Deep Science Publishing
https://doi.org/10.70593/978-93-49307-53-7_5

Chapter 5: Ethical considerations and data privacy in artificial intelligence

Ebenezer Gaisie, Obeng Owusu-Boateng, Issifu Yidana, Benjamin Ghansah

Abstract: The integration of Artificial Intelligence (AI) in education offers unprecedented opportunities to personalize learning experiences, automate routine tasks, and derive data-driven insights. However, this technological evolution is inextricably linked with significant ethical considerations and data privacy challenges. This chapter provides a comprehensive exploration of the critical ethical dimensions and data protection imperatives that arise from the increasing adoption of AI within educational tools and systems. It examines the key ethical issues, including algorithmic bias and fairness, transparency and accountability, the impact on student autonomy, and the evolving role of human educators. Furthermore, it explores the data privacy risks associated with the extensive collection and processing of sensitive student information, highlighting the importance of robust data protection protocols and compliance with legal and regulatory frameworks. The chapter also explores strategies for fostering ethical AI development and implementation, emphasizing the importance of inclusive design methodologies, stakeholder engagement, and the integration of ethical considerations into AI education curricula. Ultimately, the chapter highlights the imperative for a balanced approach that harnesses the transformative potential of AI in education while upholding ethical principles and safeguarding data privacy. It calls for collaborative efforts among policymakers, educators, and technology developers to establish comprehensive ethical guidelines and ensure the responsible and equitable deployment of AI in educational contexts.

Keywords: Artificial Intelligence (AI), Education, Ethical Considerations, Data Privacy, Algorithmic Bias, Transparency, Accountability

Ebenezer Gaisie
Department of ICT Education, University of Education, Winneba. Ghana

Obeng Owusu-Boateng
Department of Mathematics & ICT, E. P. College of Education, Bimbilla. Ghana.

Issifu Yidana
Department of ICT Education, University of Education, Winneba. Ghana

Benjamin Ghansah
Department of ICT Education, University of Education, Winneba. Ghana

1. Introduction

Artificial Intelligence (AI) rapidly transforming the educational landscape, playing an increasingly significant role in shaping pedagogical approaches, learning experiences, and institutional management (Ifenthaler et al., 2024). The integration of AI-driven tools and intelligent systems offers educators unprecedented opportunities to personalize learning pathways (Admane et al., 2024), automate routine tasks, and derive data-informed insights into student progress and outcomes (Greller & Drachsler, 2012). Technologies such as adaptive learning platforms, automated assessment systems, and intelligent virtual assistants are becoming more widespread in educational settings, promising to enhance both the efficiency and effectiveness of teaching and learning processes (Adams et al., 2021). The potential benefits of AI in education are extensive, ranging from addressing diverse learning needs and providing timely feedback (Ifenthaler et al., 2024) to streamlining administrative burdens and fostering innovative pedagogical practices (Adeleye et al., 2024). However, alongside these considerable advancements, a crucial imperative emerges to thoroughly examine and address the inherent ethical considerations and challenges to data privacy that are inextricably linked to the deployment of these powerful technologies (Adams et al., 2021).

Given that AI systems in education fundamentally rely on the extensive collection, processing, and analysis of substantial amounts of student data (Afzal et al., 2023), profound concerns regarding the privacy and security of this sensitive information are paramount (Ward, 2023). The potential for data breaches, the misuse of personal information, and the far-reaching implications of increasing datafication within educational contexts necessitate meticulous consideration and the establishment of robust data protection protocols (Afzal et al., 2023). Furthermore, the very design and deployment of AI algorithms in education raise fundamental ethical questions pertaining to fairness, bias, transparency, and accountability (Baker & Hawn, 2022). Unless these

critical issues are thoughtfully addressed, biases that may be embedded within the data used to train AI systems can inadvertently lead to these systems perpetuating or amplifying existing societal inequalities, ultimately resulting in discriminatory outcomes for certain student populations (Baker & Hawn, 2022; Bansal et al., 2023; Chinta et al., 2024). Ensuring the transparency and interpretability of decisions made by AI-driven systems is also of vital importance for cultivating trust among stakeholders and empowering educators and learners to comprehend and, where necessary, challenge the outputs and recommendations of these technologies (Manna & Sett, 2025). The broader impact of AI on the fundamental dynamics of the teacher-student relationship, the promotion of student autonomy and agency, and the overarching ethical landscape of education necessitates thorough investigation and the development of appropriate ethical guidelines and policy frameworks (Gouseti et al., 2024).

The purpose of this chapter is to provide a detailed exploration of the critical ethical considerations and data privacy challenges that arise from the increasing integration of AI within educational tools and systems. Specifically, it aims to offer a comprehensive overview of the key ethical dimensions that must be carefully navigated to ensure the responsible, equitable, and trustworthy deployment of AI in both K-12 and higher education contexts (Adams et al., 2021; Gouseti et al., 2024). The scope of this chapter will encompass a critical examination of prevailing ethical frameworks that are relevant to the application of AI in education, a detailed analysis of the specific data privacy risks and vulnerabilities associated with AI-driven educational technologies (Afzal et al., 2023), and a thorough discussion of strategies and best practices that can be implemented to mitigate bias and actively promote fairness in AI algorithms used in educational settings (Baker & Hawn, 2022; Chinta et al., 2024). Moreover, this chapter will explore the multifaceted implications of AI in education for a diverse range of stakeholders, including students themselves, educators at all levels, policymakers responsible for shaping educational technology governance, and the technology developers who create and deploy these AI systems, thereby highlighting the critical importance of collaborative and multi-disciplinary efforts in establishing robust ethical guidelines and ensuring accountability in their application (Ifenthaler et al., 2024).

To facilitate a structured and comprehensive exploration of these critical issues, this chapter will be organized into the following key sections. Firstly, it will offer a more in-depth exposition of the current landscape of AI in education, outlining the diverse array of its applications across different educational levels and the underlying AI technologies that power these innovations (Adams et al., 2021; Ifenthaler et al., 2024). Following this, the chapter will delve into a comprehensive analysis of the primary ethical considerations that are closely associated with the use of AI in education, including detailed discussions of issues such as fairness and bias in algorithms and outcomes (Baker & Hawn, 2022; Bansal et al., 2023; Chinta et al., 2024), the critical need for

transparency and explainability in AI-driven decision-making (Manna & Sett, 2025), the potential impact of AI on student autonomy and agency in learning (Gouseti et al., 2024), and the overarching implications for educational equity and access for all learners (Stephenson & Harvey, 2022; Stephenson et al., 2022). Subsequently, the chapter will focus specifically on the crucial aspects of data privacy and security in AI-driven education, thoroughly examining the various types of student data that are collected, the potential risks of data breaches, unauthorized access, and the misuse of this sensitive information (Afzal et al., 2023; Ward, 2023), as well as the relevant legal and regulatory frameworks that are in place to protect student data and ensure compliance (Regulation (EU) 2016/679). The chapter will then explore a range of strategies and practical recommendations for fostering ethical AI development and implementation within educational settings, underscoring the importance of principles such as inclusive design methodologies, meaningful engagement with all relevant stakeholders (Ifenthaler et al., 2024), and the critical integration of ethical considerations into AI education curricula for both current students and future educators (Schiff, 2022). Finally, the chapter will conclude by providing a concise summary of the key ethical and data privacy challenges that have been discussed and offering insightful perspectives on future directions for research, policy development, and practical implementation in the continuously evolving landscape of AI in education.

2. Ethical Implications of Using AI in Education

2.1 Understanding AI in Education

Artificial intelligence in education (AIEd) represents a dynamic and multifaceted field dedicated to employing artificial intelligence technologies to enhance and transform educational practices and outcomes (Agarwal et al., 2024; Holmes & Porayska-Pomsta, 2023; Nguyen et al., 2023). While definitions of AI itself are acknowledged to be complex and contested (Akgun & Greenhow, 2022), AIEd, at its core, involves the application of computational methods and algorithms to analyze educational data, automate tasks, personalize learning experiences, and provide intelligent support to both learners and educators (Agarwal et al., 2024). This interdisciplinary area draws upon principles from computer science, cognitive science, educational psychology, and ethics to develop and implement AI-powered tools and systems within diverse educational contexts, ranging from K-12 to higher education and lifelong learning initiatives (Holmes & Porayska-Pomsta, 2023). The overarching aim of AIEd is to leverage the capabilities of AI to foster more effective, efficient, equitable, and engaging learning environments, while also addressing the unique challenges and opportunities presented by the integration of these technologies within the educational sphere (Eden et al., 2024).

The landscape of AI applications in education is rapidly expanding, encompassing a diverse array of tools designed to address various aspects of the learning and teaching process (Akgun & Greenhow, 2022; Gouseti et al., 2024). Adaptive learning platforms, for instance, utilize AI algorithms to analyze individual student performance and tailor the pace, content, and instructional strategies to meet their specific learning needs and preferences (Gouseti et al., 2024; Lata, 2024; Spulber, 2024). Automated grading systems employ natural language processing and machine learning techniques to evaluate student assignments and provide feedback, potentially freeing up educators' time for more personalized interactions (Akgun & Greenhow, 2022; Chinta et al., 2024). Chatbots and intelligent virtual assistants can offer students immediate support by answering questions, providing resources, and guiding them through learning materials (Akgun & Greenhow, 2022). Furthermore, learning analytics tools leverage AI to analyze large datasets of student interactions and performance data, providing educators and institutions with valuable insights into learning patterns, potential areas of difficulty, and the effectiveness of pedagogical approaches (Agarwal et al., 2024; Spulber, 2024). Beyond these common applications, AI is also being explored for use in assistive technologies to support inclusive education (Lata, 2024; Spulber, 2024), for automating administrative tasks to streamline educational management (Gouseti et al., 2024; World Economic Forum, 2024), and for creating immersive learning experiences through virtual and augmented reality (Eden et al., 2024).

At the core of many AI-driven educational applications lie the fundamental concepts of machine learning and big data (Akgun & Greenhow, 2022; Spulber, 2024). Machine learning, a subfield of AI, involves the development of algorithms that enable computer systems to learn from data without being explicitly programmed (Akgun & Greenhow, 2022). In the context of education, machine learning algorithms are trained on vast datasets of student interactions, performance records, and contextual information to identify patterns, make predictions, and personalize learning experiences (Spulber, 2024). For example, adaptive learning platforms utilize machine learning to predict a student's optimal learning path based on their past performance and interactions with the system (Lata, 2024). Similarly, automated grading systems employ machine learning models trained on annotated student work to assess new submissions (Chinta et al., 2024). The effectiveness and ethical implications of these AI systems are intrinsically linked to the availability of large, diverse, and high-quality educational data, often referred to as big data (Agarwal et al., 2024; Nguyen et al., 2023). The collection, storage, and analysis of this sensitive student data raise significant ethical concerns related to privacy, security, bias, and fairness, underscoring the critical importance of responsible data stewardship in AIEd (Agarwal et al., 2024; Eden et al., 2024; Roshanaei et al., 2023; Yan & Liu, 2024).

2.2 Ethical Challenges of AI in Education

Bias and Fairness in AI Systems

A significant ethical challenge arising from the integration of artificial intelligence in education (AIEd) pertains to the pervasive issues of bias and the imperative of fairness within AI systems (Agarwal et al., 2024; Barnes & Hutson, 2024; Chinta et al., 2024; Donatus et al., 2024). The very nature of AI algorithms, particularly those relying on machine learning, means that they learn from the data they are trained on (Akgun & Greenhow, 2022; Bansal et al., 2023). Consequently, if this training data reflects existing societal, historical, or systemic biases, the AI system will inevitably inherit and potentially amplify these biases in its operations and outputs (Akgun & Greenhow, 2022; Bansal et al., 2023; Chinta et al., 2024; Donatus et al., 2024; Holstein & Doroudi, 2021; Khan, 2024; Manna & Sett, 2025; Schiff, 2022). These biases can originate from various sources, including but not limited to, underrepresentation or misrepresentation of certain demographic groups in the data (Holmes et al., 2022; Holstein & Doroudi, 2021; Porayska-Pomsta et al., 2023), the subjective preferences or prejudices embedded in human-annotated data (Chinta et al., 2024), and the very design choices and assumptions made during the development of the algorithms themselves (Eden et al., 2024; Holstein & Doroudi, 2021). As algorithms reflect the values of their creators who often hold positions of power, they can inadvertently perpetuate existing inequalities (Akgun & Greenhow, 2022).

The manifestation of bias in AI-driven educational tools can have profound and detrimental effects across various applications. In the realm of student assessments, AI-powered essay grading systems, for example, have the potential to reflect biases present in the data used to train them, which might include the stylistic preferences of human graders that favour certain writing styles or cultural backgrounds over others (Chinta et al., 2024; Donatus et al., 2024). This can lead to unfair evaluations for students whose writing does not conform to the norms established in the training data, potentially putting those from marginalized groups at a disadvantage (Chinta et al., 2024). Similarly, in college admissions, predictive algorithms trained on historical data that reflect existing disparities in access and achievement could inadvertently perpetuate race, gender, or socioeconomic biases by directly or indirectly favouring students with certain demographic characteristics (Barnes & Hutson, 2024; Schiff, 2022; Syaidina et al., 2024). The controversy surrounding the Office of Qualifications and Examinations Regulation (Ofqual)'s use of an algorithm in college admissions in the United Kingdom serves as a stark reminder of the risks associated with unjust or poorly considered applications of AI in administrative educational contexts (Schiff, 2022). In personalized learning systems, if the underlying algorithms are trained on biased data, they may offer

less effective support or recommend resources that are less enriching to certain groups of students, thereby exacerbating existing educational inequalities (Naseeb & Bhatti, 2024; Roshanaei et al., 2023).

The impact of biased AI decisions on educational equity and inclusion is substantial and far-reaching (Agarwal et al., 2024; Eden et al., 2024; Roshanaei et al., 2023). By perpetuating and amplifying existing biases, AI systems can undermine the very goals of equitable and inclusive education, leading to unfair outcomes and reinforcing patterns of discrimination against historically marginalized groups (Chinta et al., 2024; Donatus et al., 2024; Holstein & Doroudi, 2021; Khan, 2024; Lata, 2024; Leta & Vancea, 2023; Naseeb & Bhatti, 2024; Nguyen et al., 2023; Nurhasanah & Nugraha, 2024; Porayska-Pomsta et al., 2023; Sato et al., 2024; Spulber, 2024; Syaidina et al., 2024; Yan & Liu, 2024). For instance, if AI-driven tools consistently provide less challenging or fewer opportunities to students from underrepresented backgrounds, this can limit their potential and widen achievement gaps (Holmes et al., 2022; Holstein & Doroudi, 2021). Conversely, more advantaged groups might disproportionately benefit from AI systems trained on data that overrepresents their experiences (Holmes et al., 2022; Holstein & Doroudi, 2021; Porayska-Pomsta et al., 2023). Therefore, addressing bias and ensuring fairness in AIEd systems is not merely a technical challenge but a fundamental ethical imperative to ensure that these technologies serve to promote equity and inclusivity rather than becoming tools for perpetuating educational disparities. Achieving this requires a multifaceted approach encompassing careful curation of diverse and representative datasets, the development and application of algorithmic fairness techniques, continuous auditing and monitoring for bias, and the integration of ethical considerations throughout the entire lifecycle of AI system design and deployment (Barnes & Hutson, 2024; Chinta et al., 2024; Eaton, 2024; Holmes et al., 2022; Kizilcec & Lee, 2022; Lata, 2024; Leta & Vancea, 2023; Nguyen et al., 2023; Roshanaei et al., 2023; Sato et al., 2024; Syaidina et al., 2024; Yan & Liu, 2024).

Transparency and Accountability

A critical ethical dimension of artificial intelligence in education (AIEd) centres on the intertwined principles of transparency and accountability, particularly considering the inherent complexities often associated with AI decision-making processes (Agarwal et al., 2024; Barnes & Hutson, 2024; Donatus et al., 2024; Khan, 2024; Leta & Vancea, 2023; Naseeb & Bhatti, 2024; Nguyen et al., 2023; Syaidina et al., 2024). The challenge of "black box" AI arises from the sophisticated nature of many AI models, especially those employing machine learning and deep learning algorithms, which operate through intricate hyperspaces and decision-making pathways that are often opaque and difficult for humans to interpret (Agarwal et al., 2024; Akgun & Greenhow, 2022; Schiff, 2022;

Stephenson et al., 2022). This lack of inherent transparency means that the reasoning behind AI-generated recommendations, assessments, or decisions can remain obscure, making it challenging to understand how specific outcomes are achieved or why particular conclusions are reached (Bogina et al., 2022; Kizilcec & Lee, 2022; Raza, 2024; Remian, 2019). The stochastic nature of the training phase and the multifaceted weights and attributes within AI models further obfuscate the retrieval and comprehension of their underlying parameters and rationales (Manna & Sett, 2025). Consequently, the opacity of these "black box" systems poses significant hurdles for establishing both transparency and accountability in their educational applications (Agarwal et al., 2024; Donatus et al., 2024; Raza, 2024; Remian, 2019).

The lack of explainability in AI models significantly affects the trust that both students and educators place in these systems (Chinta et al., 2024; Kizilcec & Lee, 2022; Khan, 2024; Raza, 2024; Remian, 2019; Sporrong et al., 2024). If the logic behind an AI-driven assessment or recommendation remains hidden, stakeholders may struggle to understand its validity or fairness (Chinta et al., 2024; Raza, 2024). Indeed, as noted by Holmes et al. (2022) and Sporrong et al. (2024), educators might even prefer an AI system that offers greater transparency around its decision-making, even if it means sacrificing some accuracy, over a perfectly accurate but entirely opaque system. Nguyen et al. (2023) highlight that this lack of explainability can impede teachers' ability to effectively utilize AIEd tools and to promptly identify problems related to student behaviour and learning performance. For end-users lacking specialized expertise in AI, these systems can appear as inscrutable entities, fostering skepticism and apprehension towards their outputs (Manna & Sett, 2025). Kizilcec and Lee (2022) caution that an erosion of trust in algorithmic systems can lead decision-makers to disregard their predictions, potentially resulting in discriminatory actions that reinforce prevailing stereotypes when human biases come into play. Furthermore, as Khan (2024) and Nurhasanah and Nugraha (2024) suggest, students and parents may understandably distrust AI systems if they are not provided with clear insights into how these technologies arrive at conclusions that impact their educational experiences and outcomes.

Ethical concerns are particularly salient regarding AI-driven recommendations and automated grading systems in the absence of transparency (Chinta et al., 2024; Donatus et al., 2024; Raza, 2024; Sporrong et al., 2024). The "black box" nature of some AI algorithms can make it exceedingly difficult for students to understand and contest decisions made in educational evaluations, raising fundamental ethical questions about fairness and due process (Chinta et al., 2024). Sporrong et al. (2024) highlight that a lack of insight into how automated grading decisions are made by AI problematizes the accountability of teachers for the grades they assign. Educators may find themselves in a "quandary" (Sporrong et al., 2024), unable to adequately explain or justify AI-generated grades to students who seek clarification or wish to appeal. Concerns also exist

that the use of AI in assessment might inadvertently encourage "algorithm-pleasing behaviours" in students, potentially undermining genuine learning and focusing instead on strategies to game the AI system (Sporrong et al., 2024). Schiff (2022) points out that institutions employing algorithms in admissions may face ethical challenges if they cannot transparently explain why certain students were admitted or denied. Similarly, parents may lack transparent explanations from school districts regarding pedagogical or remediation decisions made by opaque AI systems (Schiff, 2022). Ultimately, ensuring the ethical use of AI in educational recommendations and assessments necessitates a move towards greater transparency and explainability to foster trust, enable accountability, and safeguard against potential bias and unfairness (Agarwal et al., 2024; Eaton, 2024; Khan, 2024; Raza, 2024; Remian, 2019).

Equity and Accessibility Concerns

A fundamental ethical challenge in the integration of artificial intelligence (AI) into education revolves around the critical issues of equity and accessibility (Adams et al., 2021; Akgun & Greenhow, 2022; Donatus et al., 2024; Eden et al., 2024; Ifenthaler et al., 2024). While AI holds the potential to enhance learning experiences and personalize instruction, significant concerns exist regarding the equitable distribution of its benefits and the potential for exacerbating existing disparities (Porayska-Pomsta et al., 2023; Roshanaei et al., 2023).

The **digital divide** stands as a primary obstacle to achieving equitable access to AI-powered educational tools and resources (Ağca, 2023; Donatus et al., 2024; World Economic Forum, 2024). This divide refers to the significant gap between individuals, households, and geographic areas with differing socioeconomic levels in terms of both their opportunities to access information and communication technologies (ICT) and their effective use of the internet (Porayska-Pomsta et al., 2023). In the context of AI in education, this manifests as unequal access to the necessary hardware (computers, tablets), reliable internet connectivity, and the digital literacy skills required to effectively engage with AI-powered learning platforms and resources (Ağca, 2023; Donatus et al., 2024; Eden et al., 2024; Lata, 2024; Nurhasanah & Nugraha, 2024; Roshanaei et al., 2023; Syaidina et al., 2024). Students from low-income families, those with disabilities, and individuals in underserved communities, particularly in less developed countries, often lack the fundamental technological infrastructure and digital competencies to benefit from AI in education (Ağca, 2023; Donatus et al., 2024; Eden et al., 2024; Khan, 2024; Lata, 2024; Nurhasanah & Nugraha, 2024; Roshanaei et al., 2023; Syaidina et al., 2024; World Economic Forum, 2024). This uneven accessibility can lead to a new kind of digital divide, one based on the use of data-based knowledge to inform intelligent decision-making, further disadvantaging marginalized populations

(Ağca, 2023; Donatus et al., 2024). As AI-driven education becomes more prevalent, those without access risk falling further behind, exacerbating existing achievement gaps and creating new technological, economic, and social divides (Ağca, 2023; Donatus et al., 2024; Lata, 2024).

Consequently, there are significant risks of AI amplifying existing social and educational inequalities if these accessibility issues are not addressed proactively (Donatus et al., 2024; Porayska-Pomsta et al., 2023; Syaidina et al., 2024; World Economic Forum, 2024). Well-resourced schools and students are more likely to have the necessary infrastructure and digital literacy to fully leverage the advantages of AI-powered tools, potentially widening the gap between them and under-resourced counterparts (Naseeb & Bhatti, 2024; Roshanaei et al., 2023; Syaidina et al., 2024). Moreover, as discussed previously, biases embedded in AI algorithms can disproportionately affect students from marginalized groups, leading to unfair assessments, less effective personalized learning recommendations, and the perpetuation of negative stereotypes (Akgun & Greenhow, 2022; Donatus et al., 2024; Holstein & Doroudi, 2021). For instance, AI systems trained primarily on data from Western, educated, industrialized, wealthy, and democratic nations might not adequately consider cross-cultural variations or the diverse physical characteristics of learners from other backgrounds, potentially leading to misidentifications or culturally insensitive educational experiences (Mouta et al., 2024). The promise of personalized learning can be undermined if access to the technologies enabling it is unequal, resulting in a scenario where only privileged students benefit from tailored educational experiences, while others are left with generic or less effective resources (Donatus et al., 2024; Holstein & Doroudi, 2021; Roshanaei et al., 2023).

Designing **inclusive AI-driven education systems** necessitates careful consideration of several key factors to mitigate these equity and accessibility concerns (Barnes & Hutson, 2024; Donatus et al., 2024; Yan & Liu, 2024). Firstly, accessibility must be a central tenet of AI tool development, ensuring that these technologies are usable by students with diverse learning needs and abilities (Donatus et al., 2024; Eaton, 2024; Lata, 2024; Roshanaei et al., 2023; Sato et al., 2024; Syaidina et al., 2024). This includes incorporating features compatible with assistive technologies, providing multiple means of engagement, representation, and action and expression, and adhering to universal design principles (Eaton, 2024; Lata, 2024; Sato et al., 2024). Secondly, addressing algorithmic bias is crucial for ensuring fairness and equity. This requires employing diverse and representative training datasets, implementing bias-aware machine learning algorithms, conducting regular audits for bias, and establishing mechanisms for identifying and rectifying discriminatory outcomes (Barnes & Hutson, 2024; Chinta et al., 2024; Eaton, 2024; Holmes et al., 2022). Furthermore, AI-powered educational resources should be culturally relevant and sensitive to the diverse backgrounds of students (Lata, 2024; Mouta et al., 2024; Roshanaei et al., 2023). Thirdly, bridging the

digital divide requires a multifaceted approach involving investments in technological infrastructure, promoting digital literacy skills, and developing affordable and accessible AI solutions (Ağca, 2023; Donatus et al., 2024; Lata, 2024; Nurhasanah & Nugraha, 2024; Roshanaei et al., 2023; Syaidina et al., 2024; World Economic Forum, 2024). Collaboration between policymakers, educators, technology developers, and communities is essential to create sustainable and equitable conditions for digital rights and internet access (Ağca, 2023; Donatus et al., 2024; Lata, 2024; Roshanaei et al., 2023; Syaidina et al., 2024). Finally, ensuring fair access to AI resources for students from diverse backgrounds and perspectives must be a guiding principle in the design and development of AI in education (Ağca, 2023; Lata, 2024; Roshanaei et al., 2023; Syaidina et al., 2024; Yan & Liu, 2024). Ultimately, stakeholders can work towards harnessing the transformative potential of AI in education in a manner that promotes equity and inclusivity for all learners, rather than exacerbating existing inequalities.

Autonomy and the Role of Human Educators

The integration of artificial intelligence into education engenders significant ethical considerations pertaining to autonomy and the evolving role of human educators (Ağca, 2023; Barnes & Hutson, 2024). Navigating the relationship between AI-powered assistance and the indispensable judgment of human educators, addressing the ethical implications of AI potentially supplanting traditional teaching methodologies, and ensuring the preservation of teacher and student agency within AI-supported learning contexts represent critical aspects of this ethical terrain (Ağca, 2023; Barnes & Hutson, 2024; Donatus et al., 2024; Holmes et al., 2022; Ifenthaler et al., 2024;; Porayska-Pomsta et al., 2023).

Balancing **AI assistance with human judgment** is paramount in maintaining an ethical and effective educational environment (Donatus et al., 2024; Ifenthaler et al., 2024;). While AI offers considerable potential in automating administrative tasks, personalizing learning experiences, and providing data-driven insights, it cannot fully replicate the nuanced understanding, emotional intelligence, and ethical considerations inherent in human pedagogical expertise (Ağca, 2023; Agarwal et al., 2024; Akgun & Greenhow, 2022; Donatus et al., 2024; Habibulloh, 2025; Ifenthaler et al., 2024; Khan, 2024; Lata, 2024). Agarwal et al. (2024) highlight that humans should ultimately be responsible for educational outcomes and thus must retain appropriate oversight of AI systems. The emotional intelligence, mentoring capabilities, and in-depth understanding of human behaviour that educators possess cannot be fully matched by AI, even though AI technology can offer immediate feedback (Ağca, 2023). Therefore, AI should be viewed as a supplementary tool that augments and enhances the skills of educators, providing them with data-driven insights and personalized tools rather than a replacement for their fundamental role. The development and integration of AI-enabled tools must be a participatory process, designed to support educators' needs rather than the perceived

needs of technologists. Teachers will ultimately decide how and when AI-enabled tools are appropriately used.

The **ethical question of AI replacing traditional teaching methods** raises significant concerns about the very nature and purpose of education (Porayska-Pomsta et al., 2023; Sporrong et al., 2024). While AI can automate repetitive tasks and provide personalized learning pathways, it risks neglecting the crucial creative and socio-emotional aspects of teaching that extend beyond mere knowledge transmission (Habibulloh, 2025; Khan, 2024; Lata, 2024; Leta & Vancea, 2023; Naseeb & Bhatti, 2024). Human educators play an irreplaceable role in fostering critical thinking, emotional intelligence, ethical awareness, and holistic development (Barnes & Hutson, 2024; Donatus et al., 2024; Habibulloh, 2025) and the direct interaction between teachers and students is essential for social skill development and the cultivation of humanistic values in education (Habibulloh, 2025). While AI can assist in delivering personalized instruction, it cannot replace the emotional and social development, mentorship, and meaningful human connections that educators provide (Khan, 2024). The focus should be on augmenting teacher roles through automation of routine tasks, empowering them to concentrate more on personalized instruction and mentorship, rather than viewing AI as a substitute for educators (Ifenthaler et al., 2024; Khan, 2024).

In conclusion, upholding ethical considerations surrounding autonomy and the role of human educators necessitates a balanced approach that leverages the benefits of AI while preserving the indispensable contributions of human judgment, resisting the notion of complete replacement of traditional teaching methods, and actively fostering teacher and student agency within AI-supported learning environments. A human-centred perspective must guide the design, development, and implementation of AI in education to ensure that technological advancements align with ethical principles and empower both educators and learners.

3. Ensuring Data Privacy and Security in AI-Driven Educational Tools and Platforms

3.1 The Importance of Data Privacy in Education

The integration of artificial intelligence (AI) into educational tools and platforms has ushered in an era where vast quantities of student data are routinely collected and processed (Adams et al., 2021; Ağca, 2023; Chinta et al., 2024; Donatus et al., 2024; Mouawad, 2020; Naseeb & Bhatti, 2024; Nguyen et al., 2023; Ward, 2023; Yan & Liu, 2024). This datafication of education, driven by the capabilities of learning analytics and educational data mining techniques, allows for the tracking of learner behaviour and

performance (Adams et al., 2021). Learning platforms readily facilitate the collection of data, which can then inform teaching practices and personalize learning experiences (Adams et al., 2021; Ağca, 2023; Chinta et al., 2024; Eden et al., 2024; Khan, 2024; Lata, 2024; Naseeb & Bhatti, 2024; Nurhasanah & Nugraha, 2024; Roshanaei et al., 2023; Spulber, 2024). AI algorithms often rely on these extensive datasets to deliver personalized learning experiences, predict student performance, automate administrative processes, and provide data-driven insights (Khan, 2024; Naseeb & Bhatti, 2024; Nurhasanah & Nugraha, 2024; Roshanaei et al., 2023; Syaidina et al., 2024). The types of data collected can be wide-ranging, including academic records, learning habits, performance data, behavioural patterns, biometric information, and even inferred emotional states (Habibulloh, 2025; Holmes et al., 2022; Khan, 2024; Naseeb & Bhatti, 2024; Nguyen et al., 2023; Nurhasanah & Nugraha, 2024; Syaidina et al., 2024). This information is then processed and analyzed to tailor learning materials, provide guidance, offer real-time feedback, and assess student performance (Ağca, 2023; Chinta et al., 2024; Donatus et al., 2024; Eden et al., 2024; Gouseti et al., 2024; Khan, 2024; Naseeb & Bhatti, 2024; Nurhasanah & Nugraha, 2024; Roshanaei et al., 2023; Spulber, 2024).

The extensive collection, storage, and use of student data by AI in education raise significant ethical concerns (Syaidina et al., 2024; Yan & Liu, 2024). A primary concern revolves around the protection of personal information and the potential for privacy breaches (Donatus et al., 2024; Eden et al., 2024; Yan & Liu, 2024). The very act of gathering and analyzing student data for tailored learning can compromise privacy if not handled with utmost care (Ağca, 2023). Ethical theories of privacy emphasize the importance of safeguarding individuals' autonomy and personal information (Donatus et al., 2024). Questions arise regarding data ownership, access, security, and the potential for misuse beyond instructional purposes (Ağca, 2023; Donatus et al., 2024). Concerns also exist about the lack of awareness among students and parents regarding how their data is being utilized (Khan, 2024; Mouawad, 2020; Naseeb & Bhatti, 2024; Syaidina et al., 2024), highlighting the need for informed consent and transparency in AI-driven educational systems (Khan, 2024; Syaidina et al., 2024). Striking a balance between data granularity for successful AI insights and minimizing data exposure is a significant challenge (Ağca, 2023). Furthermore, the potential for algorithmic discrimination and surveillance through AI-powered proctoring systems creates a tension between academic integrity and student privacy (Donatus et al., 2024; Khan, 2024; Naseeb & Bhatti, 2024). The continuous monitoring of student interactions can lead to feelings of surveillance, potentially causing stress and anxiety and reducing trust in educational institutions (Khan, 2024; Naseeb & Bhatti, 2024).

Past data privacy breaches in the EdTech sector serve as stark reminders of the vulnerabilities associated with the collection and storage of sensitive student information

(Khan, 2024; Naseeb & Bhatti, 2024; Roshanaei et al., 2023; Syaidina et al., 2024). There is an increasing risk of such breaches due to the increasing volume of student data handled by AI systems (Adams et al., 2021; Ağca, 2023; Donatus et al., 2024). The potential consequences of data breaches in education are profound, including identity theft, academic fraud, and the exposure of sensitive personal information (Eden et al., 2024). The increased risk of data breaches with the adoption of AI technologies has been noted in empirical studies, emphasizing the need for robust data protection measures (Roshanaei et al., 2023; Syaidina et al., 2024). Educational institutions may lack the necessary resources or knowledge to properly secure this sensitive data, further exacerbating the risk (Naseeb & Bhatti, 2024; Roshanaei et al., 2023; Syaidina et al., 2024). Therefore, ensuring data security and establishing transparent data usage regulations are critical to preventing misuse and maintaining the trust of students and parents (Khan, 2024; Mouawad, 2020; Naseeb & Bhatti, 2024; Nguyen et al., 2023; Yan & Liu, 2024).

3.2 Key Data Privacy and Security Challenges

One key area of concern involves the **risks associated with student data protection** due to the large-scale collection and analysis of student information. The advent of learning platforms has made it easier to track learner behaviour and performance, generating vast amounts of data. This data, which can include academic history, learning habits, behavioural patterns, biometric information – for instance, through facial recognition for attendance (Pattnaik & Mohanty, 2020) – and even insights into learning difficulties, is crucial for AI systems to provide tailored learning experiences (Admane et al., 2024). However, the sheer volume and sensitive nature of this data create significant ethical challenges regarding how it is handled responsibly (Pammer-Schindler & Rosé, 2022). Concerns arise about data ownership (Rainie et al., 2019), access, consent (Mouawad, 2020), and the potential for misuse, such as targeted advertising or other non-educational purposes. Striking a balance between the granularity of data needed for effective AI-driven insights and minimizing data exposure to protect student privacy is a significant challenge.

Furthermore, **privacy risks are intrinsically linked to the use of AI-driven learning analytics and monitoring tools**. AI tracking systems, through algorithms and machine-learning models, not only monitor student activities but can also determine their future preferences and actions. Surveillance mechanisms can be embedded in AI's predictive systems to foresee students' learning performances, strengths, weaknesses, and learning patterns (Ifenthaler et al., 2024). While monitoring might be seen as part of a teacher's responsibility, such actions can also be perceived as problematic surveillance systems that threaten students' privacy (Mouawad, 2020). The use of facial recognition software

(Pattnaik & Mohanty, 2020) and voice-activated platforms that capture a wide range of data without explicit consent also raises serious concerns. Continuous monitoring through AI can lead to students feeling constantly watched, potentially increasing stress and anxiety, and reducing trust in educational institutions. The ethical dilemma lies in balancing academic integrity and the benefits of data-driven insights with the fundamental right to privacy (Mitta, 2023, as cited in Gouseti et al., 2024).

Navigating the landscape of **legal and regulatory frameworks** is crucial for addressing data privacy in AI-powered education. Several major data privacy laws have a significant impact on the use of AI in educational settings (Mouawad, 2020). In the United States, the Family Educational Rights and Privacy Act (FERPA) of 1974 requires federally funded institutions to obtain parental or student consent before disclosing personal information (Mouawad, 2020). The Children's Online Privacy Protection Act of 1998 (COPPA) mandates that web hosts and content providers seek parental consent to store data about children under 13 (Mouawad, 2020). Additionally, the Student Digital Privacy and Parental Rights Act of 2015 prohibits operators from selling personal information to third parties or collecting student information for purposes unrelated to educational activities (Mouawad, 2020). In Europe, the General Data Protection Regulation (GDPR) stands as a rigorous data protection law, emphasizing principles like transparency, user consent, and the right to be forgotten (Mouawad, 2020). These regulations provide a legal underpinning for data protection in educational contexts.

Educational institutions face considerable **compliance challenges** in adhering to these diverse and often complex legal frameworks. Ensuring that AI systems and data handling practices align with the specific requirements of laws like FERPA, COPPA, and GDPR can be difficult. Obtaining informed consent, especially when dealing with children, is critical yet challenging to implement effectively in the context of AI-driven data collection. Educational institutions are entrusted with drafting comprehensive privacy policies that describe data handling methods and offer recourse processes in the event of a breach. The ambiguity surrounding data usage and the extent of surveillance possible through AI tools further complicates compliance efforts. The complexity of regulations like GDPR can make their application in the educational field particularly challenging (Mouawad, 2020).

Data security risks are a significant concern in AI-powered education, particularly regarding cybersecurity vulnerabilities in AI-based learning management systems (LMS). The proliferation of AI-powered educational platforms and learning analytics tools increases the risk of data breaches, unauthorized access, and misuse of student data by malicious actors (Pammer-Schindler & Rosé, 2022). Sensitive student information stored in these systems can be a target for cyberattacks, potentially leading to identity theft, academic fraud, and the exposure of private details. Educational institutions must implement robust data protection measures, such as encryption, access controls, and

regular security audits, to safeguard student data against unauthorized access and cyberattacks. The risk of student data being hacked, misused, or even sold to third parties underscores the need for stringent data security protocols and transparent data governance policies.

Finally, **ethical concerns regarding AI-driven student surveillance and monitoring** are substantial. While AI can be used to monitor student progress and behaviour to identify potential issues, the intrusiveness of such surveillance raises ethical questions about autonomy, confidentiality, and digital rights (Mouawad, 2020). Constant monitoring can create an environment where students feel unsafe to take ownership of their ideas and may limit their participation in learning events. The use of AI to predict students' future preferences and actions based on collected data also raises concerns about manipulation and the erosion of independent thought. Ensuring transparency and obtaining informed consent are crucial to address these ethical concerns and maintain trust between students, educators, and AI systems. Educational institutions must carefully consider the ethical implications of deploying AI surveillance tools and strive to strike a balance between maintaining a safe and productive learning environment and respecting students' fundamental rights to privacy and autonomy.

3.3 Strategies for Ethical AI and Data Protection

Ethical AI Development for Education

Developing AI for educational purposes demands a commitment to ethical principles throughout the design and implementation process. This includes proactive measures to mitigate bias in AI algorithms and a strong reliance on established ethical guidelines and principles.

One critical aspect of ethical AI development is addressing and reducing algorithmic bias (Baker & Hawn, 2022; Chinta et al., 2024). AI algorithms learn from the data they are trained on, and if this data reflects existing societal biases, the AI system can perpetuate or even amplify these inequalities. Several approaches can be employed to mitigate AI bias in education. The use of diverse training data that accurately represents the student population across various socio-economic, cultural, and demographic factors is essential (Baker & Hawn, 2022). Over-reliance on data from specific groups can lead to AI systems that do not fairly serve all learners. Furthermore, regular bias audits should be conducted to identify and rectify any discriminatory patterns within the AI algorithms. This involves systematically evaluating the AI's outputs for different student groups to ensure equitable outcomes. The US Department of Education (2023, as cited in Baker & Hawn, 2022) recommended reviewing the quality of foundational data in AI

models to ensure accurate and contextually appropriate information is used for fair and unbiased pattern recognition. Kizilcec and Lee (2020) highlight the importance of considering measurement (data input), model learning (algorithm), and action (output) to mitigate bias.

The development and adherence to ethical AI guidelines and principles play a crucial role in ensuring responsible innovation in education. Various organisations and researchers have proposed frameworks to guide the ethical use of AI in education. These guidelines often emphasize principles such as fairness, transparency, accountability, privacy, and inclusivity (Chinta et al., 2024; Richardson & Gilbert, 2021). Agarwal et al. (2024) identified several ethical norms for AI in education, including the need for AI systems to promote equity and avoid discrimination. They also highlight the importance of education about algorithmic bias and the need for algorithmic transparency. A framework for ethical AI development should prioritize human well-being, equity, transparency, accountability, fairness, bias mitigation, privacy protection, and collaborative development (Lata, 2024). The "Ethical Framework for AI in Education" by The Institute for Ethical AI in Education (2021, as cited in Agarwal et al., 2024) provides a comprehensive tool to monitor AI technologies throughout their adoption in education, addressing ethical design, privacy, equity, transparency, and accountability. The United Nations Educational, Scientific and Cultural Organization (UNESCO) has also adopted a "Recommendation on the Ethics of AI" to ensure AI aligns with human rights and values, including considerations for data governance, privacy, and reducing biases (Gouseti et al., 2024).

Best Practices for Data Privacy and Security

Protecting the privacy and security of student data is fundamental to the ethical integration of AI in education. Educational institutions and technology providers must implement the best practices to ensure responsible data handling and compliance with privacy regulations.

Transparency and informed consent are essential elements of data privacy in AI-driven education (Bogina et al., 2022). Students and their parents or guardians should be clearly informed about what data is being collected, how it is being used by AI systems, and who has access to it. Educational institutions should practice transparency about student data privacy to address any misconceptions and concerns related to data use (Agarwal et al., 2024). Obtaining informed consent before collecting and using student data is crucial for respecting user autonomy and privacy rights. This includes providing clear and understandable information about the purposes of data collection and the potential risks involved.

Secure data storage and encryption methods are vital for safeguarding student data from unauthorized access and breaches. Robust security measures, including encryption both in transit and at rest, should be implemented to protect sensitive personal information. Cloud-based data storage requires clear terms for privacy, security, and trust (Mouta et al., 2024). Implementing privacy-preserving AI techniques, such as federated learning and differential privacy, can help mitigate risks associated with data exposure (Khan, 2024, as cited in Afzal et al., 2023). Federated learning allows AI models to be trained across decentralized data sources without transferring raw data, while differential privacy ensures individual data points remain indistinguishable within large datasets, protecting student identities. Techniques like k-anonymity and l-diversity can also be used to protect privacy by generalizing or suppressing identifying information, although this may come at the cost of data fidelity (Adams et al., 2021). Furthermore, generating synthetic data based on master datasets, as proposed by Adams et al. (2021), can offer an alternative unrestricted source of data without compromising student privacy.

Establishing comprehensive AI governance policies is crucial for ensuring ethical data handling within educational institutions. These policies should define clear guidelines and protocols for data collection, storage, processing, and disposal. They should also outline the responsibilities of different stakeholders, including educators, administrators, and technology providers, in maintaining data privacy and security. Regular review and updating of these policies are necessary to adapt to evolving AI technologies and privacy regulations. Educational leaders can choose to require stronger data protections or discontinue harmful contracts with EdTech products that exploit user data (Agarwal et al., 2024).

Balancing AI Innovation with Ethical and Privacy Considerations

Achieving a balance between the innovative potential of AI in education and the critical need for ethical and privacy safeguards requires a collaborative effort involving policymakers, educators, and EdTech developers.

Policymakers play a vital role in establishing regulatory frameworks that promote responsible AI usage while safeguarding student rights. Legislation and regulations should address issues such as data privacy, algorithmic bias, and transparency in AI decision-making. The EU's General Data Protection Regulation (GDPR) serves as a significant example of stringent data protection law (Leta & Vancea, 2023). Policymakers must also support open science and research on AI in education to build a robust body of knowledge and evidence (Ifenthaler et al., 2024).

Educators are essential in implementing ethical AI practices in the classroom and fostering ethical AI literacy among students. Training teachers and staff on the

capabilities, uses, and risks of AI is paramount to ensure responsible usage. Educators should engage in ongoing professional development to build expertise in effective and ethical AI use (Lata, 2024). Furthermore, incorporating ethical AI literacy into educational curricula equips students with the knowledge required to navigate AI-driven environments responsibly (Khan, 2024, as cited in Lata, 2024).

EdTech developers have a responsibility to prioritize ethical considerations during the design and development of AI systems for education. Embedding transparency, fairness, and accountability into AI models is crucial. Strategies for designing AI systems that prioritize student rights and privacy include implementing privacy by design principles and incorporating safeguards against bias and discrimination. Collaborative innovation between educators, researchers, and technology developers is essential to share best practices and develop effective and ethical AI solutions (Lata, 2024). It is important to note, as highlighted by Barocas and Selbst (2016), that the use of seemingly neutral data in algorithms can still lead to disparate impacts on certain groups, necessitating careful consideration of proxy variables and potential for discrimination.

Future directions for responsible AI integration in education should focus on a human-centred approach that places ethical considerations at the forefront. Continuous ethical evaluation of AI systems is necessary to prevent them from becoming tools of surveillance and discrimination. Research should explore human-AI collaboration in education, ensuring AI tools act as supportive agents rather than autonomous decision-makers (Khan, 2024, as cited in Lata, 2024). Developing comprehensive fairness frameworks and promoting global cooperation are crucial for ensuring a responsible and fair integration of AI technologies in educational environments (Chinta et al., 2024). The focus should be on supporting children to develop a critical understanding of how AI works, including its potential to replicate and exacerbate existing biases (Gouseti et al., 2024). Ultimately, balancing AI innovation with ethical responsibility is crucial to ensure that AI-driven educational technologies benefit all stakeholders while minimizing potential risks (Khan, 2024, as cited in Lata, 2024). The need for explainable AI (xAI) in education, as discussed by Manna and Sett (2025), is critical for building trust and understanding in AI-driven decisions, particularly in high-stakes areas. Continuous scrutiny and adaptation of ethical frameworks are essential to navigate the evolving landscape of AI in education responsibly (Agarwal et al., 2024).

4. Conclusion

The integration of artificial intelligence into education presents a transformative frontier, offering considerable potential to reshape pedagogical approaches, personalize learning experiences, and streamline administrative tasks. However, as this chapter has explored in detail, this technological evolution is inextricably linked with significant ethical and

data privacy concerns that demand careful consideration. The reliance of AI systems on the extensive collection, processing, and analysis of student data raises profound questions regarding privacy, security, and the potential for misuse of sensitive information. Furthermore, the inherent risks of algorithmic bias can perpetuate or even amplify existing educational inequalities, undermining the fundamental principles of fairness and equity. Beyond these core issues, concerns surrounding transparency, accountability, the impact on student autonomy and the teacher-student relationship, and the potential for surveillance necessitate a comprehensive and ethically informed approach to AI adoption in educational settings.

The imperative for a balanced approach between the undeniable innovation offered by AI and the critical need to uphold ethical principles and safeguard data privacy cannot be overstated. While AI promises enhanced personalisation, automated assessments, and data-driven insights (Akgun & Greenhow, 2022), these benefits must not come at the expense of fundamental rights and values (Khan, 2024). As Akgun and Greenhow (2022) highlight, the ethical challenges of AI in education are often not fully considered, underscoring the need for proactive identification and addressal of these issues. The development and deployment of AI in education must be guided by robust ethical frameworks that prioritise human well-being, equity, transparency, accountability, fairness, bias mitigation, privacy protection, and collaborative development (Lata, 2024; Nguyen et al., 2023). As noted by Holmes et al. (2021), there is a need to differentiate between doing ethical things and doing things ethically, emphasizing the importance of pedagogical choices that are inherently ethical. Ignoring these considerations risks the erosion of trust, the perpetuation of inequalities, and the potential for AI to become a tool of surveillance and discrimination rather than an enabler of inclusive and equitable education.

Therefore, a clear call to action is essential to ensure the responsible and ethical integration of AI in education, underpinned by stronger data privacy safeguards. Policymakers must establish comprehensive regulatory frameworks that address data privacy, algorithmic bias, and transparency, drawing inspiration from existing models such as the EU's GDPR (Leta & Vancea, 2023). Educational institutions must adopt and enforce clear ethical guidelines and data privacy protocols, prioritising transparency and informed consent regarding the collection and use of student data (Agarwal et al., 2024; Khan, 2024, as cited in Afzal et al., 2023). Educators require comprehensive training to develop ethical AI literacy, enabling them to use AI responsibly and to equip students with the knowledge to navigate AI-driven environments critically (Khan, 2024, as cited in Lata, 2024; Naseeb & Bhatti, 2024). EdTech developers bear the responsibility to embed ethical considerations, privacy by design principles, and safeguards against bias into the very fabric of AI educational tools (Donatus et al., 2024; Lata, 2024; Roshanaei et al., 2023). Continuous ethical evaluation, interdisciplinary collaboration among

stakeholders, and a human-centred approach are crucial to ensure that AI serves as a supportive agent in education, enhancing learning without compromising ethical values and fundamental rights. A future of AI in education that prioritises ethics and data privacy will allow the harnessing of AI's transformative potential to create more equitable, inclusive, and effective learning environments for all students

References

Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2021). Artificial Intelligence Ethics Guidelines for K-12 Education: A Review of the Global Landscape. In I. Roll, D. McNamara, S. Sosnovsky, R. Luckin, & V. Dimitrova (Eds.), *Artificial Intelligence in Education* (Vol. 12749, pp. 24–28). Springer International Publishing. https://doi.org/10.1007/978-3-030-78270-2_4

Afzal, M. U., Abdellatif, A. A., Zubair, M., Mehmood, M. Q., & Massoud, Y. (2023). Privacy and Security in Distributed Learning: A Review of Challenges, Solutions, and Open Research Issues. *IEEE Access*, *11*, 114562–114581. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3323932

Agarwal, B., Urlings, C., van Lankveld, G., & Klemke, R. (2024). *Identifying the Ethical Values and Norms of Artificial Intelligence in Education: A Systematic Literature Review*. https://files.de-1.osf.io/v1/resources/e7t3f/providers/osfstorage/6720f920e0a27b40ee629611?format=pdf&action=download&direct&version=1

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, *32*(4), 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bansal, C., Pandey, K. K., Goel, R., Sharma, A., & Jangirala, S. (2023). Artificial intelligence (AI) bias impacts: Classification framework for effective mitigation. *Issues in Information Systems*, *24*(4), 367–389.

Barnes, E., & Hutson, J. (2024). Navigating the ethical terrain of AI in higher education: Strategies for mitigating bias and promoting fairness. *Forum for Education Studies*, *2*(2). https://digitalcommons.lindenwood.edu/faculty-research-papers/648/?&target=_blank

Bogina, V., Hartman, A., Kuflik, T., & Shulner-Tal, A. (2022). Educating Software and AI Stakeholders About Algorithmic Fairness, Accountability, Transparency and Ethics. *International Journal of Artificial Intelligence in Education*, *32*(3), 808–833. https://doi.org/10.1007/s40593-021-00248-0

Chen, Y. (2020). IoT, cloud, big data and AI in interdisciplinary domains. *Simulation Modelling Practice and Theory*, *102*, 102070. https://doi.org/10.1016/j.simpat.2020.102070

Chinta, S. V., Wang, Z., Yin, Z., Hoang, N., Gonzalez, M., Quy, T. L., & Zhang, W. (2024). *FairAIED: Navigating Fairness, Bias, and Ethics in Educational AI Applications* (No. arXiv:2407.18745). arXiv. https://doi.org/10.48550/arXiv.2407.18745

Donatus, U. O., Obinna, V. O., Samuel, U. O., Odera, C. U., & Nkechi, F. O. (2024). The Ethical Implications of Artificial Intelligence in Education. *AJAP-AMAMIHE Journal of Applied Philosophy*, *22*(2). https://www.igwebuikeresearchinstitute.org/amamihe_journal_article.php?paper=339

Eaton, S. E. (2024). *Assessing for Integrity in the Age of AI*. https://prism.ucalgary.ca/items/f3d41824-8a4a-4adc-800e-7d63acdfba6f

Eden, C. A., Chisom, O. N., & Adeniyi, I. S. (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews*, *10*(2), 006–013.

Gouseti, A., James, F., Fallin, L., & Burden, K. (2024). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education*, 1–22. https://doi.org/10.1080/1475939X.2024.2428601

Habibulloh, M. (2025). Ethics of Artificial Intelligence Usage in Education: A Qualitative Study on Teachers' and Students' Perceptions. *International Journal of Interdisciplinary Research*, *1*(1), 1–11.

Holmes, W., & Porayska-Pomsta, K. (2023). The ethics of artificial intelligence in education. *Lontoo: Routledge*. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780429329067&type=googlepdf

Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in Education: Towards a Community-Wide Framework. *International Journal of Artificial Intelligence in Education*, *32*(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1

Holstein, K., & Doroudi, S. (2021). *Equity and Artificial Intelligence in Education: Will "AIEd" Amplify or Alleviate Inequities in Education?* (Version 1). arXiv. https://doi.org/10.48550/ARXIV.2104.12920

Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024). Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. *Technology, Knowledge and Learning*, *29*(4), 1693–1710. https://doi.org/10.1007/s10758-024-09747-0

Khan, W. N. (2024). Ethical Challenges of AI in Education: Balancing Innovation with Data Privacy. *Journal of AI in Education: Innovations, Opportunities, Challenges, and Future Directions*, *1*(1), 1–13.

Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In *The ethics of artificial intelligence in education* (pp. 174–202). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9780429329067-10/algorithmic-fairness-education-ren%C3%A9-kizilcec-hansol-lee

Lata, P. (2024). *Towards Equitable Learning: Exploring Artificial Intelligence in Inclusive Education*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4987156

Leta, F. M., & Vancea, D. P. C. (2023). Ethics in education: Exploring the ethical implications of artificial intelligence implementation. *Ovidius University Annals, Economic Sciences Series*, *23*(1), 413–421.

Manna, S., & Sett, N. (2025). *Need of AI in Modern Education: In the Eyes of Explainable AI (xAI)* (No. arXiv:2408.00025). arXiv. https://doi.org/10.48550/arXiv.2408.00025

Mouawad, G. P. (2020). *Students' Privacy in a Digital Age* (PhD Thesis, University of La Verne). https://search.proquest.com/openview/b8ab7b9b120a10971e3574a2a019c6f7/1?pq-origsite=gscholar&cbl=18750&diss=y

Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sánchez, E. M. (2024). Uncovering Blind Spots in Education Ethics: Insights from a Systematic Literature Review on Artificial Intelligence in Education. *International Journal of Artificial Intelligence in Education*, *34*(3), 1166–1205. https://doi.org/10.1007/s40593-023-00384-9

Naseeb, J., & Bhatti, N. (2024). Ethical Considerations of AI in Educational Curriculum. *Multidisciplinary Journal of Emerging Needs of Curriculum*, *1*(1), 30–38.

Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B.-P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, *28*(4), 4221–4241. https://doi.org/10.1007/s10639-022-11316-w

Nurhasanah, S., & Nugraha, M. S. (2024). The future of learning: Ethical and philosophical implications of artificial intelligence (AI) integration in education. *Indonesian Journal of Multidisciplinary Research*, *4*(2), 341–352.

Porayska-Pomsta, K., Holmes, W., & Nemorin, S. (2023). The ethics of AI in education. In *Handbook of artificial intelligence in education* (pp. 571–604). Edward Elgar Publishing. https://www.elgaronline.com/edcollchap/book/9781800375413/book-part-9781800375413-38.xml

Raza, H. (2024). AI-Driven Assessment: Reliability, Bias, and Ethical Implications. *Journal of AI in Education: Innovations, Opportunities, Challenges, and Future Directions*, *1*(2), 36–47.

Remian, D. (2019). *Augmenting education: Ethical considerations for incorporating artificial intelligence in education*. https://scholarworks.umb.edu/instruction_capstone/52/

Richardson, B., & Gilbert, J. E. (2021). *A Framework for Fairness: A Systematic Review of Existing Fair AI Solutions* (No. arXiv:2112.05700). arXiv. https://doi.org/10.48550/arXiv.2112.05700

Roshanaei, M., Olivares, H., & Lopez, R. R. (2023). Harnessing AI to Foster Equity in Education: Opportunities, Challenges, and Emerging Strategies. *Journal of Intelligent Learning Systems and Applications*, *15*(04), 123–143. https://doi.org/10.4236/jilsa.2023.154009

Sato, E., Shyyan, V., Chauhan, S., & Christensen, L. (2024). Putting AI in fair: A framework for equity in AI-driven learner models and inclusive assessments. *Journal of Measurement and Evaluation in Education and Psychology*, *15*(Special Issue), 263–281.

Schiff, D. (2022). Education for AI, not AI for Education: The Role of Education and Ethics in National AI Policy Strategies. *International Journal of Artificial Intelligence in Education*, *32*(3), 527–563. https://doi.org/10.1007/s40593-021-00270-2

Sporrong, E., McGrath, C., & Cerratto Pargman, T. (2024). Situating AI in assessment—An exploration of university teachers' valuing practices. *AI and Ethics*. https://doi.org/10.1007/s43681-024-00558-8

Spulber, D. (2024). AI in inclusive education which differences in research trend. *Geopolitical, Social Security and Freedom Journal*, *7*(1), 85–99. https://doi.org/10.2478/gssfj-2024-0007

Stephenson, B., & Harvey, A. (2022). Student equity in the age of AI-enabled assessment: Towards a politics of inclusion. In *Assessment for Inclusion in Higher Education* (pp. 120–130). Routledge.

https://library.oapen.org/bitstream/handle/20.500.12657/60125/1/9781000842722.pdf#page=135

Stephenson, B., Harvey, A., & Huang, Q. (2022). *Towards an inclusive analytics for Australian higher education*. Perth, Western Australia, National Centre for Student Equity in Higher …. https://www.ncsehe.edu.au/app/uploads/2022/03/Stephenson_LaTrobe_Final.pdf

Syaidina, M. O., Fahrudin, R., & Mutiara, I. A. (2024). Implementation of Ethics of Using Artificial Intelligence in the Education System in Indonesia. *Blockchain Frontier Technology*, *4*(1), 63–71.

Ward, L. S. (2023). Educational Technology Graduate Students' Attitudes Toward Online Privacy in Academic and Non-Academic Usage of Technologies: A Qualitative Study on Reactions and Recommended Actions. Ohio University. https://search.proquest.com/openview/d490ac9a1ec2dea761e34be16464c658/1?pq-origsite=gscholar&cbl=18750&diss=y

World Economic Forum. (2024). Shaping the Future of Learning: The Role of AI in Education (Insight Report). *World Economic Forum.*

Yan, Y., & Liu, H. (2024). Ethical framework for AI Education based on large Language Models. *Education and Information Technologies*. https://doi.org/10.1007/s10639-024-13241-6