Ethical AI in Education: Principles, Governance, and Responsible Implementation

Igor Britchenko[1]

[1]Doctor of Science (Economics), Professor, University of the National Education Commission, Krakow, Poland, ORCID: https://orcid.org/0000-0002-9196-8740

Citation:
Britchenko, I. (2025). Ethical AI in Education: Principles, Governance, and Responsible Implementation. *Pedagogy and Education Management Review*, (4(22), 17–30. https://doi.org/10.36690/2733-2039-2025-4-17-30

Received: November 27, 2025
Approved: December 29, 2025
Published: December 31, 2025

Abstract. *Artificial intelligence is increasingly embedded in education through learning analytics, adaptive learning systems, automated feedback, proctoring, student support chatbots, and generative AI tools, with implications for how decisions are justified and how responsibility is distributed. The paper aims to articulate domain-specific ethical AI principles for education that protect learner rights, reinforce equity, and preserve the integrity of assessment while enabling responsible innovation. A structured narrative review and normative synthesis are used to integrate AI ethics and governance guidance with AI-in-education research, then translate these insights into implementable principles and lifecycle governance mechanisms. The analysis shows that generic AI ethics statements are insufficient without pedagogical grounding, because educational quality depends on developmental, relational, and legitimacy conditions that are not captured by technical metrics alone. The resulting framework prioritizes human-centered educational benefit, learner agency with meaningful oversight, fairness and inclusion, privacy and data minimization, transparency proportional to decision impact, safety and well-being protections, academic integrity by design, and accountability with remedy in high-impact uses. Ethical AI in education requires institutional governance that connects values to procurement, deployment, classroom practice, monitoring, and evaluation across the AI lifecycle. Future work should strengthen measurement frameworks and empirical evidence for safeguarded AI use in high-stakes contexts, and examine implementation capacity in procurement, training, and post-deployment monitoring.*

Keywords: *ethical AI; education governance; learner rights; transparency; explainability; fairness; privacy; academic integrity; accountability; generative AI; risk management; child centered design.*

JEL Classification: I21; I28; O33; D63; K24; L86

Formulas: 0; fig. 0; tabl. 3; bibl. 30

1. Introduction. Artificial intelligence is increasingly embedded in education through learning analytics, adaptive learning systems, automated feedback, proctoring, student support chatbots, and generative AI tools used by teachers and learners. This expansion is not only technical but also institutional, because AI changes how educational decisions are made and justified, how performance is measured, and how responsibility is distributed across people, platforms, and policies. At the same time, education is a domain with distinctive ethical stakes. Decisions about learners affect life chances, identity development, and participation in society, and many learners are minors or otherwise vulnerable. For this reason, ethical AI in education cannot be treated as a generic compliance checklist. It must be anchored in educational values such as learner agency, inclusion, developmental appropriateness, and the integrity of assessment.

Recent global guidance highlights that education requires a human centered approach with strong protections for privacy, safeguards for children, and clear governance for generative AI use. UNESCO's guidance on generative AI in education emphasizes age appropriate design, data protection, and governance processes that connect ethical validation to pedagogical design. At the same time, broader AI governance frameworks such as the OECD AI Principles and the NIST AI Risk Management Framework provide a vocabulary for trustworthy AI, including fairness, transparency, safety, and accountability. Regulation is also becoming more explicit. The EU AI Act establishes a risk based regime and includes strong restrictions relevant to education, including limits on certain emotion recognition uses in educational institutions. These developments create an opportunity to articulate ethical AI principles that are simultaneously normative, operational, and pedagogically meaningful.

Literature Review. The expansion of artificial intelligence across teaching, assessment, administration, and student support has made ethical governance a central concern for education systems. In the literature, ethical AI in education is generally framed as a socio-technical challenge that involves pedagogical aims, institutional power relations, and the design choices embedded in tools and data infrastructures (Holmes et al., 2022; Selwyn, 2022). Education is treated as a distinctive domain because decisions about learners influence life chances, identity formation, and civic participation, while many learners are minors and therefore require heightened protections (UNICEF, 2025). Scholars argue that purely generic AI ethics principles are insufficient unless they are translated into education-specific requirements that address classroom practice, assessment integrity, and the distribution of responsibility among educators, institutions, and vendors (Holmes et al., 2022). A recurring theme is that ethical evaluation should not focus only on accuracy or efficiency, because educational quality includes developmental and relational dimensions that resist narrow optimisation (Selwyn, 2022). Research also notes that AI adoption can reshape what counts as legitimate knowledge and measurable success, especially when platforms prioritise standardised indicators over context-sensitive judgement (Selwyn, 2022). As a result, ethical AI principles in education are increasingly discussed as a combination of normative commitments, such as rights and inclusion, and operational controls, such as oversight, documentation, and contestability (Holmes et al., 2022). This conceptual shift supports the view that "ethics" must be enacted through governance routines that are transparent to learners and accountable to the public.

A substantial portion of the literature derives ethical principles from international norm-setting instruments that emphasise human dignity, human rights, and the responsibilities of institutions that deploy AI. UNESCO's Recommendation on the Ethics of Artificial Intelligence positions accountability, fairness, privacy, and human oversight as central elements of trustworthy AI governance, and these themes translate directly into educational contexts where vulnerability and asymmetries of power are prominent (UNESCO, 2021). The OECD AI principles similarly frame trustworthy AI around human-centred values, transparency, robustness, security, and accountability, providing a policy vocabulary that education systems increasingly adopt in national strategies and institutional guidelines (Organisation for Economic Co-operation and Development [OECD], 2019). In the European context, ethics has also become tied to enforceable obligations through risk-based regulation, which intensifies the relevance of formal governance processes for education providers (European Union, 2024). The EU AI Act is frequently discussed in educational policy analyses because it clarifies that certain education-related uses can fall into higher-risk categories and it restricts specific practices, which raises the standard for documentation, monitoring, and human oversight in schools and universities (European Union, 2024). Alongside regulation, the EU Ethics Guidelines for Trustworthy AI remain influential as a synthesis of requirements that include human agency, technical robustness, privacy and data governance, transparency, fairness, societal well-being, and accountability (European Commission, 2019). This combined normative and regulatory environment has shifted the literature from aspirational ethics statements toward implementation frameworks aligned with compliance, procurement standards, and lifecycle governance. Consequently, ethical AI principles in education are increasingly articulated as requirements that must be demonstrable through institutional policies and auditable practices, not only expressed as values.

Within AI in education research, fairness is widely discussed as a multidimensional concept that includes algorithmic bias, unequal predictive performance across groups, and differential access to AI-enabled learning opportunities (Holmes et al., 2022). The literature stresses that fairness in education cannot be reduced to a single statistical criterion because educational systems pursue plural aims, including inclusion, capability development, and formative support (Holmes et al., 2022). Transparency and explainability are commonly framed as prerequisites for legitimate educational decision-making, particularly when AI influences grading, placement, admissions support, or identification of students deemed "at risk" (Holmes et al., 2022). Authors also emphasise that explainability in education has multiple audiences, since teachers require actionable interpretations, while learners and families require understandable reasons and the possibility of contesting outcomes (Holmes et al., 2022). A central education-specific ethical issue is learner agency, because AI tools may subtly shift autonomy away from learners and educators toward platform recommendations and automated feedback loops that shape study behaviour and self-perception (Selwyn, 2022). Accountability is equally prominent because educational AI systems are commonly produced by vendors, configured by institutions, and operationalised by staff, which can generate gaps in responsibility when harms occur (Holmes et al., 2022). The literature therefore calls for governance structures that explicitly assign obligations, define escalation pathways, and ensure remedy when decisions are contested or errors produce adverse consequences (Holmes et al., 2022). Overall, these themes converge on the need for ethical principles that are actionable across design, procurement, deployment, and evaluation, rather than limited to classroom etiquette or individual user behaviour.

A large body of research argues that ethical AI in education is inseparable from data governance because many educational systems rely on learning analytics infrastructures that collect behavioural and performance data at scale. Foundational work identifies ethical dilemmas around consent, privacy, transparency, and function creep, particularly where data collected for learning support may be repurposed for monitoring, discipline, or high-stakes categorisation (Slade & Prinsloo, 2013). Subsequent studies show that students often accept data use when it is clearly linked to benefit and bounded by safeguards, but they also demand meaningful transparency, control, and institutional trustworthiness (Ifenthaler & Schumacher, 2016). In response, governance frameworks have developed practical mechanisms such as checklists and codes of practice that encourage institutions to articulate purposes, minimise data, define access, and assign roles and responsibilities for ethical oversight (Jisc, 2016). This literature highlights that privacy is not only a legal requirement but also a legitimacy condition, because learners' willingness to engage and disclose difficulties may decline under perceptions of surveillance (Slade & Prinsloo, 2013). Data governance debates also intersect with equity concerns, since monitoring systems can disproportionately target marginalised students if risk models are trained on historically biased data patterns (Holmes et al., 2022). Child-centred guidance reinforces the idea that educational data practices must be evaluated through the lens of vulnerability, developmental appropriateness, and protection from harmful profiling (UNICEF, 2025). As educational institutions adopt more AI-enabled analytics and proctoring tools, the literature increasingly frames surveillance risk as a central ethical challenge rather than an exceptional case. Accordingly, ethical principles in education must include strict data minimisation, transparent communication, participatory policy development, and robust safeguards against repurposing.

Recent literature treats generative AI as a catalyst that intensifies established ethical issues while introducing new challenges related to authorship, epistemic trust, and assessment validity. Policy-oriented scholarship notes that generative tools complicate conventional distinctions between assistance and substitution, creating pressure for institutions to revise academic integrity policies, redesign assessments, and establish disclosure norms for AI-supported work (UNESCO, 2023). UNESCO's guidance on generative AI emphasises governance, capacity building, and safeguards that address privacy, equity, and the reliability of knowledge practices in education and research (UNESCO, 2023). This focus shifts ethical principles toward operational questions, including how institutions communicate tool limitations, how teachers are trained to integrate AI responsibly, and how student evaluation can remain fair under unequal access and differing levels of support (UNESCO, 2023). The literature also highlights risks of overreliance and dependency, especially where learners use AI as an authority source rather than a fallible tool, which can undermine critical thinking and epistemic agency (Selwyn, 2022). Child-focused perspectives reinforce the need for age-appropriate design, protection from harmful content, and safeguards against manipulation or excessive data extraction, particularly in school settings (UNICEF, 2025). Another recurring argument is that quick technical fixes, such as automated detection of AI-written text, can produce injustice if reliability is limited and due process is weak (Holmes et al., 2022). As a result, the literature increasingly positions academic integrity as an institutional governance issue rather than a purely disciplinary matter. In this view, ethical AI principles must support transparent, educative, and procedurally fair responses that maintain trust in assessment while enabling responsible innovation.

Across the reviewed sources, ethical AI principles in education converge on a relatively stable set of themes: human agency and oversight, privacy and data governance, transparency and contestability, fairness and inclusion, safety and robustness, and accountability with access to remedy (European Commission, 2019; Holmes et al., 2022; UNESCO, 2021). However, the literature also suggests that implementation remains uneven, particularly in procurement practices, staff training, and post-deployment monitoring that can detect drift, bias, and unintended consequences (Holmes et al., 2022; Jisc, 2016). One gap concerns limited empirical evidence on the educational effectiveness of AI under strong ethical safeguards, especially in high-stakes contexts such as grading, admissions support, and automated proctoring (Zawacki-Richter et al., 2019). Another gap involves the political economy of educational AI, including vendor influence, platform dependency, and the redistribution of decision authority away from educators and public institutions (Selwyn, 2022). Comparative research is also needed to understand how ethical principles are interpreted and enforced across different regulatory environments, particularly as risk-based approaches become more influential in shaping institutional obligations (European Union, 2024). In addition, scholars call for participatory governance models that include students, teachers, and families in policy formation, since legitimacy depends on shared understanding of acceptable use and meaningful avenues for contestation (Ifenthaler & Schumacher, 2016; Slade & Prinsloo, 2013). Finally, there is a growing need for practical measurement frameworks that can operationalise concepts such as educational benefit, learner agency, and fairness in ways that can be monitored over time (Holmes et al., 2022). Taken together, the literature supports an integrated agenda that combines rights-based norms, risk management, and education-specific professional judgement, anchored in transparent institutional accountability.

Aims. The article aims to synthesize contemporary ethical and regulatory approaches to AI and adapt them to the specific logic of education, where human development, equity, and learner agency are central. It also aims to formulate an integrated set of ethical AI principles that can guide institutional decisions on procurement, deployment, classroom use, and evaluation, with particular attention to generative AI, data governance, transparency, fairness, child rights, and accountability in high-impact applications such as grading and student profiling.

Methodology. The study applies a structured narrative literature review combined with normative and policy analysis. It consolidates peer-reviewed AI-in-education scholarship and major international governance frameworks referenced in the manuscript, then performs thematic synthesis to identify recurring ethical requirements relevant to education. The analysis operationalizes these requirements by translating them into education-specific principles and mapping them to governance controls across selection, deployment, use, monitoring, and retirement of AI systems. Finally, the study derives gaps and future research priorities by interpreting the reported implementation challenges, including uneven procurement practice, staff training, and post-deployment monitoring, as well as limited evidence in high-stakes educational contexts.

Results. Ethical issues in educational AI extend beyond technical bias and data security, because education is a value-driven institution concerned with human development, legitimacy of decisions, and conditions for learners' autonomy and competence (Selwyn, 2022). AI systems can encourage optimization toward short-term performance indicators while weakening deeper learning, creativity, and intrinsic motivation, especially when analytics are treated as proxies for educational quality (Selwyn, 2022). A related concern is curricular narrowing, where content that is easiest to standardize, quantify, or automate becomes privileged over locally meaningful, culturally responsive, and dialogic learning practices (Selwyn, 2022). The literature also warns about "governance by numbers," in which automated classification, prediction, or ranking substitutes for professional judgement despite uncertain evidence or weak causal validity (Selwyn, 2022). From the AI in Education perspective, ethics is increasingly framed as a community-wide, practice-oriented agenda rather than an abstract declaration of values, because education depends on accountable relationships among learners, educators, institutions, and technology providers (Holmes et al., 2022). Holmes et al. (2022) emphasize the need to clarify stakeholders, harms, trade-offs, and accountability across the full lifecycle of educational AI, including design, procurement, classroom integration, and evaluation. Overall, the ethical problem is socio-technical, involving system design, institutional incentives, professional practice, and learner experience, which implies that principles should function as operational requirements guiding real decisions and oversight (Holmes et al., 2022; Selwyn, 2022).

Table 1 below synthesizes why generic AI ethics is insufficient for education and identifies typical high-risk educational pathways.

Table 1. Why education needs domain specific AI ethics: key risks and educational implications

Ethical pressure point	How it appears in education	Typical consequence	Why generic ethics is insufficient
Purpose ambiguity	AI adopted for innovation signaling or efficiency	Tool use decoupled from learning goals	Education requires alignment with developmental aims and pedagogy (Selwyn, 2022)
Metric substitution	Dashboards become proxies for learning	Short-term optimization displaces deeper learning	Learning quality is multidimensional and context-sensitive (Selwyn, 2022)
Authority shift	Automated recommendations guide decisions	Reduced teacher judgement and learner agency	Educational legitimacy depends on contestability and professional responsibility (Holmes et al., 2022)
Stakeholder asymmetry	Learners have limited negotiating power	Consent becomes formal rather than meaningful	Power imbalance is central in schooling, especially for minors (UNICEF, 2025)
Lifecycle accountability gaps	Vendor tool, institution use, teacher implementation	Responsibility diffusion when harms occur	Education needs explicit roles, review, and remedy procedures (Holmes et al., 2022)

Source: systematized by the author

The synthesis indicates that education requires ethics frameworks that explicitly connect AI use to pedagogical purposes, legitimacy, and responsibility allocation, not only to technical performance.

The principles below should be treated as an integrated set, because weaknesses in one domain often undermine the others, for example fairness failures often co-occur with transparency gaps and weak accountability (Holmes et al., 2022; NIST, 2023).

Human centered purpose and educational benefit. AI use should be justified by a clear educational purpose and evidence of benefit, rather than novelty or market pressure (OECD, 2019). This requires articulation of the learning problem, intended outcomes, and the pedagogical mechanism by which AI supports teaching and learning (OECD, 2019). Institutions should require an evidence narrative before adoption, including expected gains, risks, and explicit non-use conditions.

Learner agency and human oversight. Because education is relational and developmental, ethical AI must preserve learner agency and educators' professional responsibility (UNESCO, 2023). UNESCO's guidance on generative AI highlights human agency and age appropriateness, implying that AI should support, not replace, human judgement and educational relationships (UNESCO, 2023). Oversight should include meaningful opt-out where feasible, human review for high-impact decisions, and restrictions on fully automated decisions affecting grades, progression, or discipline (UNESCO, 2023).

Fairness, inclusion, and non-discrimination. AI can reproduce or amplify inequities through biased data, uneven access, and differential impacts on groups (NIST, 2023). NIST frames fairness and harmful bias management as core characteristics of trustworthy AI, requiring testing, monitoring, and mitigation (NIST, 2023). In education, fairness includes accessibility, linguistic inclusion, and culturally responsive content. Institutions should implement bias impact assessments, subgroup reporting, and contestation mechanisms, especially in admissions support, early warning systems, and adaptive pathways (NIST, 2023).

Privacy, data protection, and data minimization. Educational data can reveal cognitive profiles, socio-economic conditions, behavioral patterns, and sensitive signals, so strong privacy safeguards and strict minimization are essential (UNESCO, 2023). Practical requirements include clear retention limits, secure processing agreements, and restrictions on secondary use. For minors, privacy is inseparable from rights, and child-centered guidance emphasizes governance, safety, and safeguards against harmful profiling (UNICEF, 2025).

Transparency, explainability, and communication. Transparency in education has two layers: institutional transparency about what systems are used and why, and pedagogical transparency about how outputs should be interpreted (NIST, 2023). Explainability should be proportional to impact:

low-stakes writing assistance may require disclosure and guidance, while systems recommending track placement or flagging risk should provide interpretable reasons, limitations, and avenues for contestation (NIST, 2023; UNESCO, 2023).

Safety, security, and psychological well-being. Educational AI can cause harm through incorrect advice, manipulation, unsafe content, or security failures such as data breaches and prompt injection (UNICEF, 2025). Safety also includes psychological risks such as surveillance pressure, chilling effects, and dependency on automated feedback (UNICEF, 2025). Institutions should evaluate misuse scenarios, adversarial risks, and well-being impacts, and they should maintain escalation and incident response pathways (NIST, 2023).

Academic integrity and authenticity of learning. Generative AI intensifies concerns about authorship, plagiarism, and unreliable detection practices, so integrity should be built into pedagogy and assessment design rather than addressed primarily through punitive control (UNESCO, 2023). Ethical practice includes aligning permitted uses with learning objectives, requiring transparent disclosure when relevant, and designing assessments that evaluate authentic understanding and process. Institutions should communicate limitations of detection and ensure procedural fairness in academic conduct decisions (UNESCO, 2023).

Accountability, responsibility, and remedy. When AI contributes to educational decisions, accountability must remain with identifiable people and institutions, supported by documentation, auditability, and accessible remedy (Council of Europe, 2024). Remedy includes complaint mechanisms, the right to contest outcomes, and corrective actions when harms are detected. Accountability also requires procurement standards obligating vendors to disclose limitations, support audits, and cooperate with investigations (Council of Europe, 2024; NIST, 2023).

Table 2 operationalizes the principles as institutional requirements and examples of controls.

The mapping shows that principles become enforceable in education only when translated into governance artifacts, workflows, and measurable controls that can be audited and contested.

Ethical principles become effective when implemented through governance routines across the AI lifecycle, including selection, deployment, use, monitoring, and retirement (NIST, 2023). The NIST AI Risk Management Framework supports this approach by emphasizing role definition, documentation, testing, monitoring of drift, and learning from incidents (NIST, 2023). Regulatory regimes reinforce lifecycle governance through risk-based duties. The EU AI Act organizes obligations by risk level, identifies multiple education-related uses as high risk, and restricts certain practices such as emotion recognition in educational contexts except under limited conditions (European Union, 2024).

Table 2. Ethical AI principles in education: operational requirements and example controls

Principle	Operational requirement	Example institutional controls
Educational benefit	Clear purpose and evidence expectations	Pre-adoption evidence narrative; pilot with evaluation criteria (OECD, 2019)
Agency and oversight	Human review for high-impact uses	Human-in-the-loop grading review; opt-out options where feasible (UNESCO, 2023)
Fairness and inclusion	Measure and mitigate disparate impacts	Subgroup performance reporting; accessibility testing (NIST, 2023)
Privacy and minimization	Collect only necessary data, limit retention	Data minimization checklist; retention schedule; vendor processing clauses (UNESCO, 2023)
Transparency and explainability	Disclose use, provide interpretable reasons	Plain-language notices; decision rationale templates for high-impact tools (NIST, 2023)
Safety and well-being	Prevent misuse and psychological harm	Misuse scenario testing; incident reporting workflow; student support referral (UNICEF, 2025)
Academic integrity	Align AI use with learning objectives	Assessment redesign; AI disclosure norms; integrity education modules (UNESCO, 2023)
Accountability and remedy	Ensure contestation and corrective action	Appeals process; audit logs; vendor audit cooperation requirements (Council of Europe, 2024)

Source: systematized by the author

Even outside the EU, the Act functions as a reference point for procurement and governance because educational technology markets and vendors operate transnationally (European Union, 2024). Institutionally, governance can be structured through an AI ethics and safety committee with representation from educators, learners, legal and privacy officers, disability services, and technical experts, which aligns with UNESCO's emphasis on interdisciplinary capacity for evaluating long-term implications for curriculum, assessment, and social dynamics (UNESCO, 2023). Committee deliverables typically include an AI use policy, model risk classification, required impact assessments, and an approval and review process for high-impact tools (NIST, 2023; UNESCO, 2023).

The following table summarizes lifecycle governance controls aligned with risk management logic.

Lifecycle governance makes ethical principles actionable by embedding them into institutional decision points, documentation, and monitoring, which is necessary for trust and compliance.

Implementation should be treated as capacity building rather than a one-time policy announcement. Institutions should first develop a taxonomy of AI uses, separating instructional support, administrative automation, and high-impact decision systems such as grading, progression recommendations, and risk scoring, which helps align oversight intensity with potential harm (NIST, 2023).

Table 3. Lifecycle governance for educational AI: stages, controls, and outputs

Lifecycle stage	Key questions	Core controls	Typical outputs
Selection and procurement	What problem is solved and for whom	Evidence narrative; risk classification; vendor due diligence	Procurement checklist; risk register entry (NIST, 2023)
Deployment	What data and integration are required	Data minimization; security review; user training	Data processing documentation; training plan (UNESCO, 2023)
Classroom and operational use	How is AI used in pedagogy or decisions	Usage policy; disclosure and guidance; oversight for high impact	Classroom use protocol; human review steps (UNESCO, 2023)
Monitoring and audit	Are harms emerging or performance drifting	Subgroup monitoring; incident tracking; periodic audits	Monitoring reports; corrective action log (NIST, 2023)
Retirement and replacement	When should the system be stopped	Decommission criteria; data deletion; transition plan	Retirement decision record; deletion confirmation (NIST, 2023)

Source: systematized by the author

Second, procurement standards should require privacy by design, transparent documentation, bias testing evidence, security commitments, and audit support (NIST, 2023; UNESCO, 2023). Third, staff development is essential, because institutional governance depends on educators’ ability to interpret outputs, recognize limitations, and integrate tools without undermining learning objectives (U.S. Department of Education, 2023). Training should cover tool limitations, typical error patterns, responsible use practices, and procedures for escalation and human review (U.S. Department of Education, 2023). Fourth, institutions should redesign assessment using more process-based evaluation, oral defenses, project work tied to local contexts, and reflective disclosure about AI use when relevant, reducing reliance on detection as a primary enforcement strategy (UNESCO, 2023). Fifth, AI literacy should be integrated into curricula, including limitations, bias, privacy, and ethical use, reflecting the premise that governance is partly pedagogical and must be age appropriate (UNESCO, 2023; UNICEF, 2025). Finally, monitoring must be continuous through incident reporting, differential impact measurement, and iterative policy revision based on evidence and stakeholder feedback (NIST, 2023).

Discussion. Ethical AI in education involves trade offs that must be managed transparently. Personalization can conflict with privacy when it relies on extensive behavioral data. Automation can reduce workload but may also deskill educators or normalize surveillance. Transparency can be constrained by proprietary models, yet education requires reasons that learners can understand and contest. These tensions are not reasons to abandon AI, but they require explicit governance and ongoing evaluation.

Community wide frameworks in AI in education scholarship stress that ethical alignment is a continuing process of negotiation among stakeholders, not a static rule set. Critical perspectives also warn that AI adoption can be driven by commercialization and governance agendas that reshape educational priorities, which strengthens the case for institutional deliberation and democratic accountability. The most robust ethical posture therefore combines principled constraints with evidence oriented experimentation. Institutions can allow limited pilots under strong safeguards, evaluate outcomes, and scale only when educational benefit and rights protections are demonstrated.

Conclusions. Ethical AI principles in education must be grounded in the purposes of education and implemented through concrete governance mechanisms. A workable framework includes human centered benefit, learner agency and oversight, fairness and inclusion, privacy protection, transparency, safety and well being, academic integrity, and accountability with remedy. Current global guidance and regulatory developments provide strong foundations, but educational institutions must translate them into lifecycle controls, procurement standards, staff capacity, and learner literacy. Future research should focus on measurement and evaluation: how to operationalize educational benefit, how to assess equity impacts over time, how to validate explainability for educational decision making, and how to design assessments that preserve integrity while supporting responsible AI use. Research is also needed on governance at scale, including shared audit infrastructure, public reporting norms, and cross institutional learning systems that reduce duplication and improve accountability. Finally, child rights centered approaches should be expanded into practical design standards for school contexts, especially as generative AI becomes more accessible and embedded in everyday learning.

Funding. The author declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest. The author declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement. The author declare that no Generative AI was used in the creation of this manuscript.

Publisher's note. All claims expressed in this article are solely those of the author and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References:

1. Council of Europe. (2024, September 5). *Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law.* (CETS No. 225). Council of Europe. https://rm.coe.int/1680afae3c rm.coe.int
2. Council of Europe. (2024). Council of Europe Framework Convention on Artificial Intelligence and human rights, democracy and the rule of law (CETS No. 225). https://rm.coe.int/1680afae3c
3. European Commission. (2019). Ethics guidelines for trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

4. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
5. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ%3AL_202401689
6. European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)*. Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng EUR-Lex
7. European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union*. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng EUR-Lex
8. High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai digital-strategy.ec.europa.eu
9. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Santos, O. C., Rodrigo, M. M. T., Cukurova, M., Bittencourt, I. I., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education, 32*(3), 504-526. https://doi.org/10.1007/s40593-021-00239-1
10. Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. Educational Technology Research and Development, 64(5), 923-938. https://doi.org/10.1007/s11423-016-9477-y
11. Jisc. (2016). Jisc's learning analytics code of practice. https://community.jisc.ac.uk/system/files/391/LA%20and%20Ethics%20v0-16.pdf
12. Jisc. (2023). *Code of practice for learning analytics* (July 2023 ed.). Jisc. https://repository.jisc.ac.uk/9204/1/code-of-practice-for-learning-analytics.pdf repository.jisc.ac.uk
13. Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000386693
14. National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0) (NIST AI 100-1). https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf
15. Organisation for Economic Co-operation and Development. (2019). Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449
16. Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449)*. OECD. https://legalinstruments.oecd.org/api/print?ids=648&lang=en OECD Legal Instruments
17. Reiss, M. J. (2021). The use of AI in education: Practicalities and ethical considerations. London Review of Education, 19(1), Article 5, 1-14. https://doi.org/10.14324/LRE.19.1.05
18. Selwyn, N. (2022). The future of AI and education: Some cautionary notes. European Journal of Education, 57(4), 620-631. https://doi.org/10.1111/ejed.12532
19. Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. American Behavioral Scientist, 57(10), 1509-1528. https://doi.org/10.1177/0002764213479366
20. U.S. Department of Education, Office of Educational Technology. (2023). Artificial intelligence and the future of teaching and learning: Insights and recommendations. https://www.ed.gov/sites/ed/files/documents/ai-report/ai-report.pdf
21. UNESCO. (2021). AI and education: Guidance for policy-makers. https://unesdoc.unesco.org/ark:/48223/pf0000376709
22. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. https://unesdoc.unesco.org/ark:/48223/pf0000380455
23. UNESCO. (2021/2022). Recommendation on the Ethics of Artificial Intelligence (SHS/BIO/PI/2021/1). https://unesdoc.unesco.org/ark:/48223/pf0000381137.locale=en
24. UNESCO. (2023). Guidance for generative AI in education and research. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research
25. UNICEF. (2025). *Guidance on AI and children (Version 3.0): Recommendations for AI policies and systems that uphold children's rights*. UNICEF Innocenti. https://www.unicef.org/innocenti/media/11991/file/UNICEF-Innocenti-Guidance-on-AI-and-Children-3-2025.pdf

26. 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
27. 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