

Review paper

AI Ethics in Education: Integrating Pedagogy, Technology, and Governance

Aylin Karakuş 

Gazi University, TÜRKİYE

Corresponding Author: aylinkarakuş@gazi.edu.tr

Citation: Karakuş, A. (2026). AI ethics in education: Integrating pedagogy, technology, and governance. *Journal of Interdisciplinary Research in Artificial Intelligence and Society*, 2(1), Article 6. <https://doi.org/10.20897/jirais/19468>

Published: September 30, 2026

ABSTRACT

The rapid integration of artificial intelligence (AI) into education has transformed teaching, learning, and educational decision-making while simultaneously generating complex ethical challenges. Although research on AI ethics in education has expanded considerably, the existing literature remains conceptually fragmented across pedagogical, technological, and governance perspectives. This fragmentation limits our understanding of how to translate ethical principles into educational practice and policy. Addressing this gap, the present study develops an integrated theoretical framework for AI ethics in education through a critical literature review and comparative policy analysis. The analysis synthesizes international scholarship and major policy frameworks, including those developed by UNESCO, the OECD, the European Union, China, and Türkiye, to examine how different educational contexts conceptualize and govern ethical issues. The findings reveal three interconnected yet insufficiently integrated dimensions within the literature: pedagogical ethics, technological–algorithmic risks, and global governance. These dimensions are predominantly examined in isolation, resulting in limited conceptual coherence across the micro (learner–teacher), meso (technological systems), and macro (policy and governance) levels. In response, the study proposes a multi-level conceptual framework grounded in the principles of fairness, transparency, accountability, and human-centeredness. By integrating technological design, pedagogical processes, ethical theory, and governance structures, the proposed framework provides a more holistic understanding of AI ethics in education. The study contributes to the literature by advancing an interdisciplinary and theoretically grounded model that supports ethical evaluation, policy development, and the responsible governance of AI in educational contexts.

Keywords: AI ethics in education, artificial intelligence governance, algorithmic bias, multi-level framework

Artificial intelligence (AI) has emerged as a rapidly evolving technology that must be considered within the context of moral norms and societal values that shape human well-being and social cohesion. The field of education has been significantly influenced by this transformation, taking important steps to leverage the transformative potential of AI. In particular, AI-driven applications such as personalized learning platforms and adaptive assessment systems have been increasingly implemented to enhance student achievement and streamline administrative processes (Luckin et al., 2016; Holmes et al., 2019). These systems enable learners to access content tailored to their individual needs and receive real-time feedback, thereby supporting self-paced learning. Computerized adaptive testing used in large-scale assessments represents a concrete example of this potential.

However, this rapid technological advancement has also brought to the forefront critical ethical challenges, including data privacy, algorithmic bias, system transparency, and learner autonomy (Crawford & Calo, 2016). The widespread adoption of AI in education has occurred in a context where ethical governance mechanisms are not

yet fully institutionalized, leading to complex and multi-dimensional risks. The extensive collection and analysis of personal learning data by AI systems increase the risk of what has been described as “surveillance capitalism,” in which individuals’ digital activities are commodified (Zuboff, 2019). Moreover, AI applications that aim to enhance educational equity may inadvertently reproduce and reinforce existing social inequalities due to biases embedded in training data (O’Neil, 2016). These risks place ethical considerations at the center of discussions surrounding the integration of AI in education, making the incorporation of ethical principles a fundamental necessity.

Discussions on AI ethics have largely been framed within technical standards (Floridi et al., 2018) or limited regulatory perspectives (Crawford & Calo, 2016). However, ethical issues in educational contexts lie at the intersection of philosophical, pedagogical, and policy dimensions (Williamson & Eynon, 2020). This multi-layered structure transforms AI governance in education from a purely technical issue into a normative and value-based policy domain (Selwyn, 2019).

A review of the global literature reveals a clear gap between ethical principles and practical applications across different educational levels (from K–12 to higher education) and diverse national contexts (e.g., the European Union, Türkiye, China) (Holmes et al., 2021; Knox et al., 2020). Although international frameworks developed by organizations such as UNESCO (2021) and OECD (2023) provide strong normative guidance at the level of abstract principles, they often fail to adequately address the concrete algorithmic risks experienced by educational stakeholders, including students, teachers, and parents (Regan & Jesse, 2019; Sloane & Moss, 2019).

Furthermore, studies on AI-driven decision-making processes, such as learning analytics, emphasize that ethical considerations are frequently overlooked in these high-level policy frameworks (Slade & Prinsloo, 2013; Ifenthaler & Schumacher, 2016). Despite the global rise of what has been termed the “AI ethics boom” (Jobin, Ienca, & Vayena, 2019), empirical and integrative research examining how ethical principles are translated into educational practice remains limited (Zawacki-Richter et al., 2019; Holmes et al., 2021). This highlights the need for a critical and comparative analysis of the gap between international ethical principles and their practical implementation in educational contexts (Selwyn, 2022; Williamson et al., 2023).

Managing risks such as algorithmic bias, erosion of learner autonomy, and data surveillance requires not only technical solutions but also comprehensive pedagogical and regulatory approaches. In this regard, the question of how global AI ethics principles is operationalized in education has become a central concern in contemporary research. However, existing literature addresses these issues only in a limited and fragmented manner. In response, this study critically and integratively analyzes the fragmented and multi-layered literature on AI ethics in education and proposes a multi-level conceptual model that brings together micro (learner/teacher), meso (technological systems), and macro (governance and policy) dimensions.

Research problem and gap in the literature

The rapid expansion of AI use in education has made it necessary to address its ethical implications in a comprehensive and systematic manner. Existing research in the global literature can be broadly categorized into three main strands:

- (1) pedagogical studies focusing on AI ethics education,
- (2) technological studies examining risks generated by AI systems, and
- (3) policy-oriented studies addressing AI governance at national and international levels.

First, pedagogical studies investigating how AI ethics is taught—particularly in higher education—have generated a growing body of knowledge. Systematic reviews indicate that AI ethics has increasingly become an independent area of learning. While these studies contribute to the development of students’ ethical reasoning skills, they also highlight methodological limitations in measuring ethical awareness and learning outcomes, particularly regarding consistency, comparability, and theoretical grounding.

Second, studies in the technological strand focus on risks such as data privacy, learner autonomy, algorithmic bias, and uncertainty in automated decision-making. Research demonstrates that AI systems create tensions between pedagogical diversity, institutional governance, transparency, human dignity, and data ethics. In particular, the opaque nature of deep learning models weakens accountability in educational applications.

Additionally, research on human–AI interaction highlights the problem of overreliance on algorithmic systems. Experimental studies show that even when explainable AI is used, individuals have limited ability to detect algorithmic errors, increasing the risk of ethically problematic decisions in educational settings.

Third, policy-oriented studies examine global frameworks such as UNESCO’s AI Ethics Recommendation (2021), the OECD AI principles (2023), and the European Union’s AI Act (2024). While these frameworks provide comprehensive ethical guidance, studies analyzing their application in educational contexts remain limited, often focusing on specific educational levels or offering only descriptive analyses.

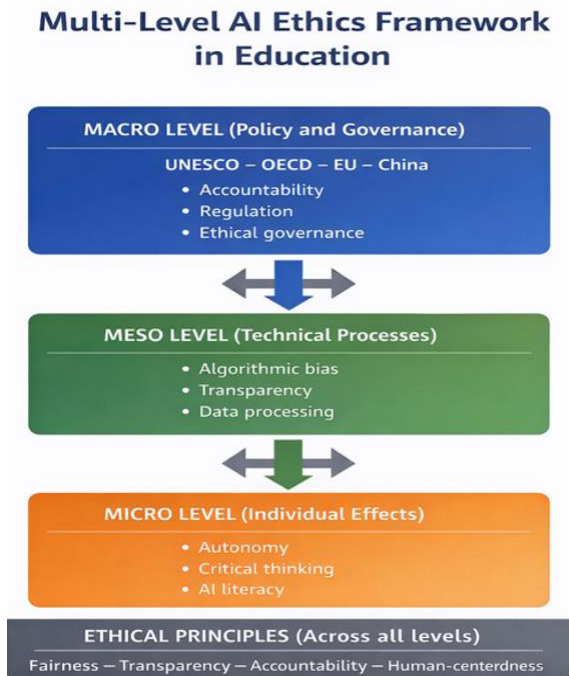
Taken together, the literature demonstrates that AI ethics in education is discussed across micro (learner/teacher), meso (technological systems), and macro (governance/policy) levels; however, integrative

studies that bring these levels together within a unified framework are scarce. Existing research tends to be fragmented, single-dimensional, and insufficiently grounded in ethical theory.

This fragmentation indicates a lack of comprehensive approaches that simultaneously address technical, pedagogical, normative, and political dimensions. To address this gap, the present study integrates ethical theories, international policy documents, and current research findings to provide a critical, comparative, and multi-layered analysis of AI ethics in education, thereby offering an original theoretical contribution to the literature. The proposed multi-level conceptual model is presented in [Figure 1](#). As illustrated, the framework is visually organized in a top-down direction, beginning with the macro level of policy and governance, moving through the meso level of technological systems, and extending to the micro level of learner and teacher outcomes, while the bidirectional relationships between these levels emphasize their reciprocal interaction.

Figure 1

Multi-level conceptual model of AI ethics in education



Note. Figure 1 visualizes a critical and comparative assessment of AI ethics in education by bringing together the technical, pedagogical, normative, and political dimensions that are commonly treated in isolation in the literature.

Purpose and contribution of the study

This study aims to examine the ethical issues arising from the increasing integration of artificial intelligence (AI) in education through a theoretical and critical perspective. In particular, it focuses on key concerns such as algorithmic bias, the erosion of learner autonomy, and the inadequacy of ethical governance mechanisms, and evaluates how these issues are defined and discussed within the framework of international policy documents and philosophical ethical theories. The primary objective of the study is to identify the implementation gaps between existing ethical principles and educational practices, to explore the possible causes of these gaps, and to provide a comprehensive evaluation of AI ethics in education.

In line with this aim, the study seeks to answer the following research questions:

- What are the main ethical risk areas associated with AI applications in education?
- How are algorithmic bias, data privacy, and learner autonomy conceptualized and interpreted within existing ethical frameworks?
- What common principles and differing priorities characterize the ethical governance approaches developed by actors such as UNESCO, OECD, the European Union, China, and Türkiye?
- What gaps and critical areas of debate does the existing literature reveal in the educational context in light of these global frameworks?

Within this scope, the study approaches discussions of AI ethics in education not only at the level of technical standards, but also through a perspective that simultaneously considers moral, pedagogical, and governance dimensions. In doing so, it moves beyond fragmented and one-dimensional approaches in the literature and proposes a multi-level conceptual framework that integrates technical, pedagogical, normative, and political dimensions. In this respect, the study does not merely review existing literature; rather, it offers an original

theoretical contribution by bringing together different ethical theories, international policy approaches, and current research findings within a holistic analytical framework.

Furthermore, the study examines ethical debates surrounding AI in education across micro (learner/teacher), meso (technological systems), and macro (governance and policy) levels, making visible the interactions between these dimensions and addressing a significant conceptual gap in the literature. This comprehensive approach provides a more coherent and applicable foundation for researchers, policymakers, and practitioners to evaluate AI ethics in education.

THEORETICAL FRAMEWORK

Use of artificial intelligence in education

AI applications in education are increasingly being developed to personalize learning experiences, deepen instructional processes, and optimize administrative decision-making. However, this expansion also makes it necessary to redefine ethical boundaries. Global research has particularly focused on personalized learning environments and adaptive assessment systems (Luckin et al., 2016; Zawacki-Richter et al., 2019). AI-based platforms can dynamically adapt content and assessment by monitoring students' knowledge and skill levels in real time. Because these systems require continuous processing of student data, they create new responsibilities regarding privacy protection and data security. In this way, learners can progress at their own pace and receive feedback aligned with their needs. For example, computerized adaptive testing used in higher education entrance examinations or international assessments provides concrete evidence of AI's potential contribution to achievement (Shute & Rahimi, 2017). While such applications support individual learning trajectories, the extent to which data-driven matching and decision algorithms comply with principles of fairness and impartiality remains contested. Indeed, Chen and colleagues (2020) emphasize that one of AI's most significant contributions to education is its capacity to provide personalized learning pathways. Notable differences also emerge across contexts when the geographical distribution and modes of AI use are examined. Educational AI research is concentrated in parts of Asia—particularly China—where studies are often supported by strong state-led initiatives and rely on the analysis of large-scale student data via supervised machine learning approaches (Knox, 2020; Cai & Yin, 2025). Such contexts facilitate access to large datasets and the rapid scaling of innovations. In contrast, in many Western contexts, data privacy and individual autonomy are among the most salient concerns due to a stronger emphasis on human rights and privacy-oriented governance (Roberts et al., 2023; Hine & Floridi, 2024; Xue et al., 2025). This divergence suggests that cultural and political conditions shape how AI is integrated into education and underscores that ethical governance efforts must remain sensitive to contextual differences.

At the pedagogical level, the growing use of AI also raises a paradox. While continuous guidance and immediate feedback may improve short-term performance, these features may constrain deep understanding and creative thinking over time (Luckin et al., 2016; Holmes et al., 2021; Selwyn, 2019). As learners increasingly rely on automated solutions, independent problem solving and critical thinking practices may weaken (Renz & Hilbig, 2020). Some studies further argue that AI-supported environments can limit “productive struggle”—including learning from errors, questioning, and sustained cognitive challenge (Zawacki-Richter et al., 2019). Therefore, the efficiency gains offered by AI must be balanced against education's core aims of cognitive and creative development, with particular attention to protecting autonomy, curiosity, and independent decision making. AI tools are also widely used in instructional support and administrative decision support (Luckin, Holmes, Griffiths, & Forcier, 2016; Holmes, Bialik, & Fadel, 2019). Intelligent Tutoring Systems (ITS) collect and analyze interaction data in real time and provide personalized feedback to support the learning of complex concepts (VanLehn, 2011; Koedinger & Corbett, 2006). Platforms such as Carnegie Learning can provide hints or guidance based on learners' solution steps and function as quasi-personal tutors (Koedinger & Corbett, 2006; Alevan, McLaren, Roll, & Koedinger, 2016). Such systems may reduce teachers' routine assessment and grading workload, allowing more time for high-quality human interaction, including guidance, empathy, and socio-emotional support (Holmes, Bialik, & Fadel, 2019; Baker & Smith, 2019). Accordingly, when appropriately designed, AI may strengthen and transform teachers' roles rather than replace them (Luckin, 2017). This transformation, however, entails new ethical and professional responsibilities: educators must understand how AI tools operate, the limits of their outputs, and potential ethical implications (Baker & Smith, 2019; UNESCO, 2021; Temur, 2025). For effective integration, continuous professional development and AI literacy are therefore critical (UNESCO, 2021; OECD, 2021; Kaman, 2025; Erdoğan & Çakır, 2024). Beyond classrooms, the use of AI in educational institutions' administrative processes is increasing. In areas such as attendance monitoring, student placement, counseling recommendations, and institutional performance tracking, data from learning management systems (LMS) and learning analytics tools are analyzed via AI to support decision-making (Rabhi, Beheshti, & Gill, 2025; Kesgin, 2025). These data-driven approaches can support early warning systems, personalized intervention plans, and more efficient resource allocation (Mishra, Mishra, & Agarwal, 2025; Beheshti et al., 2025). However, because AI-driven recommendations

may shape students' access to educational opportunities, scholarships, or determinations of support needs, algorithmic errors or biases can generate adverse educational and social consequences (Kesgin, 2025; Mishra et al., 2025). In this context, developing transparent and accountable AI systems is a key requirement for strengthening ethical governance in education (Rabhi et al., 2025; Demirtaş, 2024) and directly informs the subsequent discussions of fairness, privacy, and autonomy (Kesgin, 2025; Boztepe, 2025).

Overall, while AI offers substantial advantages in efficiency and accessibility, it also redefines the boundaries of pedagogical autonomy and ethical responsibility. As teachers' decision-making becomes increasingly dependent on algorithmic recommendations, the human element at the core of education may be weakened. Therefore, debates about AI in education should be conducted not only at the level of technological innovation but also within the framework of values education, autonomy, and human-centeredness.

The relationship between artificial intelligence and ethics

The relationship between artificial intelligence and ethics extends beyond a purely technological issue and constitutes a deeply philosophical field of inquiry. AI ethics is commonly examined through three major ethical theories: deontology, consequentialism/utilitarianism, and virtue ethics. These theoretical perspectives offer different criteria and priorities for determining what is considered "ethical," "fair," or "right" in educational AI applications (Dutta, 2025; Irnawati et al., 2024; Oluwatosin, 2025).

Deontology (duty-based ethics)

The deontological approach evaluates the ethical value of an action based on its conformity to universal moral principles and individual duties, regardless of its outcomes. Within this framework, AI applications in education must prioritize students' fundamental rights, privacy, and equality under all circumstances (Irnawati et al., 2024; Stahl et al., 2022). From a deontological perspective, AI systems should not systematically disadvantage any group of students. For instance, even a highly accurate grading algorithm would be considered unethical if it indirectly disadvantages certain socio-economic or cultural groups (Oluwatosin, 2025; Vandemeulebroucke, 2025). Similarly, the processing or sharing of student data without informed consent constitutes a violation of autonomy and privacy rights (Kesgin, 2025; Boztepe, 2025).

This perspective also emphasizes that teachers should not rely solely on algorithmic outputs but must consider students' contextual conditions. Thus, AI systems in education must be structured not only around efficiency but also in alignment with human dignity, justice, and autonomy (Stahl et al., 2022; Vandemeulebroucke, 2025). Accordingly, ethical responsibility must be shared among developers and educators throughout the design, implementation, and evaluation processes (Irnawati et al., 2024; Kesgin, 2025).

Consequentialism/utilitarianism

The utilitarian perspective evaluates the ethical value of an action based on its outcomes, particularly its ability to maximize overall benefit. In education, this approach legitimizes AI applications through measurable gains such as improved learning outcomes, increased efficiency, and enhanced equality of opportunity (Mishra, Mishra, & Agarwal, 2025; Beheshti, Rabhi, & Gill, 2025). AI-based systems that adapt to individual learning needs or reduce teachers' workload are therefore seen as ethically justified in terms of maximizing educational and institutional benefits (Luckin, 2017; Holmes et al., 2019). However, this perspective carries the risk of prioritizing collective benefits over individual rights and privacy (Dutta, 2025; Demirtaş, 2024).

While utilitarianism emphasizes efficiency, performance, and optimization (Rabhi et al., 2025; Stahl et al., 2022), it must be balanced with normative values such as justice, equal access, and autonomy. Research shows that performance-driven AI systems may unintentionally deepen inequalities, particularly for disadvantaged groups (Boztepe, 2025; Vandemeulebroucke, 2025). Therefore, ethical oversight and human-centered design principles are necessary to ensure that benefits are maximized without compromising fundamental rights (Stahl et al., 2022; Beheshti et al., 2025).

Virtue ethics

Virtue ethics evaluates ethical behavior based on the character, intentions, and moral virtues of individuals rather than rules or consequences. In this approach, AI in education is not merely a technical tool but also a medium for fostering character development, responsibility, and social awareness (Irnawati et al., 2024; Dignum, 2022). From this perspective, ethical AI should promote virtues such as justice, honesty, empathy, and wisdom (Vallor, 2016; Holmes et al., 2019). Educational AI systems should not only improve performance but also support ethical reasoning, responsibility, and moral decision-making processes.

This approach requires teachers and system designers to go beyond technical competence and actively contribute to ethical awareness and character development (Boztepe, 2025). It also emphasizes that AI systems should be designed not only for efficiency but for fostering “good human beings” (Vallor, 2016; Coeckelbergh, 2020; Dignum, 2022). Thus, virtue ethics offers a human-centered perspective that transcends purely technical and utilitarian approaches.

Integrated evaluation of ethical theories

Deontology, utilitarianism, and virtue ethics illuminate different dimensions of ethical issues in educational AI. Deontology emphasizes rights, utilitarianism focuses on collective benefit, and virtue ethics highlights character development. Their intersection underscores that AI in education must be not only functional but also morally sustainable. Therefore, ethical governance requires balancing individual autonomy and collective benefit by integrating these perspectives into a coherent framework.

Ethical principles in the use of AI in education

Various international organizations and academic frameworks emphasize that AI in education should be guided by principles such as human dignity, autonomy, justice, and social good (UNESCO, 2021; OECD, 2021; Dignum, 2022). Despite contextual differences, these frameworks converge around core principles: fairness, transparency, and accountability. In education, these principles guide not only technical system design but also teaching practices, data governance, and decision-making processes.

Fairness and non-discrimination

AI systems must ensure equal learning opportunities and avoid reinforcing existing inequalities (Barocas & Selbst, 2016; UNESCO, 2021). Bias detection, diverse datasets, and equitable validation processes are essential.

Transparency and explainability

AI systems must provide understandable explanations of their decision-making processes (Adeoye & Obaze, 2025; Herrera-Poyatos & Peláez-González, 2025). Explainable AI (XAI) enhances trust and accountability (Brohi et al., 2025; Sharma & Sharma, 2025). However, transparency must be balanced with privacy and data protection concerns (Talapina, 2025; Bannor & Baysah, 2025). Particularly in education, excessive opacity reduces trust and hinders adoption.

Accountability, human-centeredness, and AI literacy

Accountability requires clearly defining responsibility for AI-generated outcomes (Jobin, Ienca & Vayena, 2019; Hagendorff, 2020). In education, where decisions directly affect individuals, responsibility must be transparent and enforceable (Floridi et al., 2022). Mechanisms such as appeals processes, independent audits, and ethical impact assessments are essential (Bannor & Baysah, 2025; Jobin et al., 2019). The EU AI Act (2024) reinforces these requirements for high-risk systems (European Commission, 2024; Parlak, 2025).

Human-centered approach

A human-centered approach emphasizes that AI should never replace human decision-making (UNESCO, 2021; Floridi & Cows, 2021). In education, final decisions must remain under human supervision. For example, automated grading systems should be reviewed by teachers, ensuring that AI functions as a supportive tool rather than a replacement (Brohi et al., 2025; Dadson et al., 2025). This principle preserves the central role of educators and prioritizes human development (Hagendorff, 2020).

AI literacy and awareness

AI literacy is both a technical and ethical competence (UNESCO, 2021; OECD, 2023). It involves understanding how AI systems collect, process, and interpret data, as well as recognizing potential biases (Adeoye & Obaze, 2025). Teachers must integrate AI tools pedagogically while understanding their limitations (Rampilla et al., 2025). Many countries are incorporating AI ethics, data literacy, and digital citizenship into curricula (OECD, 2023). Ultimately, ethical principles must be translated into practical governance frameworks. This ensures that AI in education remains aligned with global ethical standards while being adaptable to local contexts, thereby supporting sustainable ethical governance (Bannor & Baysah, 2025).

CONCEPTUAL SCOPE ANALYSIS

Ethical risks and challenges of AI use in education

The integration of artificial intelligence into educational environments promises increased efficiency and personalization in pedagogical processes; however, it simultaneously introduces significant ethical challenges. These issues are often shaped not by abstract normative debates but by concrete risks encountered in practice. One of the most prominent ethical risk categories in this context is algorithmic bias and discrimination.

Algorithmic bias and discrimination

AI systems used in education are typically trained on historical student data or institutional performance indicators. However, such datasets often reflect existing societal inequalities. Barocas and Selbst (2016) define algorithmic bias as systematic distortions in data that lead to discriminatory outcomes against certain groups. For example, if students from socioeconomically disadvantaged schools historically demonstrate lower academic performance, AI systems trained on such data may automatically classify similar students as “high-risk,” thereby reproducing inequality (Crawford, 2021). The sources of bias are multidimensional, including unrepresentative datasets, narrow performance metrics, and implicit biases embedded by developers (Mehrabi et al., 2021). Moreover, many machine learning systems tend to conflate statistical correlation with causation, leading to the neglect of individual potential and the reduction of students to standardized profiles (Selwyn, 2019). Addressing these risks requires more than technical solutions. While fairness-oriented machine learning techniques—such as reweighting, resampling, or post-hoc adjustments—may offer short-term mitigation, long-term solutions necessitate institutional oversight (Binns, 2018). The FAT framework (Fairness, Accountability, Transparency) therefore implies not only technical but also governance-based interventions. Educational institutions must conduct periodic bias audits and evaluate their outcomes across stakeholders. In some regulatory contexts, AI-based educational tools are required to submit discrimination impact reports prior to deployment (European Commission, 2024). Thus, algorithmic bias must be understood as a multidimensional issue encompassing technical, legal, pedagogical, and governance dimensions.

Data privacy and surveillance

The expansion of AI in education has intensified ethical concerns regarding the collection, processing, and sharing of student data. Digital learning environments generate extensive datasets ranging from academic performance to behavioural and emotional indicators (Carrillo & Flores, 2025). While this enables personalized learning, it also risks eroding privacy and institutionalizing surveillance practices. The erosion of privacy is not only a matter of individual rights but also a threat to trust in educational systems (Nguyen, 2025). Continuous monitoring may lead to self-censorship and behavioural conformity, creating a “digital surveillance culture” within learning environments (Hosen, 2025). Additionally, concerns regarding data commercialization and third-party access further exacerbate ethical risks (Sambhaji, 2025). Contemporary approaches emphasize governance-based solutions beyond technical fixes. Privacy-by-Design principles advocate embedding privacy into system architecture, minimizing data collection, anonymizing data, and ensuring local processing (Demir, 2025). Ethical data governance also requires restricted access, limited retention periods, and informed consent mechanisms (Joshua, Micciancio & Gilbert, 2025). At the institutional level, regular ethical audits and cybersecurity integration are essential (Rivera & Osen, 2025). Without such measures, data breaches may undermine institutional legitimacy and lead to broader ethical crises. Therefore, AI-driven data management in education must be conceptualized as a multidimensional governance issue intersecting ethical, legal, and pedagogical domains (Putri, Salim & Inayati, 2025).

Erosion of learner autonomy and critical thinking

The increasing use of AI in education not only transforms instructional processes but also reshapes the cognitive and ethical position of learners. One major concern is the weakening of learner autonomy and critical thinking capacity (Azman, Musidi & Azhar, 2025). AI systems determine what, when, and how students learn based on algorithmic parameters, potentially reducing learners to data-driven objects (Izquierdo-Condoy & Arias-Intriago, 2025). Continuous guidance and real-time feedback may transform learning into an “automated cognitive process,” limiting reflection, questioning, and alternative problem-solving (Ninghardjanti et al., 2025). This creates a form of pedagogical heteronomy, where students become passive recipients of algorithmic direction.

Surveillance culture and cognitive conformity

AI-driven learning analytics systems contribute to an invisible surveillance culture by continuously monitoring student behavior. This environment reduces psychological safety and discourages cognitive risk-taking (Prosper, 2025). As students align their behavior with algorithmic expectations, learning environments may become conformist rather than exploratory. However, critical thinking requires space for error, experimentation, and cognitive conflict. Over time, the corrective nature of AI systems may weaken reflective thinking capacities (Pantano, Essien & Kremantzis, 2025).

Reconstructing learner autonomy

Protecting learner autonomy is a central ethical responsibility in contemporary education. Hybrid pedagogical models aim to balance AI-supported learning with human-centered elements (Meylani, 2025). Design strategies that allow learners to set their own goals and learning pathways—rather than relying solely on automated systems—help reestablish autonomy (Ismail, 2025). Such approaches reposition students as active agents rather than passive data objects.

Digital well-being and critical AI literacy

Promoting digital well-being has become an ethical necessity. Practices such as “screen-free days” and “technology breaks” help restore cognitive flexibility and awareness (Cedo et al., 2025). Moreover, learners must develop not only technical skills but also critical awareness of AI systems. Education on algorithmic bias, data ethics, and AI literacy fosters informed and reflective engagement with technology (Wu & Liu, 2025). Ultimately, AI in education should be redefined as a tool that supports—not replaces—critical thinking and learner autonomy. Otherwise, the advancement of digital pedagogy risks weakening the process of subject formation rather than deepening learning.

COMPARATIVE ANALYSIS

International AI ethics and its implications for education

The rapid global development of artificial intelligence has positioned ethical governance, human rights, and digital responsibility at the center of education policy debates. AI use in education raises multidimensional concerns—such as equity of opportunity, data privacy, learner autonomy, and algorithmic transparency—making international, comprehensive ethical frameworks increasingly necessary. In this landscape, international organizations (e.g., UNESCO, OECD, and the European Union) provide global reference standards, while countries such as Türkiye and China develop governance approaches shaped by their own cultural, legal, and institutional contexts.

UNESCO recommendation on the ethics of artificial intelligence (2021) and educational standards

UNESCO published the first comprehensive global ethics governance document in this field through the Recommendation on the Ethics of Artificial Intelligence, adopted by Member States in 2021 (UNESCO, 2021). The Recommendation places human dignity, human rights, justice, and sustainable development at the foundation of AI policies.

In education, UNESCO explicitly foregrounds awareness and AI literacy, encouraging Member States to integrate components such as algorithmic awareness, data ethics, and digital citizenship into curricula. The Recommendation further emphasizes that AI-enabled learning systems should be designed to be inclusive, ethically grounded, and subject to human oversight. UNESCO also underscores the importance of preparing teachers not only as technology users but as ethically informed digital guides, aiming to prevent AI from becoming a mechanism that reproduces—rather than mitigates—educational inequalities.

OECD AI principles and the AI literacy agenda

The OECD established one of the most influential international policy frameworks with its AI Principles, first published in 2019 (OECD, 2019). These principles emphasize human-centeredness, accountability, safety, transparency, and inclusive growth. From an education policy perspective, OECD initiatives increasingly stress that learners should become not merely technology users but ethically informed digital citizens capable of critical engagement in AI-mediated environments. Through its Artificial Intelligence and Education & Skills platform, OECD develops policy guidance on AI literacy, teacher competencies, and ethical awareness (OECD, 2025). This

agenda includes recommendations to strengthen teachers' professional development in AI pedagogy while fostering students' critical thinking, algorithmic awareness, and data ethics capacities.

The European Union: AI Act (2024) and the AILit framework (2025)

AI Act (2024): A risk-based governance model

With the adoption of the Artificial Intelligence Act (AI Act) in 2024, the European Union introduced the world's first binding AI regulatory regime (European Commission, 2024). The Act classifies AI systems into four categories based on risk level: unacceptable, high, limited, and minimal risk. Because educational AI systems can directly affect individuals' opportunities and developmental rights, the education sector is treated as high-risk. For systems used in contexts such as exam grading, scholarship assessment, or student behavior analytics, the AI Act introduces obligations including:

- Algorithmic transparency and traceability,
- Bias/Discrimination testing of datasets,
- Integration of human oversight,
- Pre-Deployment risk and conformity assessments.

In addition, the Act restricts certain practices in educational settings—particularly emotion recognition, behavioral scoring, and biometric analysis—prioritizing the protection of students' personal and affective data. This signals a distinctly human-rights-centered understanding of AI governance in Europe (European Parliament, 2024).

AILit framework (2025): Institutionalizing ethical literacy

To connect legal governance with ethical awareness and pedagogical competence, the EU—together with OECD and the European Commission—published “Empowering Learners for the Age of AI: An AI Literacy Framework for Primary and Secondary Education (AILit Framework)” in 2025 (OECD & European Commission, 2025). The framework conceptualizes AI literacy across four dimensions:

Engage: the ability to interact with AI systems consciously, safely, and ethically.

Create: the ability to use AI responsibly for innovation and production.

Manage: the ability to assess limits, risks, and societal impacts.

Design: the capacity to question and improve the ethical dimensions of algorithms.

Positioned as a bridge between the AI Act (2024) and the Digital Education Action Plan (2021–2027), the AILit Framework argues that ethical evaluation cannot be separated from technical competence. It also calls for embedding ethical decision-making, algorithmic awareness, and human-centered pedagogical design into teacher education programs—reflecting a policy shift from a primary focus on technological efficiency toward ethical consciousness.

China's ethical approach: A state-centered model and emerging updates

China's New Generation Artificial Intelligence Development Plan (2017) aims to achieve global leadership in AI by 2030 and positions education as a strategic domain for AI deployment (State Council of China, 2017). AI-based learning systems are widely used at scale, supported by strong policy coordination. China's ethical approach, however, differs from many Western rights-based models by prioritizing collective benefit and social order, which shapes the balancing of privacy and individual autonomy. The Technology Ethics Guidelines draft published in 2024 is presented as an effort to strengthen institutional ethics governance by emphasizing human-centeredness, safety, fairness, and accountability (Geopolitechs, 2024). Nevertheless, transparency in data use and surveillance-related practices remains comparatively limited. As a result, China's model may be considered advantageous in terms of speed and scale yet ethically contested due to constrained flexibility and accountability.

Türkiye: AI ethics and education policies after 2025—A governance paradigm

In recent years, Türkiye has begun institutionalizing ethical governance in the education ecosystem through strategic policy documents. The Ministry of National Education (MEB) published AI Ethics Recommendations in 2025, articulating a human-centered, values-based, and trustworthy AI orientation in education (MEB, 2025). This document functions not merely as a technical compliance guide but as a normative governance framework that jointly addresses pedagogical, ethical, and cultural dimensions of digital transformation.

Conceptual foundations

The framework is built around fairness, transparency, accountability, privacy, and human dignity, aligning closely with UNESCO's Recommendation (UNESCO, 2021) and OECD AI Principles (OECD, 2019, 2024). Türkiye positions these universal norms within the context of the "Türkiye Yüzyılı Maarif Modeli", placing the virtue–value–action triad at the center of education digitalization policy and integrating national-moral values with universal human rights principles.

Institutional structure and governance model

Türkiye's governance trajectory is shaped by the National AI Strategy (2021–2025) and the AI Policy Document and Action Plan in Education (2025–2029). These documents align with national development priorities and envision coordination mechanisms for developing AI applications that are reliable, fair, and accountable. Within this structure, YEGİTEK plays a key role in strengthening teachers' digital competencies, delivering ethics-awareness training, and ensuring pedagogical alignment audits of AI-supported learning environments. The framework adopts a multi-stakeholder governance logic by defining differentiated ethical responsibilities for students, teachers, parents, administrators, and developers (MEB, 2025).

Data protection, transparency, and accountability

MEB (2025) identifies personal data protection as a critical pillar of ethical AI in education, emphasizing data minimization, privacy by design, and privacy by default. It calls for clear definitions of data controller/processor roles and restricts the processing of sensitive student data to educational purposes. This orientation is supported by the Personal Data Protection Authority's 2025 recommendations on personal data protection in AI. In parallel with the EU AI Act's high-risk framing of educational applications (European Commission, 2024), Türkiye foregrounds data security as a priority ethical governance domain.

Stakeholder-focused ethical guidance

A distinctive contribution of the MEB (2025) document is that it provides practical guidance beyond the policy level:

- For students: privacy awareness, digital footprint literacy, and critical AI literacy;
- For teachers: transparency, algorithmic fairness awareness, content appropriateness, and human oversight;
- For administrators: data breach management, ethics audit protocols, and stakeholder participation;
- For parents: awareness of digital surveillance, data-sharing boundaries, and a child-rights perspective.

International alignment and local identity

Türkiye's model aligns with UNESCO/OECD principles while emphasizing a locally grounded ethical axis of rights–justice–virtue. The MEB (2025) emphasis on positioning technology as a tool in service of humanity resonates with the EU AI Act's human oversight requirement (European Commission, 2024), yet Türkiye embeds this principle within a broader civilizational-humanistic educational philosophy. As such, Türkiye's framework stands out as a model that targets not only technical compliance but also cultural sustainability.

Finally, Türkiye's ethical governance is likely to be further shaped by legislative initiatives, including the AI law proposal submitted to the Grand National Assembly in 2025, which addresses transparency, data security, and ethical responsibility at the public-institution level and indirectly informs educational practice. In the coming period, strengthening this framework through formal legal regulation (e.g., an AI law draft under preparation) and institutionalizing educators' ethical awareness will be decisive for sustainability.

Toward a multi-level era of ethical governance

Taken together, UNESCO's universal principles (UNESCO, 2021), OECD's accountability-oriented policy agenda (OECD, 2019; OECD, 2025), the EU's binding regulatory framework and literacy institutionalization (European Commission, 2024; European Parliament, 2024; OECD & European Commission, 2025), China's state-centered approach (State Council of China, 2017; Geopolitechs, 2024), and Türkiye's value-based governance model (MEB, 2025) demonstrate that AI ethics in education has become a multi-level governance field rather than a single-center policy domain. This diversity suggests that ethical governance must balance universal principles with local cultural and institutional dynamics. Education constitutes one of the most visible "laboratories" of this transformation, as AI directly shapes learning opportunities, data practices, and autonomy conditions. Therefore, future education policies should be grounded not only in innovation but also in ethical awareness, human dignity, and social justice.

Future research should empirically examine how these governance frameworks affect teacher education, school policies, and student participation. Such work would clarify how normative principles operate in practice and support evidence-based policy design for integrating national and international ethics standards into educational systems.

METHOD

Methodological approach: Critical literature and policy analysis

This study employs document analysis and a theory-driven critical integrative review within the framework of a qualitative research approach. Document analysis is a qualitative method based on the systematic examination, classification, and interpretation of written and digital materials relevant to the research purpose (Bowen, 2009). This method is widely used as an effective research strategy, particularly in the analysis of normative frameworks, policy documents, and theoretical texts.

This research does not rely on empirical data collection; rather, it provides a critical and integrative conceptual analysis based on theoretical studies, international policy documents, and peer-reviewed academic publications in the field of AI ethics in education. In this respect, the study adopts a critical literature review approach that goes beyond mere description, questioning the assumptions, limitations, and implicit normative positions of existing studies. According to Grant and Booth (2009), a critical review aims to generate new conceptual insights by filtering the literature through an analytical lens and to make a theoretical contribution to the field. Following this approach, the present study evaluates the literature on AI ethics in education from a holistic and critical perspective.

Within the scope of the study, international academic articles, ethical guidelines, policy reports, and regulatory documents published between 2010 and 2025 and directly related to AI ethics in education were examined. The literature search was conducted using the keywords “artificial intelligence,” “AI ethics,” “education,” “algorithmic bias,” “student autonomy,” and “AI governance” in the Web of Science, Scopus, and Google Scholar databases. In addition to peer-reviewed academic literature, national and international policy and grey literature documents were identified through the official institutional repositories and websites of the relevant organizations and public authorities, including UNESCO, the OECD, the European Commission and European Parliament, the State Council of China, and the Republic of Türkiye Ministry of National Education (MEB). English-language policy documents and Turkish-language official documents relevant to the Türkiye case were included; where official English versions were available, these were prioritized to facilitate cross-national comparison.

The selection of studies was based on predefined inclusion and exclusion criteria. The inclusion criteria were: (i) direct relevance to AI ethics in education, (ii) inclusion of key themes such as algorithmic bias, data privacy, student autonomy, and governance, and (iii) providing a theoretical or empirical contribution. Studies focusing solely on technical aspects, lacking an educational context, or only indirectly related to the topic were excluded. Additionally, to enhance the representativeness of the literature, highly cited studies and those frequently referenced in the field were prioritized.

The analysis process followed a three-stage structure:

- 1. Theoretical Grounding:** In the first stage, the literature on AI ethics in education was conceptually classified based on deontological, utilitarian, and virtue ethics approaches. The normative implications of these ethical theories in the educational context were analyzed. This stage establishes the theoretical foundation of the study and provides a conceptual reference framework for subsequent analyses.
- 2. Conceptual Scope Analysis:** In the second stage, the prominent ethical risk areas identified in the literature were examined. In this context, themes such as algorithmic bias, data privacy and surveillance, student autonomy, transparency, and accountability were systematically analyzed, and their implications for educational practices were evaluated from a critical perspective. These themes were categorized based on recurring conceptual patterns.
- 3. Comparative Policy Analysis:** In the final stage, major international policy frameworks—including UNESCO’s Recommendation on the Ethics of Artificial Intelligence (2021), the OECD AI Principles (2023), the European Union Artificial Intelligence Act (AI Act, 2024), China’s recent ethical guidelines (2024), and current regulatory frameworks in Türkiye—were comparatively analyzed. Their approaches, priorities, and limitations in the educational context were examined to identify converging and diverging aspects in terms of ethical governance.

Validity and Reliability

To ensure validity, data triangulation was achieved by utilizing multiple sources, including academic publications and policy documents. In addition, the themes identified during the analysis were cross-checked with existing studies to ensure consistency. To enhance transparency, the data collection and analysis processes were reported in detail.

FINDINGS

The critical literature and policy analysis conducted in this study reveals that research on AI ethics in education is concentrated around three main analytical axes: (i) pedagogical ethics and ethics education, (ii) technological–algorithmic risks, and (iii) global governance and policy frameworks. However, the findings indicate that these axes are largely addressed in isolation in most studies, and that an integrated ethical approach linking micro, meso, and macro levels has not yet been fully developed.

Theme 1: Limited measurement and lack of theoretical depth in pedagogical ethics discussions

The pedagogical literature demonstrates that AI ethics has increasingly emerged as a visible area of learning, particularly in higher education contexts. The systematic review by Wiese, Liu, and Vosoughi shows that ethics education enhances students' awareness and ethical reasoning skills. However, the findings of this analysis indicate that the measurement approaches used to evaluate the effectiveness of ethics education are largely short-term, based on self-reports, and lack theoretical consistency. Despite the growing body of content in the pedagogical domain, this situation reveals a lack of comprehensive models capable of explaining the behavioral and long-term outcomes of ethics education.

Theme 2: Disconnection of technological and algorithmic risks from the pedagogical context

Studies within the technological domain provide detailed analyses of risk areas such as algorithmic bias, data privacy, transparency, and human oversight. The systematic review by Collin, Lepage, and Nebel demonstrates that the opacity of deep learning–based systems can weaken accountability and human oversight in educational contexts. However, the findings reveal that most of these studies fail to systematically link technological risks with pedagogical processes, student autonomy, and teacher decision-making. Experimental findings by Janssen and colleagues further show that even when explainable AI is used, users struggle to detect algorithmic errors, thereby increasing the risk of over-reliance on AI in educational settings. Nevertheless, the educational implications of these risks are often treated as secondary in the literature.

Theme 3: Structural disconnection between global ethical governance frameworks and educational practices

Policy- and governance-oriented literature provides detailed descriptions of comprehensive ethical frameworks developed by actors such as UNESCO, OECD, the European Union, China, and Türkiye. However, the findings indicate that most studies address these policy documents at a descriptive level and offer limited discussion of their concrete implications for educational institutions, teaching practices, and student experiences. While reviews such as PRISMA-Ethics shed light on specific educational levels, they do not provide a comprehensive model explaining how global ethical principles are translated into pedagogical and institutional governance processes.

Overall Evaluation of the Findings

The findings of this study demonstrate that the literature on AI ethics in education is rich yet fragmented. While pedagogical, technological, and policy-oriented studies show depth within their respective domains, theoretical and analytical integration across these areas remains insufficient. Micro-level experiences of students and teachers are not systematically linked with meso-level algorithmic and institutional processes, nor with macro-level global ethical and governance frameworks.

This fragmentation clearly indicates the need for a multi-layered analytical model that integrates ethical theories, technological processes, and policy frameworks within the literature on AI ethics in education. In this context, the conceptual model proposed in this study aims to bridge this fragmentation and address a significant gap in the literature.

DISCUSSION

Studies conducted in the pedagogical domain indicate that AI ethics education has rapidly emerged as a growing field, particularly in higher education. The systematic review by Wiese, Liu, and Vosoughi demonstrates that ethics-

related content enhances students' awareness and ethical reasoning skills; however, it also reveals widespread methodological limitations in how these effects are measured (Wiese, Liu, & Vosoughi, 2024). In particular, the lack of theoretically grounded and methodologically robust frameworks capable of longitudinally tracking changes in ethical awareness, attitudes, and behaviors represents a major gap in the pedagogical literature. This highlights the need for the development of theoretically informed, culturally responsive, and long-term measurement approaches to evaluate the impact of ethics education.

At the technological–algorithmic level, studies provide detailed accounts of the ethical risks associated with AI applications in educational contexts; however, they often address the relationship between these risks and pedagogical or managerial dimensions only to a limited extent. Findings by Collin, Lepage, and Nebel indicate that the opaque nature of deep learning–based systems weakens accountability and creates new ethical dilemmas by limiting human oversight in educational decision-making processes (Collin, Lepage, & Nebel, 2023). Similarly, the study by Janssen, van der Voort, and van Veenstra shows that even when explainability is provided, users struggle to detect algorithmic errors. This increases the risk of over-reliance on AI in educational contexts and may lead to pedagogically problematic decision-making processes (Janssen, van der Voort, & van Veenstra, 2020). Despite these insights, such risks are rarely systematically connected to student autonomy, teacher decision-making, and learning processes.

At the macro level, the policy and governance literature demonstrate that global ethical frameworks offer comprehensive guiding principles. However, the analysis by Medaglia, Milisavljević, and Ferreira reveals that the mechanisms for implementing these principles at the institutional level remain unclear in many countries (Medaglia, Milisavljević, & Ferreira, 2021). Similarly, while the principles articulated by UNESCO (2021), OECD (2023), and the European Union (2024) are grounded in strong normative foundations, guidance on how these principles should be operationalized within educational institutions remains limited. This situation points to a structural disconnect between policy frameworks and educational practices.

Findings related to learner experiences further support this multi-layered disconnection. The study by Lim, Gottipati, and Cheong shows that students are particularly sensitive to issues of fairness, transparency, and cognitive load in algorithmic assessment systems, indicating a clear mismatch between technological design and learner needs (Lim, Gottipati, & Cheong, 2022). In addition, Smallman's multi-scalar ethical framework emphasizes that AI produces different effects at individual, institutional, and societal levels, suggesting that ethical evaluations cannot be reduced to a single level of analysis (Smallman, 2023). However, the educational literature has not sufficiently internalized this multi-level perspective.

Taken together, these findings indicate that while the literature on AI ethics in education has deepened along three primary axes—pedagogical, technological, and policy-oriented—there remains a lack of theoretical and analytical integration across these domains. The tendency of studies to remain confined within their respective boundaries results in a fragmented structure across micro, meso, and macro levels.

In this context, the multi-layered conceptual model proposed in this study aims to address this fragmentation by offering an integrative framework. The model links technological risks with pedagogical processes, pedagogical outcomes with governance frameworks, and ethical theories with practical implementation, thereby enabling a more holistic analysis of AI ethics in education. This integrative approach not only contributes theoretically but also provides a conceptual foundation for educational institutions to manage AI-driven transformation processes in a more ethical, sustainable, and responsible manner.

CONCLUSION

The use of artificial intelligence in education offers significant opportunities, such as personalizing instructional processes and enriching learning experiences. However, as highlighted in the literature, this transformation also brings serious ethical challenges. The analysis conducted in this study reveals that issues such as algorithmic bias, the erosion of student autonomy, data privacy and surveillance risks, and the necessity of ethical governance principles are particularly prominent (Varol, 2025, p. 121). These challenges should be addressed not only from a technical perspective but also through pedagogical and managerial dimensions. Both international frameworks and developments in the Turkish context emphasize the necessity of implementing AI in education through a human-centered, fair, and transparent approach (UNESCO, 2021; MEB, 2025).

First, in terms of data privacy, AI-supported systems that collect large-scale data may threaten students' privacy. In particular, the collection and monitoring of children's personal data can raise serious concerns that may undermine trust if misused (Selwyn, 2021). The increasing prevalence of surveillance in educational environments may lead students to feel constantly monitored, resulting in negative psychological consequences. Therefore, it is essential to strictly comply with existing data protection laws and adopt transparent data governance policies in the development and use of AI applications. It is necessary to clearly specify who uses student data and for what purposes, ensure effective consent mechanisms, and establish accountability structures to address potential data breaches.

Second, algorithmic bias emerges as a critical threat to the principles of fairness and inclusivity in education. When AI algorithms are used as decision-support tools in various educational contexts, they may systematically disadvantage certain student groups if trained on biased datasets. Indeed, biased algorithms have been shown to create discrimination across different socio-cultural groups and deepen inequalities of opportunity (O'Neil, 2016). This poses a significant risk, particularly for disadvantaged students. Therefore, diversity and inclusivity principles must be prioritized during the development of algorithms, and systems should be regularly audited for potential bias. Transparency must be ensured to make algorithmic decisions explainable and accountable; educational institutions and developers should provide clear explanations of how AI models function, enabling stakeholders to critically evaluate these systems. Otherwise, "black box" decision-making processes may lead to a loss of trust among students and parents.

Third, student autonomy and, more broadly, the role of the human element remain critical values that must be preserved in AI-enhanced learning environments. Research indicates that excessive exposure to automation in learning processes may weaken students' sense of control over their own learning. Over-reliance on algorithms may reduce teachers' influence in curriculum design and assessment processes, while overly structured learning pathways may hinder creativity and critical thinking. While some scholars argue that AI may replace teachers, others emphasize that AI should not substitute human interaction but rather function as a supportive tool for educators (Holmes et al., 2019). Education is not merely a process of knowledge transmission; it also encompasses emotional and social learning, guidance, and inspiration. Therefore, AI applications should be designed to complement the roles of teachers and students, with humans remaining in supervisory and decision-making positions. This approach helps prevent students from being reduced to passive data points and supports the preservation of their autonomy and self-regulation skills. While AI enhances efficiency and accessibility, it also redefines the boundaries of pedagogical autonomy and ethical responsibility. Consequently, discussions on AI in education should extend beyond technological innovation and be framed within the principles of values education, autonomy, and human-centeredness.

Finally, at the intersection of all these discussions lies the necessity of placing ethical governance principles at the core of AI integration in education. When examining the frameworks developed by international organizations, principles such as transparency, accountability, fairness, respect for privacy, and human-centeredness emerge as common values (UNESCO, 2021). In line with these principles, the goal should be to ensure the use of AI in ways that protect human dignity and rights while promoting societal well-being. Reports by organizations such as UNICEF, OECD, and UNESCO, as well as frameworks proposed by Floridi and colleagues, emphasize the need for a multi-layered ethical approach to ensure that AI in education remains fair, transparent, and human-centered.

Recent policy developments in Türkiye largely align with these global trends. The principles outlined in UNESCO's Recommendation on the Ethics of Artificial Intelligence (2021) guide the development of national strategies. Accordingly, recent policy documents and action plans include significant steps toward integrating AI technologies into the education system in an ethical, reliable, and inclusive manner (MEB, 2025). This approach is noteworthy in that it adapts internationally recognized ethical principles to the national context and translates them into practical applications.

In summary, the findings of this study indicate that the success of AI in education depends not only on technological innovation but also on adherence to ethical principles and a multi-stakeholder governance approach. The responsible and human-centered adoption of AI in education requires efforts to ensure that algorithmic decisions are explainable and fair, that student privacy and autonomy are protected, and that technology is directed toward pedagogical purposes.

Recommendations

Addressing ethical challenges in AI in education requires a holistic approach involving policymakers, educators, and technology developers. Based on the findings of this study and the existing literature, the following recommendations are proposed:

For policymakers (government and regulatory bodies)

A comprehensive ethical framework for AI use in education should be established and supported by legal regulations. Ethical guidelines developed in line with international best practices should define concrete criteria for transparency, fairness, accountability, and privacy protection (UNESCO, 2021). Strong legislation and monitoring mechanisms should be implemented, particularly in the domain of data privacy, ensuring compliance with standards such as KVKK and GDPR. Ethical committees within the Ministry of National Education should review AI projects prior to implementation and conduct approval processes to mitigate algorithmic bias and security risks.

Policymakers should also set age-appropriate boundaries for AI use (e.g., restrictions for children under 13) and encourage all educational institutions to develop AI usage policies. A balanced approach that aligns universal

ethical frameworks with national values should be adopted to establish context-sensitive ethical AI standards. Furthermore, equitable resource allocation and infrastructure investments should be planned to ensure that disadvantaged schools and regions have access to AI-supported educational technologies, thereby preventing the deepening of the digital divide.

For educators (school administrators and teachers)

Teachers and school administrators should act as critical filters in integrating AI tools into educational environments. Therefore, enhancing AI literacy and ethical awareness among educators is essential. Through in-service training programs and workshops, teachers should be informed about the fundamental principles of AI, potential biases, and privacy risks, as well as how to use these technologies pedagogically (Holmes et al., 2019; Dignum, 2021).

Educators should critically evaluate AI-generated outputs and balance them with human judgment when necessary. Particularly in student assessment and guidance, teachers should integrate AI outputs with their professional expertise to act in the best interest of students. To support student autonomy, AI tools should be used in ways that enhance creativity, problem-solving, and self-regulation skills, ensuring that students do not become passive consumers in the learning process.

Additionally, school administrations should inform students and parents about how AI systems function, how data are used, and their potential implications, thereby fostering transparency and trust. Each institution should establish context-specific AI ethics principles (e.g., ethical agreements, consent forms, complaint mechanisms) and integrate them into practice.

For technology developers (AI designers and companies)

Developers of educational AI applications should integrate ethical design principles from the outset. In this regard, privacy by design and fairness by design should be fundamental principles. Developers must ensure demographic and cultural diversity in datasets and algorithms to prevent systemic advantages or disadvantages for any group.

Independent ethical audits should be conducted during the model development process, and algorithms should be tested for potential biases, discriminatory effects, and unintended consequences. In line with transparency principles, developers should provide clear and understandable explanations of how AI systems function (e.g., decision criteria, recommendation mechanisms).

The adoption of Explainable AI (XAI) techniques will facilitate understanding of algorithmic outputs by teachers and students, thereby building trust. Developers should also establish mechanisms to take responsibility for incorrect or harmful outputs (e.g., feedback channels, correction protocols, remediation processes).

Finally, ethical certification standards can be developed by professional organizations or government bodies to ensure that educational technologies meet minimum ethical requirements before being introduced to the market. In the long term, adopting an innovation approach that prioritizes societal benefit will be essential for ensuring the sustainability and acceptability of AI in education.

Acknowledgement

The author has no acknowledgements to declare.

Funding

This research received no external funding.

Ethical statement

This study was based exclusively on the analysis of published literature, policy documents, institutional reports, and publicly available materials. It did not involve human participants, animals, personal data collection, or experimental interventions. Therefore, institutional review board or ethics committee approval and informed consent were not required.

Competing interests

The author declares no competing interests.

Author contributions

The author was solely responsible for all aspects of the study, including conceptualization, research design, literature review, document selection and analysis, interpretation of findings, development of the theoretical framework, writing of the original draft, reviewing, editing, and final approval of the manuscript.

Data availability

No primary or original datasets were generated or analyzed in this study. The study is based on publicly available academic publications, policy documents, institutional reports, and other sources cited in the manuscript. All materials supporting the analysis are identified in the reference list.

AI disclosure

Generative artificial intelligence tools were used solely for formatting the reference list and adapting the references to the journal's required citation style. The scientific content, analysis, interpretations, and conclusions of the manuscript were prepared by the author.

Biographical sketch

Aylin Karakuş is a Lecturer with a PhD at the Career Planning Application and Research Center of Gazi University, Türkiye. She earned her PhD in Social Studies Education from Gazi University. Her doctoral research focused on artificial intelligence-supported teaching processes and included the development of Social AI. Her research broadly centers on artificial intelligence in education, AI ethics, AI literacy, attitudes and perceptions toward AI, scale development, educational technology, and social studies education. She also conducts studies on the pedagogical, ethical, and governance dimensions of artificial intelligence in educational contexts.

Disclaimer/Publisher's Note

The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and do not necessarily reflect the views of Lectito Publications and/or the editor(s). Lectito Publications and/or the editor(s) disclaim responsibility for any injury to persons or property resulting from any ideas, methods, instructions, or products referred to in the content.

REFERENCES

- Adeoye, A. E., & Obaze, C. A. (2025). Explainable artificial intelligence (XAI): A comprehensive review of methods, applications, and open issues. *Tech-Sphere Journal for Pure and Applied Sciences*, 2(1), 25–41. <https://doi.org/10.5281/zenodo.17237252>
- Ahmed, S., Rahman, M. S., Kaiser, M. S., & Hosen, A. S. M. S. (2025). Advancing personalized and inclusive education for students with disability through artificial intelligence: Perspectives, challenges, and opportunities. *Digital*, 5(2), Article 11. <https://doi.org/10.3390/digital5020011>
- Ahn, J. W., Han, J. B., & Choi, H. K. (2021). AI ethics in education: A framework for policy and practice. *Journal of Educational Technology Development and Exchange*, 14(2), 19–34.
- Aleven, V., McLaren, B. M., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16(2), 101–128.
- Azman, M. A. M. B., Musidi, N. B., & Azhar, N. B. B. K. (2025). The impact of artificial intelligence on language learners' critical thinking skills, language learning, and academic integrity. *British Journal of Research in Education*, 9(24), 817–824. <https://ideas.repec.org/a/bjef/journal/v9y2025i24p817-824.html>
- Baker, R. S., & Inventado, P. S. (2014). Analyzing the ethical implications of using AI in education. *International Journal of Artificial Intelligence in Education*, 24(4), 548–563.
- Baker, T., Smith, L., & Anissa, N. (2019). *Educ-AI-tion rebooted? Exploring the future of artificial intelligence in schools and colleges*. Nesta. https://media.nesta.org.uk/documents/Future_of_AI_and_education_v5_WEB.pdf
- Bannor, F. O., & Baysah, J. O. (2025). Ethical leadership challenges in the age of artificial intelligence: An in-depth analysis. *Pan-African Journal of Education and Social Sciences*, 12(2), 87–105. <https://journals.aua.ke/pajes/article/view/759>
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.15779/Z38BG31>
- Beheshti, A., Rabhi, F., & Gill, A. (2025). Business transformation through AI-enabled technologies. *Frontiers in Artificial Intelligence*, 8, Article 1577540. <https://doi.org/10.3389/frai.2025.1577540>

- Binns, R. (2018a). Fairness in machine learning: Lessons from political philosophy. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, 81, 149–159. <https://proceedings.mlr.press/v81/binns18a.html>
- Binns, R. (2018b). On the ethics of artificial intelligence in education. In *Proceedings of the IEEE International Conference on Robotics and Automation* (pp. 197–204). IEEE.
- Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies*. Oxford University Press.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. <https://doi.org/10.3316/QRJ0902027>
- Boztepe, C. (2025). Eğitimde yapay zekâ uygulamaları: Fırsatlar, sınırlılıklar ve etik tartışmalar [Artificial intelligence applications in education: Opportunities, limitations, and ethical debates]. *Dumlupınar Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi*, 9(1), 98–115. <https://doi.org/10.71272/debder.1706141>
- Brohi, S., Mastoi, Q., Jhanjhi, N. Z., & Pillai, T. R. (2025). A research landscape of agentic AI and large language models: Applications, challenges and future directions. *Algorithms*, 18(8), Article 499. <https://doi.org/10.3390/a18080499>
- Cai, C., & Yin, J. (2025). Cultural and ethical foundations of AI governance divergence: A comparative analysis of China and the West. *Revista Política Internacional*. <https://cas.fudan.edu.cn>
- Carrillo, C., & Flores, M. A. (2025). Charting new territories: Artificial intelligence in the context of teacher education. *European Journal of Teacher Education*, 48(5), 909–913. <https://doi.org/10.1080/02619768.2025.2586365>
- Carrillo, C., Flores, M. A., & Mena, J. (2025). Artificial intelligence in the context of teacher education: Emerging themes and critical issues. *European Journal of Teacher Education*, 48(5), 1143–1162. <https://doi.org/10.1080/02619768.2025.2584283>
- Cedo, A. R., Gonzales, P. X., Patawaran, M. R., & Dionisio, S. A. S. (2025). *The perceptions of psychology students in the use of artificial intelligence for academic work* [Preprint]. ResearchGate. <https://www.researchgate.net/publication/397176908>
- Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8, 75264–75278. <https://doi.org/10.1109/ACCESS.2020.2988510>
- Coeckelbergh, M. (2020). *AI ethics*. MIT Press.
- Collin, S., Lepage, A., & Nebel, L. (2023). Enjeux éthiques et critiques de l'intelligence artificielle en éducation : Une revue systématique de la littérature [Ethical and critical issues of artificial intelligence in education: A systematic literature review]. *Canadian Journal of Learning and Technology / Revue canadienne de l'apprentissage et de la technologie*, 49(4), 1–29. <https://doi.org/10.21432/cjlt28448>
- Crawford, K. (2021). *The atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.
- Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, 538(7625), 311–313. <https://doi.org/10.1038/538311a>
- Dadson, Y. A., Mohammed, Z., & Hussein, B. (2025). Ethics and privacy in AI education: Prospects and challenges in higher education. In *Artificial intelligence and ethics in education* (pp. 207–223). Springer Nature. https://doi.org/10.1007/978-981-96-7937-9_10
- Demir, I. (2025). *Data privacy and student protection in AI-enabled K–12 learning environments* [Preprint]. ResearchGate. <https://www.researchgate.net/publication/397100916>
- Demirtaş, Z. (2024). Eğitim yönetiminde yapay zekâ uygulamaları: Etik sorunlar ve yönetsel karar destek sistemleri [Artificial intelligence applications in educational administration: Ethical issues and managerial decision support systems]. *Kuram ve Uygulamada Eğitim Yönetimi*, 30(1), 87–108.
- Dignum, V. (2019). *Responsible artificial intelligence: How to develop and use AI in a responsible way*. Springer. <https://doi.org/10.1007/978-3-030-30371-6>
- Dutta, S. (2025). Ethics in the age of algorithms: Moral accountability in artificial intelligence. *International Research Journal of Humanities and Interdisciplinary Studies*, 6(8), 92–104. <https://irjhis.com/paper/IRJHIS2508009.pdf>
- Erdoğan, F., & Çakır, Ö. (2024). Öğretmen adaylarının yapay zekâ okuryazarlıklarının ve yapay zekâyâ ilişkin algılarının belirlenmesi [Determining pre-service teachers' artificial intelligence literacy and perceptions of artificial intelligence]. *Uluslararası Türk Kültür Coğrafyasında Sosyal Bilimler Dergisi (TURKSOSBİLDER)*, 9(2), 63–95.
- European Commission. (2020). *Digital education action plan 2021–2027: Resetting education and training for the digital age*. European Union. <https://education.ec.europa.eu/focus-topics/digital/education-action-plan>
- European Commission. (2024). *Artificial intelligence act*. Publications Office of the European Union. <http://data.europa.eu/eli/reg/2024/1689/oj>
- European Parliament. (2024, March 13). *European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))*.

- European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. *Official Journal of the European Union*. <http://data.europa.eu/eli/reg/2024/1689/oj>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Geopolitechs. (2024, February 26). *China releases draft technology ethics guidelines*. <https://www.geopolitechs.org/p/china-releases-draft-technology-ethics>
- Ghasemaghahi, M., Hassanein, K., & Al-Debei, M. M. (2020). Organizational value of artificial intelligence: A qualitative study. *Journal of Business Ethics*, 165(4), 915–931.
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Herrera-Poyatos, D., & Peláez-González, C. (2025). *Triadic fusion of cognitive, functional, and causal dimensions for explainable LLMs: The TAXAL framework* (arXiv:2509.05199). arXiv. <https://doi.org/10.48550/arXiv.2509.05199>
- Hine, E., & Floridi, L. (2024). Artificial intelligence with American values and Chinese characteristics: A comparative analysis of American and Chinese governmental AI policies. *AI & Society*, 39(1), 257–278. <https://doi.org/10.1007/s00146-022-01499-8>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Irnowati, I., Fatmawati, Y., Mahardhika, I., & Suratno, S. (2024). The ethics of using artificial intelligence (AI) in learning in a philosophical perspective. *International Journal of Education, Information Technology, and Others*, 7(4), 319–330. <https://jurnal.peneliti.net/index.php/IJEIT/article/view/10977>
- Ismail, A. A. M. B. (2025). Harnessing AI tools responsibly: Supporting academic writing skills in Malaysian tertiary ESL classrooms. *PASAK Conference Proceedings*, 10, 82–88. https://conference.uis.edu.my/pasak/images/eprosiding2025/pasakx_012_aina_82-88.pdf
- Izquierdo-Condoy, J. S., Arias-Intriago, M., Tello-De-la-Torre, A., Busch, F., & Ortiz-Prado, E. (2025). Generative artificial intelligence in medical education: Enhancing critical thinking or undermining cognitive autonomy? *Journal of Medical Internet Research*, 27, Article e76340. <https://doi.org/10.2196/76340>
- Kaman, Ş. (2025). Sınıf öğretmenlerinin yapay zekâ okuryazarlık düzeylerinin çeşitli değişkenlere göre incelenmesi [Examining primary school teachers' artificial intelligence literacy levels according to various variables]. *Bilgi ve İletişim Teknolojileri Dergisi*, 7(1), 63–77. <https://doi.org/10.53694/bited.1628589>
- Kesgin, K. (2025). Yapay zekâ tabanlı karar destek sistemlerinde teknik, etik ve yönetsimsel yaklaşımlar: Çok sektörlü bir derleme çalışması [Technical, ethical, and governance approaches in artificial intelligence-based decision support systems: A multi-sector review]. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 9(1), 29–48. <https://doi.org/10.33461/uybisbbd.1691218>
- Kişisel Verileri Koruma Kurumu. (2025). *Yapay zekâ alanında kişisel verilerin korunmasına dair tavsiyeler* [Recommendations on the protection of personal data in the field of artificial intelligence]. <https://www.kvkk.gov.tr/SharedFolderServer/CMSFiles/25a1162f-0e61-4a43-98d0-3e7d057ac31a.pdf>
- Knox, J. (2020). Artificial intelligence and education in China. *Learning, Media and Technology*, 45(3), 298–311. <https://doi.org/10.1080/17439884.2020.1754236>
- Knox, J., Williamson, B., & Bayne, S. (2020). Machine behaviourism: Future visions of "learnification" and "datafication" across humans and digital technologies. *Learning, Media and Technology*, 45(1), 31–45. <https://doi.org/10.1080/17439884.2019.1623251>
- Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). Cambridge University Press.
- Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, 1(3), Article 0028. <https://doi.org/10.1038/s41562-016-0028>
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson.
- 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), Article 115. <https://doi.org/10.1145/3457607>

- Meylani, R. (2025). Rethinking explainability in educational artificial intelligence: A critical systematic review of models, applications, and ethical implications. *Veredas do Direito*, 22(3), 55–79. <https://revista.domhelder.edu.br/index.php/veredas/article/view/3553>
- Millî Eğitim Bakanlığı. (2025). *Yapay zekâ etiği tavsiyeleri* [Artificial intelligence ethics recommendations]. Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü. <https://yegitek.meb.gov.tr/www/yapay-zeka-etigi-tavsiyeleri-kitapcigi-yayimlandi/icerik/3878>
- Mishra, D., Mishra, R. K., & Agarwal, R. (2025). Impacts of artificial intelligence on society. *Journal of Science Research International*, 11(1), 29–47. <https://doi.org/10.5281/zenodo.15044020>
- Nguyen, T. M. T. (2025). Beyond motivation: Students' perception of ethical and learning-related risks of AI-driven gamified language learning. *International Journal of AI in Language Education*. <https://ijaile.org/index.php/ijaile/article/view/45>
- Ninghardjanti, P., Umam, M. C., Subarno, A., & Winarno, W. (2025). Evaluating the impact of AI on the critical thinking skills among higher education students by combining the TAM model and critical thinking theory. *Frontiers in Education*, 10, Article 1719625. <https://doi.org/10.3389/feduc.2025.1719625>
- OECD. (2021). *AI and the future of skills, Volume 1: Capabilities and assessments*. OECD Publishing. <https://doi.org/10.1787/5ce71f34-en>
- OECD. (2023). *OECD framework for the classification of AI systems*. OECD Publishing. <https://oecd.ai/en/classification>
- OECD. (2025). *Artificial intelligence and education and skills*. <https://www.oecd.org/en/topics/artificial-intelligence-and-education-and-skills.html>
- Oluwatosin, A. P. (2025). Artificial intelligence in the African context: A philosophical and ethical inquiry. *Nnadiabube Journal of Philosophy*, 9(1), 51–73. <https://www.acjoi.org/index.php/NJP/article/view/7767>
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.
- Pantano, E., Essien, A., & Kremantzis, M. D. (2025). Uncovering the generative AI (GenAI) to agentic AI (AgAI) shift for business school education. *Journal of Global Education and Technology*, 4(2), 201–221. <https://www.researchgate.net/publication/396324540>
- Parlak, İ. E. (2025). Blockchain-assisted explainable decision traces (BAXDT): An approach for transparency and accountability in artificial intelligence systems. *Knowledge-Based Systems*, 304, Article 1114418. <https://doi.org/10.1016/j.knosys.2025.1114418>
- Poovey, M. (2017). [Review of the book *Weapons of math destruction: How big data increases inequality and threatens democracy*, by C. O'Neil]. *Notices of the American Mathematical Society*, 64(9), 933. <https://doi.org/10.1090/noti1561>
- Prosper, D. (2025). *Balancing automation and academic integrity: Ethical challenges of AI-generated feedback in higher education* [Preprint]. ResearchGate. <https://www.researchgate.net/publication/396563205>
- Putri, B. M., Salim, H., & Inayati, N. L. (2025). Empowering millennial Muslim families through artificial intelligence in Qur'anic learning: Opportunities and ethical challenges in Indonesia. *Iseedu: Journal of Islamic Educational Thoughts and Practices*, 9(1), 125–134. <https://doi.org/10.23917/iseedu.v9i1.10437>
- Rabhi, F., Beheshti, A., & Gill, A. (2025). Business transformation through AI-enabled technologies. *Frontiers in Artificial Intelligence*, 8, Article 1577540. <https://doi.org/10.3389/frai.2025.1577540>
- Rampilla, Y., Satyala, K., & Kailasam, S. (2025). Enhancing trust in AI-powered learning: The role of explainable AI in the education field. *International Journal on Science and Technology*, 13(3), 112–128. <https://www.ijstat.org/papers/2025/3/8093.pdf>
- Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data, and personalized learning: Twenty-first century student sorting and tracking. *Ethics and Information Technology*, 21(3), 167–179. <https://doi.org/10.1007/s10676-018-9492-2>
- Renz, A., & Hilbig, R. (2020). Prerequisites for artificial intelligence in further education: Identification of drivers, barriers, and business models of educational technology companies. *International Journal of Educational Technology in Higher Education*, 17, Article 14. <https://doi.org/10.1186/s41239-020-00193-3>
- Rivera, A. M., & Osená, F. L. (2025). Exploring ethical perceptions of AI use in academic integrity among students in teacher education and business studies. *International Journal of Research and Innovation in Social Science*, 9(3S), 3949–3971. <https://doi.org/10.47772/IJRISS.2025.903SEU0283>
- Roberts, H., Cows, J., Hine, E., & Morley, J. (2023). Governing artificial intelligence in China and the European Union: Comparing aims and promoting ethical outcomes. *The Information Society*, 39(1), 35–50. <https://doi.org/10.1080/01972243.2022.2124565>
- Sambhaji, D. T. (2025). Artificial intelligence in education: Transforming learning, teaching, and administration. In U. V. Patil & B. S. Sakate (Eds.), *Unlocking potential: The intersection of human creativity and AI* (pp. 64–70). Akshara Publication.

- Sanayi ve Teknoloji Bakanlığı. (2021). *Ulusal yapay zekâ stratejisi (2021–2025)* [National artificial intelligence strategy (2021–2025)]. <http://www.sp.gov.tr/upload/xSPTemelBelge/files/Rpbea+TR-UlusalYZStratejisi2021-2025.pdf>
- Selwyn, N. (2019). *Should robots replace teachers? AI and the future of education*. Polity Press.
- Selwyn, N. (2022). Less work for teacher? The ironies of automated decision-making in schools. In S. Pink, M. Berg, D. Lupton, & M. Ruckenstein (Eds.), *Everyday automation: Experiencing and anticipating emerging technologies* (pp. 73–86). Routledge. <https://doi.org/10.4324/9781003170884-6>
- Sharma, R., & Sharma, A. (2025). From data to insight: Decoding student cognition with explainable AI models. *Euro Vantage Journal of Artificial Intelligence*, 7(1), 59–74. <https://evjai.com/index.php/evjai/article/download/59/112>
- Shute, V. J., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning*, 33(1), 1–19. <https://doi.org/10.1111/jcal.12172>
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529. <https://doi.org/10.1177/0002764213479366>
- Solak, Y., Dal, H. O., Çinkir, G., & Onu, F. N. (2025). Yapay zekâ etiği: Eğitimde kullanımının sınırları ve sorumlulukları [Artificial intelligence ethics: Limits and responsibilities of its use in education]. *Socrates Journal of Interdisciplinary Social Researches*, 11(49), 108–116. <https://doi.org/10.5281/zenodo.14762322>
- Stahl, B. C. (2022). From computer ethics and the ethics of AI towards an ethics of digital ecosystems. *AI and Ethics*, 2, 65–77. <https://doi.org/10.1007/s43681-021-00080-1>
- Stahl, B. C., Antoniou, J., Ryan, M., Macnish, K., & Jiya, T. (2022). Organisational responses to the ethical issues of artificial intelligence. *AI & Society*, 37(3), 1189–1202. <https://doi.org/10.1007/s00146-021-01148-6>
- State Council of the People's Republic of China. (2017). *New generation artificial intelligence development plan*. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
- Talapina, E. (2025). Transparency of artificial intelligence algorithms. *Law. Journal of the Higher School of Economics*, 18(3), 301–321. <https://journals.rcsi.science/2072-8166/article/view/318080>
- Temur, S. (2025). Eğitimde yapay zekâ kullanımı: Etik sorunlar ve çözümler [The use of artificial intelligence in education: Ethical issues and solutions]. *Mehmet Akif Ersoy University Journal of Education Faculty*, 74, 568–595. <https://doi.org/10.21764/mauefd.1516576>
- UNESCO. (2021a). *AI and education: Guidance for policy-makers*. <https://doi.org/10.54675/PCSP7350>
- UNESCO. (2021b). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000377897>
- Vallor, S. (2016). *Technology and the virtues: A philosophical guide to a future worth wanting*. Oxford University Press.
- Vandemeulebroucke, T. (2025). The ethics of artificial intelligence systems in healthcare and medicine: From a local to a global perspective, and back. *AI and Ethics*, 5(1), 39–58.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Varol, N. (2025). Eğitimde yapay zekâ ve etik: Bir alanyazın taraması [Artificial intelligence and ethics in education: A literature review]. *Okul Yönetimi*, 5(1), 117–125. <https://dergipark.org.tr/en/download/article-file/4848391>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223–235. <https://doi.org/10.1080/17439884.2020.1798995>
- Williamson, B., Macgilchrist, F., & Potter, J. (2023). Re-examining AI, automation and datafication in education. *Learning, Media and Technology*, 48(1), 1–5. <https://doi.org/10.1080/17439884.2023.2167830>
- Wu, H. C., & Liu, Y. C. (2025). Development trends in transfer learning theory: Applications and challenges in the era of generative AI. *Journal of Education Research*, 15(2), 45–59. <https://search.proquest.com/openview/a894ee03f9d980abb6a3183aaf7ca37b/1>
- Xue, Y., Chinapah, V., & Zhu, C. (2025). A comparative analysis of AI privacy concerns in higher education: News coverage in China and Western countries. *Education Sciences*, 15(6), Article 650. <https://doi.org/10.3390/educsci15060650>
- Yıldırım, M. (2024). Yapay zekâ ve eleştirel düşünme paradoksu: Eğitimde YZ'nin sınırları [The paradox of artificial intelligence and critical thinking: The limits of AI in education]. *Dersler Dergisi*, 5(1), 45–59. <https://derslerdersisi.com/egitimde-yapay-zeka-kullanimi-bir-paradoks-mu-yoksa-ikilem-mi/>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, Article 39. <https://doi.org/10.1186/s41239-019-0171-0>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.