

Chapter 11

Ethical Dilemmas and Practical Strategies for Leveraging Generative AI in STEM Assessment in Higher Education



Dharel P. Acut, Sotero O. Malayao, Amelia T. Buan, Jun Karren V. Caparoso, Jivulter C. Mangubat, and Manuel B. Garcia

1 Introduction

The rapid advancements in generative artificial intelligence (AI) have significantly impacted education, reshaping teaching methodologies and assessment practices across various disciplines [1, 2]. STEM Education in higher education encompasses multidisciplinary programs that primarily apply science, technology, engineering, and mathematics—AI offers innovative tools that enhance instructional efficiency and learning outcomes [3]. Recent developments in generative AI technologies, such as ChatGPT, Codex, Gemini, and Photomath have enabled educators to design complex assessment questions, provide real-time personalized feedback, and automate labor-intensive grading processes [4–6]. These advancements are particularly relevant for STEM fields, where managing large classes and assessing higher order thinking skills and non-cognitive outcomes remain a challenge [7].

Generative AI's potential lies in its ability to personalize academic assessments by adapting to individual student needs, thus fostering engagement and deeper conceptual understanding [8]. AI's capability to analyze large datasets also aids in identifying trends in student performance, providing educators with actionable

D. P. Acut (✉) · J. C. Mangubat
Cebu Technological University, Cebu City, Philippines

S. O. Malayao · A. T. Buan · J. K. V. Caparoso
MSU Iligan Institute of Technology, Iligan City, Philippines
e-mail: sotero.malayao@g.msuiit.edu.ph; amelia.buan@g.msuiit.edu.ph;
junkarren.caparoso@g.msuiit.edu.ph

M. B. Garcia
FEU Institute of Technology, Manila, Philippines

Korea University, Seoul, South Korea
e-mail: mbgarcia@feutech.edu.ph

insights to refine their teaching approaches [9]. Moreover, automation reduces the administrative burden on educators, enabling them to focus on delivering quality instruction. However, these opportunities also introduce complexities—such as ethical considerations, technical nuances, and evolving pedagogical practices—that must be carefully managed to ensure the equitable and effective use of AI in education [10].

The integration of generative AI into STEM education raises a host of ethical challenges, including algorithmic bias, data privacy, academic dishonesty, and the transparency of AI-driven decision-making [3]. These issues have been highlighted in recent studies, emphasizing the need for caution and responsibility in utilizing AI technologies [11–13]. For STEM faculty, addressing these challenges involves striking a balance between leveraging AI's benefits and maintaining academic integrity.

Ethical dilemmas also extend to concerns about over-reliance on AI tools [14]. Questions arise about the extent to which AI should influence assessment practices and whether it could diminish the critical role of human judgment in evaluating student performance [8]. Faculty must also contend with the potential for inequities, as students with limited access to technology may be disadvantaged by AI-driven approaches [15]. These challenges highlight the need for targeted strategies to address ethical concerns and ensure that AI integration enhances, rather than undermines, the quality of STEM education [16].

While the literature attest to generative AI's role in transforming STEM assessment in higher education, the practical and ethical dimensions of its use are often overlooked. Most studies focus on technological capabilities, providing limited insight into how faculty navigate ethical dilemmas while integrating AI into academic assessments. This chapter addresses these gaps by examining the experiences of STEM education faculty in their use of AI tools for assessments. Exploring the perspectives, challenges, and strategies of STEM educators offers a nuanced understanding of generative AI's role in academic assessments and contributes to the discourse on its responsible and impactful implementation in higher education. Specifically, this chapter aims to:

1. Explore the integration of generative AI in STEM academic assessments, analyzing how educators across disciplines are utilizing these tools and the perceived benefits and challenges.
2. Investigate ethical dilemmas associated with AI-driven assessments, including fairness, bias, privacy, and transparency, to understand how faculty navigate these issues.
3. Highlight practical strategies for ethical and effective AI integration, leveraging case-specific insights from STEM educators to provide actionable recommendations for enhancing assessment practice.

2 Understanding Generative AI in STEM Education Assessments

2.1 Defining Generative AI

Generative Artificial Intelligence (AI) refers to advanced machine learning systems designed to create new content based on existing data [17]. These systems, including models such as OpenAI’s ChatGPT and Google’s Bard, utilize large language models (LLMs) to generate text, code, images, and other outputs by analyzing patterns in extensive datasets [18, 19]. Unlike traditional AI, which primarily follows programmed rules to analyze or respond, generative AI possesses the ability to “create,” offering outputs that mimic human-like reasoning and creativity [20].

In education, generative AI has gained traction for its capacity to perform complex tasks such as drafting essays, summarizing large datasets, and creating customized learning materials [21, 22]. Its potential extends even further in STEM disciplines, which emphasize rigor, precision, and problem-solving. Beyond content creation, generative AI provides adaptive solutions for assessments, addressing key challenges such as scalability, accessibility, and personalization [3, 23].

2.2 How AI Is Poised to Transform STEM Assessments

Generative AI has introduced innovative opportunities to transform how STEM assessments are designed, delivered, and evaluated [3, 24]. In traditional academic environments, assessments often rely on standardized tests, lab reports, and projects [24]. Traditional assessments in STEM education, such as written exams, problem sets, and practical demonstrations, have long served as reliable measures of student understanding [25]. These methods emphasize structured problem-solving, procedural knowledge, and direct evaluation of student outputs. However, they often face challenges related to scalability, personalization, and real-time feedback [26]. While these methods remain foundational, integrating generative AI offers potential advancements in several key areas.

AI-driven assessment methods, in contrast, leverage adaptive learning technologies, automated grading, and generative AI models to create more dynamic, responsive, and scalable evaluation frameworks [27]. AI-driven tools can instantly evaluate coding assignments, create personalized problem sets tailored to student performance, and gauge conceptual understanding through natural language interactions [3].

A key distinction between the two approaches lies in their flexibility and efficiency. Traditional assessments require significant manual effort from instructors, whereas AI-driven approaches can automate grading, detect patterns in student learning, and tailor feedback accordingly [2]. Despite these advantages, concerns

about reliability, transparency, and ethical implications necessitate careful consideration before full-scale adoption.

Personalization of Assessments

Generative AI enables the creation of assessments tailored to individual learning needs [28]. For example, AI tools can adapt questions based on a student's proficiency, ensuring that assessments are neither too difficult nor too easy [29]. In STEM education, this could involve generating dynamic problem sets for physics or chemistry, where complexity adjusts based on prior responses. This adaptive assessment design supports differentiated instruction, catering to diverse student populations and learning styles.

Automation and Efficiency

One of the most significant transformations AI brings to STEM assessments is the automation of repetitive tasks [30]. Grading lab reports, mathematical problem sets, or engineering blueprints often consumes substantial instructor time. Generative AI tools can analyze submissions, provide instant feedback, and assign scores with high accuracy [31]. For instance, Malik et al. [32] illustrate how AI-powered platforms can evaluate essays or scientific explanations, identifying errors in reasoning or misapplications of principles. This efficiency allows instructors to focus on deeper pedagogical goals rather than administrative burdens.

Enhanced Feedback Mechanisms

Providing meaningful, timely feedback is essential for STEM learning, yet it is often constrained by time and resource limitations [25]. As per Mahapatra [33], Generative AI can analyze student responses and offer targeted, constructive feedback almost instantaneously. In disciplines like biology or chemistry, AI could identify gaps in conceptual understanding and recommend specific resources or strategies for improvement [34]. This immediate feedback loop fosters active learning and engagement, particularly in subjects requiring cumulative knowledge building.

Simulation and Modeling Capabilities

STEM fields frequently rely on practical, hands-on learning experiences, such as lab experiments or engineering design projects [35]. Generative AI can support these

activities by simulating scenarios or generating virtual experiments that mirror real-world conditions [36]. For example, students studying earth sciences might interact with AI-generated climate models to predict environmental outcomes under various conditions [37]. Similarly, physics students could experiment with AI-generated virtual apparatus to test hypotheses [38]. These simulations enhance accessibility for institutions with limited physical resources and provide safe, cost-effective alternatives to traditional practices.

Broadening Assessment Modalities

Traditional assessments often emphasize written responses or multiple-choice questions, which may not fully capture the multidimensional nature of STEM expertise. Generative AI allows for more creative and interactive assessment formats [39]. For instance, Yilmaz & Yilmaz [40] reported that AI tools can generate complex case studies for students to analyze or create coding challenges in technology education. Additionally, AI-driven platforms can enable students to present findings or solutions in innovative formats, such as AI-designed visualizations or multimedia presentations [41].

Addressing Equity and Accessibility

Generative AI holds promise in addressing educational inequities by creating opportunities for inclusive learning [42]. For students in underserved regions or institutions, AI-driven assessments can democratize access to high-quality resources and standardized testing frameworks [10]. For example, an AI tool could generate localized examples and problem sets for students in different geographical contexts, making content more relevant and engaging [43]. Furthermore, AI technologies can accommodate students with disabilities by offering features like text-to-speech, alternative formats, or simplified language options [44].

Generative AI is poised to revolutionize STEM education assessments through personalization, automation, and expanded modalities [28, 30, 40]. Addressing inefficiencies and fostering more engaging, equitable assessment experiences, AI presents promising pathways for advancing STEM pedagogy [3]. However, its adoption requires careful navigation of ethical and practical challenges to ensure that it serves as a complement to, rather than a replacement for, traditional educational methods [45]. This transformative journey calls for collaborative efforts among educators, researchers, and policymakers to shape the future of STEM in higher education.

3 Ethical Dilemmas in Generative AI for STEM Education Assessment

As generative AI becomes more integrated into STEM education, its potential to streamline and enhance academic assessments is undeniable. However, these innovations are accompanied by significant ethical dilemmas that challenge the fairness, accountability, and integrity of AI-driven systems [46]. This section explores key ethical concerns that arise in the use of generative AI for STEM education assessments.

3.1 Bias and Fairness in AI Algorithms

Generative AI relies on large datasets for training, and these datasets often carry the biases of the societies or systems they were sourced from [47]. In STEM assessments, these biases can lead to inequitable outcomes. For instance, an AI system trained predominantly on data from English-speaking students may struggle to evaluate the work of non-native English speakers fairly, disadvantaging students in diverse linguistic and cultural contexts [48]. Similarly, biases in historical datasets may favor students from certain socioeconomic backgrounds or educational systems while marginalizing others [47].

In STEM fields such as biology or engineering, where assessments may require creative solutions or non-standard approaches, biased AI could unfairly penalize students whose submissions deviate from the norm. This raises questions about whether AI systems can adequately and fairly assess diverse learning targets [49].

3.2 Transparency and Accountability

A major ethical concern in AI-driven assessments is the opacity of decision-making processes. Generative AI operates as a “black box,” meaning its inner workings are often opaque, even to the developers [50]. In STEM education, this can create significant issues as students need to be informed how complex assignments such as physics simulations, mathematical proofs, or engineering designs are graded.

When an AI system generates a grade or evaluation, it is unclear to both educators and students how a particular decision was made. This lack of transparency undermines trust and accountability, especially in cases where the assessment results are contested [8, 14, 51]. For instance, a chemistry instructor who uses AI in grading a chemistry lab report may find it difficult to explain how a mark was made to any dissatisfied student.

3.3 Data Privacy and Security

AI systems rely on vast amounts of data to operate efficiently, raising concerns about data collection, storage, and security [52]. In STEM assessments, sensitive student information—ranging from personal details to academic submissions—may be stored in centralized databases that could be vulnerable to breaches or misuse [53].

For example, an AI tool used to evaluate earth science projects might store detailed data about students' geographical locations, research topics, and environmental measurements. Without robust data protection measures, this information could be accessed by unauthorized parties or repurposed for commercial gain. These concerns become even more critical in the context of cross-border education, where different regions have varying regulations on data privacy [54].

3.4 Academic Integrity and AI Misuse

Generative AI presents new challenges in upholding academic integrity in STEM education. On the one hand, students may misuse AI tools to complete assessments, undermining the learning process [55]. For example, a student might use an AI platform to generate answers to biology quizzes or chemistry problem sets, bypassing the need to engage with the material. This kind of misuse can diminish the educational value of assessments and hinder the development of essential skills [15].

On the other hand, educators may misuse AI by relying too heavily on it for grading and evaluation [56]. For instance, an instructor in a mathematics course might use AI exclusively to grade problem sets without cross-checking the results, potentially overlooking errors or nuances in the AI's assessment. Such over-reliance can reduce the quality of feedback students receive, limiting their opportunities for growth and improvement [14].

The integration of generative AI into STEM education assessments is a double-edged sword, offering unprecedented opportunities while introducing complex ethical dilemmas. Bias and fairness issues challenge the equity of assessments, while transparency concerns undermine trust and accountability [47]. Data privacy and security risks raise questions about the protection of sensitive information, and the potential for academic integrity violations highlights the need for careful oversight [52]. As these dilemmas come to the forefront, they underscore the importance of ongoing scrutiny and ethical reflection to ensure that AI-driven assessments align with the values of fairness, accountability, and educational excellence.

4 AI-Integrated Assessment Experiences of Higher Education Faculty

This section explores the real-world applications of generative AI in assessment practices across various STEM disciplines in higher education, guided by a qualitative case study methodology. Drawing upon Yin's [57] principles for case study research—specifically the value of contextualized, in-depth inquiry into contemporary phenomena within real-life settings—this chapter aims to understand how faculty members in higher education institutions experience and integrate AI-driven assessment tools into their pedagogical practices.

The chapter adopts a multiple-case study design, focusing on seven faculty members from selected higher education institutions in the Philippines. These participants were purposively selected based on their practical engagement with AI in STEM-related assessment tasks. Each case represents a distinct STEM sub-discipline, including biology, chemistry, mathematics, earth science, physics, engineering, and technology education. The diversity of these cases enables comparative insights into disciplinary nuances while uncovering shared ethical, pedagogical, and institutional themes.

Data were collected in December 2024 through a structured qualitative questionnaire administered via Google Forms. This asynchronous mode of data collection allowed participants to respond at their own pace and convenience, facilitating thoughtful and reflective answers. Participants were asked to share narrative accounts of their use of AI in assessment—highlighting specific tools used, pedagogical intentions, perceived benefits, encountered challenges, and institutional or ethical concerns.

All participants provided informed consent, with the study procedures adhering to the ethical principles outlined in the Declaration of Helsinki and the Philippine Data Privacy Act of 2012. Anonymity and confidentiality were ensured throughout data handling and reporting processes.

The resulting narratives reflect rich, contextually embedded insights into AI's role in reshaping assessment methodologies. The cases detail the application of AI in tasks such as personalized grading of student research papers in biology, AI-enhanced feedback mechanisms for chemistry lab reports, and tailored mathematics problem sets that respond dynamically to individual learning profiles. Other cases involve the use of AI in analyzing complex geophysical data, implementing simulations to improve conceptual understanding in physics, automating the evaluation of engineering design projects, and facilitating project-based assessments in technical-vocational education.

While a thematic analysis approach was employed to identify patterns across the cases, the findings are presented as individual narratives to highlight disciplinary contexts and preserve the depth of faculty experiences. This strategy allows for an appreciation of both shared themes and the uniqueness of each disciplinary integration of AI tools.

Beyond merely cataloging technological affordances, the analysis also addresses critical issues surrounding AI integration—including concerns over algorithmic bias, data privacy, interpretability of AI-generated results, and readiness of institutions and educators. Faculty narratives reveal both optimism about AI’s potential to enhance assessment validity and caution about ethical and logistical barriers that need to be addressed.

These diverse yet interconnected cases offer rich, experience-based insights into the evolving relationship between AI and assessment in higher education. The findings demonstrate that while generative AI holds significant potential for transforming STEM education, its integration must be anchored in sound pedagogical principles, supported by ethical safeguards, and enabled through institutional readiness. Such insights contribute to a deeper understanding of both the transformative possibilities and the complex challenges involved in embedding AI within academic assessment frameworks.

4.1 Carla’s Case: AI for Personalized Grading of Research Papers in Biology

Carla, an associate professor of biology education with 8 years of experience, has found great value in incorporating AI tools into her assessment practices. She uses AI platforms like ChatGPT, Bing AI, and Bard to create customized quizzes, projects, and problem sets, and utilizes Gradescope and Turnitin AI for evaluating student responses. Her assessments span quizzes, exams, lab reports, projects, and performance-based tasks such as experiments and demonstrations. Carla appreciates AI for its ability to “save time” by automating grading and administrative tasks, which has been particularly beneficial in managing large classes. Additionally, AI’s accessibility and “cost-effectiveness” make it easier for her to keep up with technological trends in education while ensuring assessments align with learning outcomes and curriculum goals.

Carla defines generative AI as “artificial intelligence systems that can autonomously create new content, such as written material, visualizations, or simulations, based on existing data or patterns.” In the context of biology education, she believes AI can “revolutionize academic assessments” by generating customized quizzes, lab exercises, and interactive models that test students’ understanding of biological concepts. Moreover, AI can create “dynamic case studies or problem-solving scenarios” that allow students to apply their knowledge to real-world situations, promoting deeper learning and critical thinking. She also highlights that AI’s ability to provide “personalized feedback” can help students quickly identify areas for improvement, facilitating a more individualized learning experience.

Carla is convinced that “generative AI offers several advantages for assessing students’ work in the discipline of science education”. She explains that AI enables the creation of personalized assessments that adapt to the individual needs and

progress of students, ensuring that evaluations are “more relevant and targeted”. Furthermore, AI can automate the generation of complex problem sets, simulations, and interactive case studies that foster “deeper learning” by challenging students to apply their knowledge in critical ways. AI also provides “instant, data-driven feedback” that helps students improve rapidly. Analyzing patterns in student responses, AI enables educators to identify trends in understanding and adjust instructional strategies accordingly.

One of the primary advantages Carla sees in AI is its ability to improve grading accuracy. She notes that AI can “automate the evaluation of complex answers”, ensuring consistency and objectivity in scoring. AI also enables “personalized feedback”, tailoring responses based on individual student answers to target specific areas of weakness or confusion. In terms of administrative efficiency, Carla emphasizes that AI “streamlines tasks like creating and organizing assessments”, saving educators significant time and allowing them to focus more on substantive teaching.

However, Carla acknowledges the limitations of AI in biology education. She points out that AI struggles with evaluating the “depth of understanding in open-ended or highly complex biological concepts”, as it may not recognize the nuances in students’ reasoning. She also warns that AI-generated content may lack the “creativity” needed to capture the full range of student thinking, particularly in biology, where critical applications of knowledge are essential. She further expresses concerns about the potential for bias in AI algorithms, which could affect the fairness of assessments if the AI is trained on biased data. Moreover, Carla notes that AI “may not fully account for contextual factors”, such as a student’s background knowledge or learning style, which significantly influence their approach to biology tasks.

Looking ahead, Carla believes that “AI will significantly shape the future of assessment practices in STEM education” by enabling dynamic, personalized, and efficient evaluation methods. She envisions AI facilitating the creation of “adaptive assessments” that respond to individual student progress, making evaluations both challenging and supportive. She also highlights AI’s potential in supporting “formative assessments”, offering real-time feedback that guides student growth. AI’s capability to incorporate complex simulations and problem-solving tasks enables students to showcase their understanding of biological concepts in “real-world, applied contexts,” shifting the emphasis from rote memorization to fostering critical thinking and problem-solving skills.

Carla emphasizes that AI-generated assessments should complement traditional forms of assessment rather than replace them. While AI can “efficiently generate adaptive, personalized quizzes” and provide immediate feedback, it cannot capture the complexity of “critical thinking and problem-solving” that traditional assessments, such as lab reports or research projects, are designed to evaluate. These traditional assessments allow students to demonstrate their ability to “synthesize knowledge, engage in in-depth analysis, and communicate their understanding”. Carla believes that AI can support traditional assessments by providing “continuous,

formative feedback”, but it should not replace the holistic skills these assessments encourage.

In the context of research papers and lab reports, Carla envisions AI as a tool to assist in the evaluation of certain aspects, such as “grammar, syntax, and adherence to formatting standards”. AI could also help assess the “coherence and structure” of a paper, ensuring that students have logically developed their arguments and clearly articulated their methods and results. However, Carla insists that AI should “complement the human aspect of assessment”, where educators assess the “depth of analysis, creativity, and the application of biological principles” in students’ work.

Ultimately, Carla advocates for a balanced approach where AI and human judgment work together. AI can automate routine tasks like grading quizzes and analyzing lab report data, providing “quick and objective evaluations”, but human instructors are essential for assessing more complex elements, such as “scientific methods, the interpretation of results, and the creativity involved in experimental design.” Carla believes that this balance will offer a “more comprehensive and effective approach to assessing student learning in biology.” Combining the strengths of AI and human judgment allows educators to provide a richer, more personalized learning experience for students.

4.2 Margaret’s Case: AI-Enhanced Feedback on Lab Reports in Chemistry

Margaret, an assistant professor of chemistry at a public university with 5 years of teaching experience, has integrated artificial intelligence (AI) into her assessment practices in innovative and effective ways. In her role, Margaret teaches chemistry courses to undergraduate students, with a focus on ensuring that assessments are not only efficient but also provide meaningful feedback that supports student learning. She has leveraged a variety of AI tools to enhance both her grading processes and her interactions with students, while striving to maintain the integrity of her assessments.

Margaret’s use of AI spans multiple aspects of assessment. To generate personalized and adaptive assessment questions, she utilizes AI tools like ChatGPT, Bing AI, and Bard. These platforms allow her to create complex and diverse chemistry problems that challenge students at various levels of understanding. Margaret explains, “AI has allowed me to quickly generate questions that match the varying levels of student understanding. Whether it’s a simple conceptual question or a complex problem involving chemical equations, AI makes it easy to create tailored assessments.” These tools can tailor questions to match the student’s learning pace, ensuring that each student receives the appropriate level of challenge. Additionally, AI assists Margaret in grading by automating repetitive tasks, particularly for

multiple-choice and fill-in-the-blank assessments, enabling her to provide faster feedback without compromising accuracy.

One of the most significant benefits that Margaret has observed in using AI is the personalized feedback it offers to students. For example, she uses AI tools like Grammarly and ChatGPT to provide real-time, tailored feedback on students' written responses. She notes, "AI allows me to give detailed, immediate feedback on their conceptual understanding. It identifies errors, suggests improvements, and provides more personalized guidance than I could give manually for every student." These tools not only identify grammatical and syntactical errors but also offer suggestions for improving clarity and coherence in students' explanations. By using AI-generated feedback, Margaret can help students identify their mistakes more quickly, reinforcing key concepts and guiding them to more accurate and deeper understanding. She believes that the combination of AI and personalized feedback promotes a more engaged learning experience, enabling students to track their progress and focus on areas that need improvement.

Another area where AI has proven invaluable to Margaret is in the analysis of student performance data. Tools like Edulastic and Learning Analytics AI allow Margaret to track trends in student performance over time, providing insights into which topics or concepts are more challenging for her students. She explains, "The data I get from AI tools helps me track performance patterns. I can quickly spot topics where students consistently struggle and adapt my teaching methods accordingly." This data helps her identify patterns and adapt her teaching strategies to meet the needs of her students. For instance, if a significant number of students struggle with a particular concept, Margaret can revise her teaching approach, develop additional resources, or offer supplementary materials to help them better understand the topic. The AI-generated data also helps Margaret track individual student progress, providing her with a clear understanding of each student's strengths and weaknesses.

In terms of assessments, Margaret primarily uses AI for quizzes, exams, lab reports, and performance-based assessments such as experiments and demonstrations. The AI tools help her design assessments that are both time-efficient and innovative. For example, AI allows Margaret to create dynamic simulations that engage students in virtual experiments, enabling them to practice and apply their knowledge in a controlled, digital environment. These simulations allow students to engage with content in new and interactive ways, providing them with opportunities to reinforce their learning while developing practical problem-solving skills. Margaret notes, "Simulations generated by AI offer my students a chance to experiment in virtual labs, which is a great alternative to traditional lab sessions. They can experiment with complex reactions safely and in real-time, deepening their understanding of chemistry."

Despite the many benefits Margaret has seen from using AI, she acknowledges its limitations, particularly when it comes to assessing more nuanced, creative problem-solving approaches. While AI tools excel at grading surface-level answers and objective questions, they often struggle with evaluating open-ended tasks that require complex reasoning or creativity. For example, in chemistry, many problems

require students to develop hypotheses, design experiments, or think critically about scientific concepts in novel ways. Margaret admits, “AI tools do a great job assessing factual accuracy, but when it comes to evaluating a student’s unique approach to a problem or their creativity in experimental design, it’s not as effective. The AI may miss subtleties that are important in chemistry.” AI systems may not fully capture the depth of students’ understanding in these areas, as they rely heavily on pre-programmed algorithms and patterns in the data. Margaret believes that this is a key limitation of AI, as it may overlook the creative and critical thinking aspects of student work that go beyond standard solutions.

Moreover, Margaret is aware of the risks associated with over-relying on AI in assessment. While AI can certainly enhance the efficiency of grading and feedback, she cautions against using it as the sole tool for evaluation. She says, “While AI helps with grading efficiency, it shouldn’t replace human judgment, especially in evaluating creativity and critical thinking.

There are areas in chemistry that require nuanced understanding, and that’s where human intuition still plays a crucial role.” Margaret believes that human judgment remains essential for assessing the more complex aspects of chemistry, such as experimental design, the application of theoretical knowledge to new situations, and the ability to think critically and creatively. In her view, AI should complement traditional assessment methods, not replace them.

To balance these strengths and weaknesses, Margaret employs a hybrid approach to assessment. She uses AI for tasks that benefit from automation, such as grading objective questions and providing quick feedback on technical aspects of assignments like lab reports. However, for more complex tasks that require critical thinking, creativity, and problem-solving, Margaret still relies on traditional methods of assessment. For example, in assessing experimental design, Margaret uses peer reviews and oral presentations, which allow her to evaluate students’ ability to explain their thought processes and defend their methodologies. These traditional methods provide a more holistic understanding of a student’s knowledge and creativity, ensuring that AI does not overshadow the human judgment that is essential for evaluating complex, higher-order thinking.

Furthermore, Margaret recognizes the potential for AI to reduce workload and make her teaching more efficient. Automating administrative tasks, such as grading quizzes and analyzing performance data, enables Margaret to devote more time to interacting with students, providing individualized feedback, and refining her teaching methods. This approach enables her to create a more personalized learning experience for each student, while also ensuring that her assessments are rigorous, fair, and reflective of students’ true abilities.

As a result, Margaret sees AI as a transformative tool that can significantly enhance the assessment process in chemistry education. By automating repetitive tasks, providing personalized feedback, and offering data-driven insights into student performance, AI can help instructors create more dynamic, engaging, and effective learning experiences. However, Margaret also believes that AI should be used judiciously and in conjunction with traditional methods of assessment to ensure that it complements rather than replaces the critical thinking, creativity, and human

judgment that are central to the learning process. Through this balanced approach, Margaret is able to harness the power of AI while preserving the elements of teaching that foster deeper learning and intellectual growth.

4.3 Percy's Case: AI in Personalized Mathematics Problem Sets

Percy, an assistant professor in Mathematics Education at a state university with 6 years of teaching experience, emphasizes the critical role AI plays in grading and providing personalized feedback. He uses AI tools such as Turnitin AI to evaluate and grade student responses, ensuring accuracy and fairness. Percy values AI's ability to assist with grading, particularly in large classes, as it enhances consistency and efficiency. Additionally, he leverages tools like ChatGPT and Grammarly to provide personalized feedback, enriching the learning experience for his students.

Percy emphasizes the importance of prompt quality when it comes to generative AI. He notes, "Generative AI is as good as how prompts are constructed." Drawing from a discussion by Prof. Marek Kowalkiewicz, Percy highlights the need for caution when using AI, describing it as "an eccentric colleague who knows a lot but can spout weird stuff from time to time." He sees AI as a tool that can generate scenarios for problem-solving or forecasting from an educator's perspective, while on the student side, AI can help integrate ideas or explore variations of thought.

In the past, Percy used Turnitin to check for potentially AI-generated content in essays and research papers, aiming to ensure that students engaged in analytic and synthesis work. This practice reflects his concern that AI-generated content, while potentially useful, may lack the depth and critical thinking required in academic tasks.

Percy is cautious about the limitations of AI in assessment. He believes that AI's effectiveness is tied to the quality of its training data and the prompts it receives. "Grading accuracy? I haven't been teaching in my field in the last year. So my idea is theoretical at this point," he says. He foresees challenges if AI training shifts from human-generated to AI-generated data, which could result in unrealistic or flawed content. Percy is optimistic, however, about AI's potential impact on assessment practices, noting that its role will depend on AI literacy among both educators and students. "AI can have a positive role if both educators and students alike have sufficient AI literacy," he states.

Percy also believes that AI will not replace all forms of assessment in math education, especially in performance-based assessments. "Math education is a professional field that has performance-based assessments," he explains. He envisions AI complementing, rather than replacing, traditional assessment practices, particularly by helping to generate complex problems. However, he is aware that students could use AI to solve those same problems, potentially undermining the integrity of assessments.

Looking forward, Percy is cautious but open-minded about the advanced applications of AI in mathematics education. On an optimistic note, he imagines AI being used to anticipate students' thinking processes when analyzing their solutions to mathematical problems. However, he acknowledges that this would require sophisticated applications of machine learning and programming. On a more moderate perspective, AI could be useful for generating complex problems, but this too presents a challenge as students could use AI to solve these problems, potentially circumventing traditional problem-solving methods.

Regarding grading, Percy emphasizes the importance of a tailored approach to AI's use in assessment. He acknowledges that AI's ability to consistently assess solutions is dependent on the diversity and accuracy of the dataset it is trained on. As such, AI tools may struggle to consistently grade solutions from diverse student populations with varying skill levels. "The quality and ability of AI to consistently assess solutions is relative to the accuracy and diversity of its dataset trained in," he says. This suggests that while AI can be a valuable tool, its effectiveness depends on the context in which it is used.

Percy also discusses the ethical considerations of AI in assessment, such as the need for transparency and ensuring equity in assessments for students of varying skill levels. He recalls an example of a physics teacher who used an AI-generated syllabus and asked students to critique it, which he sees as a good practice for ensuring transparency in AI usage. He adds that as the user of AI-driven assessments, it is his responsibility to ensure equity and use AI tools with nuanced parameters that account for students' diverse skill levels.

In terms of accountability, Percy firmly believes that the responsibility for AI-driven assessments lies with the educator. He sees AI as a tool, similar to how data visualization tools are used by teachers. "The culpability rests primarily on the user and designer of AI-driven assessments, i.e., the teachers," Percy explains. He is a proponent of a balanced approach, where AI complements traditional assessment methods rather than replacing them. Ultimately, Percy believes that AI's effectiveness depends on the literacy of both educators and students, just as information and communication technology (ICT) tools had varying levels of impact depending on how they were used.

In sum, Percy's approach to AI in mathematics education is thoughtful and cautious, emphasizing the importance of understanding AI's limitations while exploring its potential to complement and enhance assessment practices.

4.4 Ben's Case: AI-Assisted Analysis of Geophysical Data in Earth Science

Ben, an instructor in Science Education at a public technological university for 2 years with 6 years of prior experience teaching in basic education, integrates AI as an essential part of his assessment practices in Earth Science Education. He utilizes

AI tools for diverse tasks, including creating personalized assessment questions, providing detailed feedback to students, and designing innovative or gamified assessments that engage learners. Ben highlights the transformative potential of generative AI in STEM education, particularly in automating problem generation and delivering tailored feedback to promote deeper learning and critical thinking. He describes generative AI as “artificial intelligence systems capable of creating new content, such as text, code, or simulations, based on given inputs,” underscoring its role in reshaping academic assessments in STEM disciplines through enhanced efficiency and innovation.

Ben values AI’s capacity for creating diverse and adaptive assessments tailored to individual student needs. He notes, “Generative AI offers the advantage of creating diverse, adaptive assessment tasks tailored to students’ individual needs, enabling a more personalized learning experience.” He also highlights the importance of AI in providing instant, detailed feedback, which helps students in his field develop critical thinking and problem-solving skills more effectively.

In terms of grading, Ben appreciates how AI enhances grading accuracy by ensuring consistency and fairness. He adds, “Generative AI can enhance grading accuracy by ensuring consistent and unbiased evaluation of students’ work while also personalizing feedback to address individual learning gaps.” AI’s role in administrative tasks is also notable, as it automates repetitive processes, freeing up more time for teaching and mentoring. This aligns with his broader view of AI as a tool for enhancing the efficiency and scalability of assessment practices, while allowing educators to focus on instructional roles.

However, Ben acknowledges the limitations of AI, particularly its potential to misinterpret nuanced answers in open-ended tasks. He points out that AI may struggle to evaluate complex, context-specific reasoning or creative problem-solving that requires human judgment. “The limitations of generative AI in academic assessments include its potential to misinterpret nuanced answers, especially in open-ended tasks, and its reliance on training data, which may introduce biases or inaccuracies,” he notes. He also highlights the risk of AI generating false-positive assessments when over-reliance occurs, cautioning that excessive dependence on AI tools might compromise the accuracy and fairness of evaluations. Furthermore, Ben is wary of AI’s tendency to oversimplify complex problems or datasets, potentially leading to inaccurate conclusions, particularly in fields like Earth Science, which often involve highly intricate and multifaceted data.

Despite these limitations, Ben sees a promising future for AI in STEM assessments. “Generative AI will play a pivotal role in shaping future STEM assessment practices by enabling adaptive, real-time evaluations that cater to diverse student needs and learning styles.” He foresees AI enabling more dynamic, inquiry-based assessments that encourage critical thinking and innovation, while streamlining the assessment process for educators.

Ben advocates for a blended approach to assessment, where AI complements traditional methods rather than replacing them. He believes AI-generated assessments are particularly useful for tasks like data analysis, trend identification, and providing real-time feedback, but human judgment remains essential for evaluating more

complex and creative aspects of student performance. “AI-generated assessments should complement traditional forms of assessment rather than replace them, as they excel in efficiency and personalization but may lack the depth needed to evaluate complex, human-centered skills like creativity and ethical reasoning.”

In Earth Science specifically, AI has the potential to assist with the analysis of complex datasets, such as seismic wave patterns and climate models. Ben suggests, “AI could assist in Earth Science assessments by automating the analysis of complex datasets . . . to provide accurate and timely feedback.” Furthermore, AI can simulate real-world scenarios, enabling students to practice and refine their analytical skills in a dynamic, data-driven environment. However, Ben is aware of the ethical challenges AI presents in interpreting sensitive data. He warns that AI may misinterpret environmental data or oversimplify results, leading to inaccurate conclusions. “The ethical challenges of using AI to interpret complex environmental or geophysical data include the risk of misinterpreting context due to limited domain-specific training and potential biases in the algorithms.”

To mitigate these risks, Ben intends to guide his students in using AI responsibly. “I would guide students by emphasizing the importance of critical thinking and cross-verification when using AI tools, encouraging them to compare AI-generated interpretations with established scientific principles and empirical data,” he explains. This approach ensures that students use AI as a tool to support their learning rather than rely on it as a definitive source of truth.

Ultimately, Ben believes that a balanced approach, combining AI-driven assessments with human judgment, will be essential for fostering a deeper, more nuanced understanding of Earth Science and other STEM disciplines. “A balanced approach would involve using AI-driven assessments for tasks like data analysis, trend identification, and real-time feedback, while relying on human judgment for evaluating complex reasoning, creativity, and ethical implications.”

4.5 Rico’s Case: AI-Based Simulations for Physics Conceptual Understanding

In Rico’s teaching experience as an associate professor in physics education at a state university with 20 years of teaching experience, AI-based tools have significantly enhanced his assessment practices, particularly by enabling him to analyze student performance data more effectively. He explains, “AI will prompt me to the latest pool of interest in my field; it will allow me to give the most recent concerns or constructs widely discussed. It will also allow me deeper analysis into divergent answers that will provide rich contexts locally.” Using AI to generate questions and process responses allows Rico to keep his students updated with global trends while gaining valuable insights into the depth of their understanding. His extensive teaching background allows him to integrate AI tools strategically,

fostering a balance between modern technological applications and traditional educational approaches.

Rico emphasizes the value of AI in analyzing open-ended questions, stating that it allows him to “categorize the responses into emerging or clustered themes thus giving me a very rich visualization on the thought processes of my learners.” Traditional assessments, such as multiple-choice questions, fail to provide the same level of insight into students’ mental reasoning. Using AI to process open-ended responses enables Rico to gain a clearer understanding of his students’ cognitive processes, particularly when combined with computational tasks. “Letting my learners map every part of their computation with explanation will give me the specific details of their computational skills,” he notes, highlighting the AI’s ability to organize and synthesize complex responses.

However, Rico also acknowledges the limitations of AI, particularly when it comes to specific, factual questions. He explains, “AI is best when the question is divergent, requiring the richness of argument that reflects the extent of understanding of the learners. It works well for teachers to ask general questions accepted at the global scope, but it will break down when the instruction provided to learners is not of global standard.” AI can also handle computation, but for questions requiring precise or factual answers, Rico recommends caution, as the AI may not always align with the specific expectations of the instructor.

Despite these limitations, Rico believes that AI can elevate the assessment of higher-order thinking skills. “Generative AI can elevate the assessment of the STEM education by being able to gather functional information of the learners. This means that the upper taxonomy of Bloom’s taxonomy can be greatly measured in such a short period of time.” He further notes that traditionally, assessing higher-order thinking skills would require ample time, but AI can streamline this process, making it easier to evaluate students’ ability to synthesize, analyze, and evaluate information.

In Rico’s view, AI-generated assessments complement traditional assessments. He suggests, “AI-generated assessment complements so well with traditional forms of assessment, be it in generating the questions or in processing the answers.” While AI offers immediate visualization and efficiency, it is crucial for the teacher to maintain control over the final evaluation. “The task of having compact and immediate visualization is greatly reduced by AI, but the final input on the numerical equivalent to be given to learners must be your responsibility.”

Rico acknowledges that AI is pervasive in today’s learning environment, but he stresses the importance of balance and teacher judgment. “As to the balance, it will be entirely in the hands of the teacher in striking that balance. Unless, of course, there is an institutional stand on how AI is allowed in the institutional practice. But in the absence of higher-level guidance in the use of AI, the teacher must decide, provided there is corresponding reflection and documentation so that the teacher can very well elucidate the use of AI in one’s classroom.” Finally, Rico believes that AI is a powerful tool for enhancing assessment, but it requires careful oversight to ensure its effective and appropriate use in the classroom.

4.6 Roy's Case: AI for Assessment of Engineering Design Projects

Engr. Roy, an associate professor in computer engineering with 12 years of teaching experience, has integrated AI tools like ChatGPT and Grammarly into his academic practices to enhance the personalization and efficiency of assessments. In his view, AI's primary strength lies in its ability to offer personalized feedback, automate grading processes, and generate dynamic, individualized problem sets, including coding tasks and algorithmic challenges, tailored to each student's proficiency level. He emphasizes, "I use AI tools, such as ChatGPT, Grammarly, and specialized coding assistants like GitHub Copilot and Replit, to provide personalized feedback to students. These tools help improve the quality of their written work and coding assignments by offering real-time suggestions, identifying potential errors, and enhancing their coding skills through tailored recommendations."

One of the key advantages Engr. Roy sees in generative AI is its potential to simulate real-world systems and generate complex engineering problems, fostering a deeper understanding of theory and practice. He explains, "Generative AI in computer engineering can generate coding challenges and system simulations, tailored to the student's ability, providing them with opportunities to solve real-world engineering problems through new solutions and simulations." AI can also evaluate the quality of students' code, providing real-time feedback on issues such as efficiency, correctness, and overall code structure. According to Engr. Roy, this instant feedback helps students refine their problem-solving skills in a timely and effective manner.

AI's role in automating routine administrative tasks has also proven invaluable in Engr. Roy's teaching. By automating grading and generating problem sets, AI reduces the burden of administrative tasks, allowing instructors to dedicate more time to engaging with students on complex topics. "AI helps to reduce workload by automating grading and administrative tasks," he notes, "allowing me to focus on providing deeper insights and mentorship to students."

However, Engr. Roy is mindful of the limitations of AI in academic assessments. One of the significant concerns he highlights is AI's inability to understand the context or nuances of complex problem-solving approaches. He cautions, "AI may struggle to fully comprehend complex solutions or unconventional approaches, which can lead to potential oversights in evaluating creative or non-traditional answers." Furthermore, AI systems may fail to assess essential soft skills such as teamwork, communication, and collaboration, which are vital in engineering projects. "AI cannot evaluate soft skills that are critical in group projects and real-world engineering tasks," he notes.

Another limitation Engr. Roy points out is the potential for bias in AI algorithms, which can result in unfair assessments. "AI systems are only as good as the data they are trained on," he remarks. "If the training data is biased or unrepresentative, AI may inadvertently favor certain solutions or students, which could affect the fairness of assessments." He also expresses concern about the over-reliance on AI tools,

fearing that they might reduce opportunities for personalized mentorship. “AI can provide useful feedback, but it cannot replace the deeper, context-driven mentorship and guidance that human instructors can offer,” he says.

Despite these limitations, Engr. Roy remains optimistic about the future of AI in academic assessments, particularly in computer engineering. He foresees a hybrid assessment model, where AI handles the more technical and objective tasks, such as grading coding assignments or evaluating system designs, while human instructors focus on assessing creativity, problem-solving approaches, and the practical application of engineering principles. He explains, “Generative AI will play a pivotal role in shaping future assessments by generating diverse coding challenges and offering detailed feedback on technical skills. But human judgment will still be necessary for evaluating creativity and real-world applicability.”

Ethical concerns also arise when using AI in assessing engineering design projects. Engr. Roy acknowledges that AI might overlook creative problem-solving in favor of technical correctness, potentially undervaluing students’ innovative approaches. “One concern is that AI may not appreciate creativity in engineering design,” he admits. “It might focus on technical accuracy but overlook unique or unconventional solutions that deviate from traditional models.” Moreover, AI’s reliance on historical data can lead to biases that disadvantage students proposing novel ideas. “AI may not recognize innovative approaches that don’t fit within its trained data,” he explains. “This can lead to undervaluing creative solutions that break from established norms.”

To address these concerns, Engr. Roy advocates for a balanced, hybrid approach to assessment. “I would supplement AI-generated assessments with human judgment to ensure that creativity and innovation are appropriately recognized,” he quips. “For example, I could include open-ended questions or design challenges where students can share the thought process and reasoning behind their solutions.” This approach would allow instructors to assess not only the final product but also the creativity, problem-solving, and thought processes that led to the solution.

Engr. Roy also emphasizes the importance of continuously updating AI systems to mitigate bias and ensure fairness. “AI systems should be trained on diverse, real-world data to ensure that they can recognize innovative and unconventional ideas,” he asserts. This ensures that AI can handle the technical aspects of assessment, such as verifying design accuracy, while still allowing for human evaluation of creativity and real-world applicability.

Overall, Engr. Roy believes that AI has the potential to revolutionize assessment practices in computer engineering by enhancing efficiency, accuracy, and personalization. However, he stresses the importance of maintaining a hybrid approach that combines AI’s technical strengths with human judgment to ensure a comprehensive evaluation of both the technical and creative competencies necessary in engineering education. “The future of assessment in computer engineering will be driven by a combination of AI and human insights,” he concludes. “AI will handle the technical tasks, while educators will focus on nurturing creativity, critical thinking, and problem-solving skills in students.”

4.7 Mark's Case: AI in Project-Based Assessment in Technical-Vocational Teacher Education

Mark, an associate professor with 24 years of experience in drafting technology, acknowledges the growing impact of AI in education, particularly in assessment practices. Having taught for over two decades, Mark has witnessed significant technological advancements and embraces AI tools like ChatGPT, Bing AI, and Bard for creating assessment questions and providing personalized feedback. He believes that AI's use in assessment is a game-changer in classroom assessment, especially that its use significantly reduces preparation time and increases grading accuracy. Mark explains, "Generative AI for assessment is a good tool to generate accurate data based on set parameters," as it helps minimize personalized evaluation, giving fair and reasonable ratings. He notes that AI-driven assessment tools can reduce the workload for professors, especially when dealing with large classes, making grading more efficient and objective.

In terms of assessment types, Mark uses AI for various formats, including quizzes, exams, projects, presentations, and performance-based assessments like experiments and demonstrations. He emphasizes the utility of AI in automating administrative tasks and grading, improving both efficiency and accuracy. According to Mark, AI's assistance can help educators keep up with technological trends in education while also enhancing student engagement and motivation. "AI can make the educational process as efficient and effective as an assessment tool to lessen and fast-track evaluation and data analytics," he adds. With AI, the amount of time spent on grading and administrative tasks can be significantly reduced, allowing instructors to focus more on engaging with students.

While Mark is a strong advocate for the benefits of AI, he also recognizes its limitations. AI has not yet reached the point where it can fully replace human instructors for tasks requiring creativity or non-electronic outputs. As he explains, "It limits the creativity and personalized assessment to specific areas, such as non-electronic outputs." For instance, while AI can assess quizzes or projects submitted digitally, it may struggle with evaluating hands-on, creative tasks that require human judgment and interpretation. For now, Mark asserts that human intervention remains crucial for these types of personalized assessments. "For me, AI is a good tool in assessing parametric data defined specifically to interpret. But as for personalized tasks, still we need human intervention for the moment," he states.

Mark is also mindful of the evolving capabilities of AI. He sees the potential for AI to assess hands-on activities, which have traditionally been difficult to evaluate objectively. He notes, "As technology progresses, the picture of the hands-on activities can be detected and interpreted easily by AI." AI's ability to assess video demonstrations or operational procedures could transform how practical skills are evaluated in education. However, Mark remains cautious about the rapid development of AI. He acknowledges that while AI is improving exponentially, there are still flaws in its algorithms that need to be addressed. "As of today, AI algorithms are still learning, flaws may be observed, but along the way, AI can

correct itself for better assessment,” he explains. While he recognizes that AI has tremendous potential to assist in grading and administrative tasks, he is unsure whether AI can ever fully replace human evaluators for more nuanced educational processes.

To wrap up, Mark emphasizes that AI tools can significantly enhance the educational experience by making assessment more efficient, accurate, and data-driven. However, he stresses that AI should not replace human educators entirely. Instead, he suggests that AI should serve as an assistant, enhancing the academic experience rather than replacing the personal touch that human instructors provide. He advocates for “a robust and intense immersion” into AI’s capabilities to help educators understand its power and potential. While AI is still in its early stages and cannot replace human involvement in personalized tasks, it has already proven to be an invaluable tool for improving efficiency in educational settings. Mark’s experience highlights the promising future of AI in education, especially when used to complement, rather than replace, human expertise.

5 Practical Strategies for Integrating AI in STEM Education Assessments

The integration of AI into STEM education assessments offers significant potential for enhancing both teaching and learning experiences [3]. However, to fully harness the benefits of AI in assessment, it is essential to employ a range of practical strategies that involve faculty training, foster AI-human collaboration, encourage ethical AI use, and raise student awareness.

5.1 Faculty Training and Development

Successful integration of AI tools into STEM education assessments begins with comprehensive faculty training and professional development. Educators need the know-how to effectively utilize AI technologies, enabling them to enhance assessment practices while maintaining the quality of learning experiences [58]. Faculty training programs should cover the following areas:

Understanding AI Tools and Their Applications

Faculty members need to familiarize themselves with various AI tools available for assessment purposes, such as automated grading software, plagiarism checkers, adaptive learning platforms, and AI-driven feedback systems [59]. Training should

focus on how these tools can support formative and summative assessments, create personalized learning pathways, and offer real-time feedback.

AI Integration Strategies

Instructors should be guided on how to strategically integrate AI into their existing assessment methods. This includes using AI to automate routine tasks, such as grading quizzes or analyzing large data sets, while maintaining a balance with traditional, subjective evaluations for more complex tasks, like problem-solving or project-based assessments [6].

Continuing Pedagogical Support and Collaboration

Regular workshops, webinars, and collaboration among faculty members can ensure that educators remain up-to-date with the latest advancements in AI technologies. Furthermore, fostering a community of practice where instructors can share their experiences, successes, and challenges with AI in assessment can strengthen the overall integration process [60].

5.2 AI-Human Collaboration in Assessment

While AI can assist in various aspects of assessment, it is crucial to view it as a tool that complements, rather than replaces, human involvement [8]. The collaboration between AI and human judgment is particularly vital in STEM education, where the cultivation of critical thinking, creativity, and the application of knowledge are of prime importance. To promote AI-human collaboration in assessment, consider the following strategies:

AI for Routine Tasks, Human for the Evaluation of Higher-Order Thinking

AI can handle repetitive and administrative tasks, such as grading objective questions or analyzing large amounts of data, allowing instructors to focus on more nuanced and complex aspects of assessment, like evaluating creative problem-solving and assessing students' ability to apply knowledge in novel contexts [58].

Hybrid Assessments

Combining AI-driven assessments with traditional evaluation methods can ensure a more comprehensive approach. For example, AI can be used for quizzes and

instant feedback, while more complex assignments like lab reports, projects, and presentations can still be evaluated by instructors based on criteria that require human expertise, such as creativity, reasoning, and understanding [6].

Human Oversight and Ethical Considerations

Faculty must maintain oversight in all aspects of assessment, particularly when AI algorithms may overlook context or fail to recognize subtle differences in student responses. Faculty should use AI-generated results as a guide but rely on their professional judgment to make final assessments, ensuring fairness and consistency [61].

5.3 Encouraging Ethical AI Use

As AI becomes more integrated into STEM education assessments, it is essential to encourage ethical use of these technologies, both from faculty and students. The potential for bias, inequity, and privacy concerns in AI systems necessitate a clear ethical framework for its implementation [3]. To promote ethical AI use in STEM assessments, institutions can adopt the following strategies:

Transparency in AI Tools

Educators should be transparent about the AI tools they use in assessments. This includes informing students about how AI is being used, what data is being collected, and how the results will be applied. Transparency builds trust and helps students understand the role of AI in their learning journey [21].

Bias Detection and Mitigation

Faculty must ensure that the AI systems they use are designed to minimize bias. AI tools should undergo regular audits and evaluations to detect and address any algorithmic biases that may disproportionately impact specific student groups [29].

Data Privacy and Security

It is crucial to safeguard student data when using AI tools. Faculty should work closely with institutions' system administrator to ensure that AI systems comply with privacy regulations (e.g., FERPA) and that students' personal information is protected [62].

Clear Ethical Guidelines

Institutions should establish clear ethical guidelines for the use of AI in assessment. This includes setting boundaries for AI's role, such as ensuring it supports learning rather than replacing human judgment, and establishing a framework for addressing any ethical dilemmas that arise, such as AI-generated errors or inappropriate feedback [63].

5.4 *Fostering Student Awareness*

Students are central to the success of AI integration in STEM assessments, and fostering their awareness of AI's role in their learning is key to ensuring its effectiveness. Educating students about AI's capabilities, limitations, and ethical implications can enhance their engagement with AI tools and encourage responsible usage [7]. To foster student awareness of AI in assessment, the following strategies can be employed:

AI Literacy Programs

Institutions should offer workshops or courses that help students understand how AI works, its applications in STEM fields, and how it is used in assessment. Teaching students the basics of AI can demystify the technology and empower them to use it responsibly. Topics could include how AI algorithms are designed, how biases can emerge, and how to interpret AI-generated feedback [3].

Promoting Responsible AI Usage

Students should be educated on how to use AI tools ethically in their studies. This includes understanding the importance of academic integrity, recognizing when AI-generated content is appropriate, and learning how to avoid over-reliance on AI for assessments. Promoting responsible use can help ensure that students use AI as a complement to their own knowledge and skills, rather than as a shortcut [14].

Encouraging Critical Thinking

While AI can provide immediate feedback, it is important to encourage students to critically evaluate AI-generated responses. Instructors can promote critical thinking by asking students to reflect on the feedback they receive from AI tools and consider how it aligns with their own understanding of the subject. This process helps

students develop deeper insights and learn to apply their knowledge more effectively [7].

Open Dialogue About AI's Role in Education

Providing a platform for students to discuss their experiences and concerns with AI can foster a deeper understanding of its role in STEM assessments. This open dialogue can also help educators refine their approaches and address any issues that may arise, ensuring that the integration of AI enhances, rather than detracts from, the educational experience [27].

Integrating AI into STEM education assessments offers exciting possibilities for enhancing learning experiences, providing personalized feedback, and streamlining assessment processes. However, its successful integration requires strategic planning and collaboration among faculty, students, and institutional leaders. Table 11.1 outlines key AI integration strategies in STEM assessments that highlight essential approaches for achieving this balance. Emphasizing faculty training, promoting AI-human collaboration, encouraging ethical AI use, and fostering student awareness allow institutions to establish a balanced and effective approach to AI in STEM assessments. These strategies will ensure that AI remains a supportive tool that enriches the educational process, rather than overshadowing the critical thinking, creativity, and human judgment that are essential to STEM learning.

6 Implications for Policy and Practice

The integration of generative AI tools in STEM education assessments presents significant implications for policy and practice, particularly in higher education. Policymakers must prioritize the standardization of AI tools to ensure their transparency, accessibility, and ethical application within educational settings [45]. Establishing clear guidelines on data privacy, algorithmic fairness, and bias mitigation is crucial to prevent inequities and ensure trust in AI-driven systems [47]. Moreover, professional development programs for educators are essential to equip them with the technical skills and ethical awareness needed to effectively incorporate AI tools in assessment processes [60]. Such policies will ensure that AI complements the nuanced judgment of human educators rather than replacing it, reinforcing the value of human expertise in the assessment process.

Importantly, technologies such as generative AI should be adopted only if their use demonstrably enhances assessment practices. This includes improving the authenticity, appropriateness, validity, reliability, flexibility, timeliness of feedback, interactivity, administrability, and discriminatory power of assessments. Educators must leverage AI not merely for efficiency but to design assessment tasks that encourage original thought and minimize opportunities for academic dishonesty.

Table 11.1 AI integration strategies in STEM assessments

Strategy	Key actions	Description	Suggested readings
Faculty training and development	Understanding AI tools and applications	Educate faculty on AI tools (automated grading, feedback systems, etc.) for assessment purposes.	Kumar [59]. Faculty members' use of artificial intelligence to grade student papers: a case of implications.
	AI integration strategies	Provide guidance on integrating AI into existing assessment methods and balancing AI with traditional evaluation techniques.	Xia, et al. [6]. A scoping review on how generative artificial intelligence transforms assessment in higher education.
	Ongoing support and collaboration	Offer workshops, webinars, and foster collaboration among faculty to share experiences and challenges with AI tools.	Ding, et al. [60]. Enhancing teacher AI literacy and integration through different types of cases in teacher professional development.
AI-human collaboration in assessment	AI for routine tasks, human for higher-order evaluation	Use AI for repetitive tasks (grading quizzes) and humans for complex, subjective assessments (creative problem-solving).	Alkhouk & Khalif [58]. AI-resistant assessments in higher education: practical insights from faculty training workshops.
	Hybrid assessments	Combine AI-driven assessments with traditional methods, such as using AI for quizzes and humans for projects or presentations.	Xia et al. [6]. A scoping review on how generative artificial intelligence transforms assessment in higher education.
	Human oversight and ethical considerations	Ensure faculty maintain oversight and apply professional judgment to AI-generated assessments to avoid biases and errors.	Chai et al. [61]. Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness.
Encouraging ethical AI use	Transparency in AI tools	Be transparent with students about the AI tools being used, data collection, and feedback mechanisms.	Chan [55]. A comprehensive AI policy education framework for university teaching and learning.
	Bias detection and mitigation	Regularly audit AI tools for biases and ensure they are designed to minimize unfair judgments.	Swiecki et al. [29]. Assessment in the age of artificial intelligence.
	Data privacy and security	Adhere to privacy regulations (e.g., FERPA) and safeguard students' personal information when using AI tools.	Yang & Bei [62]. Ensuring data privacy in AI/ML implementation.

(continued)

Table 11.1 (continued)

Strategy	Key actions	Description	Suggested readings
Fostering student awareness	Clear ethical guidelines	Establish clear ethical guidelines for AI use in assessments, ensuring fairness and transparency in all AI-driven processes.	Balasubramaniam et al. [63]. Transparency and explainability of AI systems: From ethical guidelines to requirements.
	AI literacy programs	Provide students with AI literacy training, helping them understand how AI works and its role in assessments.	Xu & Ouyang [3]. The application of AI technologies in STEM education: a systematic review from 2011 to 2021.
	Promoting responsible AI usage	Educate students on the ethical use of AI, promoting academic integrity and avoiding over-reliance on AI.	C. Zhai et al. [14]. The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review.
	Encouraging critical thinking	Encourage students to critically evaluate AI feedback and integrate it with their own understanding.	Walter [7]. Embracing the future of Artificial Intelligence in the classroom: the relevance of AI literacy, prompt engineering, and critical thinking in modern education.
	Open dialogue about AI's role in education	Foster open discussions between students and faculty to address concerns and improve AI integration in assessments.	Park et al. [27]. Integrating artificial intelligence into science lessons: teachers' experiences and views.

Echoing the guidance of e-learning scholar Gilly Salmon, the focus should always be on what the pedagogy requires, rather than on what technology can offer.

From a practical standpoint, the adoption of a hybrid assessment model that integrates AI and human evaluation is indispensable. Generative AI can efficiently handle technical tasks, such as grading coding assignments or solving numerical problems, streamlining administrative workloads for educators [10]. However, human input remains vital for assessing complex competencies such as creativity, critical thinking, and real-world applications, which are integral to higher education outcomes [4, 64]. Moreover, STEM educators must approach AI-generated assessments with a critical mindset. Constructing assessments that accurately measure student understanding is a fundamental responsibility of educators, and no technology can rectify an invalid assessment. As emphasized by Phil Race, the fundamental principles of assessment—validity, reliability, transparency, and authenticity—remain paramount despite technological advancements.

Furthermore, we envision that technologies like generative AI should not only enhance but transform STEM assessment practices. Drawing from the SAMR model of technology integration, true transformation involves Modification and Redefinition of tasks, not just Substitution or simple Augmentation. Thoughtful AI integration should thus aim to create novel assessment practices that were previously inconceivable without the technology, opening pathways for deeper student engagement and innovative demonstrations of learning.

Additionally, AI facilitates personalized learning experiences by providing students with tailored feedback and adaptive problem sets that address their individual strengths and weaknesses, thereby fostering continuous improvement and higher engagement [6]. Yet, as Brown et al. [65] emphasized, nothing educators do is more consequential than assessing student work and providing feedback—decisions that can influence students for a lifetime. Hence, responsible use of AI in assessment is critical to uphold educational integrity.

Equity remains a pressing concern in the integration of AI tools, as disparities in infrastructure and socio-economic status may limit access for some students. Educational institutions must address the digital divide by investing in infrastructure, providing access to necessary tools, and implementing inclusive policies that ensure all students benefit from AI-enhanced assessments [43]. Bridging these gaps is critical to creating an equitable learning environment where students from diverse backgrounds have equal opportunities to succeed.

Ultimately, the integration of AI in STEM assessments holds the potential to better prepare students for future STEM careers. Exposing students to AI-driven technologies helps them develop key twenty-first century skills such as adaptability, problem-solving, and innovation, which are increasingly in demand in the global workforce [66]. The transformative potential of AI in education requires careful planning and thoughtful implementation, underpinned by robust policies that ensure fairness, accessibility, and effectiveness. Through this balanced approach, AI can serve as a powerful tool to enhance and transform STEM education, equipping students with the skills needed to thrive in a rapidly evolving technological landscape.

7 Conclusion

AI's transformative potential in STEM education assessments lies in its ability to personalize feedback, automate tasks, and enhance creativity in assessment design. Tools such as ChatGPT, Grammarly, and domain-specific coding platforms offer opportunities to streamline processes and elevate learning experiences. However, these advancements should complement rather than replace the pedagogical expertise of educators, ensuring that assessments balance technical proficiency with creativity and critical thinking.

This chapter highlighted both the benefits and ethical dilemmas associated with AI-driven STEM assessments. Although AI enhances efficiency and adaptability, significant concerns persist regarding data privacy, algorithmic bias, and equitable access. A key finding is that effective AI integration depends on institutional policies and faculty preparedness. Without proper training and ethical safeguards, the risk of over-reliance on AI-generated assessments could undermine essential educational outcomes.

One limitation of this chapter is its reliance on qualitative insights rather than empirical data obtained from experimental or quantitative methods. Future research could explore comparative studies between AI-driven and traditional assessment methods, incorporating student performance metrics and faculty perceptions to validate claims about AI's impact on learning outcomes.

To maximize AI's benefits while mitigating risks, policymakers and educational institutions must develop clear guidelines, ethical frameworks, and professional development programs. These measures will empower educators to effectively harness AI tools while ensuring that assessments remain inclusive, transparent, and aligned with educational goals. With thoughtful implementation, AI can support an educational landscape that is both technologically advanced and deeply human-centered, helping students succeed in an evolving digital world.

References

1. Giannakos, M., Azevedo, R., Brusilovsky, P., Cukurova, M., Dimitriadis, Y., Hernandez-Leo, D., Järvelä, S., Mavrikis, M., & Rienties, B. (2024). The promise and challenges of generative AI in education. *Behav. Inf. Technol.*, 1–27. <https://doi.org/10.1080/0144929x.2024.2394886>
2. Wang, S., Wang, F., Zhu, Z., Wang, J., Tran, T., & Du, Z. (2024). Artificial intelligence in education: A systematic literature review. *Expert Syst. Appl.*, 252, 124167. <https://doi.org/10.1016/j.eswa.2024.124167>
3. Xu, W., & Ouyang, F. (2022). The application of AI technologies in STEM education: A systematic review from 2011 to 2021. *Int. J. STEM Educ.*, 9, 59. <https://doi.org/10.1186/s40594-022-00377-5>
4. Bozkurt, A., Xiao, J., Farrow, R., Bai, J. Y. H., Nerantzi, C., Moore, S., Dron, J., Stracke, C. M., Singh, L., Crompton, H., Koutropoulos, A., Terentev, E., Pazurek, A., Nichols, M., Sidorkin, A. M., Costello, E., Watson, S., Mulligan, D., Honeychurch, S., . . . Asino, T. I. (2024). The Manifesto for Teaching and Learning in a Time of Generative AI: A Critical Collective Stance to Better Navigate the Future. *Open Praxis* 16, 487–513. <https://doi.org/10.55982/openpraxis.16.4.777>

5. Lahby, M. (Ed.). (2024). Empowering digital education with ChatGPT: From theoretical to practical applications (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/9781032716350>
6. Xia, Q., Weng, X., Ouyang, F., Lin, T. J., & Chiu, T. K. (2024). A scoping review on how generative artificial intelligence transforms assessment in higher education. *Int. J. Educ. Technol. High. Educ.*, 21, 40. <https://doi.org/10.1186/s41239-024-00468-z>
7. Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *Int. J. Educ. Technol. High. Educ.*, 21, 15. <https://doi.org/10.1186/s41239-024-00448-3>
8. Kizilcec, R. F., Huber, E., Papanastasiou, E. C., Cram, A., Makridis, C. A., Smolansky, A., Zeivots, S., & Radulescu, C. (2024). Perceived impact of generative AI on assessments: Comparing educator and student perspectives in Australia, Cyprus, and the United States. *Comput. Educ. Artif. Intell.*, 7, 100269. <https://doi.org/10.1016/j.caeai.2024.100269>
9. Adewale, M. D., Azeta, A., Abayomi-Alli, A., & Sambo-Magaji, A. (2024). Impact of artificial intelligence adoption on students' academic performance in open and distance learning: A systematic literature review. *Heliyon* 10, e40025. <https://doi.org/10.1016/j.heliyon.2024.e40025>
10. Kamalov, F., Calonge, D. S., & Gurrub, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, 15(16), 12451. <https://doi.org/10.3390/su151612451>
11. Acut, D. P., Malabago, N. K., Malicoban, E. V., Galamiton, N. S., & Garcia, M. B. (2024). "ChatGPT 4.0 Ghosted Us While Conducting Literature Search:" Modeling the Chatbot's Generated Non-Existent References Using Regression Analysis. *Internet Ref. Serv. Q.* 1, 1–26. <https://doi.org/10.1080/10875301.2024.2426793>
12. Mennella, C., Mascalco, U., De Pietro, G., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, 10(4), e26297. <https://doi.org/10.1016/j.heliyon.2024.e26297>
13. Sharma, S. (2024). Benefits or concerns of AI: A multistakeholder responsibility. *Futures*, 157, 103328. <https://doi.org/10.1016/j.futures.2024.103328>
14. Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: A systematic review. *Smart Learn. Environ.*, 11, 28. <https://doi.org/10.1186/s40561-024-00316-7>
15. Al-Zahrani, A. M. (2024). Unveiling the shadows: Beyond the hype of AI in education. *Heliyon* 10, e30696. <https://doi.org/10.1016/j.heliyon.2024.e30696>
16. Parviz, M. (2024). AI in education: Comparative perspectives from STEM and non-STEM instructors. *Comput. Educ. Open*, 6, 100190. <https://doi.org/10.1016/j.caeo.2024.100190>
17. Lv, Z. (2023). Generative artificial intelligence in the metaverse era. *Cogn. Robot.*, 3, 208–217. <https://doi.org/10.1016/j.cogr.2023.06.001>
18. Chiarello, F., Giordano, V., Spada, I., Barandoni, S., & Fantoni, G. (2024). Future applications of generative large language models: A data-driven case study on ChatGPT. *Technovation* 133, 103002. <https://doi.org/10.1016/j.technovation.2024.103002>
19. Imran, M., & Almusharraf, N. (2024). Google Gemini as a next-generation AI educational tool: A review of emerging educational technology. *Smart Learn. Environ.*, 11, 22. <https://doi.org/10.1186/s40561-024-00310-z>
20. Yusuf, A., Pervin, N., & Román-González, M. (2024). Generative AI and the future of higher education: A threat to academic integrity or reformation? Evidence from multicultural perspectives. *Int. J. Educ. Technol. High. Educ.*, 21, 21. <https://doi.org/10.1186/s41239-024-00453-6>
21. Chan, C. K. Y., & Hu, W. (2023). Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int. J. Educ. Technol. High. Educ.* 20, 43. <https://doi.org/10.1186/s41239-023-00411-8>
22. Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Educ. Sci.*, 13(11), 1109. <https://doi.org/10.3390/educsci13111109>

23. Nixon, N., Lin, Y., & Snow, L. (2024). Catalyzing equity in STEM teams: Harnessing generative AI for inclusion and diversity. *Policy Insights Behav. Brain Sci.*, 11(1), 85–92. <https://doi.org/10.1177/23727322231220356>
24. Brown, G. T. L. (2022). The past, present and future of educational assessment: A transdisciplinary perspective. *Front. Educ.* 7. <https://doi.org/10.3389/educ.2022.1060633>
25. Gao, X., Li, P., Shen, J., & Sun, H. (2020). Reviewing assessment of student learning in interdisciplinary STEM education. *Int. J. STEM Educ.*, 7, 24. <https://doi.org/10.1186/s40594-020-00225-4>
26. Saputra, I., Kurniawan, A., Yanita, M., Putri, E. Y., & Mahniza, M. (2024). The evolution of educational assessment: How artificial intelligence is shaping the trends and future of learning evaluation. *Indones. J. Comput. Sci.*, 13(6). <https://doi.org/10.33022/ijcs.v13i6.4465>
27. Park, J., Teo, T. W., Teo, A., Chang, J., Huang, J. S., & Koo, S. (2023). Integrating artificial intelligence into science lessons: Teachers' experiences and views. *Int. J. STEM Educ.*, 10, 61. <https://doi.org/10.1186/s40594-023-00454-3>
28. Arslan, B., Lehman, B., Tenison, C., Sparks, J. R., López, A. A., Gu, L., & Zapata-Rivera, D. (2024). Opportunities and challenges of using generative AI to personalize educational assessment. *Front. Artif. Intell.* 7. <https://doi.org/10.3389/frai.2024.1460651>
29. Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Comput. Educ. Artif. Intell.*, 3, 100075. <https://doi.org/10.1016/j.caeai.2022.100075>
30. Seo, K., Tang, J., Roll, I., Fels, S., & Yoon, D. (2021). The impact of artificial intelligence on learner–instructor interaction in online learning. *Int. J. Educ. Technol. High. Educ.*, 18, 54. <https://doi.org/10.1186/s41239-021-00292-9>
31. Escalante, J., Pack, A., & Barrett, A. (2023). AI-generated feedback on writing: Insights into efficacy and ENL student preference. *Int. J. Educ. Technol. High. Educ.*, 20, 57. <https://doi.org/10.1186/s41239-023-00425-2>
32. Malik, A. R., Pratiwi, Y., Andajani, K., Numertayasa, I. W., Suharti, S., Darwis, A., & Marzuki, N. (2023). Exploring artificial intelligence in academic essay: Higher education students' perspective. *Int. J. Educ. Res. Open*, 5, 100296. <https://doi.org/10.1016/j.ijedro.2023.100296>
33. Mahapatra, S. (2024). Impact of ChatGPT on ESL students' academic writing skills: A mixed methods intervention study. *Smart Learn. Environ.*, 11, 9. <https://doi.org/10.1186/s40561-024-00295-9>
34. Chen, X., Xie, H., Zou, D., & Hwang, G. (2020). Application and theory gaps during the rise of Artificial Intelligence in Education. *Comput. Educ. Artif. Intell.* 1, 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
35. Acut, D. P. (2024). From classroom learning to real-world skills: an autoethnographic account of school field trips and STEM work immersion program management. *Discip. Interdiscip. Sci. Educ. Res.* 6, 20. <https://doi.org/10.1186/s43031-024-00111-x>
36. Petersen, G. B., Stenberdt, V., Mayer, R. E., & Makransky, G. (2023). Collaborative generative learning activities in immersive virtual reality increase learning. *Comput. Educ.*, 207, 104931. <https://doi.org/10.1016/j.compedu.2023.104931>
37. Zhao, T., Wang, S., Ouyang, C., Chen, M., Liu, C., Zhang, J., Yu, L., Wang, F., Xie, Y., Li, J., Wang, F., Grunwald, S., Wong, B. M., Zhang, F., Qian, Z., Xu, Y., Yu, C., Han, W., Sun, T., . . . Wang, L. (2024). Artificial intelligence for geoscience: Progress, challenges and perspectives. *The Innovation*, 5(5), 100691. <https://doi.org/10.1016/j.xinn.2024.100691>
38. Dai, C., & Ke, F. (2022). Educational applications of artificial intelligence in simulation-based learning: A systematic mapping review. *Comput. Educ. Artif. Intell.*, 3, 100087. <https://doi.org/10.1016/j.caeai.2022.100087>
39. Runge, A., Attali, Y., LaFlair, G. T., Park, Y., & Church, J. (2024). A generative AI-driven interactive listening assessment task. *Front. Artif. Intell.*, 7. <https://doi.org/10.3389/frai.2024.1474019>
40. Yilmaz, R., & Yilmaz, F. G. K. (2023). The effect of generative artificial intelligence (AI)-based tool use on students' computational thinking skills, programming self-efficacy and motivation. *Comput. Educ. Artif. Intell.*, 4, 100147. <https://doi.org/10.1016/j.caeai.2023.100147>

41. Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., Liu, J., Yuan, J., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity*, 1, 8812542. <https://doi.org/10.1155/2021/8812542>
42. Ulla, M. B., Advincula, M. J. C., Mombay, C. D. S., Mercullo, H. M. A., Nacionales, J. P., & Entino-Señorita, A. D. (2024). How can GenAI foster an inclusive language classroom?: A critical language pedagogy perspective from Philippine university teachers. *Comput. Educ. Artif. Intell.*, 7, 100314. <https://doi.org/10.1016/j.caeai.2024.100314>
43. United Nations Educational, Scientific and Cultural Organization (UNESCO). (2023). How generative AI is reshaping education in Asia-Pacific. Retrieved from <https://www.unesco.org/en/articles/how-generative-ai-reshaping-education-asia-pacific>
44. Marino, M. T., Vasquez, E., Dieker, L., Basham, J., & Blackorby, J. (2023). The future of artificial intelligence in special education technology. *J. Spec. Educ. Technol.*, 38(3), 404–416. <https://doi.org/10.1177/01626434231165977>
45. Funa, A. A., & Gabay, R. A. E. (2024). Policy guidelines and recommendations on AI use in teaching and learning: A meta-synthesis study. *Soc. Sci. Humanit. Open*, 11, 101221. <https://doi.org/10.1016/j.ssaho.2024.101221>
46. Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Comput. Educ. Artif. Intell.*, 5, 100152. <https://doi.org/10.1016/j.caeai.2023.100152>
47. Ferrara, E. (2023). Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. *Sci.*, 6(1), 3. <https://doi.org/10.3390/sci6010003>
48. Myers, A. (2023). AI-detectors biased against non-native English writers. *Stanford Univ. Hum.-Cent. Artif. Intell.* Retrieved from <https://hai.stanford.edu/news/ai-detectors-biased-against-non-native-english-writers>
49. Karan, B., & Angadi, G. R. (2023). Potential risks of artificial intelligence integration into school education: A systematic review. *Bull. Sci. Technol. Soc.*, 43(3–4), 67–85. <https://doi.org/10.1177/02704676231224705>
50. Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M., & Hussain, A. (2023). Interpreting black-box models: A review on explainable artificial intelligence. *Cogn. Comput.*, 16, 45–74. <https://doi.org/10.1007/s12559-023-10179-8>
51. Garcia, M. B. (2023). Using AI tools in writing peer review reports: Should academic journals embrace the use of ChatGPT? *Ann. Biomed. Eng.*, 52(2), 139–140. <https://doi.org/10.1007/s10439-023-03299-7>
52. Mestari, S. Z. E., Lenzini, G., & Demirci, H. (2023). Preserving data privacy in machine learning systems. *Comput. Secur.*, 137, 103605. <https://doi.org/10.1016/j.cose.2023.103605>
53. Ulven, J. B., & Wangen, G. (2021). A systematic review of cybersecurity risks in higher education. *Future Internet*, 13(2), 39. <https://doi.org/10.3390/fi13020039>
54. Liu, Q., & Khalil, M. (2023). Understanding privacy and data protection issues in learning analytics using a systematic review. *Br. J. Educ. Technol.*, 54(6), 1715–1747. <https://doi.org/10.1111/bjet.13388>
55. Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *Int. J. Educ. Technol. High. Educ.* 20, 38. <https://doi.org/10.1186/s41239-023-00408-3>
56. Giray, L. (2024). The problem with false positives: AI detection unfairly accuses scholars of AI plagiarism. *Serials Librarian*, 1–9. <https://doi.org/10.1080/0361526x.2024.2433256>
57. Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). Thousand Oaks: Sage.
58. Alkhouk, W. A., & Khlaif, Z. N. (2024). AI-resistant assessments in higher education: practical insights from faculty training workshops. *Front. Educ.* 9. <https://doi.org/10.3389/educ.2024.1499495>
59. Kumar, R. (2023). Faculty members' use of artificial intelligence to grade student papers: A case of implications. *Int. J. Educ. Integr.*, 19, 9. <https://doi.org/10.1007/s40979-023-00130-7>

60. Ding, A. E., Shi, L., Yang, H., & Choi, I. (2024). Enhancing teacher AI literacy and integration through different types of cases in teacher professional development. *Comput. Educ. Open*, 6, 100178. <https://doi.org/10.1016/j.caeo.2024.100178>
61. Chai, F., Ma, J., Wang, Y., Zhu, J., & Han, T. (2024). Grading by AI makes me feel fairer? How different evaluators affect college students' perception of fairness. *Front. Psychol.* 15. <https://doi.org/10.3389/fpsyg.2024.1221177>
62. Yang, E., & Beil, C. (2024). Ensuring data privacy in AI/ML implementation. *New Dir. High. Educ.*, 207, 63–78. <https://doi.org/10.1002/he.20509>
63. Balasubramaniam, N., Kauppinen, M., Rannisto, A., Hiekkänen, K., & Kujala, S. (2023). Transparency and explainability of AI systems: From ethical guidelines to requirements. *Inf. Softw. Technol.* 159, 107197. <https://doi.org/10.1016/j.infsof.2023.107197>
64. Garcia, M. B. (2024). The paradox of artificial creativity: Challenges and opportunities of generative AI artistry. *Creat. Res. J.*, 1–14. <https://doi.org/10.1080/10400419.2024.2354622>
65. Brown, S., Race, P., & Smith, B. (2005). *500 Tips on Assessment*. Routledge, 1–256. <https://doi.org/10.4324/9780203307359>
66. Jaramillo, J. J., & Chiappe, A. (2024). The AI-driven classroom: A review of 21st century curriculum trends. *Prospects*. <https://doi.org/10.1007/s11125-024-09704-w>