BHARATI INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & DEVELOPMENT (BIJMRD)

(Open Access Peer-Reviewed International Journal)

DOI Link : https://doi.org/10.70798/Bijmrd/04100001

Available Online: www.bijmrd.com | BIJMRD Volume: 4 | Issue: 10 | October 2026 | e-ISSN: 2584-1890

Ethical Principles for Implementing Artificial Intelligence in Education: A Human-Centred Framework and Future Agenda

Mrs. Rohini Bhattarai

Assistant Professor, Dept. of Computer, Mahatma Gandhi University, Daramdin, Email id:rohini.bhattarai@gmail.com

Abstract: *Artificial intelligence (AI), including generative AI, is now embedded in personalized learning, tutoring, assessment, learning analytics, accessibility, and educational administration. These uses may support teaching and learning, but they raise concerns about privacy, bias, transparency, accountability, academic integrity, learner autonomy, and unequal access. This conceptual paper synthesizes selected peer-reviewed studies and international guidance on ethical AI in education, updates the evidence base with 2026 surveys, policy reports, and regulatory developments, and proposes a lifecycle-oriented framework for translating ethical principles into institutional practice. The literature is organized around ten connected principles: privacy and data protection; fairness and non-discrimination; transparency and explainability; accountability; meaningful human oversight; academic integrity; inclusion and accessibility; safety and security; learner autonomy and well-being; and ethical governance with continuous monitoring. The framework has seven stages: planning and needs analysis, ethical risk assessment, ethical design and selection, implementation, human oversight, monitoring and evaluation, and continuous improvement, with stakeholder participation across all stages. Current evidence shows near-universal student use and rapidly growing teacher use, while formal institutional policy, training, and assessment redesign lag behind. The paper also identifies implementation challenges and a future agenda centred on governance, AI literacy, assessment redesign, equity, auditing, and evidence-based adoption. Because the study is conceptual, lacks a recorded database-by-database screening log, and does not empirically test the framework, the model should be treated as provisional rather than validated.*

Keywords: *Artificial Intelligence; AI in Education; Generative AI; Educational Technology; AI Ethics; AI Governance; Human-Centred Framework.*

1. Introduction

1.1 Background: AI has become part of the infrastructure through which education is designed, delivered, assessed, and administered. Applications include intelligent tutoring systems, adaptive learning platforms, automated feedback, learning analytics, generative AI assistants, accessibility tools, and administrative decision support. These technologies can help educators respond to learner needs, provide timely feedback, support content creation, and make services more accessible. Their educational value, however, depends on the purpose for which they are used, the evidence supporting them, and the safeguards governing them (Miao & Holmes, 2023; Miao et al., 2024).

Educational settings raise distinctive ethical concerns. AI may process sensitive learner information and influence assessment, placement, support, and perceptions of ability. A system that performs well on average

may still disadvantage particular groups because of algorithmic bias and unequal representation in training data. Automated recommendations may be difficult to challenge when their logic is unclear, and generative AI may produce plausible but inaccurate information or obscure a learner's own understanding. Unequal access to devices, connectivity, accessible interfaces, and AI literacy may distribute benefits unevenly. These risks concern educational values, professional judgement, learner rights, and institutional responsibility, not only technical performance (Miao & Holmes, 2023; Miao & Cukurova, 2024).

1.2 Problem statement: International guidance and research emphasize human dignity, fairness, transparency, accountability, privacy, and human oversight (UNESCO, 2021). Education-focused reviews have examined ethical principles in school settings, fairness and accountability in higher education, and wider social and pedagogical concerns (Adams et al., 2023; Bond et al., 2024; Collin et al., 2024; Memarian & Doleck, 2023; Mouta et al., 2024). Yet institutions still need practical ways to translate principles into decisions before procurement, during deployment, and after systems are in use. Recent surveys show the gap is widening: use is far ahead of policy, training, and evidence (Section 4).

1.3 Research questions and contribution

The paper addresses three questions:

1. What ethical concerns recur in the literature on AI in education?
2. Which principles should guide responsible educational AI?
3. How can those principles be connected to an institution's AI implementation lifecycle?

Its contribution is a practical conceptual bridge between ethical principles and institutional action, supported by an updated picture of current use and regulation. It does not empirically test the framework or claim that it is validated.

1.4 Structure of the paper: Section 2 describes the methodology. Section 3 reviews the literature. Section 4 summarizes the current landscape of AI use, evidence, policy, and regulation. Section 5 presents ten ethical principles. Section 6 proposes the seven-stage framework. Sections 7 to 10 present the discussion, future research agenda, limitations, and conclusion.

2. Methodology

2.1 Research design: This paper uses a qualitative conceptual design to synthesize literature on AI in education and develop a lifecycle-oriented ethical implementation framework. It reports no primary data, participant recruitment, experiments, or statistical analysis.

2.2 Literature identification: The original synthesis drew on a structured search using combinations of terms related to AI in education, ethical and responsible AI, generative AI, academic integrity, AI governance, algorithmic bias, privacy, and human-centred AI. It included peer-reviewed reviews and conceptual studies alongside guidance from UNESCO and the OECD. For this updated edition, a supplementary targeted search (October 2026) identified recent surveys of students and teachers, major policy reports, and regulatory developments. The update search was purposive, not systematic, and relied on publicly reported findings; several 2026 sources are survey reports or press releases that have not been peer reviewed and are labelled as such.

2.3 Selection and thematic synthesis: Sources were considered relevant when they addressed educational applications of AI or ethical concerns such as privacy, fairness, transparency, accountability, oversight, inclusion, literacy, or governance. The literature was examined thematically. Recurring concerns were grouped into ten principles, then connected to decisions across an AI system's lifecycle: identifying an

educational need, assessing risk, selecting or designing a system, implementing it, maintaining human oversight, monitoring outcomes, and improving practice. Stakeholder participation was treated as cross-cutting because students, teachers, administrators, developers, researchers, and policymakers hold different knowledge and responsibilities.

2.4 Methodological scope and transparency: The review was not conducted as a fully reproducible systematic review. The original source material did not preserve a database-by-database search log, exact search dates, result counts, or a formal screening flow. This paper therefore reports no invented search counts and claims no exhaustive coverage. A future systematic or scoping review should document databases, date limits, exact search strings, screening stages, and reasons for exclusion.

3. Literature Review

3.1 Ethical AI and education: Ethical AI research spans technical performance, human rights, institutional governance, and the social consequences of automated systems. UNESCO's Recommendation on the Ethics of Artificial Intelligence establishes a broad normative foundation centred on human rights, dignity, fairness, transparency, accountability, and oversight (UNESCO, 2021). In education, these values must be interpreted in relation to learners' developmental needs, teachers' professional responsibilities, assessment practices, and institutions' duty of care. Ethical performance therefore cannot be reduced to model accuracy.

3.2 Evidence from education-focused reviews: Reviews identify a recurring set of interdependent issues. Memarian and Doleck (2023) organize higher-education concerns around fairness, accountability, transparency, and ethics. Adams et al. (2023) examine ethical principles in K-12 education. Collin et al. (2024) discuss privacy, surveillance, bias, autonomy, and inequality. An et al. (2024) examine AI ethics through users' perspectives, highlighting the gap between abstract principles and the concerns experienced by people using educational systems. Mouta et al. (2024) identify blind spots that include social, cultural, emotional, and teacher-related dimensions. Bond et al. (2024), in a meta-systematic review, call for greater ethical attention, collaboration, and methodological rigour.

This matters because institutions may adopt technologies faster than they develop policies, staff expertise, evaluation procedures, or mechanisms for redress. Ethical implementation should connect pedagogical purpose, technical suitability, stakeholder participation, and evidence of educational value, and should distinguish between using AI to support a task and demonstrating that the use improves learning or teaching.

3.3 Generative AI, competencies, and equity: Generative AI has made the relationship between assistance, authorship, and learning more visible. It can support explanation, brainstorming, translation, lesson planning, and formative feedback, but its outputs may be inaccurate, biased, or poorly matched to learning objectives. UNESCO guidance emphasizes human-centred and age-appropriate use, data protection, institutional validation, and attention to educational purpose (Miao & Holmes, 2023). UNESCO's AI competency frameworks for teachers and students emphasize human-centred thinking, AI ethics, understanding AI applications, and responsible use (Miao & Cukurova, 2024; Miao et al., 2024). Teachers need to judge whether a tool supports curriculum goals, verify outputs, protect student information, and explain limitations; students need to evaluate generated content, recognize bias and misinformation, disclose assistance where required, and retain independent judgement.

The OECD Digital Education Outlook 2026 sharpened the central distinction between task performance and learning: used with pedagogical intent, generative AI can support learning, but outsourcing tasks to it can raise performance without real learning gains (OECD, 2026). Accessibility and equity must also be considered: AI can support captioning, translation, speech interfaces, and personalized support, yet may exclude learners if tools are costly, inaccessible, language-limited, or dependent on reliable connectivity.

3.4 Synthesis and research gap: Across these sources there is substantial agreement about privacy, fairness, transparency, accountability, human oversight, inclusion, and responsible governance. The challenge is less a lack of principles than a lack of practical integration across implementation stages. Principles may appear in policy but not be translated into procurement criteria, staff responsibilities, monitoring indicators, or processes for challenging harmful decisions. This paper responds by linking ten principles to a seven-stage lifecycle and treating stakeholder participation as a cross-cutting requirement. It is a conceptual synthesis, not an exhaustive systematic review.

4. The Current Landscape of AI in Education

4.1 Applications and their risk profiles: AI applications vary in purpose and risk. Intelligent tutoring systems and adaptive platforms tailor practice or feedback to learner responses. Generative AI assists students with explanation, idea generation, language support, and revision, and assists teachers with lesson planning and materials. Automated assessment may support scoring or formative feedback; learning analytics may identify patterns that educators use to offer support; chatbots answer routine questions; and administrative systems assist with scheduling, enrolment, and resource planning. AI also supports accessibility through speech recognition, captions, text-to-speech, translation, and alternative ways to interact with content.

Risks depend on context: a low-stakes brainstorming assistant does not carry the consequences of a system contributing to grading, admission, student profiling, or access to services. Institutions should assess tools by purpose, users, data, potential impact, and available alternatives. The central question is not whether a tool can perform a task, but whether it supports a defined educational objective, whether its performance is reliable enough for that use, who may be harmed or excluded, and what human review and accountability are needed.

4.2 Student use: Student use of generative AI has become close to universal in several higher-education settings. The HEPI Student Generative AI Survey 2026 of 1,054 UK full-time undergraduates found that 95% use AI in some way and 94% use generative AI for assessed work, compared with 51% the previous year; the share including AI-generated text directly in assessed work rose from 3% in 2024 to 12% in 2026 (Stephenson & Armstrong, 2026). Institutional support lagged: 38% of students said their institution provides generative AI tools (up from 9% in 2024), only about a third felt encouraged to use AI, and 15% reported using AI to address loneliness, seek advice, or for companionship. A 16-country survey of 3,839 students found 86% used AI but only 23% felt prepared and many cited inadequate guidance (Metzgar, 2026). At a selective US college, adoption exceeded 80% within two years of ChatGPT's release, with augmentation (explanations, feedback) more common than automation of coursework, and policies had uneven effects (Contractor & Reyes, 2026). The HEPI findings also show that use differs by gender and affluence, a concern for equity.

4.3 Teacher use: Teacher use is also rising, mainly for preparation and administration. In the UK, a YouGov survey of 1,033 teachers (June 2026) reported that 80% use AI in teaching-related activities and weekly use doubled from 26% to 52% in a year, yet only 8% use it for grading, 41% had received no school training, and only 33% reported a staff-use policy (Bett UK, 2026). In the US, an NPR/Ipsos poll found 54% of teachers use AI at least weekly for planning or administrative work, with 69% of users reporting greater efficiency, and most teachers were concerned about effects on students' critical thinking (Ipsos, 2026). Studies of classroom practice describe AI largely as a productivity tool rather than one transforming real-time teaching (Angelone & Burton, 2026). Across the 2026 surveys, teachers' individual decisions are filling the vacuum left by institutions.

4.4 Evidence on learning: The OECD Digital Education Outlook 2026 concludes that general-purpose generative AI can raise task performance without reliably producing learning, whereas tools designed with

pedagogical principles (guided questioning, hints, scaffolding) can support learning; it also reports uneven adoption, strongest in high-income countries (OECD, 2026). Brookings' year-long study concluded that, at this point, the risks of generative AI in children's education overshadow its benefits, while identifying opportunities for inclusion and recommending action to prosper, prepare, and protect (Burns et al., 2026). A review of that report found it useful for anticipating benefits and harms but less decisive about whether benefits can outweigh risks (Penuel, 2026). Together, these sources reinforce the need for evidence-informed adoption rather than adoption by availability.

4.5 Policy and training gaps: Policy remains thin. In a Gallup/Walton Family Foundation survey of 2,069 US teachers, 82% reported no formal guidance on applying AI to their work (Langreo, 2026). NPR/Ipsos found only 33% of teachers reporting a formal student-use policy and 23% a teacher-use policy (Ipsos, 2026). In England, 49% of surveyed National Education Union members reported no school AI policy and 66% no student-specific policy (National Education Union, 2026). In Canada, only 34.8% of surveyed grade 6-12 teachers reported a staff-use policy and 42.3% a student-use policy (MacPherson, 2026). An October 2026 K-12 Dive survey for Superhuman reported that only 21% of educators feel very prepared to manage AI in classrooms and 70% lack formal policies that are well understood or applied. These are mostly commissioned surveys with varying sampling; they indicate direction, not precise comparable prevalence.

4.6 Regulatory developments: The EU AI Act (Regulation (EU) 2024/1689) classifies AI used for access to education and vocational training among its high-risk categories. AI-literacy duties have applied since 2 February 2025 and general-purpose AI obligations since 2 August 2025. The Digital Omnibus on AI (Regulation (EU) 2026/1744), in force from 27 July 2026, deferred obligations for stand-alone Annex III high-risk systems, including education, from 2 August 2026 to 2 December 2027, while leaving general-purpose AI and several transparency obligations on their existing schedule; digital-rights groups have criticized the delay for leaving people exposed to consequential automated decisions in the interim (European Union, 2026; Mondaq/K&L Gates, 2026). Institutions inside and outside the EU should treat the deferral as planning time, not a reprieve, because documentation, human-oversight, and bias-monitoring practices take time to build. Other jurisdictions are moving through state, national, and sector guidance, for example US state requirements for district AI policies cited in classroom studies (Angelone & Burton, 2026).

5. Ethical Principles for Implementing AI in Education

The following ten principles synthesize recurring concerns in the literature and international guidance. They are interdependent: transparency supports accountability, privacy supports learner autonomy, and fairness depends on continuing monitoring rather than one-time testing.

5.1 Privacy and data protection: AI should collect and process only information necessary for a clearly defined educational purpose. Institutions should assess what data are collected, where they are stored, who can access them, whether they are shared with third parties, how long they are retained, and how correction or deletion requests are handled. Notices and consent procedures should be understandable and appropriate to learners' ages and circumstances. Privacy also requires secure systems, access controls, vendor review, and special care with sensitive or identifiable student data. The growing use of AI for advice and companionship (Stephenson & Armstrong, 2026) makes the sensitivity of conversational data an added concern.

5.2 Fairness and non-discrimination: AI may reproduce bias in data, design choices, or deployment conditions. Fairness requires testing across relevant learner groups, checking data representativeness, monitoring outcomes after deployment, and providing a route to contest decisions. Aggregate accuracy is not sufficient if particular groups experience systematically worse results. Institutions should consider language, disability, socioeconomic circumstances, gender, culture, and other locally relevant factors without assuming that one fairness measure suits every setting.

5.3 Transparency and explainability: Learners and educators should be told when AI is used, what role it plays, what data it may use, and what its limitations are. When outputs influence consequential decisions, affected people should receive an understandable explanation of the system's contribution and the basis for a human decision. Transparency also includes disclosing appropriate AI assistance in academic work and avoiding claims that AI outputs are necessarily accurate or objective.

5.4 Accountability: Responsibility should be assigned before deployment. Institutions should identify who approves use, manages data and vendor relationships, monitors performance, responds to complaints, and can suspend or withdraw a system. Developers and suppliers should provide information about capabilities, limitations, updates, and risks. Accountability requires records of important decisions and a clear process for correction, redress, and learning from failures.

5.5 Meaningful human oversight: Human oversight must be substantive rather than a nominal approval step. Teachers and qualified staff should be able to question AI recommendations, consider context, override outputs, and explain decisions. High-impact decisions, such as grading, admission, progression, disciplinary action, or allocation of support, require particularly careful review. Reviewers need sufficient time, expertise, information, and authority to exercise independent judgement.

5.6 Academic integrity and responsible AI assistance: Institutions should explain what AI assistance is permitted for different tasks and how it should be acknowledged. Policies should distinguish acceptable support, such as brainstorming or language feedback when allowed, from misrepresentation of generated work as unaided work. Blanket prohibitions can overlook legitimate learning uses, while unrestricted use can obscure whether intended outcomes have been achieved. Assessment may need to include process evidence, oral explanation, practical work, reflection, and authentic tasks. Institutions should also protect students from unfair misconduct accusations, a concern students themselves raised in 2026 (Stephenson & Armstrong, 2026).

5.7 Inclusion, accessibility, and digital equity: AI adoption should not widen educational inequalities. Institutions should evaluate accessibility, language support, cultural relevance, device and connectivity requirements, affordability, and the availability of non-AI alternatives. Learners should not be penalized for lacking access to a particular tool. Consultation with diverse learners and accessibility specialists should occur before procurement and continue during evaluation.

5.8 Safety, security, and reliability: AI systems may produce inaccurate, harmful, misleading, or inappropriate outputs and may be vulnerable to security failures. Institutions should test tools for the intended context, establish safeguards appropriate to learner age and task risk, and provide procedures for reporting incidents. Users should verify important claims against reliable sources. Security and reliability require ongoing attention because vendors, systems, and model behaviour change.

5.9 Learner autonomy and well-being: AI should support learners' capacity to think, question, create, and make informed choices rather than replace those activities. Excessive dependence may reduce opportunities to practise problem-solving, writing, critical thinking, and independent judgement. Educators should make clear when AI supports a learning goal and when independent performance is needed. Institutions should also consider surveillance, stress, excessive monitoring, emotional reliance on conversational AI, and effects of automated feedback on learner confidence.

5.10 Ethical governance and continuous monitoring: Ethical governance brings the other principles into institutional practice through policy, procurement criteria, risk assessment, staff development, stakeholder participation, incident response, documentation, monitoring, and periodic review. Governance should be proportionate to risk: a low-impact classroom tool may need lighter review than a system used for high-stakes decisions. Institutions need processes to reassess systems when their purpose, data, users, or technical characteristics change.

6. Proposed Ethical AI Implementation Framework: The framework connects the ten principles to seven stages of implementation. It is a continuous cycle rather than a one-time approval checklist. Stakeholder engagement, particularly with teachers and learners, operates across all stages. The framework is intended to guide institutional deliberation and is not a validated assessment instrument.

6.1 Planning and needs analysis: Define the educational problem and the outcome sought. Consider whether AI is necessary, whether a non-AI alternative may be more appropriate, and how success will be judged. Identify intended users, affected groups, accessibility requirements, and consequences of failure. Involve teachers, learners, administrators, and support staff before procurement.

6.2 Ethical risk assessment: Assess likely benefits and risks, including privacy, bias, security, transparency, academic integrity, learner autonomy, and unequal access. Determine the impact of errors and whether decisions are reversible or contestable. Identify data flows, third-party access, vendor terms, and relevant institutional and legal obligations (for example, whether a use falls within a high-risk category under applicable regulation). Record the assessment so that it can be reviewed.

6.3 Ethical design and selection: Use the educational purpose and risk assessment to define selection criteria. Examine evidence of educational value, accessibility, language coverage, explainability, privacy, security, update practices, and vendor accountability. Distinguish general-purpose tools from education-specific tools designed with pedagogical principles, and prefer the latter where stakes are high (OECD, 2026). Require clear information about limitations and acceptable use. Where evidence is insufficient, use a limited pilot or defer adoption rather than treating market availability as proof of suitability.

6.4 Implementation with safeguards: Introduce the system with clear policies, training, user information, and technical controls. Explain capabilities and limitations, what information users should not submit, and how outputs should be checked. Define acceptable use for staff and students, provide accessible guidance, and establish channels for questions or incident reports. Use phased implementation when uncertainty or risk is substantial.

6.5 Human oversight: Assign roles for reviewing outputs and decisions. Ensure staff can inspect relevant information, challenge recommendations, correct errors, and escalate concerns. For high-impact uses, specify which decisions must not be automated and how learners can request review. Oversight must be supported by training, time, and authority.

6.6 Monitoring and evaluation: Evaluate educational outcomes and ethical performance. Potential indicators include learning outcomes, feedback quality, teacher workload, accessibility, user trust, privacy incidents, error rates, disparities between groups, complaints, and frequency of human overrides. Indicators should be selected for the specific use case and interpreted with stakeholder input; they are proposed options, not validated measures. Record adverse events and examine whether benefits justify continued use.

6.7 Continuous improvement: Use monitoring results, complaints, audit findings, and user feedback to update policies, training, procurement decisions, and system configuration. Reassess when a vendor changes the model, new data are introduced, the system is used for a new purpose, or evidence reveals unexpected harms. Restrict, replace, or discontinue a tool when safeguards fail or educational value is not demonstrated.

6.8 Stakeholder participation across the lifecycle: Students can explain effects on learning, autonomy, privacy, and fairness; teachers can judge pedagogical fit; administrators coordinate policy and accountability; developers explain system capabilities and limitations; researchers support evaluation; and policymakers establish broader safeguards. Participation should be accessible and ongoing, and consultation should influence decisions rather than merely communicate them after they have been made.

6.9 Summary tables

Table 1. Ethical principles and implementation responsibilities

Stage	Primary questions and actions	Relevant principles
1. Planning and needs analysis	Define the educational need; consider non-AI alternatives; identify users and success criteria.	Educational purpose, autonomy, inclusion
2. Ethical risk assessment	Assess privacy, bias, safety, impact, contestability, data flows, and regulatory status.	Privacy, fairness, safety, accountability
3. Ethical design and selection	Review evidence, accessibility, explainability, security, vendor terms, and limitations.	Transparency, inclusion, accountability
4. Implementation	Provide policies, training, accessible guidance, and incident-reporting channels.	Integrity, literacy, privacy, safety
5. Human oversight	Assign reviewers with authority to question, override, and correct outputs.	Human oversight, accountability, autonomy
6. Monitoring and evaluation	Track educational outcomes, errors, disparities, complaints, and unintended effects.	Fairness, safety, transparency, governance
7. Continuous improvement	Update safeguards or suspend/replace tools when evidence warrants.	All principles; stakeholder participation throughout

Table 2. Illustrative indicators for monitoring ethical AI implementation

Dimension	Possible indicator(s)	Interpretation note
Learning value	Evidence of learning outcomes and quality of feedback	Interpret alongside learning objectives and baseline evidence; separate task performance from durable learning.
Fairness	Differences in error rates or outcomes across relevant learner groups	Use appropriate, lawful, and context-sensitive group analysis.
Privacy and security	Data incidents, access violations, retention compliance	Record incidents and verify vendor and institutional controls.
Transparency	Users informed about AI use; explanations available where needed	Measure whether information is understandable to users.
Human oversight	Human review, overrides, appeals, and correction rates	A low override rate alone does not prove good oversight.
Accessibility and equity	Tool usability, access barriers, availability of alternatives	Include feedback from disabled and underserved learners.
Trust and accountability	Complaints, response times, and resolution outcomes	Review patterns and whether redress is effective.

Staff capacity	Training completion and educator-reported preparedness	Pair participation data with evidence of practical competence.

These indicators are illustrative and should be selected according to the educational purpose, risk level, available data, and local requirements. They are not validated scales, and institutions should avoid collecting additional learner data solely to populate a dashboard.

Figure 1. The framework is a seven-stage cycle (planning, risk assessment, design and selection, implementation, human oversight, monitoring, continuous improvement) with stakeholder participation spanning every stage. The stages are iterative: monitoring and stakeholder feedback may require institutions to revisit the initial risk assessment or reconsider whether a tool should remain in use. *(The original figure was not included in the supplied text and should be reinserted here.)*

7. Discussion

7.1 From principles to implementation: The framework's central proposition is that ethical principles become useful when attached to decisions, responsibilities, and review points. Privacy should influence needs analysis, procurement, data configuration, user guidance, and monitoring. Fairness requires pre-deployment testing and post-deployment checks. Human oversight requires decision rights, training, and appeal mechanisms. This lifecycle approach complements high-level guidance by asking institutions to specify who acts, when they act, and what evidence informs judgement.

7.2 Relationship to existing frameworks: The framework is consistent with guidance that places human dignity, rights, fairness, transparency, and accountability at the centre of AI governance (UNESCO, 2021), and draws on education-specific concerns documented in reviews of school and higher-education contexts (Adams et al., 2023; Bond et al., 2024; Collin et al., 2024; Memarian & Doleck, 2023; Mouta et al., 2024). The Brookings 'prosper, prepare, protect' agenda (Burns et al., 2026) and the OECD's emphasis on pedagogical intent (OECD, 2026) point the same way. Its intended contribution is organizational rather than normative: it does not claim new ethical values, but arranges established values across implementation stages and connects them to evaluation and continuous improvement.

7.3 Governance, procurement, and evidence: Surveys in 2026 show that many educators use AI without formal guidance (Section 4.5), so governance is the most pressing gap. Institutions should establish policies identifying approved and restricted uses, data-handling requirements, procurement review, incident reporting, human-review obligations, and responsibility for policy maintenance. Because student use outpaces institutional provision (Stephenson & Armstrong, 2026), institutions should also consider providing vetted, data-protective tools so that equitable access does not depend on students' ability to pay for commercial products. Procurement should weigh educational evidence and total implementation requirements, not only features or cost. Vendor documentation should be examined for data reuse, retention, model updates, security, accessibility, and the ability to export or delete institutional data. Small pilots can expose operational problems before scale-up.

7.4 AI literacy and assessment: AI literacy should be embedded in teacher development and student learning. Teachers need support to evaluate tools against learning objectives, verify generated content, design suitable assessments, and respond to ethical incidents. Students need to practise critical evaluation, disclose assistance where required, protect personal information, and recognize when independent work is expected. Assessment can include drafts, process logs, oral explanation, demonstrations, reflective commentary, supervised tasks, and authentic projects, chosen according to discipline and outcomes. These approaches do not eliminate misconduct but can make reasoning and development more visible. The

distinction between augmenting learning and automating coursework (Contractor & Reyes, 2026) and between performance and learning (OECD, 2026) should guide policy design.

7.5 Equity and local context: Institutions should assess differences in connectivity, devices, language, disability access, cost, and digital confidence. Evidence of uneven uptake across income levels and between student groups (OECD, 2026; Stephenson & Armstrong, 2026) suggests that unmanaged adoption may widen gaps. Learners should be able to participate without being forced to disclose sensitive data or purchase particular services. Local curriculum, cultural context, and language needs also matter, and governance should be context-sensitive and informed by affected users.

8. Future Research Agenda

Future work should test the framework through case studies, expert-consensus or Delphi studies, institutional pilots, and longitudinal evaluation. Priority questions include:

1. **Feasibility and usefulness:** Can institutions with different resources apply the framework, and does it improve decision quality, trust, equity, or learning-related outcomes?
2. **Indicators:** Which monitoring indicators are valid, proportionate, and privacy-preserving?
3. **Causal evidence:** What do well-designed studies show about when AI supports learning rather than only performance, and for which learners?
4. **Assessment redesign:** Which assessment designs maintain validity and fairness in AI-rich settings, and how can unfair misconduct accusations be avoided?
5. **Regulatory alignment:** How do requirements such as the EU AI Act's high-risk obligations for education map onto institutional practice as they come into force?
6. **Wellbeing:** What are the effects of emotional reliance on conversational AI among students?
7. **Comparative contexts:** How do schools and higher education, disciplines, languages, and regulatory settings differ?

Evaluations should report unintended effects as well as benefits. A future systematic or scoping review should document its search fully (Section 2.4).

9. Limitations: This paper is conceptual and based on a selective synthesis of published research and institutional guidance. It includes no primary data, systematic-review screening flow, or empirical evaluation of the proposed framework. Because the original source material did not preserve exact database-level search dates, hit counts, and exclusion records, literature identification is not fully reproducible, and the October 2026 update search was purposive. Several 2026 sources are non-peer-reviewed surveys or press releases with different samples, wording, and national contexts, so their figures should not be compared directly. The field changes rapidly, and findings about one system, model version, educational level, or jurisdiction may not generalize. Regulatory dates, in particular, may change and should be verified against current official sources. The principles and framework are a provisional organizing model, not a validated standard or compliance instrument, and the proposed indicators require refinement and testing.

10. Conclusion: AI offers educational opportunities, but responsible adoption depends on more than technical capability or immediate task performance. Current evidence shows that students and teachers have adopted generative AI far faster than institutions have provided policy, training, tools, or assessment redesign. Institutions must therefore address privacy, fairness, transparency, accountability, human oversight,

academic integrity, inclusion, safety, learner autonomy, and continuous governance as connected requirements.

This paper has proposed a seven-stage Ethical AI Implementation Framework linking those principles to planning, risk assessment, system selection, implementation, human oversight, monitoring, and continuous improvement, with stakeholder participation throughout. Its contribution is a practical structure for considering ethical requirements across the implementation lifecycle rather than treating ethics as a policy statement added after adoption. The framework remains conceptual and requires empirical validation. Its value should be judged by whether it helps institutions make more transparent, inclusive, evidence-informed, and accountable decisions while protecting the human purposes of education. AI should support educators and learners, not displace professional judgement or weaken learner agency.

References

- Adams, C., Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence, 4*, Article 100131. https://doi.org/10.1016/j.caeai.2023.100131

- An, Q., Yang, J., Xu, X., Zhang, Y., & Zhang, H. (2024). Decoding AI ethics from users' lens in education: A systematic review. *Heliyon, 10*(20), Article e39357. https://doi.org/10.1016/j.heliyon.2024.e39357

- Bett UK. (2026). *AI in education 2026: Working it out* [Survey of 1,033 teachers conducted by YouGov in partnership with Lenovo, 16-29 June 2026; non-peer-reviewed]. Bett UK. Reported at https://brasil.bettshow.com/en/bett-blog/

- Bond, M., Khosravi, H., De Laat, M., Bergdahl, N., Negrea, V., Oxley, E., Pham, P., Chong, S. W., & Siemens, G. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *International Journal of Educational Technology in Higher Education, 21*, Article 4. https://doi.org/10.1186/s41239-023-00436-z

- Burns, M., Winthrop, R., Luther, N., Venetis, E., & Karim, R. (2026). *A new direction for students in an AI world: Prosper, prepare, protect*. Center for Universal Education, Brookings Institution. https://www.brookings.edu/wp-content/uploads/2026/01/A-New-Direction-for-Students-in-an-AI-World-FULL-REPORT.pdf

- Collin, S., Lepage, A., & Nebel, L. (2024). Ethical and critical issues of artificial intelligence in education: A systematic review of the literature. *Canadian Journal of Learning and Technology, 49*(4), 1-29. https://doi.org/10.21432/cjlt28448

- Contractor, Z., & Reyes, G. (2026). *Generative AI in higher education: Evidence from an elite college* [Working paper, April 2026; not peer reviewed]. arXiv. https://arxiv.org/pdf/2508.00717

- European Union. (2026). Regulation (EU) 2026/1744 amending Regulation (EU) 2024/1689 (Digital Omnibus on AI). *Official Journal of the European Union*, 24 July 2026 (in force 27 July 2026). [Dates as reported by secondary legal sources; verify against EUR-Lex.]

- Ipsos. (2026, June 5). *Teachers concerned about the impact of AI on students' critical thinking* [NPR/Ipsos poll of K-12 teachers]. https://www.ipsos.com/en-us/teachers-concerned-about-impact-ai-students-critical-thinking

- K-12 Dive & Superhuman. (2026, October 6). *Survey of high school educators on AI adoption and policy* [Press release; non-peer-reviewed]. https://www.k12dive.com/press-release/20261006-new-survey-reveals-high-school-educators-are-adopting-ai-but-policies-and-1/
- lAngelone, L., & Burton, B. (2026). More MagicSchool: A follow-up study on the use of generative AI by southwest Ohio teachers. *Ohio Journal of Teacher Education, 41*(1), Article 2. https://www.exhibit.xavier.edu/ojte/vol41/iss1/2
- Langreo, L. (2026, May 28). Teachers say lack of AI guidance is a major problem [Reporting on the Gallup/Walton Family Foundation survey of 2,069 teachers]. *Education Week*. https://www.edweek.org/technology/teachers-say-lack-of-ai-guidance-is-a-major-problem/2026/05
- MacPherson, P. (2026, August). *Survey of Canadian grades 6-12 teachers on AI use*. Fraser Institute. https://www.fraserinstitute.org/sites/default/files/2026-08/survey-of-canadian-grades-6-12-teachers-on-ai-use.pdf
- Memarian, B., & Doleck, T. (2023). Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: A systematic review. *Computers and Education: Artificial Intelligence, 5*, Article 100152. https://doi.org/10.1016/j.caeai.2023.100152
- Metzgar, M. (2026). Student use of generative AI in higher education: Patterns, gaps, and institutional readiness (pp. 41-52). DTI University. https://reference-global.com/article/10.2478/atd-2026-0004
- Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/ZJTE2084
- Miao, F., & Cukurova, M. (2024). AI competency framework for teachers. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000391104
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535
- Miao, F., Shiohira, K., & Lao, N. (2024). *AI competency framework for students*. UNESCO. https://doi.org/10.54675/JKJB9835
- Mondaq / K&L Gates LLP. (2026, August 3). *EU Digital Omnibus on AI enters into force*. https://www.mondaq.com/eu-digital-omnibus-on-ai-enters-into-force/1825722
- Mouta, A., Pinto-Llorente, A. M., & Torrecilla-Sanchez, E. M. (2024). Uncovering blind spots in education ethics: Insights from a systematic literature review on artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 34*, 1166-1205. https://doi.org/10.1007/s40593-023-00384-9
- National Education Union. (2026). *State of education: AI* [Survey of 9,408 teacher members, 5-16 February 2026]. https://neu.org.uk/latest/press-releases/state-education-2026-ai
- OECD. (2026). *OECD digital education outlook 2026: Exploring effective uses of generative AI in education*. OECD Publishing. https://doi.org/10.1787/062a7394-en
- Penuel, W. R. (2026, March 12). *Review of 'A new direction for students in an AI world: Prosper, prepare, protect'*. National Education Policy Center. https://nepc.colorado.edu/review/generative-ai

- Stephenson, R., & Armstrong, C. (2026). *Student generative AI survey 2026* (HEPI Report 199). Higher Education Policy Institute / Kortext. https://www.hepi.ac.uk/reports/student-generative-ai-survey-2026/
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Citation: **Bhattarai. R.,** (2026) "Ethical Principles for Implementing Artificial Intelligence in Education: A Human-Centred Framework and Future Agenda", *Bharati International Journal of Multidisciplinary Research & Development (BIJMRD)*, Vol-4, Issue-10, October-2026.