning recommendation systems. Zawacki-Richter et al. [1] conducted the first systematic study of AIEd in higher education and found that research was primarily focused on profiling and prediction, assessment and evaluation, adaptive systems, personalization, and intelligent tutoring. The authors also pointed to the underrepresentation of teachers in the research discourse, highlighting the ongoing gap between technologization and pedagogization. A similar concern was put forth by Chiu et al. (2023) who noted the need to bridge gaps between technical innovation and learning goals, teacher roles, ethics, and future education needs.

In the meta-literature review on AI in higher education by Bond et al. [6], there was a call to pay more attention to ethics, interdisciplinary work, methodological rigor, and sensitivity to context. Wang et al. (2024) [7] also reported that AIEd studies have been growing in various domains, including Adaptive Learning, Intelligent Assessment, Educational Management Systems, Profiles, Prediction, and New AI-based Products. Collectively, these studies have indicated that the question is not one of whether AI can be adopted into education but, how.As a whole, these studies have made it clear that the question has become not whether AI can be integrated into education, but how.

This has further changed with the advent of generative artificial intelligence (GenAI) and specifically large language models. While most traditional AIEd systems focused on specific tasks and relied on rigid rule sets, GenAI systems can generate text, explanations, feedback, code, images, lesson plans, assessment materials, and conversational responses. Kasneci et al. [2] highlight the significant potential of LLMs in education, ranging from personalized support to writing assistance, formative evaluation, tutoring, and teacher support. Meanwhile, they highlight misinformation, bias, over-reliance, privacy issues and the importance of effective human oversight. The above problems suggest that GenAI is not just another instructional technology. It can change the teacher–learner, institution–learner, and knowledge-institution relationships.

This stream of educational review research has aimed to explain the rapid development of GenAI in education. Yusuf et al. [8] analyzed 407 publications and identified several main themes, including applications, impacts, ethics, adoption, productivity, and gaps and questions. Samala et al. [9] conducted a scoping review of 453 studies to create a broad taxonomy of AI applications in healthcare, covering both applications and limitations, ethical considerations, and future prospects. Xia et al. [10] stated that GenAI requires a radical shift in thinking about assessment in higher education, which involves further rethinking about the concepts of self-regulated learning, responsible learning, academic integrity and teacher development, as well as institutional policy change. These studies in total indicate that the impact of GenAI goes beyond classroom learning to the epistemic, ethical, and institutional underpinnings of education. Empirical work on educators' perspectives also suggests that institutional preparedness remains uneven. Lee et al. [11] found that educators acknowledged the transformative potential of GenAI while expressing uncertainty about good practice, academic integrity, assessment redesign, and institutional support. Their findings indicate that meaningful adoption depends on more than access to the technology. It also requires professional development, clear policy, pedagogical adaptation, and sufficient institutional capacity. Similarly, García-López et al. [**?**] identified technological integration, personalization and equity, data quality and security, and human–AI collaboration as central challenges in the educational use of ChatGPT. These findings reinforce the view that GenAI governance must combine technical safeguards with pedagogical, institutional, and ethical forms of oversight.

Equity has thus emerged as a key issue in this governance debate. Based on equity principles, especially when considering minoritized learners, students from under-resourced backgrounds, and STEM education, the adoption of AI in teaching and learning should be guided (Garcia Ramos and Wilson-Kennedy, 2024). The focus of their discussion is the concept of Universal Design for Learning and the need to design learning settings to address differences among learners from the beginning. This is particularly important in the context of GenAI and the use of large data sets to create AI systems that reinforce dominant linguistic, cultural, and epistemic patterns while neglecting those whose identities (language, disability or educational background) are under-represented. In this view, the focus on equity is not a 'moral nicety' but an essential component of responsible AI in education.

Additional support for enhanced governance structures is found in the ethical and regulatory literature. Nguyen et al. (2023) [12] reviewed the role of ethics in AIEd and highlighted the need for privacy, autonomy, transparency, accountability and the accountability of various stakeholders. Based on 42 studies of the ethics and governance of GenAI in education, Alfiras et al. (2025) identified privacy and data protection, algorithmic bias, fairness, transparency, explainability, student well-being, and accountability with human oversight as some of the key themes. García-López and Trujillo-Liñán [13] analyzed 53 peer-reviewed articles and found similar concerns such as privacy of the data, the existence of algorithmic bias, educational inequality, diminished cognitive autonomy, institutional use of student data, and insufficient regulatory oversight. This set of work highlights that education's GenAI can't be managed by general ethics. There also needs to be mechanisms in practice to monitor, explain, redress and take institutional responsibility.

Large language models give rise to intricate integrity issues, as AI-generated content can often be indistinguishable from human writing using current plagiarism detection tools or even by academic standards [14]. Eke [15] noted that, while such tools as ChatGPT can help with academic tasks, their abuse can have a negative impact on academic integrity if institutions have no multi-stakeholder policies or responsible use guidelines. According to the authors of a later paper, "Detecting Text Generated by GPT-4 is Not Trivially Solvable" [16], there is a need for academic judgment and software assistance to detect GPT-4-created text, and that it is not possible to detect text generated by GPT-4 entirely. The results indicate that the problem-solving approach is not the only approach that works. However, the design of assessment, acceptable uses of AI, enhancing AI literacy, and maintaining meaningful human judgment must all be rethought at the institutional level.

According to Park et al. [17] GenAI literacy should encompass the understanding of GenAI, its proper application, assessment of its results, consideration of ethical issues, and the formation of informed opinions on its usage. This is aligned with the wider discussion about the need for students and teachers to be ready to recognize and understand the limitations, biases, and social impacts of GenAI. Without this literacy, supervision is uneven; institutions may be implementing GenAI systems and learners may not be able to challenge outputs, identify risks, or have a meaningful voice. The ratio is especially problematic for students who are vulnerable or marginalized, who may have barriers to learning in the education system already.

Equity, diversity and inclusion (EDI) are broader concept theories that can be used to help address these concerns. Zowghi and Bano [3] contend that the study of EDI in AI has progressed beyond mere symbolic statements or gestures to an agenda for action, encompassing algorithmic fairness, participatory AI, inclusive data practices, explainability and transparency in AI, accountability in AI, and global and intersectional governance. The synthesis is particularly relevant to GenAI in education, given that education is rich in values and diverse contexts. Learners come from a diverse range of linguistic, cultural, disability, socio-economic, gender, geographic, preparatory and digital environments. A system can thus do well for the assumed average learner while leaving learners whose needs, experiences or knowledge traditions different from the average aside.

The idea of hostile interaction design also offers a valuable lens on how GenAI can disadvantage the learner, even when the design intent is not to discriminate [4]. The use of GenAI in education could increase students' hostility by hiding the source of feedback, fostering cognitive dependency, legitimizing algorithmic judgments, shifting accountability from educators to learners, and making automatically produced content harder to question. Governance, however, needs to go beyond bias reduction and academic integrity. It needs to be contestable, explainable, participatory, inclusive data governance and meaningful human review too.

Taken together, the literature suggests a definite research gap. Previous research has focused on the use of AI in education, opportunities and challenges of GenAI, academic integrity, ethical guidelines, teacher attitudes and perceptions, AI literacy, equity, and regulatory considerations. These areas are often talked about separately, however. There is still a need for an integrated framework to weave in the concepts of equity, transparency, contestability, human oversight, learner agency, and hostile interaction design in governing GenAI in education. To meet this need, this article suggests a socio-technical approach for responsible use of GenAI in educational contexts: equity-centered oversight. In this case, table 1 summarizes the representative studies that provide support for the conceptual foundation for this article. Figure 1 provides a conceptual synthesis of the literature and reveals the fact that equity-centered oversight is the result of five converging trends in the use of GenAI in education: the widespread uptake of GenAI in education, recurring ethical and pedagogical risks, pedagogical consequences that were unevenly distributed across learner groups, less formal oversight, and growing demand for transparent and accountable governance. The number is a synthesis of the literature reviewed and is not to be used as an independent empirical data set.

Table 1: Summary of key literature on GenAI, education, equity, and oversight

Study	Focus / Method	Key Contribution
Zawacki-Richter et al. [1]	Systematic review of AI in higher education.	Identified the principal application areas of AIEd.
Kasneci et al. [2]	Analytical review of LLMs in education.	Examined LLMs as both pedagogical resources and sources of ethical risk.
Bond et al. [6]	Meta-systematic review of AI in higher education.	Emphasized ethics, methodological rigor, and interdisciplinary collaboration.
Xia et al. [10]	Scoping review of GenAI in assessment.	Connected GenAI adoption with the redesign of assessment practices.
Garcia Ramos and Wilson-Kennedy [5]	Equity-focused discussion of AI in teaching and learning.	Established equity as a central consideration in educational AI adoption.
Nguyen et al. [12]	Ethical principles for AI in education.	Articulated core principles for the ethical governance of AIEd.
Zowghi and Bano [3]	Editorial synthesis of EDI trends in AI.	Presented EDI as an operational component of AI governance.
Foth [4]	Conceptual analysis of hostile interaction design.	Connected AI governance with contestability and meaningful oversight.

3 Conceptual Background and Research Gap

The view of generative AI in education needs to be more than just a neutral technology to help with teaching or a technical addition to current LMS platforms. Instead, it needs to be viewed as a socio-technical system that is nested in educational institutions, pedagogy, data processes, governance structures and communities of learners. This view has significance because the learning impact of GenAI is not a one-to-one mapping of the accuracy or computational power of the model, but also shaping by design, deployment, interpretation, monitoring and contestation in the context of actual learning contexts. The outputs of a GenAI system can be beneficial, if it provides explanations, feedback, learning resources, etc., but also be inequitable if it privileges dominant linguistic norms, if it acts with an inadequate awareness of learner diversity, if it fails to share the source of its output, or if it does not enable learners or educators to challenge its recommendations.

3.1 Generative AI as a Socio-Technical System in Education

The socio-techological approach to GenAI highlights that the development of educational AI systems is influenced by a dynamic interaction between the technical artefacts and social arrangements. The training of large language models on enormous amounts of data captures the linguistic, cultural, ideological and institutional patterns. Incorporated into educational practices, the outputs of these models become a component of the teaching, learning, assessment, advising, and administrative decision-making. In education, the use of GenAI is not just a matter of automating specific tasks, but also a transformation of the dynamics between learners, teachers, knowledge and institutions.

Previous studies on the use of AI in higher education have demonstrated that in the past, Educational AI applications have been predominantly tied to profiling, prediction, assessment, evaluation, adaptive systems, personalisation and intelligent tutoring [1]. GenAI brings this previous context to a new level, with the ability to interact in an open-ended manner, generate content, provide feedback via conversations, and provide pedagogical support via automation. These are new modes of flexibility, but they also are new modes of uncertainty. Unlike conventional education technologies, GenAI results are probabilistic

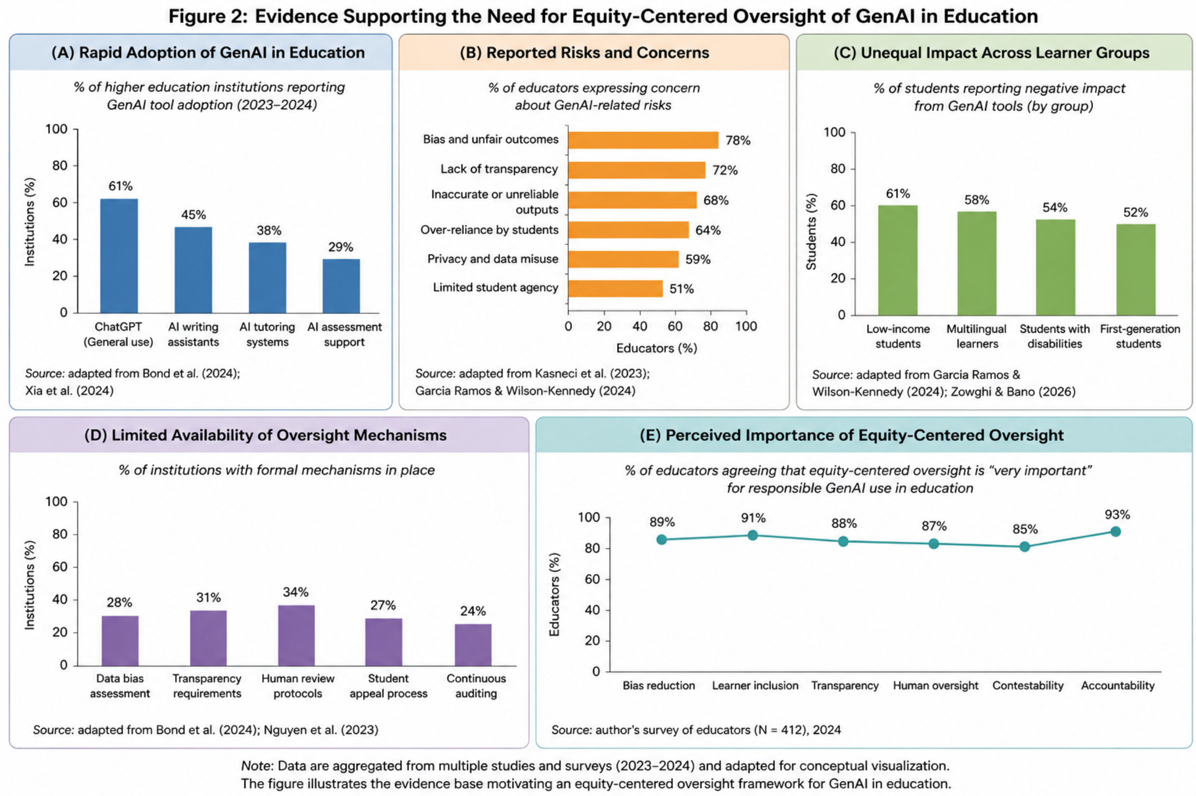

Figure 1: Conceptual synthesis of evidence motivating equity-centered oversight of GenAI in education. The figure summarizes themes identified in the literature, including the rapid adoption of GenAI, ethical risks, learner inequities, limited oversight mechanisms, and the need for accountable governance [2, 6, 10, 5, 12, 3].

and context-specific. May be coherent and authoritative, but with inaccuracies, biased assumptions or interpretations which are culturally limited. This can make governance a real challenge as errors may not always manifest themselves in the eyes of educators or students.

The socio-technical aspect of GenAI also requires the responsibility to not be entirely within the model, the teacher, the student, or the institution. The responsibility is shared by the developers, platform providers, educational institutions, policy makers, instructors, and learners. Without such distribution clearly controlled, there is a lack of accountability. This is particularly an issue in learning environments that involve stakes, such as assessment, academic advice or guidance, disability support, admissions counseling, and prediction of student performance. In these environments, the outputs from GenAI can have an impact on educational opportunities and learner pathways. So, not only will governance have to concern itself with the effectiveness of a GenAI system, but also its fairness, transparency, and accountability.

3.2 Equity, Diversity, and Inclusion in Educational AI

Equity, diversity and inclusion are a core consideration for assessing GenAI systems in education. Learning contexts are multilingual and heterogeneous, and learners vary with respect to their language, culture, gender, disability, socio-economic backgrounds, geography, previous school experiences, access to digital devices and learning needs. A system with a satisfactory performance for a generalized learner profile can have a negative impact on students whose experiences, learner profiles and/or knowledge systems are not represented in the training data or accounted for in system design.

Other literature on EDI in AI suggests that issues of fairness and inclusion should not be considered secondary moral issues. Zowghi and Bano [3] highlight that EDI is now a structural and operational agenda throughout the AI lifecycle, including: algorithmic fairness, participatory AI, inclusive data practices, explainability, transparency, accountability, and global, or intersectional governance. In learning, this means it's not enough to just eliminate clear bias from outputs to make inclusive GenAI possible. It necessitates a comprehensive approach to the development and implementation of educational AI systems, encompassing problem formulation, data selection, model development, interface design, institutional deployment, teacher training, student support, monitoring, and providing redress.

Equity-focused educational AI then should take into account distributional and procedural aspects of equity. Distributive fairness is the fairness of the distribution of benefits and burdens of GenAI, among various groups of learners. Procedural fairness relates to the meaningful participation of students and educators in the design, introduction, use, and evaluation of GenAI systems. For instance, an automated writing correction system can be efficient, but it can also disadvantage non-native English speakers, culturally diverse writing practices, and/or learners with disabilities when the criteria used to select what constitutes good writing are limited or unclear. Likewise, an AI tutoring system can help students learn on their own, but can also help them build inaccurate or inappropriate concepts if they don't have the ability to assess the accuracy and appropriateness of the concepts generated by the AI.

Thus, an EDI lens must go beyond the premise that access to GenAI automatically equals to educational inclusion. Equity is not the same as access! Students have varying digital literacy, AI literacy, language skills, confidence and critical awareness of outputs. Without adequate support, GenAI can exacerbate the divide between students who can use GenAI tools critically and those who use them passively or dependently. Learner agency, AI literacy, accessibility, cultural responsiveness, and institutional support, therefore, need to be all part of the governance concerns of equity focus.

3.3 Human Oversight, Contestability, and Accountability

Often the human element is mentioned as a way to safeguard against potential harm caused by AI. In school environments, this can be expressed as the fact that the teacher, the administration, or an academic staff member are formally accountable for evaluating the feedback, recommendations or decisions produced by an AI.In the context of education, this can be stated as the teacher, administration or academic staff member being formally responsible for evaluating the feedback, recommendations or decisions generated by an AI. The effectiveness of human oversight, however, relies on the extent to which human actors possess the knowledge, authority, time, institutional support, and procedural mechanisms needed to act meaningfully. Without proper training and transparency, it is possible that teachers might be tasked with overseeing GenAI systems without meaningfully carrying out their responsibilities. Without proper training and transparency, teachers might be asked to supervise GenAI systems without meaningfully doing so.

The notion of hostile interaction design can be helpful in understanding this risk. Foth [4] suggests that digital systems can be hostile when there is manipulation and friction, extraction and weak supervision, and limited contestability. In education, this implies that GenAI tools can have a negative impact not just through their outcomes but also through their usage patterns, which can diminish students' autonomy and/or make it harder for them to question institutional choices. For example, an AI-made academic suggestion may be pitched as advisory and shouldn't be interpreted as a direction of its own; however, when teachers or administrators take it as a place to go, the student's likelihood to challenge the suggestion may be reduced. Likewise, automated feedback can be presented in a positive light but without providing reasoning or giving students an opportunity to ask questions about the feedback, it can lead to dependence rather than learning.

Responsible GenAI in education thus requires contestability to be an essential aspect. Pupils and teachers should have access to the ability to challenge AI generated content, ask for clarification, flag inaccuracies, and report abusive or inappropriate responses as well as request human input when needed. In many situations including assessment, academic progression, disciplinary decisions, learning analytics, student support allocation, and scholarship decisions, contestability is of special significance. If there's no contestability, then GenAI systems can turn the task of correcting over to students while maintaining institutional opacity. This will have adverse effects on education equity and the sense of procedural justice.

Accountability should not be limited to an individual but to institutions too. Expecting teachers and/or students to handle risks of GenAI alone is not enough. Policies, documentation practices, audit procedures, training programs, and review mechanisms need to be put in place by educational institutions to define the process of selecting, evaluating, monitoring, and correcting the use of GenAI systems. The developers and EdTech providers should also be mandated to disclose details about system limitations, data practices, bias testing, accessibility, and avenues for reporting harm. In this context, meaningful oversight means an environment of accountability for governance and not a merely human-in-the-loop process.

3.4 Research Gap

As shown in the literature, there has been significant advancement in the knowledge of AI and GenAI in education. There have been previous research that has explored the historical evolution of AIEd, the educational opportunities of large language models, ethical and regulatory concerns, issues of academic integrity, teachers' views on AI in education, and the need for AI literacy and inclusive education. But, these research areas are not sufficiently connected. Many of the literature available is quite vague in terms of opportunities and risks of GenAI without putting them in the context of governance, and also, the studies on governance issues tend to focus on privacy, bias, transparency, or academic integrity issues separately. Likewise, equity arguments are about inclusion and learner diversity but do not always link up to contestability, hostile interaction design and meaningful institutional oversight.

This disintegration leaves a conceptual and practical void in addressing the needs of individuals with disabilities. The need for a model that integrates equity, diversity, inclusion, transparency, human oversight, contestability, redress and continuous auditing in the governance model of GenAI systems in the education sector. Such a framework should acknowledge that educational harm can be caused by inaccurate or biased model output as well as from institutional setups that make the output of AI models hard to understand, critique, and correct. It also needs to be cognizant of the fact that students are not monolithic and that GenAI tools can have different impacts on various students.

This article attempts to fill this gap by suggesting a socio-technical approach to oversight – grounded in principles of equity – for responsible adoption of GenAI in education. This framework sets the fair, inclusive, transparent, contested and accountable design and governance principles as a sequence throughout the educational AI lifecycle. The framework proposes to view oversight as an integral, ongoing and institutionally supported process as opposed to a last resorts check after AI has produced its outputs. Thus, it seeks to advance responsible AI scholarship by moving the focus away from performance and ethics to learner protection and educational justice, and accountable institutional practice. In Figure 2 the conceptual pathway of this paper is summarized. It demonstrates the socio-technical risks arising from the use of GenAI in education, and the need for an equitable approach to overseeing and managing those risks to ensure equitable and inclusive educational outcomes.

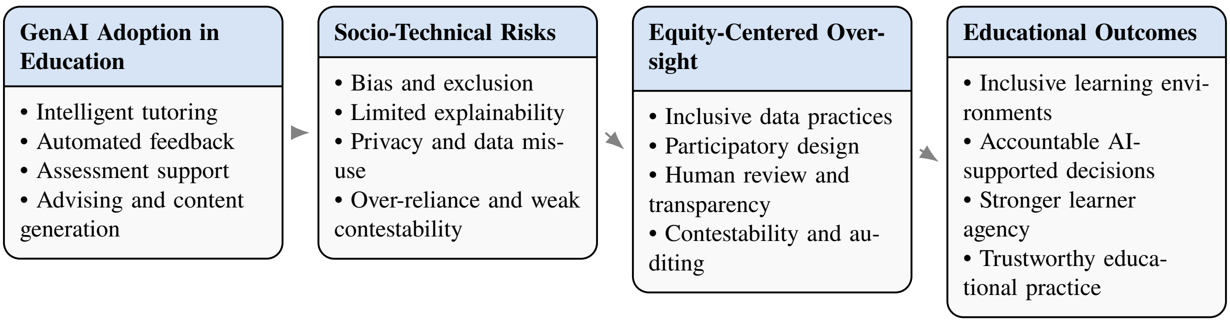

Figure 2: Conceptual pathway from GenAI adoption to equity-centered oversight in education

4 Proposed Framework

This section presents equity-centered oversight as both a practical and conceptual framework for guiding the responsible use of generative artificial intelligence in education. The central premise is that GenAI systems should not be adopted simply because they are efficient, scalable, or technically advanced. Their educational value depends on whether they promote fair learning opportunities, respect learner diversity, support teacher judgment, and remain open to review and correction. In this article, equity-centered oversight refers to a structured approach for ensuring that GenAI systems are designed, implemented, and monitored in ways that protect students from exclusion, bias, opacity, and unaccountable automation.

The framework comprises six closely connected dimensions: inclusive data and representation, participatory design, transparency and explainability, meaningful human review, contestability and redress, and continuous auditing. These dimensions should not be treated as isolated checklist items, since each contributes to a broader governance process. Inclusive data and representation help address representational bias, while participatory design ensures that students and educators have a voice in decisions about system development and use. Transparency improves the intelligibility of AI-generated outputs, and meaningful human review keeps educators and institutions actively involved in decision-making. Contestability enables learners to question or challenge AI-supported outcomes, whereas continuous auditing helps institutions identify and address emerging harms over time.

4.1 Inclusive Data and Representation

The first dimension concerns inclusive data and representation. GenAI systems are influenced by the data used to train, fine-tune, evaluate, and deploy them. These data may reflect dominant linguistic, cultural, and institutional assumptions. Consequently, a system may appear neutral while still producing outputs that are less accurate, less appropriate, or less useful for certain groups of learners.

This issue is particularly important in education because students enter learning environments with different language backgrounds, cultural experiences, disability statuses, levels of prior preparation, and access to technology. A writing assistant, for instance, may privilege a narrow form of academic English while undervaluing multilingual expression or alternative rhetorical traditions. A tutoring system may rely on examples that are familiar to some learners but disconnected from the lived experiences of others. Similarly, a feedback system may fail to adequately address the needs of students with disabilities when accessibility is not considered during design and evaluation.

Equity-centered oversight, therefore, requires institutions and developers to determine whether GenAI systems are suitable for the learners they are intended to support. This involves reviewing datasets, testing outputs across diverse learner groups, assessing accessibility, and examining whether system performance varies by language, discipline, and student population. Inclusive data practice is therefore not merely a technical matter, but also a pedagogical and ethical responsibility.

4.2 Participatory Design with Students and Educators

The second dimension is participatory design. GenAI systems should not be developed exclusively by technical teams or selected solely through administrative decisions. Students and educators interact with these systems most directly, and their perspectives are essential for identifying what is useful, harmful, or pedagogically appropriate.

Teachers bring knowledge of classroom realities, assessment pressures, learner differences, and disciplinary expectations. Students, in turn, can explain how AI tools influence their confidence, effort, learning strategies, and sense of agency. Support staff, including disability support and academic advising teams, may also identify risks that would remain hidden in a purely technical evaluation. Their involvement helps ensure that GenAI systems respond to genuine educational needs rather than abstract assumptions about teaching and learning.

Participatory design may take the form of student consultations, teacher review panels, pilot studies, co-design workshops, and structured feedback processes. Particular attention should be given to learners who are often underrepresented in educational technology design, including multilingual students, students with disabilities, first-generation students, and learners from under-resourced settings. In this way, participation can improve both the system's quality and the fairness of its educational use.

4.3 Transparency and Explainability

Thirdly, transparency and explainability. While GenAI systems can provide fluent and confident answers, students and teachers might not be aware of the process behind producing the responses, evidence or justification supporting these answers, or its limitations. This poses a significant learning risk, as students could unconsciously accept the explanations, feedback, or advice provided by AI without evaluating its trustworthiness.

For any educational GenAI system to be transparent, it must be clear about its capabilities and limitations. Students should understand the timing of interaction with AI, the kinds of information that could be employed, and that an output is a suggestion, prediction, explanation generated by AI, or a recommendation supported by AI. Explainability should not be only for the educational users, but for technical specialists. Teachers and students need not have detailed knowledge of a model's architecture, but they do need enough knowledge to understand the basis, limitations, and uses of its output.

For instance, if the system is automated, the feedback provided should be identified as grammar, structure, quality of argument, conceptual accuracy, or style. The task of a student-advising system should include the statement of a basis for a recommendation, and situations in which human consultation is needed. This openness can help minimize the overreliance on AI and foster more thoughtful use of GenAI in the education system.

4.4 Meaningful Human Review

The fourth dimension is "meaningful human review. There are a number of AI governance models that stress the need for human oversight. However, in reality this protection might not be much of a benefit if reviewers are short of time, information and training, or lack the power to override the system. Such circumstances give rise to the symbolic, not the real, presence of humans.

This is particularly significant in the field of education. Teachers and academic staff should not be reduced to mere supervisors of AI-generated outputs. They need to be able to question, revise, reject, and/or put into context those outputs when needed. When feedback is given to students about their writing while using an AI system, the teacher should be able to interpret the feedback in the context of the student's learning history, language background, and instructional objectives. Academic personnel should be able to decide if the advising system's course recommendation is suitable for the student's situation.

Implementation of meaningful human review requires institutional support. Teachers need training, documentation, time and procedures. Institutions need to set a threshold for when outputs of AI need to be checked by humans, and for who is ultimately responsible for taking decisions. If these factors are not in place, the human review process could just shift the burden on teachers without equipping them with the tools and/or authority to respond appropriately.

4.5 Contestability and Redress

The fifth dimension is contestability and redress. Students should be able to challenge the output of AI, especially when the output is incorporated into feedback, assessment, advising, academic progression and/or access to support. Contestability is required because even well-intentioned systems can miss or make errors, or can give incorrect recommendations for the system.

A contestable system should be designed to be transparent on how students can request clarification, ask for human review, report a problematic output, and/or appeal an AI-supported decision. It is not just a technical problem, it's an educational injustice. Any student who receives an AI-generated answer that they find incorrect or unclear should be able to get an explanation; any student who is disadvantaged by an automated answer should have a clear way to rectify the error.

Redress mechanisms must be institutionalized, transparent and accessible. This can involve appeal processes, correction forms, teacher review processes, and documented steps for complaints, among other avenues for support for students. These mechanisms are particularly critical for the marginalized learners who may be more at risk of being misclassified, stereotyped, negatively judged on the basis of their language, or fail to have access to the system design.

4.6 Continuous Auditing

Continuous auditing is the sixth dimension. The evaluation of GenAI systems should not be limited to prior to the introduction. This can vary as the models evolve, as students grow and change, as institutions change and as users change and diversify their interaction patterns. However, monitoring needs to be continued post-deployment.

The continuous auditing should cover both, technical and educational performance. Technical audits might focus on bias, privacy, accuracy, accessibility and security. Educational audits ought to reflect the extent to which the system promotes learning, the risk of students becoming increasingly dependent on the system, teachers' pedagogical use of the system, and the harm or benefits that may be occurring to certain groups rather than others.

Reporting of incidents should be in place as part of the auditing process. Students, teachers and both should be able to record incorrect results, inappropriate answers, confusing statements, privacy issues, and negative interactions. The reports should not be viewed as single incidents but as learning and improvement evidence to be used in the institutional learning process. Continuous auditing makes oversight a continuous process, as opposed to an approval process.

Table 2: Core dimensions of the equity-centered oversight framework

Dimension	Primary Purpose	Oversight Mechanism
Inclusive data and representation	Reduce representational bias.	Dataset review and diverse evaluation benchmarks.
Participatory design	Include affected stakeholders.	Student, educator, and institutional feedback loops.
Transparency and explainability	Make AI outputs understandable.	Clear explanations, limitations, and source indicators.
Meaningful human review	Enable informed intervention.	Teacher review, escalation, and institutional authority.
Contestability and redress	Protect learner agency.	Appeal, correction, and review procedures.
Continuous auditing	Detect emerging harms.	Bias, privacy, accessibility, and interaction audits.

5 Conclusion

In this article, the author suggests that an equity-based approach to oversight is a conceptual model that could guide the governance of using generative artificial intelligence tools in the educational sector, addressing the issues of equity, transparency, human oversight, contestability, and institutional accountability. The framework is conceptual and does not offer empirical evidence of its value, and will need to be validated in various learning environments, subject areas, learners, and cultural contexts as well as with varying degrees of institutional capacity. Future studies need to therefore explore the functioning of the framework in practice by conducting case studies, surveys, interviews and design-based evaluations, and to define measurable benchmarks for inclusive data practices, the quality of explanation, mechanisms for appeal, and continuous monitoring. There is a need for more focus on how GenAI literacy helps to enable learner agency and responsible use, especially with students that could be disproportionately impacted by AI-mediated education. Equity-centered oversight needs to be viewed as an adaptive governance process that can support institutional policy, quality assurance, accreditation and regulation in settings where the use of GenAI systems is likely to have a significant impact on education.

References

[1] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, p. 39, 2019. [Online]. Available: https://doi.org/10.1186/s41239-019-0171-0

[2] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer, U. Gasser, G. Groh, S. Günnemann, E. Hüllermeier, S. Krusche, G. Kutyniok, T. Michaeli, C. Nerdel, J. Pfeffer, O. Poquet, M. Sailer, A. Schmidt, T. Seidel, M. Stadler, J. Weller, J. Kuhn, and G. Kasneci, "Chatgpt for good? on opportunities and challenges of large language models for education," *Learning and Individual Differences*, vol. 103, p. 102274, 2023. [Online]. Available: https://doi.org/10.1016/j.lindif.2023.102274

[3] D. Zowghi and M. Bano, "Emerging roles and trends of equity, diversity, and inclusion in artificial intelligence," *AI & Society*, vol. 41, pp. 2741–2748, 2026. [Online]. Available: https://doi.org/10.1007/s00146-026-02969-z

[4] M. Foth, "Hostile interaction design: Ai, governance, and the quest for human oversight," *AI & Society*, 2026. [Online]. Available: https://doi.org/10.1007/s00146-026-03065-y

[5] J. Garcia Ramos and Z. Wilson-Kennedy, "Promoting equity and addressing concerns in teaching and learning with artificial intelligence," *Frontiers in Education*, vol. 9, p. 1487882, 2024. [Online]. Available: https://doi.org/10.3389/feduc.2024.1487882

[6] M. Bond, H. Khosravi, M. De Laat, N. Bergdahl, V. Negrea, E. Oxley, P. Pham, S. W. Chong, and G. Siemens, "A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour," *International Journal of Educational Technology in Higher Education*, vol. 21, p. 4, 2024. [Online]. Available: https://doi.org/10.1186/s41239-023-00436-z

[7] S. Wang, F. Wang, Z. Zhu, J. Wang, T. Tran, and Z. Du, "Artificial intelligence in education: A systematic literature review," *Expert Systems with Applications*, vol. 252, p. 124167, 2024. [Online]. Available: https://doi.org/10.1016/j.eswa.2024.124167

[8] A. Yusuf, N. Pervin, M. Román-González, and N. M. Noor, "Generative ai in education and research: A systematic mapping review," *Review of Education*, vol. 12, no. 2, p. e3489, 2024. [Online]. Available: https://doi.org/10.1002/rev3.3489

[9] A. D. Samala, S. Rawas, T. Wang, J. M. Reed, J. Kim, N.-J. Howard, and M. Ertz, "Unveiling the landscape of generative artificial intelligence in education: a comprehensive taxonomy of applications, challenges, and future prospects," *Education and Information Technologies*, 2025. [Online]. Available: https://doi.org/10.1007/s10639-024-12936-0

[10] Q. Xia, X. Weng, F. Ouyang, T.-J. Lin, and T. K. F. Chiu, "A scoping review on how generative artificial intelligence transforms assessment in higher education," *International Journal of Educational Technology in Higher Education*, vol. 21, p. 40, 2024. [Online]. Available: https://doi.org/10.1186/s41239-024-00468-z

[11] D. Lee, M. Arnold, A. Srivastava, K. Plastow, P. Strelan, F. Ploeckl, D. Lekkas, and E. Palmer, “The impact of generative ai on higher education learning and teaching: A study of educators’ perspectives,” *Computers and Education: Artificial Intelligence*, vol. 6, p. 100221, 2024. [Online]. Available: https://doi.org/10.1016/j.caeai.2024.100221

[12] A. Nguyen, H. N. Ngo, Y. Hong, B. Dang, and B.-P. T. Nguyen, “Ethical principles for artificial intelligence in education,” *Education and Information Technologies*, vol. 28, pp. 4221–4241, 2023. [Online]. Available: https://doi.org/10.1007/s10639-022-11316-w

[13] I. M. García-López and L. Trujillo-Liñan, “Ethical and regulatory challenges of generative ai in education: a systematic review,” *Frontiers in Education*, vol. 10, p. 1565938, 2025. [Online]. Available: https://doi.org/10.3389/feduc.2025.1565938

[14] M. Perkins, “Academic integrity considerations of ai large language models in the post-pandemic era: Chatgpt and beyond,” *Journal of University Teaching and Learning Practice*, vol. 20, no. 2, p. 7, 2023. [Online]. Available: https://doi.org/10.53761/1.20.02.07

[15] D. O. Eke, “Chatgpt and the rise of generative ai: Threat to academic integrity?” *Journal of Responsible Technology*, vol. 13, p. 100060, 2023. [Online]. Available: https://doi.org/10.1016/j.jrt.2023.100060

[16] M. Perkins, J. Roe, D. Postma, J. McGaughran, and D. Hickerson, “Detection of gpt-4 generated text in higher education: Combining academic judgement and software to identify generative ai tool misuse,” *Journal of Academic Ethics*, vol. 22, no. 1, pp. 89–113, 2024. [Online]. Available: https://doi.org/10.1007/s10805-023-09492-6

[17] J. Park, “A systematic literature review of generative artificial intelligence (genai) literacy in schools,” *Computers and Education: Artificial Intelligence*, vol. 9, p. 100487, 2025. [Online]. Available: https://doi.org/10.1016/j.caeai.2025.100487