WJAETS

World Journal of
Advanced
Engineering
Technology
and Sciences

World Journal Series
INDIA

(REVIEW ARTICLE)

Check for updates

# Ethical AI in schools: Balancing automation, privacy, and human oversight

Nafisat Zajime Bako *

*Data Systems and Reporting Analyst, Department of Graduate School, University of North Carolina, Wilmington, NC, USA.*

## Abstract

The incorporation of Artificial Intelligence (AI) technologies in educational settings offers significant advantages such as individualized instruction, automated grading, and enhanced administrative efficiency. Despite its advantages, it introduces ethical dilemmas related to student confidentiality, algorithmic impartiality, and the role of human oversight. This paper explores how to balance technological advancements with ethical safeguards in schools. It focuses on data privacy regulations, the potential for AI-induced biases, and the significance of maintaining human involvement in decision-making. The central argument is that ethical AI implementation requires a robust regulatory framework that merges technical innovation with human-centered policies. The paper concludes with strategies for responsible AI deployment, prioritizing openness, accountability, and inclusiveness.

**Keywords:** Ethical AI; Education Technology; Privacy; Algorithmic Bias; Human Oversight

## 1. Introduction

Artificial Intelligence (AI) involves machines emulating human intelligence, utilizing tools like machine learning and natural language understanding (Oseji, et al, 2021). AI has significantly transformed education by supporting individualized learning paths, adaptive assessments, and efficient administrative systems. AI-powered tutoring systems adjust content in real-time to each learner's unique requirements (Zawacki-Richter et al., 2019), and automated grading reduces educators' workload (Luckin et al., 2016).

One of the key contributions of AI in education is its role in fostering inclusive learning environments. AI tools, such as speech-to-text applications and automated translators, assist learners facing physical, cognitive, or linguistic challenges, making education more accessible (Holmes et al., 2021). AI-driven analytics help educators identify students who may be struggling, enabling timely interventions to improve learning outcomes. However, concerns regarding data privacy, ethical considerations, and the digital divide persist, highlighting the need for careful implementation of AI in educational settings (Selwyn, 2019).

Despite its advantages, AI in education is still continuously advancing, as researchers strive to improve its effectiveness and address ethical concerns. The integration of AI into education requires collaboration among educators, policymakers, and technology developers to ensure responsible use. Future advancements in AI can significantly reshape learning practices further by enabling fully adaptive learning experiences and intelligent assessment systems (Schiff, 2021). A balanced approach is necessary to harness AI's full capabilities while addressing challenges related to bias, transparency, and the moral considerations surrounding AI-driven decisions in education.

* Corresponding author: Nafisat Zajime Bako

## 1.1 Ethical AI in schools

Ethical Artificial Intelligence (AI) in schools refers to the responsible design, deployment, and regulation of AI systems to ensure fairness, transparency, and accountability in educational settings. Ethical AI seeks to align technology with human values, ensuring that AI-driven tools, such as automated grading systems, personalized learning platforms, and predictive analytics, do not perpetuate biases or compromise student privacy (Jobin, Ienca, & Vayena, 2019). This involves designing AI with clear ethical principles, including inclusivity, data security, and fairness, to prevent discrimination and enhance learning outcomes for all students (Floridi et al., 2018).

The scope of ethical AI in schools extends to various areas, including student assessment, administrative decision-making, and personalized learning experiences. AI can support teachers by automating administrative tasks and providing insights into student progress, however, ethical issues related to data protection and biased algorithms have emerged (Cowls & Floridi, 2018). AI-based grading systems must be transparent and regularly audited to prevent unfair evaluation, especially for marginalized students who may not fit predefined learning patterns. Additionally, ethical AI should prioritize informed consent, ensuring that students, parents, and educators understand how AI tools collect and use personal data (Aiken, 2019).

To ensure ethical AI integration in schools, policymakers and educators must establish guidelines that promote responsible AI use while mitigating risks. This includes implementing regulatory frameworks that oversee AI deployment, promoting AI literacy among educators and students, and fostering collaborations between educational institutions and AI developers (Selbst et al., 2019). Schools must also establish mechanisms for accountability, such as AI ethics committees, to assess the effect of AI machinery on learning and student well-being. Ethical AI in education should ultimately focus on augmenting human capabilities rather than replacing them, ensuring that technology serves as a tool for inclusive and equitable learning.

## 1.2 Importance of Balancing Automation, Privacy, and Human Oversight

The integration of automation into various sectors, from healthcare to finance, has significantly improved efficiency and productivity. However, excessive reliance on automation without human oversight can lead to errors, bias, and unintended consequences. According to Brynjolfsson and McAfee (2017), automation enhances decision-making speed and accuracy, but human judgment remains crucial in mitigating algorithmic biases and ethical dilemmas. For instance, in the medical field, AI-powered diagnostics assist doctors in identifying diseases, yet final decisions must involve human professionals to prevent misdiagnoses and ensure patient safety. Balancing automation with human oversight is essential for maintaining accountability and ethical decision-making.

Privacy concerns also arise as automation increasingly relies rely on extensive collections of individual data to operate effectively. Automated systems in social media, e-commerce, and surveillance collect and analyze user data, raising ethical and legal issues related to data protection. According to Zuboff (2019), surveillance capitalism has led to the commodification of personal information, often without users' explicit consent. Governments and organizations must implement robust privacy frameworks, such as the General Data Protection Regulation (GDPR), to safeguard individuals' data while allowing businesses to leverage automation responsibly. Balancing automation and privacy ensure that technological advancements do not compromise individuals' fundamental rights.

Human oversight plays a critical role in maintaining trust and accountability in automated systems. While automation can streamline operations and reduce human error, unchecked systems may lead to unintended social and economic consequences. Algorithm-driven financial trading has been linked to market fluctuations and economic instability when not properly supervised (Dhar, 2016). A well-balanced approach involves regulatory frameworks, ethical AI principles, and active human intervention to ensure that automation aligns with societal values and priorities. By integrating human oversight, privacy safeguards, and responsible automation, societies can make the most of technology while minimizing risks.

## 1.3 Statement of central argument

The integration of artificial intelligence (AI) in schools presents a significant ethical dilemma, as it necessitates a balance between automation, privacy, and human oversight. While AI-driven tools enhance personalized learning and administrative efficiency, they also raise concerns about student data security and the potential erosion of human decision-making in education (Selwyn, 2020). The ethical challenge lies in ensuring that AI systems respect student privacy while complementing, rather than replacing, educators' roles in learning environments. Without careful

governance, AI implementation in schools could lead to unintended consequences, such as biased algorithms influencing grading or surveillance tools infringing on student rights (Williamson & Eynon, 2020). A responsible approach to AI adoption must prioritize ethical standards that safeguard student autonomy and ensure transparency in data usage.

While automation can optimize educational processes, excessive reliance on AI may diminish the essential human connection in teaching and learning. Research indicates that students benefit from empathetic interactions with teachers, which AI lacks the capability to fully replicate (Luckin, 2018). AI-driven tutoring programs may improve efficiency, but they cannot substitute the nuanced understanding and emotional intelligence of human educators. Moreover, ethical concerns arise when AI systems make high-stakes decisions, such as predicting student performance or disciplinary actions, without adequate human oversight (Aiken, 2019). To ensure fairness and accountability, educational institutions must implement policies that ensure AI is leveraged to complement human decision-making, reinforcing educators' roles rather than displacing them.

To achieve ethical AI integration in schools, policymakers and educators must establish clear regulatory frameworks that balance innovation with responsibility. The development and deployment of AI in education should involve continuous human oversight, ensuring that algorithms align with ethical and educational values (Holmes et al., 2021). Transparent data governance policies must be in place to protect students' personal information and prevent misuse. Furthermore, encouraging active participation from students, parents, and educators in ethical discussions surrounding AI can enhance transparency and foster trust. By adopting a cooperative approach, educational institutions can responsibly integrate AI, maximizing its benefits while minimizing ethical challenges—ensuring it remains a supportive tool rather than a controlling force.

## 2 The Benefits of AI in Education

### 2.1 Personalized Learning

Artificial Intelligence (AI) has significantly enhanced customized instruction that adapts to each student's unique learning requirements. Traditional classrooms often employ a uniform method of instruction that often fails to support varying learning speeds and preferences. AI-driven learning platforms leverage data analytics and machine learning to assess students' strengths, weaknesses, and learning habits, facilitating the creation of personalized resources and guidance (Luckin et al., 2018). For instance, AI-powered adaptive platforms like Carnegie Learning's MATHia modify task difficulty in response to students' real-time progress, helping deliver targeted support and challenges suited to their capabilities (Holmes et al., 2019).

AI enables personalized feedback, helping students to learn more efficiently. Intelligent tutoring systems, like IBM's Watson Tutor, provide real-time assistance and personalized guidance by analyzing students' responses and predicting areas where they may struggle (Chen et al., 2020). These systems use Natural Language Processing (NLP) to facilitate interactive learning experiences, making education more engaging and effective. Additionally, AI-powered chatbots and virtual assistants offer round-the-clock support, answering student queries instantly and allowing for a more flexible learning experience (Zawacki-Richter et al., 2019). This level of personalization fosters self-directed learning, where students can progress at their own pace and revisit challenging concepts as needed.

AI contributes to differentiated instruction by assisting educators in curriculum planning and student assessment. Learning management systems (LMS) integrated with AI, such as Coursera and Khan Academy, generate personalized learning paths based on user performance and engagement metrics (Roll & Wylie, 2016). This not only helps teachers identify at-risk students but also enables them to design targeted interventions to bridge learning gaps. As AI continues to evolve, it holds great promise in making education more inclusive, ensuring that learners of all backgrounds and abilities receive tailored support for academic success.

### 2.2 Administrative Efficiency

Artificial Intelligence (AI) enhances administrative efficiency in education by automating routine tasks, allowing educators and administrators to focus on more critical responsibilities. AI-powered systems can manage student enrollment, timetable scheduling, and record-keeping with greater accuracy and speed compared to traditional manual processes (Luckin et al., 2018). This automation reduces administrative burdens and minimizes errors, leading to more efficient institutional management. AI chatbots and virtual assistants provide instant responses to student inquiries, streamlining communication between students and school administration (Holmes et al., 2019).

AI-powered analytics support informed decision-making by delivering timely insights into student achievement, attendance trends, and the distribution of resources. Through the analysis of large datasets, machine learning models detect patterns and forecast possible issues, allowing school administrators to implement preventative strategies effectively (Zawacki-Richter et al., 2019). For instance, predictive analytics can help schools identify students at risk of dropping out and implement timely interventions. This data-driven approach not only enhances institutional efficiency but also contributes to improved student outcomes and retention rates.

AI optimizes resource management by automating budgeting, payroll, and inventory tracking in educational institutions. Smart AI systems can forecast financial needs, monitor expenditure, and suggest cost-saving measures, ensuring sustainable resource allocation (Selwyn, 2020). Such innovations reduce administrative workload and operational costs while improving transparency and accountability. By integrating AI into administrative functions, educational institutions can enhance efficiency, improve service delivery, and create a more effective learning environment.

## 2.3 Equity and Accessibility

Artificial Intelligence (AI) has the potential to bridge educational gaps by making learning more accessible to diverse populations, including students with disabilities, those in remote areas, and individuals facing economic hardships. AI-driven tools, such as speech-to-text applications, adaptive learning platforms, and virtual tutors, enable personalized education that accommodates different learning needs (Luckin et al., 2018). For instance, AI-powered platforms can provide real-time translations and text-to-speech features for visually impaired or non-native speakers, ensuring that language and physical limitations do not hinder access to quality education (Zawacki-Richter et al., 2019). These advancements democratize education, making learning opportunities more inclusive and equitable.

AI reduces disparities in education by providing cost-effective learning solutions that can supplement traditional teaching. Automated grading systems and intelligent tutoring applications allow educators to focus on personalized instruction rather than administrative tasks, benefiting students who require additional support (Holmes et al., 2019). In low-income regions, AI-based mobile learning applications provide access to quality educational resources without the need for expensive infrastructure (Selwyn, 2020). By eliminating barriers related to geography and financial constraints, AI promotes a more equitable distribution of knowledge, ensuring that education is no longer limited to privileged groups.

However, for AI to fully enhance equity and accessibility, it must be implemented responsibly. Issues such as algorithmic bias, digital literacy, and unequal access to technology must be addressed to prevent AI from reinforcing existing educational inequalities (Williamson & Eynon, 2020). Policymakers and educators must work together to develop ethical AI frameworks that prioritize marginalized communities, ensuring that AI benefits all learners, regardless of their socio-economic background. When effectively integrated, AI can serve as a powerful tool to level the educational playing field and promote lifelong learning opportunities for diverse populations.

## 3 Ethical Concerns and Challenges

## 3.1 Privacy and Data Security

### 3.1.1 *Student data is highly sensitive and vulnerable to misuse*

Student data, which includes personal information, academic records, and behavioral insights, is highly sensitive and requires stringent protection against misuse. AI-driven educational tools collect vast amounts of data to personalize learning, assess performance, and improve teaching methods. However, without proper safeguards, this data can be exploited for unethical purposes, such as targeted advertising, unauthorized profiling, or even identity theft (West, 2019). Schools, edtech companies, and policymakers must ensure compliance with data protection laws like the General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) to mitigate risks associated with data breaches and improper handling (Schmid et al., 2021).

The integration of AI in schools raises ethical concerns regarding student privacy and consent. Many students and parents remain unaware of the extent to which AI systems process their data, potentially leading to manipulation or bias in decision-making (Zuboff, 2019). For instance, AI-driven predictive analytics could unfairly label students based on historical data, reinforcing existing inequalities (Binns, 2020). To prevent such issues, educational institutions must implement transparent policies, enforce strict access controls, and prioritize human oversight in AI-driven decisions.

Ethical AI governance frameworks should guide the responsible use of student data, ensuring that technological advancements do not come at the expense of student privacy and autonomy (Floridi & Cowls, 2022).

### 3.1.2    *Regulations like GDPR and FERPA highlight the need for data protection*

Regulations such as the General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA) underscore the critical importance of data protection, especially in educational settings where AI systems process vast amounts of student information. GDPR, enforced in the European Union, mandates strict data privacy measures, emphasizing transparency, user consent, and accountability in handling personal data (European Parliament and Council, 2016). Similarly, FERPA, a U.S. law, safeguards students' educational records by granting parents and eligible students control over their data access and disclosure (U.S. Department of Education, 2020). These laws highlight the ethical imperative for schools to adopt AI responsibly, ensuring that automation does not compromise privacy rights or expose students to unauthorized data exploitation.

As AI-driven tools increasingly support student assessments, personalized learning, and administrative tasks, adherence to GDPR and FERPA becomes essential to maintain trust and compliance. AI applications in schools must incorporate robust data protection mechanisms, such as encryption, anonymization, and restricted access controls, to align with regulatory requirements. Failure to comply can lead to legal consequences and ethical concerns, reinforcing the need for human oversight in AI deployment (OECD, 2021). By adhering to these regulations, educational institutions can strike a balance between leveraging AI for efficiency and safeguarding student data, fostering an ethical framework that prioritizes privacy and security.

## 3.2  Bias and Fairness

### 3.2.1    *AI systems can reinforce biases if not properly designed*

AI systems in schools, when not carefully designed, can reinforce biases, leading to unfair treatment of students and educators. Bias in AI arises from skewed training data, flawed algorithms, or systemic prejudices embedded in educational institutions. For instance, predictive analytics used in student assessments may disproportionately favor certain demographics while disadvantaging others, perpetuating inequalities (Mehrabi et al., 2021). Additionally, automated decision-making in disciplinary actions could mirror existing societal biases, resulting in unfair targeting of marginalized students (Zhang & Daumé, 2022). Without proper oversight, these biases can become deeply embedded, negatively impacting students' opportunities and academic outcomes.

To mitigate these risks, ethical AI deployment in schools must prioritize transparency, accountability, and continuous auditing. AI developers and educators should collaborate to ensure that training datasets are diverse and representative of the student population (Holstein et al., 2019). Regular bias detection mechanisms should be incorporated to identify and correct any discriminatory patterns in AI-driven decisions. Moreover, human oversight is essential to contextualize AI recommendations and prevent unintended consequences (Binns, 2020). By implementing these safeguards, schools can harness AI's potential while upholding fairness and inclusivity in education.

## 4    The Need for Human Oversight

## 4.1  Ensuring Ethical Decision-Making

Ethical decision-making in AI-driven educational environments requires a structured approach that prioritizes fairness, transparency, and accountability. Schools must implement policies that ensure AI systems align with ethical guidelines to prevent bias, discrimination, and the violation of student privacy. According to Floridi and Cowls (2019), ethical AI frameworks should emphasize principles such as beneficence, non-maleficence, and justice to protect students from potential harm. By integrating human oversight into AI decision-making processes, educators can mitigate unintended consequences and ensure that automated systems enhance, rather than replace, human judgment. Transparency in algorithmic decision-making, including explainable AI (XAI) models, further supports accountability by allowing stakeholders to understand how and why AI reaches specific conclusions (Molnar, 2022).

Ethical decision-making in AI applications for schools necessitates continuous evaluation and adaptation to emerging challenges. As AI technologies evolve, regular audits and stakeholder engagement including input from students, parents, and educators—can help refine policies that govern their use. Implementing ethical review boards within educational institutions can ensure compliance with ethical standards and regulatory frameworks, such as the General Data Protection Regulation (GDPR) for student data protection (European Commission, 2020). Ethical AI education for teachers and administrators can empower them to recognize and address ethical dilemmas effectively. By fostering an

AI-literate school community, institutions can balance the benefits of automation with ethical safeguards that prioritize student welfare and rights.

## 4.2 Transparency in AI Operations

Transparency in AI operations is crucial for ensuring ethical AI deployment in schools, particularly as it relates to automation, privacy, and human oversight. By providing insight into their decision logic, transparent AI systems make it easier for educators, students, and parents to comprehend the role of algorithms in shaping academic performance and managing school tasks. According to Floridi et al. (2018), transparency fosters trust and accountability by making AI operations interpretable and auditable. Without transparency, biases in AI models may go unnoticed, potentially leading to unfair outcomes, such as discriminatory grading or unequal access to educational resources. To address these concerns, explainable AI (XAI) techniques have been proposed to enhance interpretability, allowing stakeholders to question and refine automated decisions (Samek et al., 2019).

Transparency is essential for safeguarding student data privacy, a major ethical concern in AI-driven education. AI systems often collect vast amounts of sensitive information, including academic records and behavioral patterns, necessitating clear disclosure of data usage policies (Leslie, 2020). Transparent AI governance frameworks help institutions comply with legal and ethical standards, such as the General Data Protection Regulation (GDPR) and the Family Educational Rights and Privacy Act (FERPA), ensuring that AI respects students' rights (Veale & Binns, 2017). Schools must adopt robust transparency measures, such as open-source AI models and regular audits, to maintain public confidence in AI applications in education. By prioritizing transparency, educational institutions can strike a balance between automation and human oversight while upholding ethical AI practices.

## 4.3 Maintaining the Human Element in Education

As artificial intelligence (AI) becomes more integrated into education, maintaining the human element is essential to preserving the emotional, social, and ethical dimensions of learning. While AI-driven tools can personalize instruction and automate administrative tasks, they lack the empathy and moral judgment that teachers provide (Selwyn, 2019). Human educators play a crucial role in fostering creativity, critical thinking, and ethical reasoning—skills that AI cannot fully replicate. The presence of teachers ensures that students receive guidance, encouragement, and mentorship, which are vital for their holistic development (Williamson & Eynon, 2020). Without this human interaction, students may struggle with motivation and emotional well-being, emphasizing the need for a balanced approach where AI supports rather than replaces educators.

Ethical concerns surrounding AI in schools, including data privacy and bias, reinforce the necessity of human oversight. Teachers and administrators must actively monitor AI systems to ensure fair and responsible use, mitigating potential risks such as algorithmic discrimination or excessive surveillance (Zawacki-Richter et al., 2019). By maintaining strong human oversight, schools can leverage AI's benefits while ensuring that ethical principles remain at the core of education. The role of educators extends beyond knowledge delivery; they shape students' values, interpersonal skills, and ethical perspectives, which AI alone cannot achieve. Therefore, a well-balanced integration of AI and human oversight will ensure that technology improves instead of undermines the fundamental aspects of education.

## 5 Case Studies of AI in Education: Global Perspectives

While ethical principles provide a foundation for AI integration in schools, real-world applications reveal both the transformative potential and the ethical pitfalls of educational AI. Examining how different countries are implementing AI in education offers insight into diverse challenges, policy approaches, and outcomes.

## 5.1 United States: Emphasis on Personalization and Ethical Guidelines

In the U.S., AI has been widely adopted to enhance personalized learning. Tools such as Knewton and DreamBox Learning use real-time data to adapt educational content to individual learning styles. These technologies have shown promise in improving math proficiency and engagement, especially in underperforming districts (Holmes et al., 2019). However, privacy concerns have emerged around the use of student data by third-party vendors. The lack of uniform national standards has led to inconsistent applications of ethical frameworks, prompting states like California to introduce stronger student privacy laws (West, 2020).

## 5.2 China: AI for Surveillance and Assessment

China leads in large-scale implementation of AI in classrooms. Facial recognition is used to monitor student attentiveness, and AI-powered grading tools provide real-time academic feedback. These technologies aim to boost

performance but have sparked concern over surveillance and psychological impact (Zuboff, 2019; Williamson & Eynon, 2020). Research suggests that constant behavioral monitoring may hinder creativity and student autonomy, reinforcing the need for informed consent and mental health safeguards (Floridi & Cowls, 2022).

### 5.3 Finland: Ethical Innovation and Equity

Finland takes a more cautious, ethics-centered approach to AI integration. The government-supported Elements of AI initiative promotes nationwide AI literacy while ensuring schools adopt technology in ways that reflect Finnish values of inclusivity and democratic participation (Floridi & Cowls, 2019). AI pilot programs focus on transparency and co-design with educators, avoiding invasive monitoring and emphasizing ethical design principles (Williamson & Eynon, 2020).

### 5.4 Kenya: Bridging the Digital Divide

In Kenya, AI initiatives aim to overcome infrastructural challenges and enhance educational accessibility. For instance, the government has recognized the potential of AI in transforming education and has included AI strategies in national plans such as the Kenya National Digital Master Plan 2022 to 2032. This plan outlines the transformative effects of AI on the Kenyan economy and service delivery, including education (Waweru, 2023). However, challenges persist, including limited infrastructure, a shortage of qualified AI professionals, and concerns about data privacy and security. Addressing these issues requires investment in infrastructure, capacity building, and the development of ethical guidelines to ensure equitable and effective AI integration in Kenyan education (Waweru, 2023).

### 5.5 Comparative Insights

These global case studies illustrate a spectrum of AI integration: personalization in the U.S., behavioral surveillance in China, inclusive ethics in Finland, and access-driven innovation in Kenya. Each case highlights the necessity of aligning AI use with local values, legal structures, and equity goals. Ultimately, they show that the success of AI in education depends not only on the technology itself but on the ethical governance, community engagement, and cultural sensitivity behind its deployment.

## 6  Recommendations for Ethical AI Implementation

### 6.1 Stronger Regulations and Policies

Ensuring compliance with global data protection laws is critical when implementing AI systems in schools, as these regulations safeguard students' personal data from misuse and unauthorized access. Key frameworks such as the General Data Protection Regulation (GDPR) in Europe and the Children's Online Privacy Protection Act (COPPA) in the U.S. establish strict guidelines on data collection, processing, and storage (European Parliament, 2016; Federal Trade Commission, 1998). Schools and AI developers must prioritize data minimization, ensuring that only essential student data is collected while ensuring clarity in how data is collected and used. Informed consent from parents or guardians is crucial when AI systems process minors' personal information (Goggin et al., 2021). Failure to adhere to these regulations can result in legal penalties and a loss of public trust, highlighting the need for rigorous compliance measures.

Beyond legal compliance, ethical considerations necessitate that AI-driven educational technologies incorporate robust privacy-by-design mechanisms to protect students' digital identities. For instance, AI algorithms should be designed to anonymize or pseudonymize student data, reducing risks associated with data breaches (Binns, 2018). Furthermore, educational institutions must establish clear data governance policies, including regular audits and impact assessments, to ensure continuous adherence to global privacy standards. As AI adoption in schools grows, collaboration between educators, policymakers, and technology providers is essential to align AI applications with both regulatory and ethical frameworks, ensuring that student data is handled responsibly and securely (Floridi & Cowls, 2019).

### 6.2 Bias Mitigation Strategies

Ethical deployment of AI technologies in education depends on regular evaluations and the use of representative datasets. AI audits involve systematic reviews of AI systems to assess biases, fairness, and compliance with ethical guidelines (Raji et al., 2020). These audits help identify unintended discriminatory patterns in AI-driven student assessments, disciplinary actions, or personalized learning tools. Without such oversight, AI systems could reinforce existing inequalities or make decisions that lack transparency and fairness (Binns, 2018). Schools must implement periodic AI audits to evaluate performance and adjust algorithms, accordingly, ensuring they align with ethical standards and educational equity principles.

Diverse training datasets play a critical role in mitigating biases and ensuring that AI models are inclusive of all student demographics. AI models trained on homogeneous or unrepresentative datasets risk reinforcing systemic biases, leading to unfair outcomes, particularly for students from marginalized backgrounds (Buolamwini & Gebru, 2018). By incorporating diverse datasets, AI systems can provide more equitable and culturally sensitive educational support. Educational institutions must work with AI developers to ensure that training datasets represent a wide range of ethnicities, socio-economic backgrounds, and learning styles. This approach enhances fairness and fosters trust in AI applications used in schools, ultimately promoting ethical AI practices (Hao, 2020).

## 6.3 Human-in-the-Loop Models

Artificial Intelligence (AI) could transform the education sector by supporting individualized learning, streamlining routine administrative processes, and offering insights based on data analysis. However, AI should function as an assistive tool rather than a replacement for educators, as human interaction remains central to effective teaching and learning. Research suggests that while AI-driven systems can improve student engagement and assessment efficiency, they do not possess emotional awareness, creativity, and adaptability of human teachers (Luckin, 2018). Educators are essential in developing students' critical thinking skills, moral development, and social-emotional learning elements that AI cannot fully replicate. Therefore, AI should be integrated to support educators in addressing diverse learning needs while preserving the irreplaceable human connection in education.

Ethical considerations in AI implementation emphasize the need for human oversight to ensure fairness, inclusivity, and data privacy. AI systems, if not carefully designed, may reinforce biases and create ethical dilemmas regarding student data usage (Selwyn, 2020). Educators must remain at the center of decision-making, guiding AI tools to align with educational values and pedagogical goals. Instead of replacing teachers, AI should be integrated as a collaborative aid that supports pedagogical innovation and reduces administrative workload, thereby enabling educators to concentrate on their essential roles as mentors and facilitators of learning (Williamson & Eynon, 2020). By maintaining this balance, schools can harness AI's benefits while upholding ethical standards in education.

## 7 Summary and Conclusion

The ethical integration of AI in schools requires a delicate balance between automation, privacy, and human oversight to ensure equitable and responsible educational practices. While AI-powered tools enhance personalized learning and administrative efficiency, concerns about student data privacy and algorithmic biases necessitate strong ethical guidelines (Williamson & Eynon, 2020). Schools must implement AI with transparency, ensuring that data collection aligns with privacy regulations and that automated decisions remain interpretable and fair (Selwyn, 2019). Without proper oversight, AI risks reinforcing inequalities, making it crucial for educators and policymakers to establish accountability measures that safeguard student rights.

A human-centered approach to AI in education emphasizes the irreplaceable role of teachers and administrators in decision-making processes. AI should be employed to support, not supplant, human judgment—particularly in contexts requiring empathy, ethical reasoning, and nuanced decision-making, ensuring that ethical dilemmas, such as data misuse or biased assessments, are addressed through continuous human oversight (Ames, 2021). Furthermore, ethical frameworks should be developed in collaboration with educators, students, and policymakers to align AI implementation with pedagogical goals and societal values (Zawacki-Richter et al., 2019). By fostering transparency, fairness, and human involvement, schools can harness the benefits of AI while mitigating its ethical risks.

The implementation of AI in schools presents transformative opportunities for personalized learning, administrative efficiency, and enhanced student engagement. However, ethical considerations must be prioritized to prevent biases, ensure transparency, and uphold student privacy. Unethical AI deployment can lead to algorithmic discrimination, reinforcing social inequalities and limiting opportunities for marginalized students (Binns, 2018). Furthermore, unchecked automation in education can erode human oversight, diminishing the role of educators in making context-based decisions that technology cannot fully grasp (Selwyn, 2019). Therefore, ethical AI implementation requires clear guidelines on data usage, fairness in algorithmic decision-making, and accountability measures to prevent harm while maximizing educational benefits.

Schools must adopt AI policies that align with ethical frameworks such as fairness, accountability, and transparency to foster trust and inclusiveness in education. Ethical AI implementation should involve all stakeholders' teachers, students, parents, and policymakers to ensure that technology complements human judgment rather than replacing it (Floridi & Cowls, 2019). Privacy concerns must be addressed through robust data protection strategies, preventing the misuse of students' personal information by third parties (West, 2020). By embedding ethical principles into AI-driven

educational tools, schools can create an environment where technology enhances learning while safeguarding students' rights and well-being.

The integration of artificial intelligence (AI) in education presents opportunities for personalized learning, administrative efficiency, and enhanced student engagement. However, ethical concerns such as data privacy, algorithmic bias, and the diminishing role of human oversight necessitate a call for responsible AI use. According to UNESCO (2023), AI applications in schools must align with ethical guidelines that prioritize transparency, accountability, and inclusivity. Without these safeguards, AI-driven decision-making in grading, admissions, and student monitoring risks reinforcing biases and infringing on students' rights (Selwyn, 2022). Ensuring responsible AI use requires a collaborative effort among educators, policymakers, and technologists to create regulatory frameworks that promote fairness while harnessing AI's potential to improve learning outcomes.

Responsible AI implementation in schools also involves fostering digital literacy and critical thinking skills among students and educators to mitigate the risks associated with AI dependency. Studies emphasize the importance of human oversight in AI-driven education systems to prevent over-reliance on automation (Luckin, 2021). For instance, AI-powered tutoring systems should enhance, not displace, the role of teachers to retain the relational depth and contextual relevance of the learning experience. Data protection policies must be strictly enforced to safeguard student information from misuse (Floridi & Cowls, 2019). By adopting ethical AI policies and investing in continuous monitoring, schools can strike a balance between technological innovation and the fundamental principles of education, ensuring that AI serves as an enabler rather than a disruptor.

## References

[1] Aiken, W. (2019). AI in education: Ethical implications and challenges. Springer.

[2] Aiken, W. (2019). The ethics of AI in Education: Balancing Innovation and Responsibility. Oxford University Press.

[3] Ames, M. G. (2021). The charisma machine: The life, death, and legacy of one laptop per child. MIT Press.

[4] Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency, 149–159.

[5] Binns, R. (2020). Human Judgement in Algorithmic Personalization. The Information Society, 36(1), 1-12.

[6] Binns, R. (2020). Human oversight and the ethics of AI. AI & Society, 35(1), 17-29.

[7] Brynjolfsson, E., & McAfee, A. (2017). Machine, Platform, Crowd: Harnessing Our Digital Future. W. W. Norton & Company.

[8] Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of the Conference on Fairness, Accountability, and Transparency, 77-91.

[9] Chen, X., Xie, H., Zou, D., & Hwang, G. J. (2020). Application and theory of artificial intelligence in education. Educational Technology & Society, 23(3), 1-9.

[10] Cowls, J., & Floridi, L. (2018). Prolegomena to a White Paper on an Ethical Framework for a Good AI Society. Minds and Machines, 28(4), 689–707.

[11] Dhar, V. (2016). When to trust robots with decisions, and when not to. Harvard Business Review, 94(3), 114-123.

[12] European Commission. (2020). General Data Protection Regulation (GDPR). Retrieved from (https://gdpr.eu)

[13] European Parliament and Council. (2016). General Data Protection Regulation (GDPR). Official Journal of the European Union.

[14] Federal Trade Commission. (1998). Children's Online Privacy Protection Act (COPPA).

[15] Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. Harvard Data Science Review, 1(1).

[16] Floridi, L., & Cowls, J. (2019). The ethics of artificial intelligence: Mapping the debate. AI & Society, 34(1), 59-72.

[17] Floridi, L., & Cowls, J. (2022). The Ethics of Artificial Intelligence in Education. AI & Society, 37(2), 245-259.

[18] Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., & Schafer, B. (2018). AI4People: An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689-707.

[19]  Floridi, L., et al. (2018). AI4People: An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689–707.

[20]  Goggin, G., Hjorth, L., & Khoo, O. (2021). AI in education: Critical perspectives and practices. Routledge.

[21]  Hao, K. (2020). AI still doesn't get people of color. MIT Technology Review.

[22]  Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign.

[23]  Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign.

[24]  Holmes, W., Bialik, M., & Fadel, C. (2021). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. Center for Curriculum Redesign.

[25]  Holstein, K., Wortman Vaughan, J., Daumé, H., Dudik, M., & Wallach, H. (2019). Improving fairness in AI-powered education systems. Proceedings of the ACM Conference on Fairness, Accountability, and Transparency, 17(1), 371-381.

[26]  Jobin, A., Ienca, M., & Vayena, E. (2019). "The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence, 1(9), 389–399.

[27]  Leslie, D. (2020). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.

[28]  Luckin, R. (2018). Machine learning and human intelligence: The future of education for the 21st century. UCL Press.

[29]  Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence Unleashed: An Argument for AI in Education. Pearson.

[30]  Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2018). Artificial Intelligence and Future of Learning. London: UCL Knowledge Lab.

[31]  Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2018). Machine learning and human intelligence: The future of education for the 21st century. UCL Institute of Education Press.

[32]  Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1-35.

[33]  Molnar, C. (2022). Interpretable Machine Learning. Retrieved from (https://christophm.github.io/interpretable-ml-book)

[34]  OECD. (2021). Artificial Intelligence in Education: Challenges and Opportunities for Data Protection. Paris: OECD Publishing.

[35]  Oseji, N.A., Olayinka, M.A., Bako, N.Z., & Sani, J.O. (2021). Integration of Artificial Intelligence into library and information workflow for enhanced services in developing economies. Proceedings of 2021 Ahmadu Bello University Library Complex International Conference, Zaria 1st – 3rd November 2021

[36]  Raji, I. D., Smart, A., White, R. N., Mitchell, M., & Gebru, T. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 33-44.

[37]  Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), 582-599.

[38]  Samek, W., Wiegand, T., & Müller, K. R. (2019). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models.

[39]  Schiff, D. (2021). Education and AI: Opportunities and Challenges for the Future. AI & Society, 36(4), 1027–1040.

[40]  Schmid, P., Smith, J., & Taylor, L. (2021). Data Privacy in AI-Driven Education: Challenges and Solutions. Journal of Educational Technology, 18(4), 215-230.

[41]  Selwyn, N. (2019). Should Robots Replace Teachers? AI and the Future of Education. Polity Press.

[42]  Selwyn, N. (2020). Should robots replace teachers? AI and the future of education. Polity Press.

[43]  U.S. Department of Education. (2020). Family Educational Rights and Privacy Act (FERPA). Washington, D.C.

[44] UNESCO. (2023). Guidance for AI in Education: Ethical Considerations and Policy Recommendations. United Nations Educational, Scientific and Cultural Organization.

[45] UNESCO. (2023). Guidance for AI in Education: Ethical Considerations and Policy Recommendations. United Nations Educational, Scientific and Cultural Organization.

[46] Veale, M., & Binns, R. (2017). Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. Big Data & Society, 4(2), 1-17.

[47] Waweru, A. (2023). AI in Kenyan Education. UNESCO AIED. Retrieved from https://aiedchair.bnu.edu.cn/wp-content/uploads/2023/12/Kenya.pdf

[48] West, S. (2019). The Datafication of Education: Risks and Implications. International Journal of Digital Ethics, 14(3), 112-128.

[49] West, S. M. (2020). Data Capitalism: Redefining the Logics of Surveillance and Privacy. Business & Society, 59(1), 213-238.

[50] Williamson, B., & Eynon, R. (2020). Algorithmic education: Critical perspectives on AI and learning analytics. Learning, Media and Technology, 45(1), 87-103.

[51] Williamson, B., & Eynon, R. (2020). Automating Education: AI and the Future of Learning. Routledge.

[52] Williamson, B., & Eynon, R. (2020). Education and technology: Key issues and debates. Bloomsbury Academic.

[53] Williamson, B., & Eynon, R. (2020). Understanding the Datafication of Education. Learning, Media and Technology, 45(2), 107-120.

[54] Williamson, B., & Eynon, R. (2020). Understanding the Role of AI in Education: Ethics, Equity, and Challenges. Learning, Media and Technology, 45(2), 110-126.

[55] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education. International Journal of Educational Technology in Higher Education, 16(1), 1-27.

[56] Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic Review of Research on Artificial Intelligence Applications in Higher Education. International Journal of Educational Technology in Higher Education, 16(1), 39.

[57] Zhang, B. H., & Daumé, H. (2022). Understanding bias in automated decision-making in education. Journal of Artificial Intelligence Research, 73, 145-168.

[58] Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Public Affairs.